

1. Sociale wenselijkheid

H. 'T HART¹

SAMENVATTING

Onder sociale wenselijkheid wordt begrepen de tendentie van mensen om bij het beantwoorden van vragen in enquêtes antwoorden te geven die de werkelijkheid 'mooier' voorstellen dan zij is. 'Mooier' wil zeggen dat de antwoorden in overeenstemming zijn met normen die ten aanzien van gedrag en opvattingen in de maatschappij gelden. Zulke normen gelden evenwel niet alleen bij de communicatie die in het kader van onderzoek plaatsvindt maar in het 'werkelijke leven' evenzeer. Mensen nemen deel aan verschillende verbanden en conformeren zich dus aan verschillende, soms tegenstrijdige normen. Bovendien is het van belang dat, al naar hun aard en al naar de omstandigheden, respondenten zich op verschillende manieren in communicatie met anderen zullen presenteren. Sociaal wenselijke antwoorden kunnen soms meer valide zijn dan antwoorden die correct weergeven wat iemand 'eigenlijk' denkt of doet.

1. NORMEN

Bijna 40 jaar geleden kwam Bert Haanstra's film *Alleman* in de bioscoop. Het was een spiegel die Nederland werd voorgehouden en Nederland kon er erg om lachen. Ik herinner mij een hilarische scène waarin mensen op het strand probeerden hun zwemkleden aan te trekken zonder dat anderen op het strand daar aanstoot aan zouden kunnen nemen. Het filmen daarvan gebeurde met een verborgen camera. Ik denk dat degenen die gefilmd werden helemaal van hun moeizame verkleedpartij hadden afgezien als zij geweten hadden dat zij door een camera werden bespied. Maar het zou natuurlijk ook kunnen dat sommige filmspelers in dat geval juist openlijk hun kleren hadden uitgetrokken. Tegenwoordig lijkt het er in ieder geval op dat sommige mensen hun intiemste gedragingen en hun intiemste gevoelens liever voor een camera etaleren dan in de nabijheid van bekenden of vreemden.

Het is duidelijk dat het gedrag van mensen wordt beïnvloed door normen. In kleding, gedrag en uitspraken zijn zij er zich van bewust dat er op hen gelet wordt. Soms zijn die normen geïnternaliseerd, zelfs als men alleen is werken ze door. Die beïnvloeding is ingewikkeld. Is het de vrouw die de handdoek om meneer heen houdt die vindt dat hij niet zo maar zijn broek uit mag trekken of is het meneer zelf of misschien allebei? En als men tegenwoordig vindt dat iets niet aan de openbaarheid mag worden prijsgegeven, waarom dan wel aan een televisie- of filmpubliek? Blijkbaar zijn er verschillen-

1 Dit is een bewerking van mijn college, gehouden op 5 oktober 2001, ter gelegenheid van mijn afscheid van de Universiteit.

de vormen van openbaarheid waarbinnen weer andere regels gelden. Het kan ook zijn dat iemand die zijn emoties toont in de nabijheid van zijn vrienden wordt bewonderd omdat hij zich kwetsbaar opstelt. Datzelfde gedrag geldt als een poging om aandacht te trekken als het zich in het openbaar afspeelt. De maatschappij stelt andere normen dan de werkvloer en daar worden weer andere dingen van mensen verwacht dan in de kring van intieme vrienden.

Wat ik hier vertel is in de sociologie en de sociale psychologie allang bekend. Alleen zal men in die disciplines aan de normen die menselijk gedrag reguleren niet zo gauw het begrip 'sociale wenselijkheid' verbinden. Dat is typisch een begrip uit de literatuur over interviews en enquêtes. Een uitzondering vormen Van der Maele-Vaernewijk, Van Raaij, & Verhallen (1980). Zij beschrijven een onderzoek naar het bevorderen van zuinigheid met energie in het huishouden. Zij noemen dit het bevorderen van sociaal wenselijk gedrag.

Onderzoekers denken vaak te rechtlijnig over sociale wenselijkheid. Zij stellen vragen aan een groot aantal mensen. Zij zijn er zich van bewust dat die mensen de werkelijkheid soms mooier voorstellen dan zij is. 'Mooier' is in hun ogen een vertekening in de richting van wat algemeen in de maatschappij als wenselijk wordt beschouwd. Zij gaan er van uit dat alle respondenten dezelfde normen hebben als de onderzoekers zelf, normen die bepalen wat wel en wat niet gepast is.

Het onderscheid tussen gedrag in overeenstemming met de normen van kleine groepen in de directe omgeving: het microniveau, de grotere collectiviteiten waarvan mensen deel uit maken, het mesoniveau en de maatschappij, het macroniveau, wordt in de literatuur over het interview zelden gemaakt. Wel door Jan van Dijk (1988). Uiteindelijk zijn het de ondervraagden zelf die bepalen aan welke normen zij zich in hun antwoorden conformeren. Zij kunnen deel uitmaken van een subcultuur waarbinnen andere wenselijkheden heersen dan in de grote maatschappij.

Door welke normen worden de mensen die Haanstra heeft gefilmd beïnvloed? Haanstra's film is leuk omdat mensen niet vertellen waarom ze sommige dingen doen en andere juist laten en de kijker kan daar dus zijn eigen verklaring voor geven. Onderzoekers daarentegen, kunnen met hun onderzoekssubjecten praten, of zij kunnen lezen wat die subjecten schrijven. We kunnen de mensen uit de film waarschijnlijk nu niet meer vragen waarom zij deden wat ze deden maar bij eigentijdse observaties kan dat wel. Dat brengt ons wat dichterbij het begrip sociale wenselijkheid, want dat is verbonden met het beantwoorden van vragen.

2. ONDERZOEK NAAR SOCIALE WENSELIJKHEID EN DE INVLOED VAN DERDEN OP HET VERLOOP VAN VRAAG-GESPREKKEN; DE SAMENWERKING TUSSEN BINNEN- EN BUITENUNIVERSITAIRE ONDERZOEKERS

Mijn universitaire werkzaamheden ben ik begonnen met een aanstelling bij het toenmalige Seminarium voor Massapsychologie, Propaganda en Openbare Mening van de Gemeente-Universiteit van Amsterdam. Ik werd daar speciaal belast met de dagelijkse werkzaamheden voor de Steinmetz Stichting voor het Opslaan en Toegankelijk Maken

van Sociaal Onderzoek. Marten Brouwer, die mij daar aanstelde en de Steinmetz Stichting had opgericht, vond dat universiteiten en hogescholen en particuliere instellingen die markt- en opinieonderzoek deden elkaar wederzijds konden aanvullen en bevruchten. Aan de term 'sociaal onderzoek' werd dan ook de voorkeur gegeven boven 'sociaal-wetenschappelijk onderzoek' opdat niemand zou denken dat alleen grondmateriaal van universitair onderzoek de moeite van het bewaren waard zou zijn. Dat betekent overigens niet dat hij dacht dat markt- en opinieonderzoek per definitie onwetenschappelijk zou zijn en universitair onderzoek uitsluitend wetenschappelijk. In die visie kon ik mij, als oud-medewerker van de Nederlandse Stichting voor Statistiek, goed vinden en dat kan ik nog steeds. Die samenwerking tussen privaat en universitair onderzoek gold overigens niet alleen voor de verzameling van grondmateriaal van afgesloten onderzoek maar ook voor het nieuwe onderzoek waarbij het Seminarium, later omgedoopt in Baschwitz Instituut was betrokken. En zo werd het onderzoek naar meningen over homoseksualiteit (Meilof-Oonk, 1969) bedacht door mensen die aan de Universiteit van Amsterdam en de Vrije Universiteit werkzaam waren en uitgevoerd door Attwood Statistics.

Toen ik het plan opvatte om een proefschrift te schrijven stond mij dus het materiaal van dat onderzoek ter beschikking. Ik had ook toegang tot een rijke bron van onderzoeksmateriaal dat in het kader van ander onderzoek door instellingen voor markt- en opinieonderzoek was verzameld.

In 1974 promoveerde ik op een dissertatie, getiteld *Selectie en zelf-selectie van informanten in enquêtes* (let op de spelling uit de jaren '70). Bij mijn speurwerk stuitte ik op onderzoekingen waarvan de steekproef een aantal paren van gezinsleden omvatte. Zulke steekproeven kwamen namelijk tot stand door een selectie van adressen maar als er op zo'n adres veel personen woonden die aan de kenmerken van de beoogde populatie voldeden had elk van die personen minder kans in de steekproef te komen dan wanneer er op een adres maar weinig in aanmerking komende personen woonden. Om daar iets aan te doen werden er soms twee personen op een adres in het onderzoek betrokken, meestal betrof het een ouder en een inwonend kind, een jongere. De efficiency van zo'n clustering staat of valt natuurlijk met de mate waarin de beide leden van hetzelfde gezin dezelfde kenmerken hebben. Komen ze volledig overeen dan levert de tweede ondervraagde geen nieuwe informatie, verschillen de twee net zoveel van elkaar als twee willekeurige ondervraagden dan kunnen we beide ondervraagden beschouwen als onafhankelijk.

In eerste instantie vond ik meer overeenstemming in de beantwoording van attitudevragen door paren van gezinsleden dan men op kansbasis zou verwachten maar dat bleek voornamelijk het gevolg van overeenstemming in zogenaamde achtergrondkenmerken. Zo bleek bijvoorbeeld de overeenstemming in kerkelijke meelevendheid tussen gezinsleden groot. Dat had te maken met hun gemeenschappelijke kerkelijke gezindheid en met overeenstemming in de frequentie van kerkbezoek. Dat wijst erop dat beide gezinsleden zich conformeren aan normen die gelden voor de wijdere verbanden waaraan beiden deelnemen. Als mensen burens hebben die geregeld naar dezelfde kerk gaan of er is een dominee wie het opvalt als gezinsleden een dienst niet bijwonen kan er sprake zijn van sociale controle. Mensen kunnen bij het beantwoorden van enquêtevragen ook het gevoel hebben dat ze worden gecontroleerd: ze geven op vaker naar de kerk te gaan dan juist is (of juist minder vaak als ze denken dat vaak naar de kerk gaan tegenwoordig een beetje belachelijk is). Met het beschikbare materiaal was het niet mogelijk zulke veronderstellingen te toetsen.

Met de gegevens van het onderzoek naar meningen over homoseksualiteit kon ik wel nagaan wat er met de antwoorden van de respondenten gebeurde als een lid van het gezin dat zelf deel uitmaakte van de onderzoeksgroep aanwezig was bij het gesprek met de ander. Ik vond enige aanwijzingen dat er binnen het gezin sociale controle werd uitgeoefend. Als paren van gezinsleden in elkaars aanwezigheid werden ondervraagd was de overeenstemming in attitudes groter dan wanneer dat niet het geval was. Dat was niet zo voor de kerkelijke meelevendheid. Daarvoor gelden kennelijk normen van andere maatschappelijke verbanden. Let wel, het ging mij toen om de kenmerken van de steekproef en niet om het vaststellen van sociale wenselijkheid. Zelfs was het mij in eerste instantie niet te doen om het vaststellen van de invloed van de aanwezigheid van derden op de beantwoording van enquêtevragen. Wel werd deze zijweg nadien een hoofdroute van mijn belangstelling.

In mijn verdere loopbaan aan de Amsterdamse universiteit was ik nauw betrokken bij het zogenaamde kontinu-onderzoek van de FSW-A (Politologie) dat eerst door de N.V., voorheen Nederlandse Stichting voor Statistiek en later door het NIPO werd uitgevoerd. Die uitvoering werd trouwens gesteund door een vruchtbare gedachteswisseling tussen marktonderzoekers en universitaire onderzoekers, waarbij de ervaring van de marktonderzoekers ons ook zeer heeft geholpen bij de opzet van het onderzoek. In het kader van dat kontinu-onderzoek kon ik nader ingaan op de invloed van anderen, aanwezig bij het vraaggesprek ('t Hart, 1978). Een onverwachte bevinding was dat mannen in aanwezigheid van de vrouw des huizes die zich met het gesprek bemoeit, vaker milde vormen van agressie toegeven (bijvoorbeeld met de deuren slaan) dan als er geen vrouw is die zich met het gesprek bemoeit. Ik heb dat het 'leugendetector-effect' genoemd. Als ik daarin gelijk heb dan betekent sociale controle tijdens een gesprek soms dat sociaal ongewenst gedrag juist wordt toegegeven.

Samenwerking tussen onderzoekers binnen en buiten de universiteit bestond niet alleen aan de Universiteit van Amsterdam. Toen ik eind 1981 aan de Rijksuniversiteit te Utrecht ging werken trof ik in mijn groep onderzoekers die bezig waren aan een onderzoek in samenwerking met het Sociaal en Cultureel Planbureau. Materiaalverzamelaars van hun onderzoek waren de N.V. voorheen Nederlandse Stichting voor Statistiek en het NIPO. Die onderzoekers zouden daar in 1988 op promoveren (De Goede, & Maassen, 1988, Maassen, & De Goede, 1988). In het voorbijgaan ontdekten zij hoe een verandering van de volgorde van enquêtevragen tussen twee golven van een longitudinaal onderzoek een schijnbare verandering in de verdeling van opinies over werklozen en arbeidsongeschikten tot gevolg kan hebben. En er was materiaal om door te gaan met het onderzoek naar de invloed van derden op het vraaggesprek. Ook in het onderzoek Wendingen in de Levensloop dat in 1991, 1994 en 1997 onder jongeren werd gehouden was er een vruchtbare samenwerking tussen universitaire onderzoekers en marktonderzoekers, de laatsten van het Bureau Intomart (Meeus, & 't Hart, 1993). Ook in die enquête zijn vragen gesteld naar de aanwezigheid van en de bemoeienis met het vraaggesprek door andere aanwezigen, waarover op een internationaal methodologencongres al eens is gerapporteerd ('t Hart, 1992-II).

De samenwerking tussen onderzoekers binnen de Universiteit en marktonderzoekers daarbuiten is ook terug te vinden in een bron van belangrijke publicaties over onderzoeksmethoden, het Jaarboek van de Nederlandse Vereniging van Marktonderzoekers, later de Nederlandse Vereniging voor Marktonderzoek en Informatiemanagement, en nu kortweg Marktonderzoekassociatie. De jaarboeken van 1975 tot en met 1999 zijn vastgelegd op een CD-Rom. De trefwoorden 'sociale wenselijkheid' en 'sociaal wense-

lijk', al of niet verbonden door een streepje, leveren 15 publicaties op die in die 25 jaar zijn verschenen. Die publicaties zijn verspreid over de hele periode. In sommige van die 15 wordt alleen maar zorg uitgesproken over het verschijnsel. Andere schrijvers noemen sociale wenselijkheid in hun beschouwing, voorafgaand aan hun onderzoeksvragen of hypothesen maar zij verrichten geen onderzoek naar het verschijnsel sociale wenselijkheid zelf. Niet alle 15 publicaties zullen daarom in de hier volgende beschouwing terugkomen. Soms vindt men in de jaarboekartikelen een aarzelend begin van theorievorming. In mijn eigen aanvankelijke publicaties over sociale wenselijkheid en de invloed van de aanwezigheid van derden ontbrak een degelijke theorie. Die theorie werd gevonden in het theoretisch werk dat is verricht naar het begrip 'zelfpresentatie'. De nog te noemen proefschriften van Den Boon, Hofhuis en Luijpers zijn geheel of gedeeltelijk gebaseerd op gegevens die door bureaus voor markt- en opinieonderzoek zijn verzameld. Over de meeste van die gegevensverzamelingen is hierboven iets gezegd.

3. ZELFPRESENTATIE

Juist om wat meer te weten te komen over de complexiteit van sociale wenselijkheid ben ik deze uiteenzetting begonnen bij gedrag, geobserveerd door een filmmaker. Dit gedrag staat schijnbaar ver af van het beantwoorden van enquêtevragen. Iets dichter bij het beantwoorden van vragen van onderzoekers staat de neerslag van verbaal gedrag 'in het werkelijke leven', dus juist niet ten behoeve van onderzoek.

Als voorbeeld neem ik een gedicht uit de poëzie kalender (Van der Graft, 1985). Dit gedicht spreekt ons gezin erg aan. Er wordt een herberg in beschreven waar we met de hele familie gegeten hebben.

HOTEL DU MAL-AIMÉ, STAVELOT

In de herberg waar Apollinaire
zeventig jaar geleden
op een vroege morgen uitkneep,

het was in de herfst, nog in October,
het moet er geroken hebben
naar houtvuur en leerlooierijen
en paardetuig in de straten,

aten wij, de gelagkamer
nog weinig veranderd, de ramen
uitziende op een cour

en ik vroeg haar of ze een ander
liever gehad zou hebben
omdat dichters van oudsher
onbetrouwbaar zijn, ook
in de stabiele zomer,

maar zij lachte zonder argwaan,
verklaarde mij teder voor gek.

Wat weet zij, zelfs na een lente,
van zoveel jaar, hoe de jachthoorn
maar moeilijk in de
herfstwind sterft.

GUILLAUME VAN DER GRAFT

Ik heb het gedicht gekozen omdat er vormen van sociale wenselijkheid in voorkomen: “en ik vroeg haar of ze een ander liever gehad zou hebben omdat dichters van oudsher onbetrouwbaar zijn, ook in de stabiele zomer.”

Schutz (1998) noemt deze wijze van ‘zelfpresentatie’ “cautious self-description”. Mensen die zich op deze wijze voordoen, beschrijven zichzelf op een bescheiden manier. Het komt zelfs voor dat zij over zichzelf geringschattend praten, zegt Schutz in navolging van anderen. Hier komen we het begrip ‘zelfpresentatie’ tegen, een begrip dat Hannelore Hofhuis (1995) in verband heeft gebracht met het interview. Zij spreekt over assertieve en defensieve zelfpresentatie. Wat Van der Graft laat zien is een vorm van defensieve zelfpresentatie. Hij doet het slim, hij doet niet geringschattend over zichzelf maar over dichters in het algemeen en laat zijn tafelenote zo de ruimte om te zeggen dat dichters best te vertrouwen zijn. Maar zelfs als zij dichters in het algemeen onbetrouwbaar vindt kan zij voor Van der Graft een uitzondering maken. Uiteraard formuleert de dichter haar reactie subtieler, zoals dat in een goed gedicht hoort: “maar zij lachte zonder argwaan, verklaarde mij teder voor gek.” Volgens de geijkte sociale normen is ‘gek’ minder wenselijk dan ‘onbetrouwbaar’ maar omdat we weten dat zijn vriendin hem *teder* voor gek verklaart krijgt dat woord een sociaal wenselijke betekenis. Zelfs aan ‘onbetrouwbaar’ wordt in het slot een positieve draai gegeven: de jachthoorn sterft maar moeilijk in de herfstwind. Zo zien we dat iemand zichzelf positief kan presenteren door zogenaamd geringschattend te spreken over zichzelf of over de groep waartoe hij behoort. Hij kan ook in het algemeen negatief gewaardeerde eigenschappen in een positief daglicht plaatsen. Dat is een vorm van assertieve zelfpresentatie.

Guillaume van der Graft of eigenlijk dominee Wilhelmus Barnard moet een jaar of 65 zijn geweest toen hij dit gedicht schreef. Hij suggereert dat zijn vriendin een stuk jonger was dan hij. Misschien heeft hij haar wel bedacht. Schutz spreekt in dit verband van ‘zelfpromotie’. Van der Graft claimt dat hij het voor elkaar gekregen heeft een jonge vrouw mee uit eten te nemen.

Ik ben begonnen met de constatering dat observeerbaar gedrag vaak in overeenstemming is met normen die in groepen van verschillende omvang gelden. Men kan er vrijwillig voor kiezen om aan zulke normen te voldoen maar het kan ook zijn dat iemand uit onze naaste omgeving daar prijs op stelt. Vervolgens heb ik vastgesteld dat een onderzoeker dichter bij de motivatie van gedrag komt als gebruik wordt gemaakt van het feit dat mensen kunnen spreken en schrijven. Door het analyseren van een gedicht werd duidelijk dat iemand zich defensief of assertief kan presenteren. Het gaat er steeds om een ongunstige, dus sociaal verwerpelijke indruk te vermijden en een gunstig, dus sociaal gewenst beeld van zichzelf op te roepen. Maar voor zo’n zelfpresentatie zijn verschillende strategieën mogelijk. Het vereist een analyse om vast te stellen *of* iemand zich op een assertieve of een defensieve manier presenteert maar ook *hoe* hij of zij dat doet.

4. GEWENSTE MENINGEN OF GEWENSTE IDENTITEIT?

Dovidio, & Fazio (1992) brengen sociaal wenselijke uitingen in verband met attitudes. De neerslag van een attitude is een uiting. Men kan zich uiten tegenover interviewers maar ook schriftelijk of via de computer. Dovidio en Fazio onderscheiden twee niveaus van presenteren. Op het eerste niveau gaat het om het publiekelijk uiten van sociaal

gewenste opvattingen, al weet de ondervraagde dat hij stiekem andere gedachten over het attitudeobject heeft. Op het tweede niveau gaat het om het vestigen van een positief zelfbeeld dat het ideale zelf weergeeft en niet het werkelijke zelf. Mensen willen soms zelf niet weten dat er verschil bestaat tussen wie men is en wie men graag zou willen zijn. Dan is het moeilijk om werkelijke van sociaal wenselijke opvattingen te scheiden. Trouwens ook gedrag waar men zich voor schaamt wordt soms echt vergeten.

Schutz (1998) probeert de verschillende zelfpresentatie-stijlen in verband te brengen met karaktereigenschappen. Zo legt ze zelfs verbanden tussen stijlen van zelfpresentatie met DSM4, de vierde editie van de standard Diagnostic and Statistical manual of Mental Disorders. Een excessief gebruik van de assertieve variant zou bijvoorbeeld gerelateerd zijn aan de narcistische persoonlijkheid. Maar zij zegt ook dat dezelfde mensen verschillende stijlen op verschillende tijdstippen kunnen vertonen. Dat hangt van de situatie af. Degene die de geënquêteerde voor ogen heeft bij het doen van uitspraken bepaalt hoe iemand zich presenteert. Als iemand bijvoorbeeld denkt een expert te zijn op een bepaald terrein zal de presentatie wellicht assertief zijn: "Neem dat nou maar van mij aan, ik zit dagelijks in de trein: ze hadden de spoorwegen nooit moeten privatiseren." Maar als men denkt dat anderen er veel meer van weten zal de presentatie defensief zijn. "Ik ben maar een eenvoudige stadsbewoner maar volgens mij denken die boeren meer om de centen dan om de beesten, denkt u ook niet?"

Als het gaat om het vestigen van een positief zelfbeeld kan iemand nog iets anders doen dan opvattingen verkondigen die algemeen aanvaard zijn. Men kan graag worden gezien als waarheidlievend. Men kan zichzelf als een onder alle omstandigheden eerlijk persoon beschouwen. Iemand kan niet wenselijke opvattingen ventileren of antisociaal gedrag rapporteren omdat hij of zij eerlijk wil zijn. Dat kan worden bevorderd als iemand anders bij een vraaggesprek betrokken is, die de ondervraagde kan betrappen op leugens, zoals Hofhuis (1995) in haar proefschrift heeft aangetoond.

Ook kan een geënquêteerde zich onder bepaalde omstandigheden voordoen als iemand die onafhankelijk van geest is, zich juist niet conformeert aan wat algemeen aanvaard is. Maar Verhage, & Richins (1987) vonden bij Amerikaanse respondenten geen verband tussen sociale wenselijkheid (het verhaal halen als een aankoop niet bevalt) en de positie op assertiviteit- en agressiviteitschalen. Dat zou men wel verwachten als het zich voordoen als onafhankelijke geest een vorm van assertiviteit is. Dat mensen zich graag als non-conformist voordoen is overigens een vrij algemeen verschijnsel. Schwartz, Hippler, Deutsch, & Strack (1985) hebben aangetoond dat Duitse ondervraagden in meerderheid denken dat zij minder vaak televisie kijken dan de meeste van hun landgenoten. Billiet, Loosveldt en Waterplas (1988; zie ook 't Hart, 1992-I) hebben hetzelfde gevonden voor de Belgen. Fields, & Schuman (1976) vonden dat inzake rassenscheiding, inwoners van Detroit aan anderen vaak dezelfde mening toeschreven als zij zelf hadden. Maar als er een verschil was dacht men dat anderen conservatiever waren dan zijzelf. Dat gold zelfs voor de naaste burens en de partner. Weinig TV-kijken en progressieve opvattingen worden kennelijk als sociaal gewenst beschouwd. Kalfs (1993) vond in een onderzoek naar de validiteit van de antwoorden op vragen naar tijdbesteding dat hoger opgeleiden ten overstaan van een interviewer vaak zeggen minder televisie te kijken dan zij blijkens metingen via een knoppenmeter doen. Zij meent dat deze ondervraagden de interviewer een gunstig beeld van zichzelf willen geven. Interessant is overigens in dit verband dat Ganzeboom (1984) in een publicatie over deelname aan culturele activiteiten, zegt gebruik gemaakt te hebben van tijdbestedingon-

derzoek omdat men bij dit type onderzoek minder risico loopt sociaal gewenste antwoorden te verkrijgen dan in surveys waarin is gevraagd naar het bezoeken van culturele instellingen. Als in dat tijdbestedingonderzoek gebruik is gemaakt van enquêtrices behoeft dat dus niet waar te zijn, maar de publicatie van Kalfs kon Ganzeboom nog niet kennen.

Hoe dan ook, het is vooral ook wenselijk dat men zich in deze onderscheidt van de meerderheid. Dat is opmerkelijk. Het zou interessant zijn om uit te zoeken of dit het gevolg is van de interviewsituatie: ondervraagden willen zich vooral ten opzichte van de enquêtrice voordoen als anders dan de massa. Ook dit kan zijn beïnvloed door een ideaal-ik: mensen zien zichzelf graag als anders dan de meesten. Deze redenering wordt ook gevolgd door De Kruijk, Pieters, & Van Raaij (1994). Hun veronderstelling is, dat hoe meer sociaal gewenst een bepaald gedrag is, hoe beter men zichzelf vindt ten opzichte van anderen. Dat komt door een positief zelfbeeld. De theorie die aan deze voorspelling verwant is, betreft de fundamentele attributiefout: eigen succes is het gevolg van eigen inspanning, eigen falen komt door de omstandigheden. Bij anderen is het juist andersom. Zij voorspellen dat hoe meer men bij een bepaald onderwerp betrokken is, hoe groter het verschil tussen eigen gedrag en dat van anderen. Die hypothese wordt niet bevestigd.

Mensen kunnen dus sociaal wenselijke opvattingen ventileren of sociaal wenselijk gedrag rapporteren ook al weten ze dat ze daarmee bezijden de waarheid zijn. Zij kunnen dat ook doen omdat zij graag zo willen zijn als zij zich voordoen. Zij kunnen zich op een minder gewenste manier uiten als zij menen dat zij eerlijk moeten zijn. Of zij kunnen, zelfs in strijd met de waarheid, juist minder gewenste opvattingen naar voren brengen of antisociaal gedrag rapporteren om te laten zien dat zij non-conformistisch zijn. Zij beschouwen dat kennelijk als wenselijk. Als iemand zichzelf ziet als een non-conformist hoeft hij geen lid te zijn van een non-conformistengroep. Toch zal hij liever met andere buitenbeentjes verkeren dan met grijze muizen. Een non-conformist beschouwt zichzelf als behorende tot een bepaalde categorie in de maatschappij en conformeert zich aan de normen van die categorie.

5. VERDEDIGING VAN DE EIGEN ORGANISATIE

Tot nu toe vatten we sociaal wenselijk gedrag op als iets van een individu. Mensen maken evenwel deel uit van kleinere en grotere groepen. Zij hebben niet graag dat die verbanden in een kwaad daglicht worden gesteld. Martha Luna Rubio (1996) heeft een proefschrift geschreven dat heet 'Leerlingbegeleiding: wens en werkelijkheid'. Haar onderzoeksgroep bestond uit scholen, die zelf een informant over de leerlingbegeleiding mochten aanwijzen. Deze informanten hebben schriftelijk vragenlijsten beantwoord waarin een groot aantal items was opgenomen over het onderwerp. Het viel haar op dat de informanten een wel heel rooskleurig beeld schetsten van de leerlingbegeleiding op hun school. Wens en werkelijkheid vielen bijna samen. Bij een tweede afname heeft zij daarom een instrument opgenomen om sociale wenselijkheid te meten. Een typisch item van die schaal is: "De persoonlijke ontwikkeling van de leerlingen is op onze school net zo belangrijk als hun leerprestaties". Het scheppen van een evenwicht tussen persoonlijke ontwikkeling en leerprestaties is sociaal wenselijk. De kans is groot dat informanten die het met deze uitspraak geheel eens zijn sociaal wenselijke antwoor-

den geven. Zij zeggen daarmee iets over het samenvallen van de feitelijke en de gewenste situatie in de organisatie. Alleen indirect zeggen ze ook iets over zichzelf: ik werk op een school waar dat evenwicht in orde is.

6. MEETINSTRUMENTEN EN KWALITATIEF ONDERZOEK

Daarmee zijn we gekomen bij pogingen om sociale wenselijkheid te meten. Edwards (1953) is daarmee begonnen. Hij gebruikte in 1957 ook het eerst het begrip ‘sociale wenselijkheid’ (Edwards, 1957). Crowne en Marlowe (1964) leveren kritiek op de schaal van Edwards omdat deze vooral persoonlijkheidskenmerken zou meten. Zij komen met een eigen schaal die niet is afgeleid uit persoonlijkheidsvragenlijsten: de MC-schaal. Arie den Boon (1987) heeft in zijn proefschrift een verkorte vorm van deze schaal opgenomen. Een typisch item is: “Als ik een fout maak ben ik altijd bereid om dat toe te geven”. Respondenten konden op de schaal van Marlowe, & Crowne aangeven of dit waar of onwaar is. Het is onwaarschijnlijk dat iemand alle fouten altijd toegeeft. Volgens Crowne en Marlowe tonen degenen die hoog scoren op de schaal een gebrek aan assertiviteit. Het gaat dus om defensieve zelfpresentatie. Den Boon constateert dat de positie op de schaal voor sociale wenselijkheid niet correleert met eigen opvattingen over een aantal politieke strijdpunten. Misschien is de schaal toch vooral geschikt voor het vaststellen van sociaal wenselijke antwoorden op tests. Maar er is wel een verband met antwoorden op de vraag hoe de meeste andere mensen erover denken. Er is een schaal te construeren voor het aantal malen dat de ondervraagde denkt dat de meeste mensen in Nederland over een onderwerp “net als ik” denken, onafhankelijk van het onderwerp en onafhankelijk van de richting van de eigen mening. Fields & Schuman (1976) noemen dit ‘spiegelperceptie’. Welnu, hoe hoger de score op de schaal voor sociale wenselijkheid, hoe vaker de respondenten (in dit geval studenten) denken dat anderen over een onderwerp net zo denken als zij. Studenten die niet kunnen toegeven ooit wel eens iets te doen wat eigenlijk niet in de haak is, die zichzelf dus defensief presenteren, menen dat hun opvattingen gesteund worden door anderen. “Niet alleen ik denk er zo over, heel Nederland is die mening toegedaan.”

De schaal van Crowne en Marlowe zou ook een rol kunnen spelen in internationaal vergelijkend onderzoek. Steenkamp (1992) beveelt aan om instrumenten die gebruikt worden in marktonderzoek op Europees niveau te correleren met deze schaal. Op die manier krijgt men inzicht in de verschillen tussen landen voor wat betreft de mate waarin respondenten sociaal wenselijke antwoorden geven op items van een instrument.

Ook Luna Rubio (1996) heeft een sociale wenselijkheidschaal opgenomen, ditmaal één waarin een rooskleurig beeld wordt gegeven van de organisatie waarin de informant werkt. Zij gebruikt dit instrument om te controleren of de structuur van haar instrumentarium niet wordt beïnvloed door sociale wenselijkheid. Dat blijkt niet het geval. Ook volgt uit haar analyses niet dat de informanten de vragenlijst op een sociaal gewenste manier invullen. Wel hangen antwoorden die blijken gegeven van het meer of minder centraal stellen van de leerling het meest samen met sociale wenselijkheid, zoals gemeten. Juist bij die antwoorden is voorzichtigheid geboden. Op statistische wijze controleren voor sociale wenselijkheid kan betekenen dat het kind met het badwater wordt weggegooid. Luna Rubio geeft dit ook zelf aan. In dit geval betekent dit dat de verschillen tussen de scholen in de mate waarin wens en werkelijkheid met elkaar overeenstemmen, buiten beschouwing worden gelaten. En die verschillen tussen scholen zijn in het

onderzoek naar de kwaliteit van de leerlingenbegeleiding nu juist zo belangrijk. Echter een vergelijking tussen de structuur van het instrument met en zonder controle voor sociale wenselijkheid heeft wel degelijk inzicht opgeleverd.

Niet tevreden met een uitkomst die zou wijzen op het ontbreken van sociaal wenselijke antwoorden heeft Luna Rubio een aantal halfgestructureerde vraaggesprekken met leerlingbegeleiders gehouden. En daaruit bleek dat de werkelijkheid achter hun antwoorden op de vragenlijsten soms, maar niet altijd, minder rooskleurig was dan die begeleiders in eerste instantie zelf wilden toegeven. Vaak deugde de opzet van de begeleiding wel. Andere begeleiders maar ook zichzelf schoten echter nog wel eens tekort in het begeleiden zelf. De leerling stond niet altijd zo centraal als de bedoeling was. Het vraaggesprek bracht het verschil tussen feitelijke en gewenste situatie naar boven. Dat was ook, en dat is belangrijk, voor sommige leerlingbegeleiders zelf onverwacht.

7. NIET MAAR ÉÉN WERKELIJKHEID

Een laatste punt. Wat er tot nu toe is gezegd is niet strijdig met de gedachte dat er één duidelijke feitelijke situatie is. Die situatie kan de omstandigheden van een informant betreffen of iemands gedrag of de meningen die iemand erop na houdt. Wat daarover door iemand wordt gerapporteerd kan van die feitelijke afwijken. Die afwijking gaat in de richting van wat in de groep waartoe iemand behoort of zou willen behoren als wenselijk wordt geacht. Eric Luijpers (2000, p. 69) geeft in zijn proefschrift aan dat in 1991 51,5 % van de ondervraagde jongeren van 15-17 jaar een delict heeft gepleegd. Junger (1987, p. 155) concludeert in een overzicht betreffende de validiteit van zelfrapportage van delinquent gedrag: "Sommigen rapporteren dat in self-report studies naar delinquent gedrag nauwelijks *over*rapportage plaatsvindt" (mijn cursivering). Kennelijk komt overrapportage van crimineel gedrag wel voor. Het is mogelijk dat het in sommige kringen stoer is om iets uitgevreten te hebben. In het algemeen is er onderrapportage en dan conformeren de jongeren die over hun delicten moeten rapporteren zich wellicht aan de heersende normen in de maatschappij. Dit is een systematische fout. Er is misschien een verschil tussen de 51,5 % die Luijpers heeft uitgerekend op grond van een enquête en de 'ware waarde in de populatie' die we zouden krijgen als we iedereen hadden ondervraagd en als niemand sociaal wenselijke antwoorden had gegeven.

De Kruijk, & Hermans (1998) menen op grond van hun onderzoek naar stemintentie inzake het referendum over de Noord-Zuid metrolijn in Amsterdam, dat het verschil tussen stemintentie enerzijds en werkelijke opkomst anderzijds niet zozeer het gevolg is van sociaal gewenste antwoorden in een telefonische enquête als wel van het feit dat niet-stemmers ook minder vaak meedoen met de enquête. Hetzelfde zou gelden voor het verschil tussen het totale gerapporteerde stemgedrag betreffende het referendum over IJburg in de enquête en de werkelijke opkomst bij dat referendum. Die conclusie is echter gebaseerd op de veronderstelling dat niet-opkomen niet minder wenselijk is dan wel opkomen. De opkomst is immers laag. Echter, zelfs als dat in bepaalde omstandigheden zou opgaan is het nog iets anders om toe te geven dat je niet meedoet aan een referendum tegenover iemand die je daar expliciet naar vraagt. Misschien kun je onder je vrienden opscheppen over het feit dat je aan die onzin niet meedoet, degene die je daarover opbelt (in dit geval iemand van de Gemeente of van het NIPO) vindt zoiets misschien wel stom.

Situaties en gedragingen zijn echter vanuit verschillend perspectief te bekijken: wat de één een ruzie noemt vindt de ander een gezond meningsverschil. Iemand die juist een flinke woordenwisseling had kan denken: “Nou daar hadden Jan en ik behoorlijk ruzie!” Na een avondje stappen en een paar biertjes met Jan is het oordeel heel anders “Fijne vent die Jan, ook als je hem de waarheid zegt blijft hij je vriend.” En ook meningen liggen niet voor eens en voor altijd vast. Iemand die een gesprek voert met een ander die een fervent tegenstander is van de doodstraf zal wellicht behoefte hebben ook de positieve kant van die straf naar voren te brengen. “Je moet er ook aan denken wat de familieleden en vrienden van die slachtoffers vinden als de moordenaar na een aantal jaren weer vrij rondloopt.” Dezelfde persoon zal wellicht in een discussie met een uitgesproken voorstander juist de doodstraf als niet in overeenstemming met de rechtsstaat veroordelen: “Denk eens aan al diegenen die achteraf onschuldig bleken maar toen was er niets meer aan te doen.” Een dergelijke redenering zien we ook bij Van Raaij (1976, p. 205). Hij veronderstelde dat respondenten in de loop van het beantwoorden van enquêtevragen zouden gaan begrijpen wat de onderzoeker beoogt. Om bij de onderzoeker in de gunst te komen (ingratiatie) zou de ondervraagde antwoorden in de richting van wat hij denkt dat de onderzoeker voor ogen staat. Vooral als bepaalde standpunten via de media worden verspreid zou de respondent geneigd zijn sociaal gewenste antwoorden te geven. Van Raaij heeft een aantal hypothesen uit deze redenering afgeleid en vervolgens via een ingenieus veldexperiment getoetst. In dat experiment trachtte hij onder andere het consumentenvertrouwen vast te stellen. Het instrument daarvoor bevatte negatief gestelde items. Van Raaij heeft voor een deel van de steekproef deze items omgezet in positief gestelde. Helaas werden zijn hypothesen niet bevestigd. In een beschouwing over de oorzaken van deze bevindingen stelde hij: “Een mogelijke verklaring voor dit tegen-intuïtieve resultaat is, dat een ‘tevredenheidsformulering’ zoveel aversie oproept dat de respondent, zowel ten aanzien van consumentensatisfactie als sociale satisfactie, hier tegen ingaat en reageert door ontevredenheid en klachten te rapporteren. De ‘tevredenheids’-vraagstelling prikkelt tot tegenspraak om de onredelijkheid van de overdreven tevredenheid hierin te compenseren.”

Van Raaij (1976) gaat er van uit dat de manier waarop de respondent tegen de vragen aankijkt mede wordt bepaald door de wens om bij de onderzoeker in het gevlei te komen. Kalfs (1993) denkt dat in een persoonlijk interview de kans op sociaal wenselijke antwoorden groter is dan in een telefonisch interview of bij een schriftelijke enquête, omdat ondervraagden op de interviewer een gunstige indruk willen maken. Een dergelijke veronderstelling ligt ook ten grondslag aan de beschouwing van De Leeuw en Hox (1985). Enquêtes door middel van persoonlijke interviews roepen meer sociaal gewenste antwoorden op dan schriftelijke of telefonische enquêtes. Hun onderzoek wordt ook elders gerapporteerd (De Leeuw, Hox, & Duijx (1985). Zij gaan vooral in op de zogenaamde ‘Dillman-methode’ (Dillman, 1978), een systeem van maatregelen om de response op schriftelijke enquêtes via de post zo groot mogelijk te maken. Tot dat systeem behoort het sturen van een persoonlijke brief aan de beoogde respondent om zo de bereidheid tot medewerking te verhogen. Zij stellen dat door zo’n persoonlijke benadering ook in de schriftelijke enquête meer sociaal wenselijke antwoorden zullen worden gegeven.

Met andere woorden het is soms onmogelijk om het sociaal gewenste antwoord te onderscheiden van de achterliggende feiten omdat de ondervraagde sociale normen heeft geïnternaliseerd en oprecht gelooft dat de situatie is zoals hij of zij haar weergeeft. Soms is het ook onmogelijk omdat er niet maar één werkelijkheid schuilgaat ach-

ter hetgeen een ondervraagde rapporteert. Er zijn verschillende werkelijkheden of er zijn verschillende manieren waarop je naar een werkelijkheid kunt kijken. Hoe men kijkt hangt af van de persoon of de groep waaraan men refereert: op het werk zelf kijkt men anders tegen zijn werk aan dan thuis. Met andere woorden: er is geen ware waarde in de populatie.

8. SAMENVATTING EN CONCLUSIES

Sociale wenselijkheid houdt in dat iemand voldoet aan bepaalde normen. Binnen verschillende sociale verbanden heersen verschillende normen. Als we sociale wenselijkheid willen vaststellen moeten we nagaan aan de normen van welk sociaal verband iemand probeert te voldoen. Gaat het om verbanden op micro-, meso- of macroniveau? Wie is de referentiepersoon of wie vormen de referentiegroep? De wijze van benaderen van de respondent, persoonlijk of schriftelijk via de post bijvoorbeeld, kan ook van invloed zijn op het perspectief van waaruit de vragen worden beantwoord. Hetzelfde geldt voor de omgeving waarin die beantwoording plaatsvindt, thuis of op het werk bijvoorbeeld. Bovendien is het van belang of iemand er zelf voor kiest om zich sociaal wenselijk te uiten of dat anderen druk uitoefenen om hem zo ver te krijgen.

Sociaal wenselijk gedrag, met inbegrip van het uiten van sociaal wenselijke meningen, is een vorm van zelfpresentatie. Zelfpresentatie kan offensief of defensief zijn. Ieder kan bij de presentatie van zichzelf een eigen strategie volgen, een eigen stijl hebben. Dat kan deel uitmaken van iemands karakter maar is ook situatiegebonden.

Sociaal wenselijke antwoorden kunnen worden gegeven omdat die het best worden geaccepteerd of omdat zij deel uitmaken van een ideaal beeld dat iemand van zichzelf heeft. Een respondent hoeft er zich niet van bewust te zijn dat hij of zij de werkelijkheid mooier voorstelt dan zij is.

Het is een vrij algemeen verschijnsel dat individuen aan de meeste mensen in een land hetzelfde gedrag of dezelfde mening toeschrijven als die welke zij zelf aan de dag leggen. Dat heet spiegelperceptie. Maar het komt ook vaak voor dat zij menen dat zij zich van anderen onderscheiden. Vermoedelijk vinden zij dan dat zij zich in gunstige zin van anderen onderscheiden. Kennelijk is dat onderscheid in hun ogen sociaal wenselijk. Men kan als informant die over de gang van zaken in een organisatie rapporteert, ook een sociaal wenselijk beeld schetsen van de werkelijkheid binnen die organisatie: sociale wenselijkheid heeft niet alleen betrekking op individuen.

Er zijn meetinstrumenten die sociale wenselijkheid meten. Het is niet duidelijk of een meetinstrument dat voldoet voor psychologische tests ook geschikt is om sociale wenselijkheid te meten in opinieonderzoek. Ondervraagden die hoog scoren op een schaal voor sociale wenselijkheid vertonen meer spiegelperceptie, denken vaker dat zij in hun mening worden gesteund door de meeste mensen in het land, dan anderen. Een instrument voor het meten van sociale wenselijkheid kan ons inzicht geven in de aard van de structuur van andere meetinstrumenten en in de validiteit ervan. Controle voor sociale wenselijkheid kan betekenen dat we een belangrijke bron van variantie in wat wordt gemeten buiten beschouwing laten. Dat zijn namelijk de verschillen tussen situaties waar het verschil tussen wens en werkelijkheid klein is en situaties waar dat verschil juist groot is.

Kwalitatief onderzoek kan een belangrijk inzicht geven in het verschijnsel sociale wenselijkheid.

De werkelijkheid is niet ongedeeld en ondubbelzinnig. Een beschrijving van die werkelijkheid kan vanuit een bepaald gezichtspunt een vertekening in sociaal gewenste richting lijken, terwijl dat vanuit een ander gezichtspunt niet zo hoeft te zijn.

Wat betekent dat nu voor het enquêteonderzoek? Omdat mensen zich conformeren aan de normen van groepen waartoe zij behoren of graag zouden willen behoren is het niet altijd zinvol om in enquêtes te zoeken naar ‘de ware mening’ of ‘het ware gedrag’ achter de opinies en de gedragingen die respondenten rapporteren. Als de publieke opinie voorwerp van onderzoek is, is de heimelijk gekoesterde mening van minder belang. Als een onderzoeker toch geïnteresseerd is in niet door sociaal wenselijke antwoorden vertekende feiten: ook bij schriftelijke vragenlijsten of vragen gesteld via de computer komt sociale wenselijkheid voor. Soms weten de ondervraagden zelf niet dat zij de werkelijkheid verfraaien uit zelfbescherming. ‘Zelfpresentatie’ speelt in elke communicatie een rol, dus ook bij het beantwoorden van vragen die door een onderzoeker worden gesteld. Op welke manier een ondervraagde zich zal willen presenteren is niet alleen een kwestie van karakter maar ook een kwestie van omstandigheden. Soms komt iemand het liefst als waarheidlievend naar voren, dan weer als brave burger of als non-conformist.

Wat heeft een en ander te maken met de ‘randomized response’ methode die elders in dit Jaarboek wordt behandeld? Deze methode zal er zeker toe bijdragen dat over gevoelige onderwerpen vaak meer waarheidsgetrouw zal worden gerapporteerd dan bij een directe ondervraging. Echter, niet elke respondent die denkt naar waarheid te antwoorden doet dat ook. Die vertekening wordt door de methode niet ondervangen. Én men moet zich afvragen welke reacties het gebruik van de methode zelf oplevert. Voor sommigen kan de gedachte achter de methode moeilijk te volgen zijn. Weinigen willen echter voor domoor worden aangezien en dat kan leiden tot fouten: men houdt zich niet aan de toevalsprocedure. Ook kan de ondervraagde denken dat de onderzoeker via slinkse wegen toch achter gedrag kan komen die hij liever stiekem wil houden. De ondervraagde wil in eigen ogen doorgaan voor iemand die zich door zulke onderzoekerstrucs niet in de luren laat leggen, hij geeft zelfs onjuiste antwoorden op onschuldige vragen. Met andere woorden: de vooronderstellingen achter de methode moeten worden uitgeschreven en getoetst. Die vooronderstellingen hebben ook betrekking op de wijze waarop respondenten zich in een randomized response situatie zullen presenteren.

LITERATUUR

- Billiet, J., Loosveldt, G., & Waterplas, L. (1988). *Response-effecten bij survey-vragen in het Nederlands taalgebied*. Leuven: Katholieke Universiteit.
- Boon, A.K. den (1987). *De perceptie van anderen*. Diss. Universiteit van Amsterdam.
- Crowne, D.P., & Marlowe, D. (1964). *The approval motive. Studies in evaluative dependence*. New York: Wiley.
- Davidio, J.F., & Fazio, R.H. (1992). New technologies for the direct and indirect assessment of attitudes. In J.M. Tanur (Ed.). *Questions about questions. Inquiries into the cognitive bases of surveys* (pp. 204-237). New York: Russell Sage Foundation.
- Dijk, J. van (1988). Aanzetten tot contextuele theorie van het interview. *Mens en maatschappij* 63, 277-293.
- Dillman, D.A. (1978). *Mail and telephone surveys; the total design method*. New York: Wiley.

- Edwards, A.L. (1953). The relationship between the judged desirability of a trait and the probability that that trait will be endorsed. *Journal of applied psychology* 37, 90-93.
- Edwards, A.L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden
- Fields, J.M., & Schuman, H. (1976). Public beliefs about the beliefs of the public. *Public opinion quarterly* 40, 427-448.
- Ganzeboom, H.B.G. (1984). Veranderingen in deelname aan kulturele activiteiten tussen 1955 en 1977. In: NMVI, *Jaarboek 1984* (pp. 85-104).
- Goede, M.P.M. de, & Maassen, G.H. (1988). *Beleving van werken en niet-werken: een onderzoek onder werklozen en arbeidsongeschikten en hun partner*. Diss. Rijksuniversiteit te Utrecht.
- Guillaume van der Graft (Barnard, W.) (1985). Hotel du Mal-aimé, Stavelot. In A. den Besten, P. Gillaerts & T. Govaart (Eds.): *Verzamelde gedichten*. Baarn: De Prom.
- Hart, H. 't (1978) Andere aanwezigen bij het vraaggesprek. In: Werkgroep "kwartaalonderzoek" (Red.) *Röntgenfoto van enquêteonderzoek, een ongeneeslijk ziekte?* (pp. 76-91) Amsterdam Universiteit van Amsterdam, December.
- Hart, H. 't (1992-I). Norms and responses to survey questions. In: H. 't Hart & W. Kox (Eds.), *Proceedings of the international workshop 'cognition and survey methodology' held at Utrecht University November 18-19, 1991* (pp. 127-140). Utrecht: ISOR, Methods series 92-3
- Hart, H. 't (1992-II). Interviewing youngsters in the presence of others. In: *Program and abstracts of the international conference on social science methodology* (pp. 22-26). Trento June 22-26. 1992.
- Hofhuis, E.H. (1995). *Invloed van derde aanwezigen bij interviews*. Diss. Universiteit Utrecht. Utrecht: ISOR
- Junger, M. (1987). Validiteit van zelfrapportage van delinquentie bij jongeren uit etnische minderheden. In J. de Jong-Gierveld & J. van der Zouwen (Red.). *De vragenlijst in sociaal onderzoek. Een confrontatie van onderzoekspraktijk en -methodiek* (pp. 153-174). Deventer: Van Loghum Slaterus.
- Kalfs, N. (1993). Validiteit in tijdsbestedingsonderzoek. Vergelijking van drie dataverzamelingssystemen. In: NMVI, *Jaarboek 1993* (pp. 251-266).
- Kruijk, M.D. de, & Hermans, E. (1998). Effect van non-response op representativiteit. In: NMVI, *Jaarboek 1998* (pp. 55-69).
- Kruijk, M.D. de, Pieters, R.G.M., & Raaij, W.F. van, (1994). Milieuvriendelijk gedrag: een kwestie van willen of kunnen? In: NVMI, *Jaarboek 1994* (pp. 157-173).
- Leeuw, E.D. de, & Hox, J.J. (1985). Recente ontwikkelingen bij schriftelijk onderzoek per post. In: NVMI, *Jaarboek 1985* (pp. 11-25).
- Leeuw, E.D. de, Hox, J.J., & Duijx, A.W.M. (1985). Sociale wenselijkheid bij postenquêtes; onbedoelde effecten van het Dillman systeem. *Sociale wetenschappen* 28, 151-157.
- Luijpers, E.T.H. (2000). *Intentie tot exploratie, sociale binding en delinquent gedrag*. Diss. Universiteit Utrecht. Delft: Eburon.
- Luna Rubio, M. C. (1996). *Leerlingbegeleiding: wens en werkelijkheid*. Diss. Universiteit Utrecht. Leuven-Apeldoorn: Garant.
- Maele-Vaernewijck, M.C.L. van der, Raaij, W.F. van, & Verhallen, Th.M.M. (1980). Energiegedrag in de woning: literatuuroverzicht en gedragsmodel. In: NMVI, *Jaarboek 1980* (pp. 169-189).
- Maassen, G.H., & Goede, M.P.M. de (1988). *Publieke opinie over werklozen en arbeidsongeschikten*. Diss. Rijksuniversiteit te Utrecht.
- Meeus, W., & Hart, H. 't, (1993). (Red.) *Jongeren in Nederland. een nationaal survey naar ontwikkelingen in de adolescentie en intergenerationele overdracht*. Amersfoort: Academische uitgeverij.
- Meilof-Oonk, S. (1969). *Meningen over homosexualiteit*. Amsterdam: Stichting Onderzoek Minderheden.
- Raaij, W.F. van, (1976). Ingratatie bij de beantwoording van enquêtevragen. In: NVMI, *Jaarboek 1976* (pp. 194-213).

- Schutz, A. (1998). Assertive, offensive, protective, and defensive styles of self-presentation: a taxonomy. *Journal of psychology interdisciplinary & applied*. 132 (6), 611-628.
- Schwartz, N., Hippler, H.J., Deutsch, B., & Strack, F. (1985). Response scales: effects of category range on reported behavior and comparative judgements. *Public opinion quarterly* 49, 388-395.
- Steenkamp, J.E.B.M. (1992). Enige aspecten van Europees marktonderzoek In: NMVI, *Jaarboek 1992* (pp. 77-96).
- Verhage, B.J., & Richins, M.L. (1987). Het meten van assertief consumentengedrag in de V.S. en Nederland. In: NMVI, *Jaarboek 1987* (pp. 127-137).

2. Multimediacampagnes: populair maar weinig onderzocht

A.E. BRONNER, P. NEIJENS en W.F. VAN RAAIJ

SAMENVATTING

In de reclame- en mediawereld wordt veel gesproken over multimediaeffecten. Algemeen is de gedachte in die werelden dat de inzet van verschillende mediumtypen effectiever is dan de inzet van één mediumtype. Deze inzet kan gelijktijdig of volgtijdelijk plaatsvinden. Omdat het onderwerp hoog op de agenda staat in de reclamewereld zou men verwachten dat op dit terrein ook uitgebreid onderzoek wordt gedaan en theoretische bespiegelingen plaatsvinden. Dat blijkt niet het geval, en de verklaring hiervoor is eenvoudig: multimediaeffectonderzoek is complex. In dit artikel besteden we aandacht aan een viertal thema's rond multimediaeffecten:

- 1) multimediazet in de praktijk van de mediaplanning,
- 2) beschikbaar onderzoek in Nederland en andere landen,
- 3) uiteenrafeling van het 'effect-begrip' en specificatie wat het multimediaeffect allemaal kan inhouden,
- 4) schets van een in de praktijk haalbare en valide onderzoeksopzet.

1. INLEIDING

Multimediaeffecten ontstaan bij het inzetten van verschillende mediumtypen in een campagne voor een bepaald merk of een bepaalde merkvariant, product of dienst. Een campagne is een geheel van communicatie-uitingen in één of meer media met een bepaalde communicatiedoelstelling binnen een bepaalde periode, bijvoorbeeld de introductie van nieuwe Lassie-rijstvarianten (Gouden Effie 2001). Het idee bestaat dat de inzet van verscheidene media effectiever is dan de inzet van één medium. Deze media kunnen gelijktijdig (simultaan) of volgtijdelijk (sequentieel) worden ingezet. Voorbeelden van in te zetten media zijn: printmedia zoals dagbladen en tijdschriften, ethermedia zoals televisie en radio, voorts direct mail, outdoor (buitenmedia), internet en in-store displays.

Hoewel – zoals we hierna zullen zien – (grote) adverteerders bijna altijd meer media inzetten in een campagne, is over het bereik en de effecten van multimediacampagnes weinig bekend. Dat maakt het bepalen van een adequate mediastrategie er niet eenvoudiger op. In deze bijdrage gaan we na wat er wel bekend is over multimedia, zullen we aangeven welke vragen er spelen, en op welke wijze die onderzocht kunnen worden.

We bespreken achtereenvolgens:

- Het belang van multimediacampagnes en welke combinaties van media het meest worden ingezet.

- In hoeverre voorzien de nationale mediabereiksonderzoeken die mediaplanners tot hun beschikking hebben in de behoefte aan multimediagegevens?
- Wat weten we uit de literatuur over multimediaeffecten, waarbij een onderscheid wordt gemaakt tussen internationaal en nationaal onderzoek.
- De uiteenrafeling van multimediaeffecten.
- Welk onderzoeksvragen zijn nog onbeantwoord en wat is een adequate opzet voor het onderzoek naar bereik en effecten van multimediacampagnes?

2. HET BELANG VAN MULTIMEDIACAMPAGNES

In 2001 liepen ruim 20.000 reclamecampagnes in Nederland¹. In driekwart van de campagnes werd één mediumtype ingezet. Als we ons beperken tot de (1470) grotere campagnes (met mediabestedingen die hoger zijn dan 450.000 euro; deze campagnes nemen 74% van alle mediabestedingen voor hun rekening), zien we dat 85% van de campagnes multimediacampagnes zijn: in 32% van de campagnes worden twee mediumtypen ingezet, in 24% worden drie mediumtypen ingezet, in 19% vier mediumtypen en in 10% meer dan vier.

De populairste mediumtypen voor deze grotere campagnes zijn televisie en tijdschriften (beide ingezet in 71% van de campagnes), dagbladen (58%), radio (44%), buitenreclame (29%) en bioscoopreclame (5%).

Wat betreft de combinaties van mediumtypen zien we in tabel 1 dat de combinatie televisie en tijdschriften het meeste wordt ingezet (15%), daarna de combinatie ether- en printmedia (9%), de combinatie van alle media behalve bioscoop (7%), de combinatie van printmedia (6%), de combinatie radio en printmedia (6%) en de combinatie van televisie en printmedia (5%). De overige mediumtypen worden in minder dan 5% van de gevallen gecombineerd.

Het belang van multimediacampagnes zien we ook bij de winnende Effie-cases. Uit een analyse van de Effie-cases over de jaren 1999-2001 blijkt dat de Effie-winnaars heel vaak meer mediumtypen inzetten (Giling, Neijens, Van den Putte, Sandee en Smit, 2002). Slechts in 14% van de gevallen wordt gebruik gemaakt van één mediumtype. Het vaakst komen voor de combinaties van:

- televisie en radio
- televisie en tijdschriften
- televisie, dagbladen en tijdschriften
- een combinatie van alle mediumtypen behalve bioscoop en dagbladen.

We kunnen dus concluderen dat de inzet van meer mediumtypen in een campagne zeer belangrijk is in de mediaplanning.

¹ Met dank aan John Faasse, Saskia van Oirschot en Suzanne Bruin van Initiative Communications. Gegevens gebaseerd op BBC. BBC vermeldt uitgaven aan media voor producten en diensten (dus niet bijvoorbeeld personeelsadvertenties). Dat betekent dat er onderscheid wordt gemaakt tussen een campagne van de Rabobank voor Beleggingsdiensten en een campagne van de Rabobank voor Hypotheken. De gegevens hebben betrekking op 2001. Zes mediumtypen zijn in de analyse betrokken: tv, radio, dagbladen, tijdschriften, buitenreclame en bioscoopreclame.

Tabel 1. Het belang van de verschillende media en mediacombinaties.

	% van de campagnes
<i>Single mediacampagnes</i>	
• TV	11,8
• Dagbladen	2,0
• Tijdschriften	1,2
• Radio	0,1
• Buitenreclame	0,1
• Bioscoop	0
<i>Multimediacampagnes</i>	
• TV en tijdschriften	14,9
• Ether en print	8,8
• Alle minus bioscoop	6,7
• Dagbladen en tijdschriften	6,3
• Radio en print	6,1
• TV en print	5,2
• Overige combinaties (minder dan 5%)	36,8

N=1470 (campagnes waaraan in 2001 meer dan 450.000 euro werd besteed).

3. DE MEDIAPLANNER EN MULTIMEDIACAMPAGNES

Mediastrategen en mediaplanners die een multimediacampagne willen ontwikkelen hebben geen gegevens voorhanden die hun beslissingen kunnen ondersteunen. De standaard bereiksonderzoeken die de mediaplanner in Nederland en daarbuiten ten dienste staan zijn onderzoeken waarin het bereik van slechts één mediumtype wordt onderzocht (Den Boon & Neijens, red. 2000). Dat geldt in Nederland voor televisie (het Continu Kijkonderzoek), voor radio (het Continu Luisteronderzoek), en het onderzoek naar het bereik van buitenreclame, brievenbusreclame en bioscoopreclame.

Slechts het printmedia-onderzoek van JIC NOM (voorheen SUMMO) omvat meer dan één mediumtype (dagbladen en tijdschriften). Dit betekent dat de mediaplanner slechts cijfers heeft over het bereik (de lees- en kijkkansen) van afzonderlijke media.

Om het *bereik* van een multimediacampagne vast te stellen kunnen de cijfers van de afzonderlijke media niet eenvoudig worden 'opgeteld'. In de eerste plaats omdat de bereikcijfers niet goed met elkaar te vergelijken zijn (gaat het om mediumbereik of reclamebereik?). Bovendien is de communicatiekracht van de verschillende mediumtypen verschillend. En het simpel samenvoegen van de cijfers houdt geen rekening met mogelijke synergie-effecten (zie hierna).

Maar ook al zouden deze bezwaren niet gelden, dan nog is het niet mogelijk om de cijfers uit de verschillende onderzoeken samen te nemen. Immers: het gebruik van verschillende mediumtypen tegelijk geeft ook op zich weer nieuwe effecten en daarom voldoet een statistische oplossing (bijvoorbeeld het vermenigvuldigen van leeskansen) niet.

Fusie van databestanden is in theorie ook een mogelijkheid om te komen tot een inschatting van het bereik van multimediacampagnes, maar blijkt in de praktijk nog onvoldoende betrouwbare gegevens op te leveren.

Voor het bereik en de effecten van multimediacampagnes kan wel gebruik gemaakt worden van campagne-tracking onderzoeken die sommige adverteerders voor hun campagnes laten uitvoeren. De resultaten van deze onderzoeken geven niet alleen (ex-post) inzicht in het succes van een campagne, maar kunnen ook als input dienen voor toekomstige campagnes. Op dit moment is er nog onvoldoende (openbare) informatie uit tracking onderzoeken beschikbaar om de mediaplanners voldoende te ondersteunen.

We kunnen concluderen dat mediaplanners weinig gegevens over bereik en effecten van multimediacampagnes tot hun beschikking hebben. In de praktijk wordt in de mediaplanning dan ook de keuze voor de verschillende mediumtypen op kwalitatieve gronden gemaakt en de media-inzet geoptimaliseerd binnen de verschillende gekozen mediumtypen (Franz, 2000).

Inzichten in de effecten van een multimediaaanpak kunnen ook worden verkregen door ad hoc onderzoek. We presenteren resultaten van ad hoc onderzoek zoals die in de literatuur naar voren komen.

4. WAT WETEN WE UIT DE LITERATUUR?

4.1. *Internationaal*

Het meeste onderzoek naar multimediaeffecten heeft betrekking op de combinatie televisie en print, met name tijdschriften. In de onderzoeken die Consterdine (1990) rapporteert wordt een geselecteerde groep respondenten (gedwongen) blootgesteld aan een tv-commercial voor een bepaald merk, een printadvertentie voor dat merk en opnieuw de tv-commercial. In de controlegroep werd de printadvertentie vervangen door een printadvertentie van een ander merk. Deze printadvertentie was een printadvertentie van één van de andere geteste campagnes. In het onderzoek werd aangetoond dat na het lezen van de printadvertentie de respondenten meer details opmerkten in de daarna getoonde tv-commercial. Ook werd vastgesteld dat productinformatie wel werd opgemerkt in de printadvertentie en niet in de tv-commercial.

Smith (1990; 1997) bespreekt een aantal experimenten (gedwongen blootstelling) en onderzoeken met vrijwillige blootstelling en concludeert dat multimediacampagnes effectiever zijn dan campagnes waarin één mediumtype (televisie) wordt ingezet. Hij verklaart dit niet alleen uit een beter bereik in de doelgroep dat tot stand wordt gebracht door een multimediacampagne, maar ook door een beter effect van de multimediaaanpak. Uit de onderzoeken die hij citeert blijkt onder andere:

- de combinatie van print en tv geeft betere effecten dan het inzetten van print of tv alleen;
- toevoeging van tijdschriften aan een tv-campagne geeft betere effecten dan het toevoegen van meer televisie;
- tijdschriften communiceren extra informatie bovenop de informatie van tv alleen;
- de impact is het grootste als de boodschap en de creatieve uitvoering complementair zijn.

Smith speculeert dat print mogelijk de grootste impact heeft op de linker hersenhelft en televisie op de rechter hersenhelft, waardoor de combinatie de gehele hersenen activeert.

De combinatie van televisie en radio is onderzocht door Edell & Keller (1989). Het betrof hier een laboratoriumonderzoek met studenten als proefpersonen. In het onderzoek werd het audiogedeelte van de televisiecommercial herhaald. Daarbij bleek dat de radioherhaling net zo effectief was als een extra confrontatie met de televisiecommercial.

Wilkins en Ford (1997) onderzochten de mogelijke bijdragen van posters, sponsoring en below-the-line promoties aan tv- en printreclame en vonden daarvoor aanwijzingen. Zij maakten gebruik van een database met ruim 1250 campagnes die op gelijke wijze waren getest. Ze vergeleken de resultaten van campagnes die onderling verschilden wat betreft ingezette mediumcombinaties. De auteurs geven toe dat deze onderzoeksopzet niet heel sterk is. Overigens wijzen ze op het feit dat printmedia een langere tijd nodig hebben om effecten te genereren dan televisie. Dat brengt met zich mee dat onderzoek over een langere periode moet worden uitgevoerd.

De conclusies van Vincent en Vincent (op basis van een analyse van een Franse database met campagne resultaten) waren niet eenduidig wat betreft het multimediaeffect: ‘a media mix may result in synergy, where the additional medium will boost the effectiveness of the main medium (for example, TV plus radio), or in dilution, where the additional medium works in partial substitution of the main medium. In this case, although a multimedia schedule will be more effective than using just one medium, overall effectiveness will be less than an addition of each medium’s individual effect.’ (Vincent & Vincent, 1997).

4.2. Multimediaonderzoek in Nederland

Multimediaonderzoek heeft in Nederland incidenteel en ad hoc plaats gevonden. Wij beschrijven kort enkele voorbeelden van onderzoek waarin verschillende combinaties van media aan de orde zijn geweest. In tabel 2 is een overzicht weergegeven.

Tabel 2. Multimediaonderzoek in Nederland.

<i>naam</i>	<i>opdrachtgever</i>	<i>jaar</i>	<i>media</i>
Het Multi Media Effect	Admedia/Lintas	1989	tijdschrift/tv
Effectiviteit van dagbladcampagnes	Cebuco	1994	dagbladen/andere media
Media Merkbeleving	PTT Post Media-service	1997	mail/andere media
Wisselwerking tussen tv- en instore reclame	IP/Initiative Media	1998	tv/instore
Visual Transfer	Initiative Media/ Radio 538	1998	tv/radio

Geïnspireerd door het Duitse onderzoek ‘Media Mix and Advertising Effectiveness’ (Axel Springer, 1989) en het Engelse onderzoek ‘The Media Multiplier’ (Consterdine, 1990) waarin een multimediaeffect tussen tv en print werd aangetoond, besloot Admedia in Nederland eind jaren tachtig onderzoek naar dit verschijnsel te gaan doen. In experimentele condities werd aan mensen hetzij alleen een printuiting, hetzij alleen een

tv-commercial, hetzij beide getoond. De hoofdconclusie is (van Hulst, 1989, p.3): 'uit het onderzoek blijkt dat de verschillende informatie-elementen in de tijdschriftadvertentie en de tv-commercial afzonderlijk worden opgemerkt en dat deze elementen elkaar versterken. Hierbij is er sprake van een interactie-effect waarbij beide media tezamen meer bewerkstelligen dan ieder medium afzonderlijk.'

Het onderzoek stimuleerde voor het eerst in Nederland een uitgebreide discussie over multimediaeffecten en de gedachte dat bij gecombineerd adverteren een 1+1=3-effect ontstaat. Onderzoekstechnisch was een kritiekpunt dat het onderzoek niet 'real-life' was, maar dat respondenten geforceerd aan bepaalde uitingen of combinaties van uitingen werden blootgesteld (forced exposure aanpak). Cebuco (1994) liet enkele jaren later door verschillende onderzoeksbureaus een zevental case studies uitvoeren, waarbij het primair ging om het reclame-effect van dagbladadvertenties aan te tonen, maar waarin ook multimediaeffecten tussen dagbladen en andere mediumtypen (tv, radio, brievenbusreclame, buitenreclame) aan de orde kwamen. Een van de hoofdconclusies was (p.37): 'een media-mix met dagbladen is effectiever en leidt tot meer resultaat voor hetzelfde budget.'

In 1997 liet PTT Post Mediaservice een omvangrijk project uitvoeren (Bronner, van Rooy, 1997) naar de invloed van mail op merkwaarden. Ook in dit project kan vastgesteld worden dat bijvoorbeeld tijdschriften en mail elkaar versterken. Een jaar later deden IP en Initiative Media een onderzoek naar de wisselwerking tussen tv-reclame en instore-reclame (Faasse, Santegoeds, Bronner, 1998). Een tv-commercial in de week voorafgaand aan de plaatsing van een display in de winkel bleek het effect van de instore-campagne te verhogen. Een conclusie is: 'de hypothese van Erik du Plessis: winkelreclame werkt, maar alleen omdat de boodschapper daarvoor al regelmatig met de commercial voor het merk op tv is geconfronteerd, wordt door dit onderzoek bevestigd.'

In onderzoek van Initiative Media, Radio 538 en The Audience Profile Company is ook (uitgebreid) onderzoek gedaan naar de synergie tussen televisie en radio (Initiative Media, 1998). Nagegaan is of er sprake is van visual transfer: bij het horen van het geluid van de tv-commercial op de radio ziet de luisteraar ook de bijbehorende beelden weer voor zich.

Uit het onderzoek van IM blijkt dat bij ruim de helft van de onderzochte tv- en radio-commercials bij het geluid van de commercial ook de relevante, bijbehorende beelden worden opgeroepen. Het effect is sterker voor jongeren, hoger opgeleiden, werkzamen en degenen met een positieve of neutrale houding ten aanzien van radioreclame. De Visual Transfer score is het hoogst voor de groep commercials die gebruik maakt van een 'toneelstuk/drama', gevolgd door de commercials met een 'verteller/voice over'. Vervolgens scoort 'eigen muziek' weer beter dan de 'klassieker'. Likeability is eveneens een belangrijke factor: hoe hoger men de commercial waardeert des te meer Visual Transfer (Initiative Media, 1998, 27).

4.3. Evaluatie van het beschikbare onderzoek

Als we de in paragraaf 4.1 en 4.2 gepresenteerde resultaten overzien kunnen we enkele algemene conclusies trekken en enkele conclusies die meer op de Nederlandse situatie zijn gericht. Eerst de algemene conclusies.

Op basis van dit literatuuroverzicht kunnen we concluderen dat de werking van de mogelijke mediacombinaties niet of nauwelijks is onderzocht. Over de werking van een groot aantal combinaties van mediumtypen is niets of nauwelijks iets bekend. Bovendien zijn contextfactoren, zoals doelgroep, afstemming van de boodschap, share of voice, share of market en type product, niet onderzocht. Tot slot kan worden opgemerkt dat de onderzoeksdesigns van de onderzoeken niet altijd optimaal zijn, zoals de onderzoekers zelf ook meestal toegeven.

Als we naar de Nederlandse situatie kijken kunnen we samenvattend stellen dat er de afgelopen jaren in Nederland wel onderzoek is gedaan waarin multimediaeffecten aan de orde zijn gekomen. Maar hierbij moet aangetekend worden dat de onderzoeksopzet ofwel niet real-life was ofwel niet specifiek op het vinden van multimediaeffecten. In het Cebuco-onderzoek en PTT Post Onderzoek ging het primair om de effectiviteit van één medium aan te tonen en het vinden van multimediaeffecten was een secundair doel. Er is ons bijvoorbeeld geen Nederlands onderzoek bekend waarin met testgebieden en controlegebieden is gewerkt om multimediaeffecten te traceren. In een testgebied kan men bijvoorbeeld het budget verdelen tussen tv en magazines (multimediastrategie) en in een gemacht controlegebied kan men het budget beperken tot één mediumtype (en de blootstelling aan één mediumtype aan mensen in dit gebied onthouden). Door bijvoorbeeld via Nielsen-cijfers of aankooppanelregistraties het koopgedrag tussen testgebied en controlegebied te vergelijken ontstaat inzicht in eventuele effecten van een multimediastrategie op koopgedrag, (zie bijvoorbeeld Rood, 1997). Ook is opvallend in de Nederlandse onderzoeken waarin het multimediaeffect aan de orde is geweest dat geen uiteenrafeling plaatsvindt naar de verschillende vormen die multimediaeffecten kunnen aannemen.

4.4. Naar de toekomst

De in deze paragraaf gepresenteerde resultaten maken duidelijk dat (a) veel onderzoeksvragen op multimediate gebied niet zijn beantwoord, (b) een uiteenrafeling van het multimediaeffect niet plaatsvindt, (c) de onderzoeksopzet verbetering behoeft. In paragraaf 5 stellen we de uiteenrafeling van een algemeen multimediaeffect in deeleffecten aan de orde en in paragraaf 6 geven wij aan hoe de effecten op adequate wijze kunnen worden onderzocht.

5. UITEENRAFELING VAN HET MULTIMEDIAEFFECT

Een reclamecampagne kenmerkt zich door een herhaalde aanbieding van de boodschap om hiermee de doelgroep te bereiken en door herhaling een leer- en overtuigingsproces op gang te brengen bij de individuele leden van de doelgroep. We kunnen hierbij de volgende mogelijkheden onderscheiden (gebaseerd op Dijkstra, 2002 en Edell, Keller 1989).

Bereikseffect: Er worden verscheidene media ingezet omdat met één medium niet de gehele doelgroep bereikt kan worden. De inzet van aanvullende media verschaft additioneel bereik voor de boodschap. Met een bepaald medium kan men bijvoorbeeld de hoger opgeleiden bereiken, terwijl een ander medium nodig is om ook de lager opgelei-

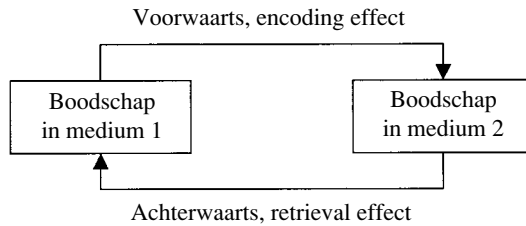
den te bereiken. Een voorbeeld van dit bereikseffect treffen we aan in het Cebuco-onderzoek (1994) bij de Gamma-case. Gamma is een overtuigde folderaar en verspreidt eenmaal per 14 dagen via PTT Post huis-aan-huis het 'Gamma-krantje'. In het kader van het adverteerdersexperiment heeft Cebuco aan Gamma de mogelijkheid geboden gedurende vier weken in dagbladen te adverteren. Iedere week bood Gamma via een advertentie vier nieuwe producten aan die ook in het huis-aan-huis krantje stonden. De advertentie en de folder bleken verschillende doelgroepen te bereiken. De advertentie bereikte vooral de hoger opgeleiden en de wat minder frequente bezoekers. De Gamma-folder scoort vooral goed bij de groepen van hoog frequente bezoekers en mensen met een beroepsopleiding. De beide media bereikten samen het gehele spectrum.

Herhalingseffect: Een boodschap wordt herhaald, in hetzelfde of in een ander medium, om verdere verwerking van de boodschap mogelijk te maken. Vooral bij bijvoorbeeld complexe boodschappen of veel activiteiten van de concurrentie geldt dat meer herhalingen nodig zijn om effect bij de doelgroep te realiseren. Het gaat hierbij om identieke boodschappen. Op een gegeven moment treedt gewenning op en daarmee slijtage (wear-out) van de boodschap, d.w.z. dat er geen additionele leereffecten meer bereikt worden en dat er irritatie over de boodschap kan optreden. Het is te verwachten dat bij gebruik van verschillende media in een campagne minder snel slijtage (wear-out) optreedt (Dijkstra, Buijtels en Van Raaij, 2002). Dit herhalingseffect kwam goed naar voren in het instore/tv-onderzoek van IP en Initiative Media (Faasse, Santegoeds, Bronner, 1998). Er was een voorbeeld van een display die niet begrepen werd door de winkelende consument en daardoor weinig aansprak. Toen later een tv-commercial met uitleg werd vertoond, sprak de display veel meer aan en traden pas effecten op op het koopgedrag (biermerk tijdens WK voetbal 1998).

Complementariteitseffect: Media hebben hun specifieke, sterke en zwakke communicatieve eigenschappen en kunnen elkaar vaak goed aanvullen. Televisie is sterk in het trekken van aandacht, terwijl gedrukte media sterk zijn in de verdere uitwerking en detaillering van de informatie. Televisie roept meer affectieve responsen op en gedrukte media meer cognitieve responsen. Dit pleit voor campagnes met televisie in de eerste fase en gedrukte media in de tweede fase. De inhoud van de boodschap wordt doorgaans aan het medium aangepast. De rijstijl, het comfort en de beleving van een auto kunnen goed worden gecommuniceerd met bewegende beelden, dus met televisie, terwijl feitelijke informatie over de prijs en het benzineverbruik van een auto beter met gedrukte media en internet kan worden overgebracht. Van Hulst (1989) geeft ook een voorbeeld: Conimex roept op tv met muziek en beelden van rijstplukkers een sfeer op die overloopt in een mooie maaltijd. In de printadvertenties wordt informatie gegeven: foto's van maaltijden, recepten.

Voorwaartse en achterwaartse overdrachtseffecten (imagery transfer): bij herhaling van boodschappen in verschillende media kunnen voorwaartse en achterwaartse overdrachtseffecten plaatsvinden. Zie het schema in figuur 1. We gaan er hierbij van uit dat de boodschap in verschillende media wordt aangeboden. Een (voorwaarts) encoding effect treedt op als de boodschap in het eerste medium de verwerking van de boodschap in het tweede medium vergemakkelijkt. Een (achterwaarts) retrieval effect treedt op als de confrontatie met de boodschap in het tweede medium de boodschap in het eerste medium uit het geheugen van de consument oproept. Een voorbeeld hiervan is

hetgeen zojuist werd gezegd over de bier tv-commercial die uitleg verschafte over de winkeldisplay.



Figuur 1. Encoding en retrieval effecten.

Het voorwaartse encoding proces kan als volgt worden beschreven. De confrontatie met de boodschap in het eerste medium laat een spoor in het geheugen achter door aandacht te trekken voor een bepaald merk, symbool of ander teken of een bepaalde persoon, handeling of situatie. De boodschap in het eerste medium verhoogt de kans dat de boodschap in het tweede medium wordt gezien en vergemakkelijkt ook de verwerking van de boodschap in het tweede medium. De boodschap in het eerste medium kan ook als 'priming' worden gezien. De boodschap in het eerste medium verhoogt de kans dat de boodschap in het tweede medium op een bepaalde manier wordt gezien. Priming creëert een focus of filter, een waarnemingsschema dat op de boodschap in het tweede medium wordt toegepast. Een advertentie waarin een bepaalde promotie-actie wordt aangekondigd, kan ertoe leiden dat consumenten die de advertentie gezien hebben, in de winkel het display van de actie sneller opmerken dan consumenten die de advertentie niet gezien hebben.

Het achterwaartse retrieval proces kan als volgt worden beschreven. De confrontatie met de boodschap in het tweede medium vergemakkelijkt het 'opdiepen' (retrieval) uit het geheugen van de boodschap in het eerste medium. Door de boodschap in het tweede medium wordt de boodschap in het eerste medium als het ware opnieuw verwerkt. Zo kunnen radiocommercials met een bepaalde jingle de eerder uitgezonden beelden van de televisiecommercials met dezelfde jingle uit het geheugen van consumenten oproepen (visual transfer). Bij het horen van de radiocommercial 'zien' consumenten de bijbehorende beelden van de televisiecommercial. In een campagne is radio hierdoor een goed follow-up medium van televisie. Edell en Keller (1989) deden uitgebreid onderzoek naar de effecten van de combinatie tussen televisiecommercials en het geluid hiervan als radiocommercial.

Bij de confrontatie met de boodschap in het tweede medium treden dus twee processen van imagery transfer op: het encoding en het retrieval proces. Het encoding proces is het verwerken van de boodschap in het tweede medium. Het retrieval proces is het uit het geheugen oproepen van de boodschap uit het eerste medium. Deze beide processen strijden om verwerkingscapaciteit en verwerkingstijd bij de tweede confrontatie. Het is te verwachten dat bij een complexe boodschap het encoding proces de overhand heeft, terwijl bij een eenvoudige boodschap het retrieval proces domineert. Bij grote verschil-

len tussen de boodschap in het eerste en in het tweede medium domineert het encoding proces, terwijl bij kleine verschillen het retrieval proces overheerst. Voorwaartse en achterwaartse overdrachtseffecten treden op (1) bij confrontaties met verschillende maar gerelateerde boodschappen in hetzelfde medium, (2) bij 'dezelfde' boodschappen in verschillende media, en (3) bij combinaties van voorgaande beide mogelijkheden.

Multiple sources effect: Het verkrijgen van gerelateerde en elkaar ondersteunende informatie uit verschillende media kan de geloofwaardigheid (believability of credibility) van de boodschap versterken. Het is alsof onafhankelijke bronnen de boodschap bevestigen. Het is dus te verwachten dat identieke boodschappen in verschillende media een sterker vertrouwen creëren dan hetzelfde aantal boodschappen in hetzelfde medium. Vertrouwen heeft niet zo zeer een effect op de verwerking van de boodschap, maar op de geloofwaardigheid ervan. Het multiple sources effect heeft dan vooral een effect op overreding en koopintentie.

Uiteraard hebben mediatitels bij consumenten al een bepaalde mate van vertrouwen. Consumenten hebben in het algemeen een groter vertrouwen in een boodschap in Het Financieel Dagblad dan in het blad Aktueel. Een bericht in de eigen vertrouwde krant roept een groter vertrouwen op dan hetzelfde bericht in een onbekend blad. Het multiple sources effect bestaat echter ook los van het al bestaande vertrouwen. Het multiple sources effect zoals we het hier beschrijven, stelt dat boodschappen in verschillende media elkaar onderling versterken.

Synergie

De mogelijkheden complementariteit en imagery transfer kunnen leiden tot synergie, d.w.z. dat het totale effect groter is dan de som van de afzonderlijke effecten. Dit wordt ook wel het 'media multiplier' effect genoemd (Consterdine, 1997): $1 + 1 > 2$.

Floor & Van Raaij (2002, p. 402) geven een voorbeeld van synergie-effecten. In het voorbeeld is de omzet van een bepaald product in de winkel afhankelijk van de factoren reclame, sales promotion en een display in de winkel. De auteurs beschrijven de weekomzetten van 19 productgroepen zonder ondersteuning (indexcijfer 100), met 10% prijsverlaging (indexcijfer 125), met advertentie en 10% prijsverlaging (indexcijfer 189), met display en prijsverlaging (indexcijfer 226), en met advertentie, display en prijsverlaging (indexcijfer 353). Hier is duidelijk sprake van synergie. Een prijsverlaging van 10% zorgt voor 25% meer omzet. Een advertentie en 10% prijsverlaging leidt tot 89% meer omzet. Een display en 10% prijsverlaging zorgt voor 126% meer omzet. Een advertentie, een display en 10% prijsverlaging geeft zelfs 253% meer omzet. De toevoeging van een advertentie aan 10% prijsverlaging leidt tot 64% meer omzet. De toevoeging van een advertentie aan een display en 10% prijsverlaging leidt tot 127% meer omzet. De toevoeging van een advertentie wordt effectiever als al enkele instrumenten worden gebruikt. De effecten van de instrumenten reclame, sales promotion en in-store communicatie zijn niet additief (de som van de afzonderlijke effecten), maar multiplicatief (het effect van reclame is ongeveer twee keer zo groot als twee andere instrumenten in plaats van een ander instrument worden ingezet).

Interferentie of verstoring kan optreden als de boodschappen in verschillende media elkaar niet ondersteunen of zelfs tegenspreken. Dan treedt een soort negatieve voorwaartse en/of achterwaartse imagery transfer op, die het totale effect lager maakt dan de som van de afzonderlijke effecten, dus een soort negatieve 'media multiplier'.

Effecten en afhankelijke variabelen

De besproken effecten kunnen betrekking hebben op de volgende communicatie-effecten (afhankelijke variabelen):

1. Herinnering (recall) van advertentie en/of merk.
2. Herkenning (recognition) van advertentie en/of merk.
3. Brand awareness.
4. Cognitieve responsen, pro- en contra-argumenten.
5. Emotionele responsen.
6. Attitude t.o.v. advertentie.
7. Attitude t.o.v. merk, sympathie voor merk.
8. Liking en disliking.
9. Begrip.
10. Koopintentie.
11. Koopgedrag.
12. Satisfactie, merktrouw.

Al deze afhankelijk variabelen kunnen in principe gebruikt worden om een multimedia-effect vast te stellen. De afhankelijk variabele dient significant groter te zijn in de multimediaconditie dan in de single-media conditie.

Volgorde van media

Ook de volgorde waarin media in een campagne worden ingezet, is bepalend voor het effect. Traditioneel werd gedacht dat televisie aandacht trekt en vooral in het begin van een campagne moet worden ingezet en dat gedrukte media vooral in het tweede deel van de campagne hun taak het beste vervullen (complementariteitseffect). Uit recent onderzoek (Dijkstra, 2002) blijkt echter dat gedrukte media en statisch internet (internet met teksten en plaatjes, vergelijkbaar met gedrukte media) vooral in het begin een functie vervullen en televisie in het tweede deel van de campagne. Er treedt dan vooral een voorwaarts overdrachtseffect (encoding) op vanuit het gedrukte medium of het statische internet.

Afstemming van de inhoud

In een adequaat geïntegreerd marketing plan wordt ook de inhoud van de campagne zodanig aangepast dat er wordt ingespeeld op de sterkten van de verschillende mediumtypen.

6. EEN ONDERZOEKSOPZET OM MULTIMEDIAEFFECTEN TE METEN

Om multimediaeffecten te onderzoeken bestaat een drietal onderzoeksopzetten: vrijwillige blootstelling (voluntary exposure), gedwongen blootstelling (forced exposure) en gedwongen onthouding (forced abstinence).

Vrijwillige blootstelling (voluntary exposure) is een populaire onderzoeksopzet, waarbij echter een zelfselectie bias kan optreden. Er wordt bijvoorbeeld in een steekproef een onderscheid gemaakt tussen consumenten die een campagne voor een bepaald merk alleen in de dagbladen hebben gezien, andere consumenten die de campagne alleen op televisie hebben gezien, en een derde groep die de campagne zowel in de dag-

bladen als op televisie onder ogen heeft gehad. We meten dan enkele afhankelijke variabelen bij de drie groepen – zoals merkbekendheid, imago, merksympathie, koopintentie – en verwachten hogere scores voor de groep die beide media heeft gebruikt (Okauchi, 2001). We vallen dan echter in de zelfselectie valkuil. De consumenten die de campagne in beide media onder ogen hebben gehad, zijn mogelijk al in de uitgangssituatie positiever gestemd over het merk en bekijken daardoor advertenties voor het merk vaker en met meer interesse.

Consumenten met een positief beeld van merk X letten meer op reclame van merk X en herkennen hierdoor meer reclame van merk X in verscheidene media. Consumenten met een neutraal of negatief beeld van merk X letten minder op reclame van merk X en herkennen ook minder reclame van merk X in de media. Hierdoor lijkt het dat reclame van merk X meer effect heeft op consumenten die meer media gebruiken. In feite is dit een zelfselectie-effect.

In de dagelijkse praktijk van het reclameonderzoek wordt veel ‘gezondigd’ op dit punt. Daarom willen wij graag (samenvattend) benadrukken dat het indelen van een steekproef achteraf in multimediablootgestelden en mono-media blootgestelden en vergelijken van deze groepen op effectvariabelen *wel* meestal verschillen te zien geeft, maar dat deze *niet* aan het multimediaeffect mogen worden toegeschreven. De respondenten stellen zichzelf bloot aan meer mediumuitingen van een merk en dat gaat vaak samen met een hogere sympathie voor dat merk vooraf. Men meet op deze wijze ‘schijn-effecten’.

Een alternatief is *forced exposure* (gedwongen blootstelling) in een experimentele proefopzet. Hierbij bepaalt de proefleider aan welke media en hoe lang de deelnemers aan deze media worden blootgesteld. Hierbij kunnen de experimentele condities goed in de hand worden gehouden, maar de externe validiteit daalt erdoor. Wellicht worden deelnemers aan het onderzoek gedwongen media te gebruiken die ze in een natuurlijke situatie niet of minder zouden hebben gebruikt. Deze aanpak wordt bijvoorbeeld nog wel eens gevolgd om mensen naar een website te laten gaan, terwijl die daar via natuurlijke weg nooit gekomen zouden zijn.

De derde weg is die van *forced abstinence* (gedwongen onthouding). Veldkamp Marktonderzoek gebruikte deze methode bij een onderzoek naar media- en merkbeleving van PTT Post (Bronner, van Rooy, 1997). Het ging hier om televisie, dagbladen en direct mail. Uit een mailbestand werd een klein deel aselekt afgezonderd aan wie gedurende een bepaalde periode geen mailings werden gestuurd. Daarna werd het verschil gemeten in de afhankelijke variabelen tussen de hoofdgroep die de mailings wel had ontvangen en de experimentele groep die de mailings niet had ontvangen. Beide groepen hadden wel de reclame op televisie en in dagbladen onder ogen gehad. Hiermee kan worden vastgesteld wat de additionele waarde van de mailings is naast televisie en dagbladen.

Een versturende factor is overigens dat de experimentele groep (zonder de mailings) minder vaak geconfronteerd is geweest met de boodschap. Hierdoor alleen al zou een verschil in de afhankelijke variabelen kunnen ontstaan.

Weglating van advertenties uit tijdschriften wordt ook wel toegepast in split-run onderzoek van mediacombinaties of advertentievarianten en kan ook een voorbeeld zijn van gedwongen onthouding. In een testregio wordt bijvoorbeeld gedurende een bepaalde periode de advertentie voor een bepaald merk doelbewust uit het blad/dagblad gehaald.

Gedwongen onthouding is relatief eenvoudig te doen voor gedrukte media, inclusief direct mail, maar moeilijk toe te passen op televisie en radio. Met gedwongen onthouding is een multimediaeffect op een natuurlijke manier vast te stellen.

In tabel 3 worden de scores van de drie onderzoeksopzetten op de genoemde criteria samengevat.

Tabel 3. Een vergelijking van de sterkten en zwakten van de verschillende onderzoeksopzetten.

	vrijwillige blootstelling	gedwongen blootstelling	gedwongen onthouding
• externe validiteit (natuurlijke situatie)	+	-	+
• lange termijn communicatie-effect	0	-	+
• eenvoud van uitvoering	+	0	0
• validiteit van conclusies	-	+	+

+ betekent methode voldoet aan dit criterium; 0 betekent methode voldoet matig aan dit criterium; - betekent methode voldoet niet aan dit criterium

Een belangrijk voordeel van gedwongen onthouding is ook dat lange termijn communicatiebehoefte gemeten kunnen worden. Veel reclameonderzoek concentreert zich rond korte termijn effecten. Maar aan de andere kant realiseren wij ons dat juist de herhaling van belang is (zie bijvoorbeeld Sutherland, Sylvester, 2000, die spreken over de kracht van het vervolg). Forced exposure kunnen we maar één of twee keer van een respondent vragen, maar forced abstinence kunnen we maanden volhouden. Al met al lijkt gedwongen onthouding de beste methode om multimedia effecten vast te stellen.

7. CONCLUSIES

Multimediacampagnes komen veel voor in de praktijk van mediaplanning. In schril contrast daarmee staat het ontbreken van adequate gegevens en inzicht in het bereik en de effecten van de inzet van meer mediumtypen in een campagne.

In dit artikel zijn we nagegaan wat mogelijke effecten van multimediacampagnes zijn en welke bevindingen uit de literatuur naar voren komen. Uit het literatuuroverzicht blijkt dat het multimediaeffect weinig onderzocht is. Bovendien voldoet de gebruikte onderzoeksopzet meestal niet aan de methodologische vereisten om een effect te kunnen vaststellen.

In de laatste paragraaf hebben we mogelijke onderzoeksopzetten voor het bestuderen van multimediaeffecten met elkaar vergeleken. Daarbij bleek dat de methode van gedwongen onthouding een veelbelovend, maar nog weinig toegepast alternatief is.

Wij bevelen aan om deze methode toe te passen op een aantal zorgvuldig geselecteerde cases.

LITERATUUR

- Axel Springer Verlag, *Media mix and advertising effectiveness*, 1989.
- Boon, A. den & P. Neijens, red. (2000). *Media en reclame. Mediaplanning en bereiksonderzoek*. Groningen: Wolters-Noordhoff.
- Bronner, F., C. van Rooy, De meerwaarde van direct mail voor het merk, *ROS Report* nr. 6, PTT Post Mediaservice, 1997.
- Cebuco, *Echt een goed product? Dat mag wel in de krant: effectiviteit van dagbladcampagnes*, interne brochure, mei 1994.
- Consterdine, G., How print and TV interact: 'The media multiplier'. *Admap*, 25, 1990.
- Consterdine, G., The Media Multiplier, *Research report*, May 1990.
- Consterdine, G. *The page and the screen nourish each other*. ESOMAR Monograph Print Media Research, 1988.
- Dijkstra, M. (2002), Universiteit van Tilburg, dissertatie, oktober.
- Dijkstra, M., H. Buijtsels & W. F. van Raaij, The separate and joint effects of medium type on consumer responses: A comparison of television, print and the Internet. *Journal of Business Research*, 2002 (in druk).
- Edell, J.A. & K.L. Keller, The information processing of coordinated media campaigns. *Journal of Marketing Research*, 26, 149-163, 1989.
- Edell, J.A. & K.L. Keller, *Analyzing media interactions: print reinforcement of television ad campaigns*. Fuqua School of Business, Duke University, Durham, NC, 1998.
- Faasse, J., A. Santegoeds, F. Bronner, *De wisselwerking tussen tv- en instore reclame*, interne brochure, november 1998.
- Floor, K. & F. van Raaij, *Marketing Communicatie Strategie*. Houten: Stenfert Kroese, 2002, vierde druk.
- Franz, G., Better media planning for integrated communication. *Admap*, 35(1), 42-44, 2000.
- Giling, A., P. Neijens, B. van den Putte, J. Sandee & E. Smit, *Effie-onderzoek*, 2002. Amsterdam: SWOCC.
- Hulst, T. van, *Het Multi Media Effect*. Admedia/Lintas, interne brochure, juni 1989.
- Initiative Media (1998). *Visual transfer. What you see is what you hear. Een onderzoek naar de Visual Transfer van 26 campagnes, in opdracht van Initiative Media en Radio 538, uitgevoerd door The Audience Profile Company*. Amstelveen: IM.
- Jacoby, J., W.D. Hoyer & M.R. Zimmer, To read, to view or to listen? A cross-media comparison of comprehension. *Current Issues and Research in Advertising*. University of Michigan, Ann Arbor, 1983, pp. 201-217.
- Okauchi, S., *Synergy brought about through duplication of contact with newspaper and internet*, Worldwide Readership Research Symposium, Venice 2001, 475-491.
- Rood, T. Instant synergy: televisie en print samen sterker, *Blad*, 10, 4, 6-10, 1997.
- Smith, A., More food for thought. It is well-known that mixed media campaigns are often more effective than television alone. *Admap*, 32, 1997.
- Smith, A. *Integrated marketing communications starts with print plus television*. Amsterdam: ESOMAR, 1997.
- Sutherland, M., A.K. Sylvester, *Reclamewerking ontrafeld*, GVR/SWOCC reeks, Kluwer, 2000.
- Vincent, M. & A. Vincent, Which medium works best? *Admap*, 31, 1996.
- Wilkins, J. & K. Ford. Detecting the effectiveness of integrated marketing. *Admap*, 32, 1997.

3. *Conversie van gestandaardiseerd vragenlijst- onderzoek naar internet: kritische succes- factoren*

E.C. DEUTSKENS, P.M. WILLEMS, S.S.M. VAN DEN BERG, P. OOSTERVELD,
J.C. DE RUYTER en M.G.M. WETZELS

SAMENVATTING

In de online marktonderzoeksliteratuur zijn de algemeen geaccepteerde richtlijnen voor het inrichten van online vragenlijstonderzoek nog volop in ontwikkeling. De best practices schieten tekort op het moment dat omvangrijke, gestandaardiseerde schriftelijke vragenlijsten worden omgezet naar internet. In een experiment is onderzocht wat de invloed is van beloningsvorm, vragenlijstlengte, presentatievorm en timing van reminders op de deelnamebereidheid van respondenten aan een uitgebreide online vragenlijst. Uit de resultaten bleek dat lange vragenlijsten weliswaar een lagere respons hebben, maar dat er toch een bevredigend responspercentage mee behaald kan worden. Daarnaast was loterij de meest effectieve beloningsvorm, aangezien met een loterij min of meer hetzelfde responspercentage werd verkregen als met waardebonnen, maar een loterij kostefficiënter is. Hoewel een vragenlijst met visuele productpresentaties een wat lagere respons had dan de tekstuele versie, was de responskwaliteit wel hoger.

1. INTRODUCTIE

Voor de ontwikkeling van internetmarktonderzoek tot volwaardig alternatief is het van belang dat mogelijkheden en grenzen van dit veldwerkkanaal expliciet worden gemaakt. Op dit moment bestaan er voor vragenlijstonderzoek via internet slechts richtlijnen of '*best practices*'. De bekende regels zoals dat een vragenlijst via internet niet langer dan 10 minuten invultijd mag vergen, zijn te weinig gebaseerd op empirische bevindingen en betekenen in de praktijk een beperking voor de inzet van internet in het brede scala van onderzoeksproducten.

Er is een toenemende behoefte om ook uitgebreidere studies via internet uit te voeren. Onderzoekers van zowel bureauijde als vanuit de internationale producenten en dienstverleners staan te trappelen om deze volgende stap te zetten. Gestandaardiseerde, omvangrijke meetinstrumenten die nu meestal nog via schriftelijke vragenlijsten worden afgenomen, roepen om een conversie naar internet. De omvangrijke handelingskosten, lange doorlooptijd, slechtere bereikbaarheid van respondenten en de afnemende respons op schriftelijke vragenlijsten doen een beroep op onderzoeksmethodologen om het instrument geschikt te maken voor internet als responsmedium.

Er zijn twee fundamentele problemen bij de conversie van uitgebreide, gestandaardi-

seerde vragenlijsten naar internet. Voor de eerste, trendbreuk, is inmiddels niemand meer zo bang. Trendbreuk is het eenmalige gevolg van de omzetting van het veldwerk naar een ander kanaal en betekent dat gemiddelde scores op vragenlijstitems niet langer rechtstreeks met benchmarks vergeleken kunnen worden. De oplossing voor trendbreuk is calibratie. In een experimenteel design waarin op aselechte wijze respondenten aan verschillende veldwerkvarianten van de vragenlijst worden toegewezen, kunnen correctiecoëfficiënten worden bepaald waarmee toekomstige resultaten alsnog met benchmarks vergeleken kunnen worden (Willems et al., 2002).

De tweede reden waarom onderzoekers op dit moment nog huiverig zijn voor het omzetten van standaardvragenlijsten naar internet bestaat uit het feit dat men onvoldoende zeker is van de kwaliteit en kwantiteit van respons via internet op dit soort uitgebreide vragenlijsten. Bestaat een ad-hoonderzoek via internet doorgaans uit vijftientwintig vragen, gestandaardiseerde paper-and-pencil vragenlijsten kennen vaak meerdere pagina's met stellingenblokken en tientallen achtergrondvragen. Het is vervolgens onduidelijk of respondenten bereid zijn om dit soort lange vragenlijsten via internet in te vullen, welke beloningsvorm moet worden gekozen als compensatie voor deelname en hoe de vormgeving van de vragenlijst moet worden aangepast voor presentatie op internet.

Experimenteel onderzoek kan ook ten aanzien van dit tweede probleem praktische richtlijnen aanreiken om conversie naar internet van gestandaardiseerde vragenlijsten te faciliteren. In dit artikel worden resultaten gepresenteerd van een studie waarin deze conversie experimenteel is onderzocht. Deze studie is het resultaat van een samenwerking tussen Centrum voor Marketing Analyses en de Universiteit Maastricht in het kader van een onderzoeksprogramma betreffende de equivalentie tussen online en traditionele marktonderzoeksmethoden. Meerdere varianten van een gestandaardiseerd, in de FMCG branche alom bekend, multiclient instrument zijn via internet aangeboden aan een uitgebreide steekproef van respondenten. De lengte van deze oorspronkelijk paper-and-pencil vragenlijst vroeg om een creatieve omzetting naar internet. Daarvoor is eerst een wetenschappelijke inventarisatie gemaakt van de belangrijkste succesfactoren voor een online vragenlijst.

2. KRITISCHE SUCCESFACTOREN VOOR EEN ONLINE VRAGENLIJST

Voor een aantal factoren bij het ontwikkelen van een online vragenlijst bestaan reeds eenduidige richtlijnen. Voorbeelden hiervan zijn onder andere het tonen van voortgangsinformatie, het bieden van zekerheid wat betreft anonimiteit en geheimhouding en het aanbieden van online help.

Feedback over de voortgang van het invullen van de vragenlijst vermindert het aantal respondenten dat voortijdig afhaakt (Couper et al., 2001). Voortgangsinformatie vormt een belangrijke motivator. De visuele presentatiemogelijkheden van een online vragenlijst maken het bij uitstek mogelijk respondenten te informeren over de voortgang van het invulproces.

Het benadrukken van anonimiteit en het opnemen van een verklaring tot geheimhouding op de vragenlijstpagina('s) vormt eveneens een onbetwiste voorwaarde voor een goede respons. Studies van onder meer Sheehan en Hoy (1999) tonen aan dat respondenten zich bewust zijn van de vertrouwelijkheid van de door hen gegeven antwoorden

en dat ze actief op zoek gaan naar informatie in de vragenlijst die wijst op een zorgvuldige behandeling van de gegeven antwoorden.

Online help wordt bij internetvragenlijsten vaak aangeboden in de vorm van een zogenaamde FAQ (Frequently Asked Questions list). Het aanbieden van dit soort informatie vergroot de slagingskans van een online onderzoek substantieel (Bauman et al., 2000). Belangrijke elementen van FAQ zijn het introduceren van de onderzoeksorganisatie, antwoorden op vragen of problemen die kunnen optreden, het verzekeren van vertrouwelijkheid van antwoorden en het geven van een contactnaam of e-mailadres voor verdere vragen. Bij voorkeur zou een link naar de FAQ op elke pagina van de vragenlijst aangeboden moeten worden.

Voor een aantal andere vragenlijstkenmerken geldt dat richtlijnen minder eenduidig zijn. Op het moment dat een vragenlijst bijzondere eisen stelt aan het invulproces, bijvoorbeeld door de lengte van de vragenlijst, lijken bevindingen uit de literatuur niet langer duidelijkheid te verschaffen over de inrichting van de vragenlijst. Vier van deze kenmerken worden hieronder belicht. Ze vormen de onafhankelijke variabelen in de experimentele studie.

2.1. Beloningsvorm

Het gebruik van financiële beloningen of incentives is effectief in zowel offline als online onderzoek (Yammarino et al., 1991; Bergmann et al., 1990; The Future Place Consultancy, 2001). Een incentive kan respondenten een reden geven om een vragenlijst in te vullen en fungeert tevens als een teken van dank namens de klant. Incentives kunnen geclassificeerd worden naar de manier waarop beloond wordt: individuele incentives belonen iedereen die deelneemt aan het onderzoek, zoals bijvoorbeeld in het geval van waardebonnen of spaarsystemen. In andere gevallen worden slechts enkele van de deelnemende respondenten beloond, bijvoorbeeld in het geval van loterijen. Daarnaast bestaat er de beloning in de vorm van donaties aan goede doelen. Hierbij ontvangt de respondent zelf niet rechtstreeks een beloning.

Resultaten van een aantal eerdere studies geven aanleiding te kiezen voor individuele, financiële incentives (Warriner et al., 1996; James en Bolstein, 1990; James en Bolstein, 1992), zelfs in situaties waarin slechts symbolische bedragen worden gegeven. Echter, Frick et al. (1999) vonden dat loterijen eveneens resulteren in een relatief lage uitval van respondenten. Daarbij werd een additioneel argument ten gunste van een loterijbeloning genoemd, namelijk dat de kosten voor de respondentbeloning onafhankelijk kunnen zijn van het aantal deelnemers in het onderzoek. Het namens respondenten doneren van geld aan goede doelen, tenslotte, doet een appèl op sociaalnormatieve waarden. Het ingaan op het verzoek deel te nemen aan een vragenlijstonderzoek heeft een gunstig effect op de beleving van eigenwaarde van de respondent (Warriner et al., 1996). Studies tonen echter geen helder verband aan tussen het gebruik van de donatiebeloningsvorm en respons op de vragenlijst.

Een vergelijking van de effecten van individuele, financiële beloningen met loterijen en donaties in schriftelijke vragenlijsten geven dan ook geen eenduidig beeld (Warriner et al., 1996). Omdat beloningen een substantiële component vormen van het onderzoeksbudget, dient de invulling ervan nauwkeurig te worden bepaald. Om die reden vormt beloning een belangrijke factor in het experiment in dit artikel.

2.2. Lengte van de vragenlijst

Resultaten van onderzoek naar de relatie tussen het aantal vragen in een vragenlijst en het responspercentage zijn allerminst eenduidig. Terwijl sommige studies suggereren dat de lengte van de vragenlijst geen invloed heeft op de respons (Brown, 1965; Bruvold en Comer, 1988; Mason et al., 1961; Fox et al., 1988; Linsky, 1975), wijzen andere onderzoeken erop dat er wel degelijk een relatie bestaat tussen vragenlijstlengte en het responspercentage, zowel in off- als online marktonderzoek (Dillman, 1991; Heberlein en Baumgartner, 1978; Steele et al., 1992; Yammarino et al., 1991; MacElroy, 2000; Sheehan, 2001).

Uit onderzoek door Marketfacts (2001) bleek dat respondenten van een online vragenlijst veelal vinden dat vragenlijsten te lang zijn. Omdat respondenten met één klik kunnen afhaken, lijkt het erop dat uitsluitend korte vragenlijsten tot een hogere respons leiden. Het is echter niet duidelijk wat precies onder kort moet worden verstaan. Volgens Rosenblum (2001) zou een online vragenlijst uit ongeveer twintig vragen moeten bestaan. Dit lijkt echter onrealistisch klein voor de meeste uitgebreide onderzoeksvragen.

Wat we kunnen verwachten bij de conversie naar internet van een vragenlijst met vragen over 190 producten is daarom niet precies te zeggen. De afname van respons als functie van vragenlijstlengte vormt dan ook een belangrijk aandachtspunt in het hieronder beschreven experiment.

2.3. Presentatie van de vragenlijst

In internetmarktonderzoek kan de vraagtekst worden aangevuld met visuele elementen, zoals kleur, afbeeldingen en interactieve elementen (bijvoorbeeld de 3D-presentatie van producten). Dit zou het aantrekkelijker kunnen maken voor respondenten om de online vragenlijsten in te vullen (Redline en Dillman, 1999). Bovendien kunnen productafbeeldingen ook effectief zijn bij het verhogen van de kans op productherkenning, aangezien respondenten meer associatiemogelijkheden krijgen aangeboden. Een probleem van bij het gebruik van afbeelding is de laadtijd (downloadtijd). Relatief eenvoudige internetvragenlijsten die weinig downloadtijd vragen, leiden dan ook in sommige gevallen tot een hoger responspercentage dan vragenlijsten met een overdaad aan grafische elementen (Dillman et al., 1998).

Het lijkt erop dat de voordelen van aantrekkelijkheid en hogere productherkenning dienen te worden afgewogen tegen nadelen als langere downloadtijd voor respondenten en de kosten die gepaard gaan met programmeren. De cruciale vraag is of grafische elementen resulteren in betere resultaten die tijd en andere investeringen rechtvaardigen. Op basis van de resultaten van eerder onderzoek verwachten we dat afbeeldingen in een online vragenlijst zowel tot hogere productherkenning als tot hogere responskwaliteit zullen leiden, maar dat het responspercentage lager zal zijn vanwege de langere downloadtijd.

2.4. Reminders

Reminders, of follow-ups, beïnvloeden het responspercentage positief bij schriftelijke vragenlijsten die per post worden verzonden (Sheehan, 2001; Solomon, 2001). Van een briefkaartreminder wordt verwacht dat die het responspercentage verhoogt om twee

redenen. Ten eerste wordt verondersteld dat de extra aandacht die blijkt uit het sturen van een reminder de medewerking stimuleert. In de tweede plaats vormt de reminder ook letterlijk een geheugensteuntje omdat respondenten nogal eens vergeten de vragenlijst in te sturen (Fox et al., 1988). Reminders in schriftelijke studies kunnen het responspercentage 8 tot 48% verhogen (Sheehan, 2001). Bij internet ligt het voor de hand vooral het tweede effect van reminders te veronderstellen, namelijk de herinneringswerking ervan. In een studie naar onderzoek via e-mail (Solomon, 2001) werd gevonden dat het mogelijk is om met reminders het responspercentage te verdubbelen. Hoewel het verzenden van meerdere reminders in online onderzoeken vrijwel kosteloos is, moet zorgvuldig met dit middel omgesprongen worden. Het valt te verwachten dat het overdadig sturen van reminders gevoelens van irritatie bij respondenten wekt en op langere termijn kan leiden tot een algemene vermindering in de bereidheid van respondenten om via internet aan onderzoek mee te doen.

Het is voorts niet duidelijk wat het beste moment is om een reminder te versturen. Het moment van versturen van reminders wordt in de onderhavige studie experimenteel gemanipuleerd. We verwachten dat wanneer een reminder vrij snel na de uitnodiging wordt verstuurd, dit een hoger responspercentage oplevert dan wanneer de reminder pas na twee weken wordt verstuurd.

3. METHODE

3.1. De vragenlijst

De vragenlijst die in dit experiment is geconverteerd naar internet, betreft een onderzoeksinstrument voor het meten van het gebruik van consumentenproducten binnen huishoudens en de bekendheid van merken. De vragenlijst wordt oorspronkelijk in schriftelijke vorm afgenomen. Resultaten worden in de vorm van multiclient participatie afgenomen door verschillende producenten. Voorbeelden van productgroepen in de vragenlijst zijn bier, shampoo, kaas, olijfolie en toast. Iedere productgroep bevat een aantal merken waarvan respondenten moeten aangeven of ze deze kennen, hoe vaak de producten gebruikt worden en hoe de prijs-kwaliteitverhouding wordt beoordeeld. De vragenlijst is een lange, uitgebreide vragenlijst die van respondenten al snel gemiddeld 30 minuten invultijd vraagt.

3.2. Steekproef

De deelnemers aan het onderzoek werden uit een database gerekruteerd. Deze bestaat uit namen en e-mailadressen van mensen in Nederland die verzameld zijn door middel van eerder telefonisch contact. De steekproef werd verdeeld naar de Minicensus, op basis van de volgende demografische gegevens:

- leeftijd (13-74) en sociale klasse
- geslacht
- regio (5).

Onze steekproef bestond uit 5.413 e-mailadressen, verkregen van 1999 tot en met 2002. We besloten om niet alle uitnodigingen in één keer uit te mailen, maar in plaats daar-

van in batches die we één voor één uitmailden. In batch 1 verdeelden we alle respondenten gelijkmatig over de verschillende cellen van ons design (zie paragraaf 3.3). De verdeling in de latere batches, batch 2, 3 en 4, werd aangepast aan de tot dan toe behaalde respons in iedere cel, om ervoor te zorgen dat de absolute respons in iedere cel aan het einde van het onderzoek ongeveer gelijk zou zijn.

3.3. De onafhankelijke variabelen

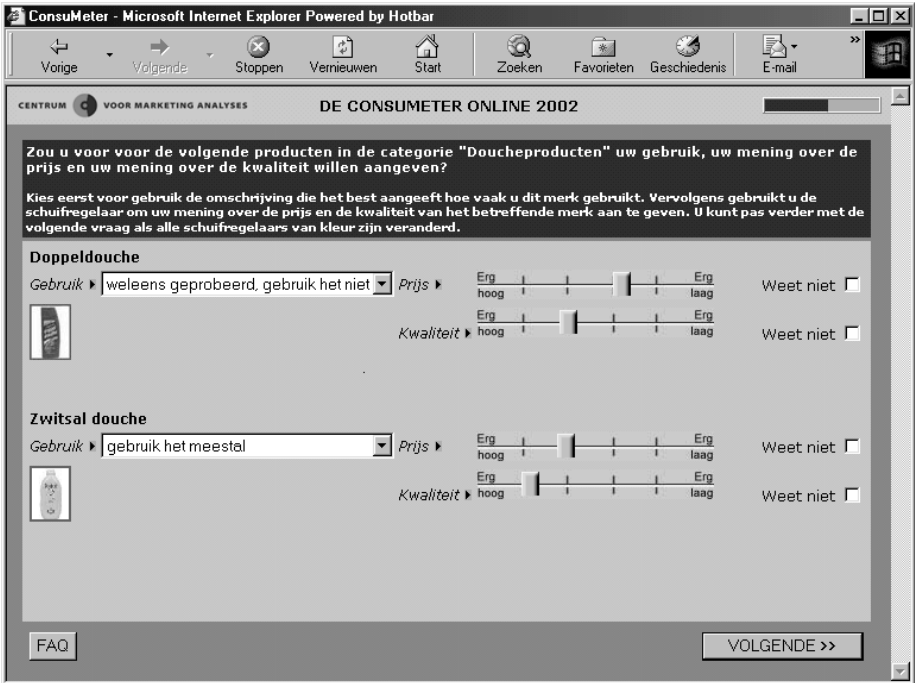
Experimentele factoren in het onderzoek waren het soort incentive, de lengte van de vragenlijst, de presentatie van de vragenlijst en de timing van de reminder (zie Tabel 1).

Beloningsvorm. Drie soorten beloningen zijn in de studie meegenomen: een waardebon, donatie aan een goed doel en een loterij. Participanten in de waardebonconditie kregen een waardebon voor een online boeken- en platenwinkel na het completeren van de vragenlijst. We kozen voor waardebonnen in plaats van contant geld aangezien het meer tijd en moeite kost om de uitbetaling van contant geld af te handelen. De hoogte van het bedrag was afhankelijk van de lengte van de vragenlijst: € 2,50 voor de korte versie, € 5,00 voor de lange versie. De donatieconditie hield in dat namens de respondenten een bepaald bedrag aan een goed doel werd gedoneerd als dank voor hun deelname. Daarbij kon men kiezen uit drie goede doelen: Wereld Natuur Fonds, Amnesty International of de Kankerbestrijding. De loterijconditie hield in dat respondenten een kans hadden een waardebon van een relatief hoog bedrag te winnen. In de korte vragenlijst maakten respondenten kans op één van de vijf waardebonnen van € 25,00, in de lange vragenlijst op één van de vijf waardebonnen van € 50,00.

Lengte van de vragenlijst. In het onderzoek zijn een lange en een korte versie van de vragenlijst gebruikt. De lange versie bevatte evenveel vragen en productcategorieën (19) als de schriftelijke vragenlijst, terwijl de korte versie de helft van het aantal vragen en productcategorieën (9) van de schriftelijke vragenlijst bevatte. De lange versie duurde ongeveer 30 tot 45 minuten om in te vullen, de korte 15 tot 30 minuten.

Presentatie van de vragenlijst. Met de vragenlijst wordt voor een groot aantal producten en merken (190 voor de lange vragenlijst, 91 voor de korte) vastgesteld of respondenten deze kennen en gebruiken en hoe men deze beoordeelt wat betreft prijs en kwaliteit. In de schriftelijke vragenlijst moeten respondenten deze vragen beantwoorden op basis van de naam van de producten en merken alleen. In de presentatie van de online vragenlijst onderscheiden we twee varianten: tekstuele presentatie van producten, waarbij respondenten alleen de naam van het product te zien krijgen en visuele presentatie van producten, waarbij respondenten naast de naam ook een afbeelding van het product zien (zie Figuur 1 en 2).

Timing van de reminder. De vierde factor die we in deze studie varieerden, was de timing van de reminder en wel op twee niveaus: vroeg en laat. Respondenten in de vroege reminderconditie ontvingen een reminder vijf werkdagen nadat ze de e-mail met uitnodiging hadden ontvangen. Respondenten in de late reminderconditie ontvingen de reminder tien werkdagen na ontvangst van de e-mail en uitnodiging. Deze reminder werd alleen naar diegenen gestuurd die de vragenlijst nog niet hadden ingevuld.



Figuur 1. Productherkenning in de visuele vragenlijst.



Figuur 2. Productherkenning in de tekstuele vragenlijst.

Tabel 1. Overzicht factoren en niveaus.

Factor	Niveau
1. Beloningsvorm	a. waardebon b. donatie (aan goed doel) c. loterij
2. Lengte van de vragenlijst	a. lang (19 productcategorieën) b. kort (9 productcategorieën)
3. Presentatie van de vragenlijst	a. tekstuele presentatie van producten b. visuele presentatie van producten
4. Reminder	a. vroeg b. laat

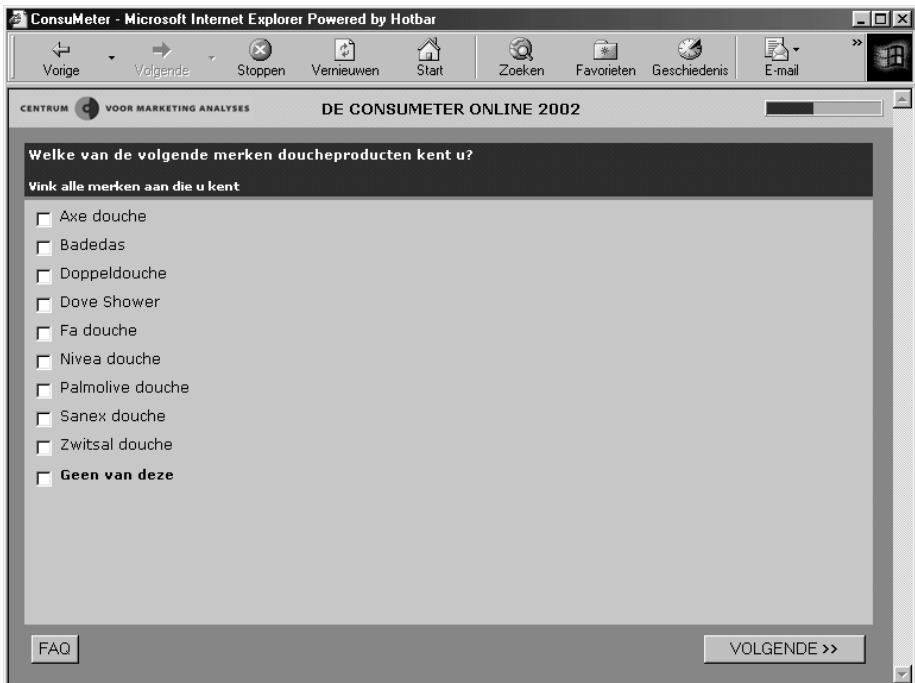
3.4. De afhankelijke variabelen

In deze studie hanteerden we vijf afhankelijke variabelen. Ten eerste het responspercentage, ofwel de respons als een percentage van de gecontacteerde steekproef (dat wil zeggen, het aantal volledig ingevulde vragenlijsten gedeeld door het aantal uitgemailde vragenlijsten). We kunnen onderscheid maken tussen bruto en netto responspercentages. Daarbij wordt het bruto responspercentage berekend op basis van het totaal aantal uitgemailde vragenlijsten. Echter, omdat een groot aantal e-mailadressen fout is of niet meer bestaat, komt het nogal eens voor dat vragenlijsten niet aankomen bij potentiële respondenten. Daarom hanteren we in deze studie het netto responspercentage, ofwel de respons op basis van het aantal werkelijk aangekomen e-mails is. E-mails die niet bezorgd konden worden, zijn bij dit responsepercentage buiten beschouwing gelaten.

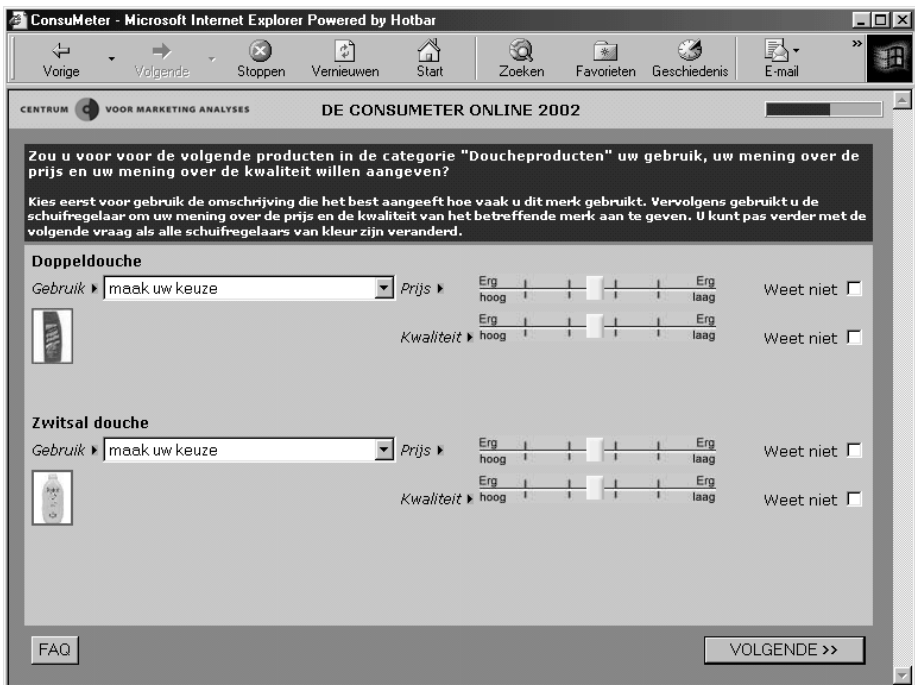
De tweede afhankelijke variabele is responstijd. Dit is de tijd tussen het moment van uitmailen van de vragenlijst en het moment dat de respondent de vragenlijst compleet heeft. De responstijd werd gemeten in dagen. De derde afhankelijke variabele, responskwaliteit, is op twee manieren onderzocht. Ten eerste door te kijken naar de gemiddelden en varianties van de antwoorden op de prijs- en kwaliteitvraag. Aangezien er 190 merken/producten in de vragenlijst voorkomen, zijn er 190 prijsvragen en 190 kwaliteitvragen. Dit zijn de vragen waarin respondenten antwoord moeten geven met behulp van een zogenaamde slider. Deze slider staat standaard in het midden (op waarde *'noch hoog/noch laag'* of *'noch goed/noch slecht'*) en is dan grijs van kleur, wat betekent dat de vraag nog niet beantwoord is (zie Figuur 3).

Respondenten moeten op de slider klikken of deze bewegen, waardoor die van kleur verandert er een antwoord wordt geregistreerd (zie Figuur 4).

Respondenten konden de slider niet standaard in het midden laten staan zonder iets te doen, dan volgde er een foutmelding waarin stond dat de vraag niet beantwoord was. Daarom is het belangrijk om naar gemiddelden en met name varianties te kijken om te achterhalen of mensen in bepaalde condities heel vaak de slider uit gemakzucht niet eens bewegen maar in het midden laten staan. Ten tweede hebben we de responskwaliteit getest door te kijken naar het aantal keer dat respondenten het *'weet ik niet'* vakje aanvinken per productcategorie. Over het algemeen wordt aangenomen dat wanneer respondenten veelvuldig *'weet ik niet'* aanklikken dit de responskwaliteit verslechtert (Singer et al., 2000; Goetz et al., 1984).



Figuur 3. Sliders in onbeantwoorde situatie.



Figuur 4. Sliders in beantwoorde situatie.

De vierde afhankelijke variabele is productherkenning. De vragenlijst bevatte negentien productcategorieën met in iedere productcategorie zo'n tien à twaalf merken waarvan respondenten aan moeten geven of ze die kennen. De variabele productherkenning is dan het aantal producten of merken waarvan respondenten aangeven dat ze die kennen per productcategorie.

Tot slot hebben we gekeken naar de beleving van de respondenten ten aanzien van het invullen van de vragenlijst, met andere woorden hoe ze het hebben ervaren. Daarbij keken we naar hun houding ten opzichte van de vragenlijst (op basis van gemak, moeite en plezier) en de intentie (of waarschijnlijkheid) om nog een keer deel te nemen aan een onderzoek via internet. Verder hebben we respondenten gevraagd welke manier van marktonderzoek ze prefereren: telefonisch, schriftelijk, via internet of geen voorkeur.

3.5. Procedure

Respondenten ontvingen een e-mail uitnodiging voor dit onderzoek met de link naar de vragenlijst in deze e-mail. Met één klik werden respondenten derhalve naar de vragenlijst geleid. Door het gebruik van een speciaal voor dit onderzoek samengestelde ID per respondent was het niet nodig dat respondenten een gebruikersnaam en password invulden.

4. RESULTATEN

4.1. Gegevensverzameling

De gegevensverzameling vond plaats van half april tot en met eind mei 2002. Na drie dagen was meer dan de helft (52,9%) van de antwoorden binnen. Na één week was dit percentage opgelopen naar 63,4%. De gemiddelde responsduur, de tijd tussen het aanklikken van de link en completering van de lijst, was ongeveer 45 minuten. De belangrijkste reden voor deze relatief lange tijd is waarschijnlijk dat sommige respondenten het invullen onderbreken om hem later verder in te vullen. 16,5% deed er een kwartier over om de vragenlijst in te vullen en 78,9% van de respondenten had de vragenlijst binnen een uur compleet. Al met al was het dus een vragenlijst van een aanzienlijke omvang.

Het profiel van de gerealiseerde steekproef is te zien in Tabel 2. De tabel laat zien dat er een goede balans is voor geslacht en een lichte oververtegenwoordiging van jongeren en hoger opgeleiden. Dit is overigens een beeld dat vaak voorkomt in online onderzoek (Lockhart en Detwiler, 1998).

4.2. Responspercentage

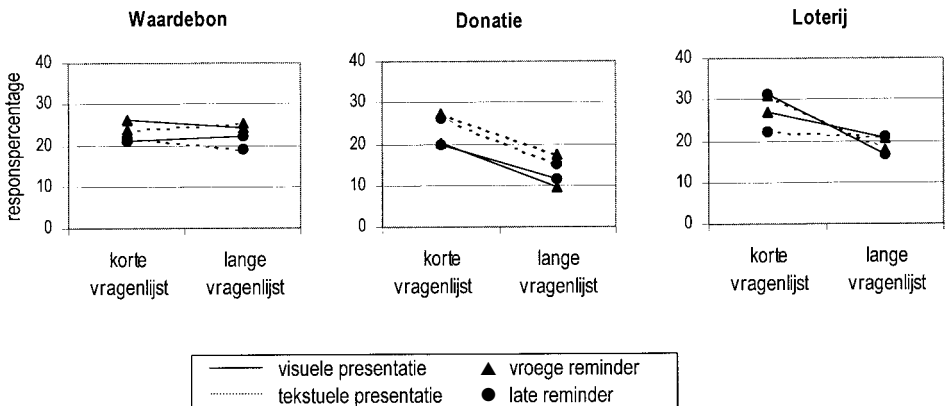
In totaal beschikten we over 5.413 e-mail adressen. Daarvan bleek ongeveer een derde (1.836) inactief te zijn, waardoor we 3.577 bruikbare e-mailadressen overhielden. Van deze 3.577 personen hebben 730 de vragenlijst gecompleteerd, wat neerkomt op een netto respons van 20,4%. Er zijn significante verschillen ($\chi^2_{23} = 74,8$, $p < 0,001$) in responspercentage tussen de cellen van het design waarbij de korte, visuele versie van de vragenlijst met loterij als incentive en een late reminder een respons van 31,4% ken-

Tabel 2. Profiel van de gerealiseerde steekproef.

Geslacht	Man	51,5%
	Vrouw	48,5%
Leeftijd	15-24 jaar	12,5%
	25-34 jaar	23,7%
	35-49 jaar	40,0%
	50-64 jaar	22,1%
	65-74 jaar	1,8%
Opleiding	Lager onderwijs	1,0%
	Lager beroepsonderwijs	4,0%
	Middelbaar algemeen voortgezet onderwijs	11,2%
	Middelbaar beroepsonderwijs	31,4%
	Hoger algemeen en voorbereidend wetenschappelijk onderwijs	10,0%
	Hoger beroepsonderwijs en wetenschappelijk onderwijs	30,0%
	Wetenschappelijk onderwijs - doctoraal	10,4%
	Wil niet zeggen	2,1%

de, terwijl de lange visuele versie met donatie als incentive en een vroege reminder bleef steken bij een respons van slechts 9,4%. In Figuur 5 staan de responspercentages van de combinaties van de factoren weergegeven.

De responspercentages voor de hoofdeffecten van de factoren staan in Tabel 3. Voor 'type incentive' vonden we een significant verband tussen incentive en responspercentage. De waardebon en de loterij hadden vrijwel gelijke responspercentages van respectievelijk 22,9% en 23,0%, terwijl de donatiegroep een responspercentage van 16,6%



Figuur 5. Interactie-effecten van beloningsvorm, vragenlijstlengte, presentatievorm en timing van de reminder.

Tabel 3. Hoofdeffecten responspercentages.

Factor		Lengte van de vragenlijst		Presentatie van de vragenlijst		Reminder	
Beloningsvorm							
waardebon	22,9%	lang	17,1%	tekstueel	21,9%	vroeg	21,2%
donatie	16,6%	kort	24,6%	visueel	19,1%	laat	19,6%
loterij	23,0%						

Noot: $\chi^2_2 = 20,9$; $p < 0,001$ (beloningsvorm); $\chi^2_1 = 30,3$; $p < 0,001$ (lengte); $\chi^2_1 = 4,4$; $p = 0,04$ (presentatie); $\chi^2_1 = 1,4$; $p = 0,23$ (reminder).

had (overigens begunstigde het merendeel, 61%, de Kankerbestrijding, 25% het Wereld Natuur Fond en 15% Amnesty International). Zoals verwacht, bleek er een significant hoger responspercentage voor de korte versie van de vragenlijst dan voor de lange versie te zijn: 17,1% voor de lange, 24,6% voor de korte versie. Ook de verwachting omtrent het lagere responspercentage voor de visuele versie van de vragenlijst in vergelijking met de tekstuele versie is uitgekomen. Het responspercentage voor de tekstuele versie was 21,9% en voor de visuele versie 19,1%. De nulhypothese omtrent de timing van de reminder kon niet worden verworpen.

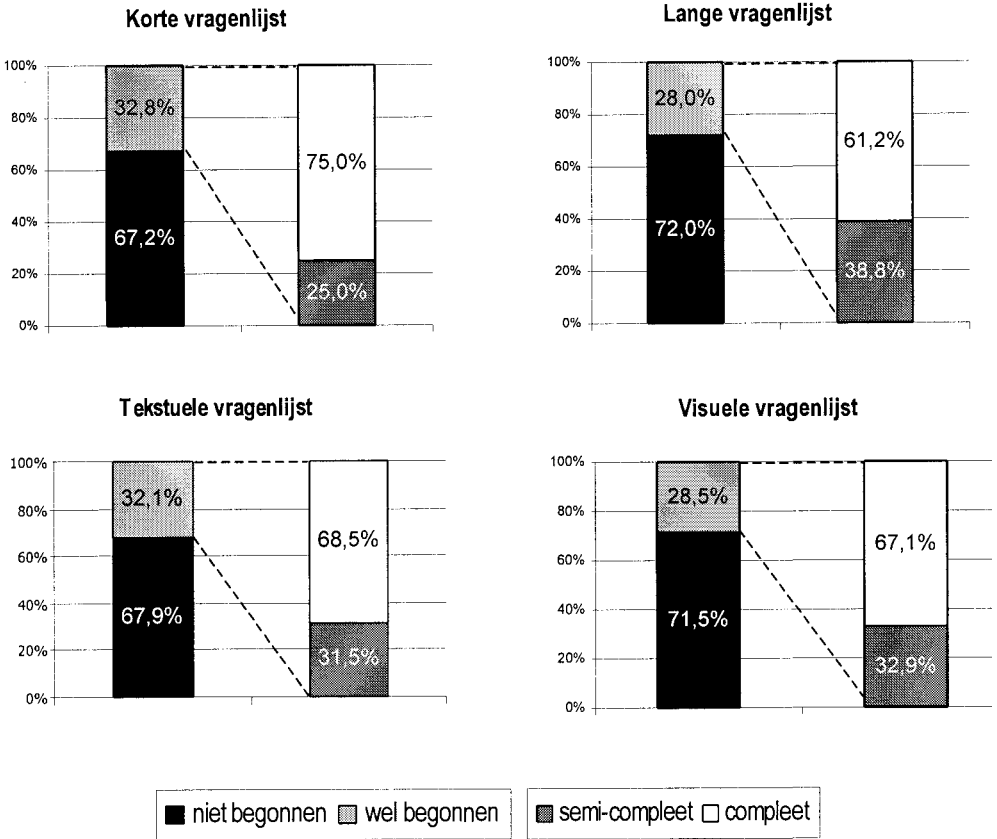
Waarom is de respons voor de lange vragenlijst zoveel lager dan voor de korte vragenlijst? Is dat omdat respondenten worden afgeschrikt door de aankondiging in de e-mail uitnodiging dat het invullen van de vragenlijst ongeveer 30 minuten zal duren? Of is dat omdat respondenten tijdens het invullen van de vragenlijst ervaren dat de vragenlijst lang en eentonig is en stoppen ze er dan mee?

Om nu te achterhalen of de uitnodiging of ervaring een belangrijkere factor is, hebben we gekeken naar hoe groot het percentage semi-completes is ten opzichte van het totaal aantal respondenten dat wel aan de vragenlijst is begonnen. Op basis van de uitnodiging begint 32,8% van de respondenten in de conditie korte vragenlijst met het invullen, in de conditie lange vragenlijst is dat marginaal lager (28,0%). Het percentage uitval laat wel een groot verschil zien, namelijk 25,0% versus 38,8% (zie Figuur 6). Er is dus een effect van ervaring.

Een soortgelijke vraag kunnen we ook stellen met betrekking tot de presentatie van de vragenlijst. Is het responspercentage voor de visuele versie van de vragenlijst lager vanwege de frustratie van de langere download tijd die gepaard gaat met de afbeeldingen of heeft het toch een andere oorzaak? Wanneer we hier kijken naar hoe groot het percentage semi-completes is ten opzichte van het totaal aantal respondenten dat wel aan de vragenlijst is begonnen, dan zien we dat ook deze percentages heel dicht bij elkaar liggen: 31,5% voor de tekstuele versie, 32,9% voor de visuele (zie Figuur 6). Het lagere responspercentage voor de visuele versie van de vragenlijst wordt kennelijk niet veroorzaakt door de langere downloadtijd van de vragenlijst vanwege de afbeeldingen.

4.3. Responstijd

De verwachting was dat de respondenten in de loterij- en in de korte vragenlijstconditie sneller de vragenlijst zouden completeren. De respondenten hadden na gemiddeld 6,6 dagen de vragenlijst ingevuld. Alleen voor type incentive vonden we een significant



Figuur 6. Effect van uitnodiging vs. ervaring op basis van non-response, semi-compleet en compleet voor vragenlijstlengte en presentatievorm.

verschil in gemiddelde responstijd ($F_{2, 727} = 3.8$, $p < 0.05$). Voor de waardebonconditie was dit 7,6 dagen, voor de donatieconditie 7,3 dagen en voor loterijconditie 5,9 dagen.

4.4. Responskwaliteit

Zoals hierboven vermeld, kijken we bij responskwaliteit enerzijds naar de gemiddelden en varianties van de vragen over prijs en kwaliteit en anderzijds naar het aantal weet niet responsen op deze vragen. Dit houdt in dat we een groot aantal toetsen moeten uitvoeren, wat de kans op het vinden van toevallige significanties vergroot. Daarom zullen we bij het vinden van significante resultaten nagaan of het op basis van het significantieniveau te verwachten aantal ligt. Bij een significantieniveau van 5% is immers de verwachting dat 5% van de nulhypoteses onterecht zal worden verworpen.

Bij de eerste benadering vonden we bij geen van de factoren die we gemanipuleerd hebben vaker significantie verschillen in gemiddelden en standaarddeviaties dan op basis van toeval kon worden verwacht. We gaan er daarom van uit dat er voor dit

aspect van responskwaliteit geen effecten zijn van lengte, presentatie en incentive. Voor het aantal maal dat respondenten ‘*weet ik niet*’ invullen, hadden we verwacht dat dit aantal groter zou zijn voor de lange versie. Er bestaat echter wel een verschil voor dit aspect van responskwaliteit voor de presentatie van de vragenlijst. Bij vijf van de productcategorieën rapporteren respondenten met de tekstuele versie van de vragenlijst significant vaker ‘*weet ik niet*’.

4.5. Productherkenning

We hadden verwacht hier een verschil te vinden bij de presentatie van de vragenlijst, waarbij productherkenning hoger is voor de visuele versie van de vragenlijst. Ook hier volgden we een zelfde toetsingprocedure als bij responskwaliteit: de toetsen worden per productcategorie uitgevoerd, de nulhypothese wordt verworpen als het aantal significante resultaten groter is dan op basis van kans kan worden verwacht. Bij vier van de negentien productcategorieën is er sprake van een significant verschil. In tegenstelling tot de verwachting is echter de productherkenning hoger voor de tekstuele versie.

4.6. Houding

In Tabel 4 zijn de gemiddelden voor houding en intentie weergegeven. De respondenten vonden het doorgaans makkelijk om de vragenlijst in te vullen en het kostte ze ook niet veel moeite. Wat betreft het plezier waren ze neutraal. De respondenten waren gematigd positief over de waarschijnlijkheid dat ze nog een keer deel zullen nemen aan een dergelijk onderzoek, waarbij men duidelijk de voorkeur had voor deelname via internet.

Wanneer we kijken of er verschillen in houding en intentie zijn voor de verschillende versies van de vragenlijst, dan blijken er alleen voor de factor lengte significante verschillen op te treden. Respondenten met de korte versie zijn positiever over gemak (26,4 vs. 30,8), moeite (33,1 vs. 45,6), plezier (43,4 vs. 53,5) en intentie (2,1 vs. 2,4). In termen van effectgrootte (het verschil ten opzichte van de standaarddeviatie) moet men hier spreken van kleine tot middelgrote effecten.

5. DISCUSSIE

In dit artikel hebben we getracht inzicht te verschaffen in de manier waarop online marktonderzoek het best kan worden ingericht en uitgevoerd wanneer het gaat om de inzet van een oorspronkelijk zeer uitgebreide schriftelijke vragenlijst.

De behaalde respons in onze studie mag een opvallend resultaat genoemd worden, aangezien de respondenten zijn geworven uit een database met personen die eenmalig telefonisch zijn benaderd. Deze personen hebben wel aangegeven dat ze opnieuw benaderd mogen worden, maar vaak was dit al zeer lang geleden. Voor ongeveer een derde van de respondenten die we voor dit onderzoek hebben uitgenodigd was dit telefonisch contact zelfs in 1999. Als je ziet dat 80 à 90% van de online onderzoeken momenteel met behulp van panels wordt uitgevoerd, dan is het resultaat van ons onderzoek waarbij we met respondenten met minder commitment dan een panel toch nog een respons van ruim 31% (max.) kunnen behalen, positief voor de toekomst van online marktonderzoek. Het behaalde responspercentage is vooral opvallend wanneer we in ogenschouw nemen

Tabel 4. Gemiddelde houding van respondenten ten opzichte van de vragenlijst en preferentie voor marktonderzoek.

Houding	Gemiddelde (sd)
Gemak (makkelijk 0 – moeilijk 100)	28,5 (16,6)
Moeite (weinig moeite 0 – veel moeite 100)	39,0 (23,3)
Plezier (plezierig 0 – onplezierig 100)	48,2 (21,6)
Intentie	Gemiddelde (sd)
Intentie (zeer waarschijnlijk 1 – zeer onwaarschijnlijk 5)	2,3 (1,1)
Welke manier van marktonderzoek prefereert u	Percentage (%)
Telefonisch	5,1
Schriftelijk	13,2
Via Internet	66,6
Geen voorkeur	16,2

dat het hier om een voor online onderzoek ongebruikelijke lange vragenlijst gaat. Uiteraard vonden we een duidelijk effect voor lengte: de korte vragenlijst had een hoger responspercentage dan de lange vragenlijst. Daar dienen twee kanttekeningen bij te worden gemaakt. In de eerste plaats was zelfs de korte versie van de vragenlijst nog relatief lang voor internetonderzoek. Met het behaalde responspercentage van bijna 25% lijkt het toch mogelijk vragenlijsten van substantiële lengte via internet af te nemen. Bovendien doet de lengte van de vragenlijst niet af aan de kwaliteit van de antwoorden en zijn respondenten na het completeren van de vragenlijst nog steeds gematigd positief en achten ze het waarschijnlijk vaker aan online onderzoek mee te zullen werken. Daarnaast blijkt de aankondiging (e-mail uitnodiging) van de lengte van de vragenlijst een belangrijke oorzaak te zijn dat mensen de vragenlijst niet invullen naast de ervaring van het invullen zelf. Een groot gedeelte van de respondenten die beginnen aan de vragenlijst maakt deze, ondanks de lengte, namelijk ook af. Een observatie die we nog niet veel hebben gezien in de bestaande literatuur. We concluderen dat het grootste gedeelte van de mensen a-priori, op basis van de e-mail uitnodiging, besluiten om niet aan de vragenlijst te beginnen.

De resultaten laten vervolgens zien dat de vragenlijst met grafische elementen een lager responspercentage vertoont dan de vragenlijst met alleen tekst. Dit is echter maar een klein effect, dat afwezig is bij respondenten die wel beginnen aan de vragenlijst maar deze niet afmaken. De aantrekkelijkheid van een vragenlijst met afbeeldingen blijkt geen motiverende factor te zijn, maar werkt tegelijkertijd ook niet demotiverend.

Belangrijker is het resultaat dat respondenten minder producten herkennen in de visuele vragenlijst dan in de tekstuele. Aanvankelijk veronderstelden we dat de visuele versie een hogere productherkenning zou opleveren, omdat de afbeeldingen een extra ondersteuning voor herkenning vormen. In de visuele variant hebben de respondenten twee aanwijzingen voor productherkenning. Gezien de lagere productherkenning moeten kennelijk zowel de naam als de afbeelding bekend voorkomen. Met andere woorden in de visuele conditie hanteren de respondenten een strenger criterium voor herkenning. Deze veronderstelling wordt gesteund door de bevindingen dat in de visuele vragenlijst minder vaak een *'weet ik niet'* wordt gescoord. Wanneer respondenten eenmaal een product hebben herkend door middel van een afbeelding, lijken ze een actievere asso-

ciatie van relevante kenmerken te hebben. Dit kan van betekenis zijn voor onderzoek naar herkenning van producten en bijvoorbeeld reclame-uitingen. Door namelijk gebruik te maken van deze kennis kunnen afbeeldingen een strengere selectie-eis teweegbrengen en een hogere associatie. Natuurlijk moet hiermee rekening worden gehouden in de interpretatie, maar het kan tot een betere kwaliteit van antwoorden en onderzoek leiden.

Deze studie laat duidelijk zien dat respondenten de beloning duidelijk zien als iets dat aan hen zelf ten goede zou moeten komen. De respons bij de donatie bleef immers achter bij de andere beloningsvormen. In dit opzicht lijkt de loterij de meest geschikte incentive, aangezien een loterij kosteffectiever is dan een waardebon. Bovendien zagen we dat de responstijd voor loterij ruim anderhalve dag sneller is dan voor de waardebon. Wellicht dat mensen sneller reageren in een loterijconditie omdat ze denken dan meer kans te maken op een prijs. Dit resultaat mag opvallend worden genoemd, aangezien contant geld meestal als optimale stimulus wordt gezien en verondersteld wordt dat contant geld het hoogste responspercentage genereert. Hierbij moeten we ons natuurlijk wel afvragen in hoeverre een waardebon voor een online boeken- en platenwinkel als contant geld kan worden beschouwd.

Wat betreft de optimale timing voor het uitmailen van de reminder geeft onze studie geen eenduidig beeld. Er is een lichte voorkeur om de reminder na vijf werkdagen te sturen in plaats van na tien werkdagen. Niet alleen vanwege het iets hogere responspercentage voor de vroege reminder, maar ook als we ons realiseren dat online onderzoek zoveel sneller is dan schriftelijk onderzoek en dat onze resultaten laten zien dat een groot gedeelte van de antwoorden na drie dagen binnen is.

Concluderend kunnen we stellen dat de resultaten van ons onderzoek als een signaal mogen worden opgevat dat de grenzen van online onderzoek minder beperkt zijn, dan toe nu toe werd aangenomen. De bevinding dat deelnemers in ons onderzoek een sterke voorkeur hebben voor marktonderzoek via internet in vergelijking met telefonisch of schriftelijk onderzoek, versterkt dit alleen maar. Het lijkt mogelijk om relatief lange vragenlijsten op succesvolle wijze via internet af te nemen, ook als dat niet met behulp van een panel plaatsvindt. Bovendien moet kritisch worden omgegaan met afbeeldingen in productherkenning. Tot slot is creativiteit vereist bij de communicatie met respondenten, met name bij de e-mail uitnodiging. Daar ligt de uitdaging voor onderzoekers om creatief te zijn en respondenten te interesseren voor deelname. Doen respondenten eenmaal mee aan een onderzoek, dan vullen de meesten de vragenlijst volledig in. Met het experiment hebben we aangetoond dat ook voor relatief complex en grootschalig onderzoek het internet een alternatief kan zijn, mits met een aantal belangrijke factoren rekening wordt gehouden.

LITERATUUR

- Bauman, S., N. Jobity, J. Airey and H. Atak (2000), "Invites, Intros and Incentives: Lessons from a Web Survey", retrieved January 7th, 2002 from <http://www.wintellitech.com>.
- Bergmann, T.J., W.J. Hannaford and J. Wenner (1990), "Amount, Timing, and Value of Financial Incentives in Mail Surveys: Does it Make a Difference?", *Marketing Research: A Magazine of Management and Applications*, 2, pp. 30-36.
- Brown, M. (1965), "Use of a Postcard Query in Mail Surveys", *Public Opinion Quarterly*, Winter, pp. 635-637.
- Bruvold, N.T. and J.M. Comer (1988), "A model for Estimating the Response Rate to a Mailed

Survey”, *Journal of Business Research*, 16 (2), pp. 101-116.

Couper, M.P., M.W. Traugott and M.J. Lamias (2001), “Web Survey Design and Administration”, *Public Opinion Quarterly*, 65, pp. 230-253.

Dillman, D.A. (1991), “The Design and Administration of Mail Surveys”, *Annual Review of Sociology*, 17, pp. 225-249.

Dillman, D.A., R.D. Tortora, J. Conradt and D. Bowker (1998), “Influence of Plain vs. Fancy Design on Response Rates for Web Surveys”, retrieved January 11th, 2002 from <http://survey.sesrc.wsu.edu/dillman/papers.htm>.

Fox, R.J., M.R. Crask and J. Kim (1988), “Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response”, *Public Opinion Quarterly*, 52, pp. 467-491.

Frick, A., M.T. Bächtiger and U.D. Reips (1999), “Financial Incentives, Personal Information and Drop-out Rate in Online Studies”, *Current Internet Science. Trends, Techniques, Results*, retrieved March 13th, 2002 from http://www.dgof.de/tband99/pdfs/a_h/frick.pdf.

Goetz, E.G., T.R. Tyler and F. Lomax Cook (1984), “Promised Incentives in Media Research: A look at Data Quality, Sample Representativeness, and Response Rate”, *Journal of Marketing Research*, 21, pp. 148-154.

Heberlein, T.A and R.A. Baumgartner (1978), “Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature,” *American Sociological Review*, 43 (August), pp. 447-462.

James, J.M. and R. Bolstein (1990), “The Effect of Monetary Incentives and Follow-up Mailings on the Response Rate and Response Quality in Mail Surveys”, *Public Opinion Quarterly*, 54, pp. 346-361.

James, J.M. and R. Bolstein (1992), “Large Monetary Incentives and Their Effect on Mail Survey Response Rate”, *Public Opinion Quarterly*, 56, pp. 442-453.

Linsky, A. (1975), “Stimulating Responses to Mailed Questionnaires: A Review”, *Public Opinion Quarterly*, 39, pp. 82-101.

Lockhart, D.C. and F.R.J. Detwiler (1998), “Internet Surveys: a Description of the Demographics”, *Quirk’s Marketing Review*, retrieved October 1st, 2001 from http://www.quirks.com/articles/article_print.asp?arg_articleid=353.

MacElroy, B. (2000), “Measuring Response Rates in Online Surveys”, *Quirk’s Marketing Research Review*, retrieved January 7th, 2002 from http://www.quirks.com/articles/article_print.asp?articleid=583.

Marketfacts (2001), “Online Consumers: Beyond Fiction to Fact Distilling Reality from the Hype”, retrieved March 13th, 2002 from <http://www.marketfacts.com/publications/pdf/r62.pdf>.

Mason, W., R. Dressel and R. Bain (1961), “An experimental study of factors affecting response to a mail survey of beginning teachers”, *Public Opinion Quarterly*, 25, pp. 296-299.

Redline, C.D. and D.A. Dillman, (1999), “The Influence of Auxiliary, Symbolic, Numeric, and Verbal Languages on Navigational Compliance in Self-administered Questionnaires”, retrieved June 18th, 2002 from <http://survey.sesrc.wsu.edu/dillman/papers/Auxiliary,Symbolic,Numeric%20paper—with%20Cleo.pdf>.

Rosenblum, J. (2001), “Give and Take: Use Online Surveys to Increase the Web’s Communication Potential”, *Quirk’s Marketing Research Review*, retrieved June 14th, 2002 from http://www.quirks.com/articles/article_print.asp?arg_articleid=705.

Singer, E., J. van Hoewyk, and M.P. Maher (2000), “Experiments with Incentives in Telephone Surveys”, *Public Opinion Quarterly*, 64, pp. 171-188.

Sheehan, K. (2001), “E-mail Survey Response Rates: A Review”, *JCMC*, 6(2), retrieved October 8th, 2001 from <http://www.ascusc.org/jcmc/vol6/issue2/sheehan.html>.

Sheehan, K.B. and M.G. Hoy (1999), “Using E-mail to Survey Internet Users in the United States: Methodology and Assessment”, *JCMC*, 4(3), retrieved January 11th, 2002 from <http://www.ascusc.org/jcmc/vol4/issue3/sheehan.html>.

Solomon, D.J. (2001), “Conducting Web-Based Surveys”, *Practical Assessment, Research & Evaluation*, 7(19), retrieved January 9th, 2002, from <http://ericae.net/getvn.asp?v=7&n=19>.

Steele, T. J., W.L. Schwendig and J.A. Kilpatrick (1992), "Duplicate Responses to Multiple Survey Mailings: A Problem?", *Journal of Advertising Research*, 37 (March/April), pp. 26-34.

The Future Place Consultancy (2001), "A Guide to Best Practices in Online Quantitative Research", retrieved May 21st, 2002 from http://www.thefutureplace.com/best_practice.htm.

Warriner, K., J. Goyder, H. Gjertsen, P. Hohner and K. McSpurren (1996), "Charities, No; Lotteries, No; Cash, Yes: Main Effects and Interactions in a Canadian Incentives Experiment", *Public Opinion Quarterly*, 60, pp. 542-562.

Willems, P.M., P. Oosterveld, J. Lakerveld en T. Wildenberg (2002), "Modaliteits- en populatie-effecten onder de loep genomen", *Onderzoek*, 2, pp. 14-16.

Yammarino, F.J., S.J. Skinner and T.L. Childers (1991), "Understanding Mail Survey Response Behavior: A Meta-Analysis", *Public Opinion Quarterly*, 55, pp. 613-639.

4. Randomized Response Technieken als instrument voor het onderzoek van sociaal gevoelige onderwerpen¹

G. LENSVELT-MULDERS

SAMENVATTING

De Randomized Response Technique (RRT) is een survey methode die speciaal ontwikkeld is voor het onderzoek van sociaal gevoelige onderwerpen. Door de privacy van de respondenten optimaal te garanderen, worden non-response en het geven van sociaal wenselijke antwoorden verminderd. Een meta-analyse van randomized response onderzoeken die zijn uitgevoerd tussen 1975 en 2000, wijst uit dat de resultaten van RRT onderzoek meer valide zijn dan de resultaten van onderzoek waarbij directe vragen worden gebruikt. Dit effect is sterker naarmate het te onderzoeken fenomeen meer sociaal gevoelig is. Tegenwoordig is het dankzij geavanceerde statistische technieken als logistische regressie niet alleen mogelijk om populatie schattingen te maken, maar is het ook mogelijk om uitspraken te doen over de verbanden tussen het gevoelige gedrag en verschillende verklarende variabelen. Hierdoor worden de resultaten van RRT-onderzoek nog beter bruikbaar bij het bepalen van bijvoorbeeld beleid. Het gebruik van de RRT heeft als nadeel dat men grotere steekproeven nodig heeft, wat resulteert in hogere onderzoekskosten en langere doorlooptijden. In dit artikel worden de voor- en nadelen van het gebruik van randomized response technieken beschreven en lijnen uitgezet naar de toekomst: het gebruik van RRT in grootschalig computergestuurd survey-onderzoek.

1. INLEIDING

Nu alweer enkele jaren geleden, werd ik op een vooravond gebeld door een dame van een onderzoeksbureau. Ik was in een vriendelijke bui, het eten stond nog niet op tafel, dus ik was wel bereid deze dame te woord te staan. Het onderzoek waarvoor zij mij een aantal vragen wilde stellen, ging over 'veiligheid in de wijk'. Eerst werden wat algemene vragen gesteld, waardoor ik lekker in het antwoordritme kwam. Langzaam kregen de vragen echter een ander karakter: patrouilleerde de politie vaak in de wijk? Was er een wijkagent actief? Had ik speciale sloten op mijn deuren? Waren er bijzondere voorzieningen op de ramen aangebracht? Ik voelde duidelijk een zekere schroom groeien om deze vragen te beantwoorden en meldde voorzichtig dat ik dacht dat we wel spe-

¹ De auteur wil op deze plek nogmaals het NIPO Amsterdam bedanken voor hun bereidwillige medewerking om het onderzoek naar de voorwaarden om de randomized response methode te implementeren in een computergestuurde survey-omgeving. Vooral de hulp van Raymond Ross en Sylvia van de Bosch waren daarbij onmisbaar.

ciale sloten hadden, omdat dit me een veilig en verstandig antwoord leek. De vraag ‘Zijn er overdag mensen thuis?’ was de druppel die de emmer deed overlopen. Ik vroeg de interviewende dame in wiens opdracht dit onderzoek eigenlijk werd gedaan. Dat was de politie vertelde ze. Ik heb toen de telefoon neergelegd en direct de politie gebeld. Na enig zoeken en navragen bleek dat er inderdaad een onderzoek naar de veiligheid van wijken liep, met als doel een betere en gerichtere inzet van het politiepotentieel in de verschillende wijken van ons dorp. Ik heb toen dus ongewild een belangrijk onderzoek versjteerd door me schuldig te maken aan partiele non-response en het geven van sociaal wenselijke antwoorden.

Maar wat mij die avond overkwam, is zeker meer mensen overkomen. Helaas zorgden al die sociaal wenselijke antwoorden, in combinatie met non-response, ervoor dat de resultaten van dit onderzoek onbetrouwbaar en niet-valide waren.

Dit onderzoek ging over een gevoelig onderwerp. Ik voelde mij door de directe manier van vragen uiteindelijk in mijn veiligheid bedreigd. Hoe hadden de interviewster of haar opdrachtgever dit responseverlies kunnen voorkomen? Hoe hadden zij het wantrouwen, dat vaker een probleem is bij telefonische interviews, kunnen wegnemen? Binnen het survey-onderzoek zijn meerdere methodes ontwikkeld om hiermee om te gaan (Houtkamp-Steenstra en van den Bergh, 1994; Pondman, 1999; Snijkers en Luppes, 2000). Er bestaan uitgebreide protocollen die erop gericht zijn om bij survey-onderzoek de non-response te verminderen en sociaal wenselijke antwoorden te vermijden. Maar soms is een onderwerp dermate lastig en gevoelig om te onderzoeken dat de inzet van ‘gewone’ instrumenten niet voldoende meer is, in dat geval moet er worden gezocht naar alternatieve onderzoekstechnieken. In dit artikel wil ik een onderzoeksmethode bespreken die speciaal ontwikkeld is voor het doen van grootschalig onderzoek naar sociaal gevoelig gedrag: de Randomized Response Techniek. Ik hoop daarmee bij te dragen aan een meer valide toekomst van gevoelig onderzoek.

2. SOCIAAL GEVOELIG ONDERZOEK

De uitdrukking ‘sociaal gevoelig onderzoek’ lijkt zichzelf te definiëren. De meeste onderzoekers kunnen zich een voorstelling maken van de soorten onderzoek die wel en niet onder de paraplu van ‘sociaal gevoelig onderzoek’ vallen. Toch is het voor de duidelijkheid van dit betoog goed om even stil te staan bij wat ik in dit artikel onder de term ‘sociaal gevoelig onderzoek’ ga verstaan.

Sociaal gevoelig zijn alle onderwerpen die psychologisch of sociaal bedreigend zijn voor respondenten. Een onderzoek wordt gedefinieerd als psychologisch bedreigend wanneer de vragen gaan over onderwerpen die zeer persoonlijk en stresserend zijn voor de respondent en waar een antwoord een negatieve invloed kan uitoefenen op het zelfbeeld van de respondent. Sociale bedreigend is een onderwerp wanneer er een maatschappelijk taboe op rust (Farberow, 1963).

Veel opdrachtonderzoek van marktonderzoekbureaus valt onder het hoofdstuk ‘gevoelig onderzoek’. We kunnen hierbij denken aan onderzoek onder mensen met een uitkering (WAO, WW of ABW), onderzoek naar personeelsproblemen binnen organisaties, onderzoek naar het voorkomen van sociaal onaanvaardbaar gedrag als vandalisme, gewelddadigheid en discriminatie, en onderzoek naar strafbaar (crimineel) gedrag, zoals fraude, diefstal en witte boorden criminaliteit al dan niet in organisaties of bedrijven.

Onderzoekers van sociaal gevoelige onderwerpen worden vaak geconfronteerd met een extra hoge non-response en het sociaal wenselijk of veilig beantwoorden van vragen (Lee, 1993; Rasinski, Willis, Baldwin, Yeh en Lee, 1999). Een reden voor het helemaal niet of niet optimaal meewerken aan gevoelig onderzoek is dat mensen zich door de onderzoeksvragen bedreigd voelen. Aan de ene kant omdat de respondent geconfronteerd wordt met het eigen sociaal niet wenselijke gedrag. Zelf voelde ik me al dom, omdat ik tegen de interviewer moest zeggen dat ik geen speciale voorzieningen had tegen inbrekers en dus min of meer om problemen zat te vragen. Aan de andere kant is het zo dat op sociaal onaanvaardbaar gedrag sancties staan. Deze sancties kunnen variëren van gerechtelijke vervolging bij illegaal gedrag tot sociale sancties, zoals het uitgestoten worden door de eigen peer-groep bij sociaal afwijkend gedrag (Sudman en Bradburn, 1982). Bij echt ernstig bedreigende onderwerpen die dicht bij de persoon van de respondent komen, zoals bijvoorbeeld onderzoek naar fraude in organisaties of met uitkeringen, kan deze bedreiging zo ernstig zijn dat de respondent van medewerking afziet, hetzij door niet aan het onderzoek te beginnen, hetzij door het geven van sociaal acceptabele antwoorden.

Een hoge non-response hoeft natuurlijk niet per definitie samen te gaan met het verminderen van de kwaliteit van de data. De uitval kan immers volledig random zijn, waardoor de beoogde steekproefsamenstelling intact blijft. Bij sociaal gevoelig onderzoek is random uitval echter niet waarschijnlijk. Het is eerder zo dat individuen die het gevoelig gedrag vertonen minder geneigd zijn om mee te werken. Als gevolg daarvan vallen de populatieschattingen van dit gedrag te laag uit, wat de resultaten minder valide maakt.

Hoe kunnen problemen met non-response en sociaal wenselijke antwoorden worden opgelost?

Daarvoor zijn door survey-onderzoekers reeds een aantal protocollen ontwikkeld (Hox en De Leeuw, 1994; De Leeuw, 1997, 1999; Stoop, 2002). Hoewel het niet in de scope van dit artikel ligt, zullen we een aantal belangrijke regels hier toch even kort aanstippen.

- Ten eerste was het beter geweest als het onderwerp uitgebreider was ingeleid. Even snel je naam zeggen en het bureau waarvoor je werkt, is niet voldoende als je onderzoek doet naar een gevoelig onderwerp en daarvoor persoonlijke vragen wilt gaan stellen. Een uitgebreidere introductie van het nut en doel van het onderzoek, waarin ook duidelijk wordt gemaakt wie de opdrachtgever is, met bijvoorbeeld een telefoonnummer dat gebeld kan worden om dit te verifiëren, zal bij beoogde respondenten veel argwaan wegnemen (Houtkoop-Steenstra en van den Bergh, 2000). Natuurlijk moet expliciet duidelijk worden gemaakt dat er maatregelen zijn genomen die de privacy van de respondent waarborgen, en maak ook duidelijk (als dat mogelijk is) dat de uitslagen niet op een voor de respondent negatieve manier gebruikt zullen worden (Singer, Von Thurn en Miller, 1995).
- Verwoord de vragen zodanig dat alle schijn van afkeur of censuur vermeden wordt, indien mogelijk kan zelfs de suggestie worden gewekt dat het gedrag sociaal acceptabel is (Dillman, 1978; Peiris, Gregor en Alm, 2000). De vraag *‘heeft u wel eens een karweitje gedaan voor iemand en daarvoor geld gekregen, zonder dit te melden bij de uitkeringsinstantie?’* is een stuk minder bedreigend dan de vraag *‘heeft u wel eens zwart bijverdiend, zonder dit te melden bij de uitkeringsinstantie?’*. Het is daarom ook belangrijk om vragen tevoren uit te testen en eventueel bij te stellen (Jobe en

Mangay, 1991; Snijkers, 2001, en Sykes en Morton-Williams, 1987).

- Bij niet telefonische interviews kunnen we ook gebruik maken van de mogelijkheden die een goede lay-out van de vragenlijst biedt en daarvan optimaal gebruik maken (Peiris et al. 2001). Tenslotte moeten we aandacht schenken aan de kwaliteit van de gebruikte antwoordcategorieën. Maak ze zo ruim mogelijk (Schwarz, 1999).

Maar ook als we alles in het werk gesteld hebben om te voldoen aan de regels van excellent survey-onderzoek, kan het voorkomen dat respondenten niet bereid zijn om mee te werken aan onderzoek, of sociaal wenselijke, veilige, antwoorden geven. Waarom? Ik denk dat dat o.a. is omdat respondenten zo'n database vol informatie een Mekka voor kwaadwillenden vinden. Het valt hierdoor te begrijpen dat een respondent het heel moeilijk vindt om vertrouwen in de bescherming van zijn gevoelige informatie te hebben en daardoor huiverig is om daarover frank en vrij met een onderzoeker te praten.

Bestaan er onderzoeksmethoden die de potentie hebben deze argwanende respondenten toch te laten meewerken aan gevoelig onderzoek? Ja, er bestaan onderzoeksmethodes die de gegevens van respondenten zodanig beschermen dat zij voldoende vertrouwen kunnen hebben om zonder terughouding mee te werken aan dergelijk onderzoek.

Deze onderzoeksmethoden zijn in beginsel ontwikkeld door Warner en de diverse onderzoeksdesigns die later zijn ontwikkeld naar aanleiding van zijn ideeën staan nu bekend onder de verzamelnaam 'Randomized Response Technieken' (Warner, 1965, 1971).

3. DE RANDOMIZED RESPONSE TECHNIEK

Randomized response technieken (RRT's) is de verzamelnaam voor een aantal designs binnen vragenlijstonderzoek die speciaal ontwikkeld zijn om gevoelige onderwerpen op een meer valide wijze te meten. Het gemeenschappelijke kenmerk van alle Randomized Response Technieken is, dat door het inbrengen van een kanselement, bijvoorbeeld met behulp van dobbelstenen of speelkaarten, de privacybescherming van de respondent optimaal gegarandeerd wordt. Een voorbeeld: de respondent krijgt twee vragen voorgelegd, een gevoelige vraag (Heb je wel eens XTC gebruikt op een feestje) en een neutrale vraag (Is /was je moeder jarig tussen 1 januari en 31 maart). Met behulp van een dobbelsteen wordt vervolgens bepaald welke vraag beantwoord moet worden. Omdat alleen de respondent weet welke vraag hij of zij heeft beantwoord, heeft het individuele antwoord 'ja' of 'nee' dus geen betekenis, het kan iets zeggen over het gebruik van XTC of over de verjaardag van de moeder. Het feit dat de onderzoeker weet hoeveel mensen er jarig zijn tussen 1 januari en 31 maart, maakt het mogelijk om een schatting te maken van het aantal mensen dat XTC heeft gebruikt op een feestje. Het inbrengen van het kanselement zorgt ervoor dat de data als het ware 'vervuild' worden waardoor individuele antwoorden hun betekenis verliezen, terwijl het wel mogelijk blijft om populatie schatters te berekenen.

Met behulp van randomized response technieken kunnen we:

- prevalentie- en omvangschattingen maken
- de relatie tussen randomized response variabelen en verklarende variabelen leggen d.m.v. logistische regressie (Van der Heijden, en van Gils, 1996; Maddala, 1983; Scheers en Dayton, 1988).

- onderzoek doen naar relaties tussen randomized response variabelen onderling
- en telvariabelen construeren (Van der Heijden, Hox en Ellfers, 2002; Van den Hout en van der Heijden, 2002a en 2002b).

De ontwikkeling van randomized response technieken

In 1965 ontwikkelde Warner een onderzoekstechniek die, door het totaal waarborgen van de privacy van respondenten, zou moeten leiden tot een vermindering van non-response en het geven van sociaal wenselijke antwoorden op gevoelige vragen. Warner's randomized response design kan nog steeds worden gebruikt om de algemene basisprincipes van randomized response technieken te demonstreren.

Een respondent krijgt twee statements voorgelegd in de vorm:

- Ik heb gevoelig gedrag A vertoond
- Ik heb gevoelig gedrag A niet vertoond

Met behulp van een kansspelletje (met bijvoorbeeld dobbelstenen, speelkaarten of muntjes) wordt nu bepaald welk statement moet worden beantwoord met 'waar' of 'niet waar'. Omdat alleen de respondent weet welk statement hij of zij heeft beantwoord, heeft het individuele antwoord ook nu geen betekenis. 'Waar' kan immers op beide statements betrekking hebben. Echter het is wel mogelijk om een zuivere (un-biased) populatie schatter van het gevoelige gedrag in de populatie ($\hat{\pi}$) te berekenen.

$$\hat{\pi} = \frac{\hat{\lambda} + (p-1)}{2p-1} \quad (1)$$

waar $\hat{\lambda}$ de geobserveerde proportie 'waar'-antwoorden in onze steekproef is, p de kans dat de respondent het eerste statement moet beantwoorden, en $p-1$ de kans dat het tweede statement moet worden beantwoord. De steekproef variantie is gelijk aan

$$Var(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \quad (2)$$

Deze variantie neemt toe als het aantal individuen (n) in de steekproef klein is en als p dichter bij 0,5 ligt.

Vergelijking 2 laat duidelijk zien dat de RRT van Warner veel minder efficiënt is dan gewone 'directe vragen'-designs (face-to-face en telefonische interviews, zelf in te vullen vragenlijsten, CAI). Er is sprake van een vergroting van de variantie met

$$\frac{p(1-p)}{n(2p-1)^2} \quad (3)$$

waardoor het noodzakelijk wordt met grotere steekproeven te werken, om de betrouwbaarheidsintervallen rondom de gemiddelden vergelijkbaar te maken met die van 'directe vragen'-designs. Wanneer we de methode van Warner gebruiken kan dit betekenen dat de steekproef tot acht maal zo groot moet worden om vergelijkbare betrouwbaarheidsintervallen te verkrijgen.

Na 1965 zijn er dan ook een aantal nieuwe Randomized Response Technieken ontwikkeld om het randomized response onderzoek meer efficiënt te maken (dat wil zeggen

dat de standaardfouten kleiner worden). De op dit moment meest gebruikte RRT is de Geforceerde Response procedure van Boruch (1971; Lensvelt-Mulders, Hox en Van der Heijden, submitted). Aan de respondenten wordt alleen de gevoelige vraag voorgelegd, i.p.v. twee statements. Voordat de respondent antwoord geeft op de vraag, worden er twee dobbelstenen gegooid en de resultaten van beide stenen worden bij elkaar worden opgeteld. Daarna krijgen de respondenten de instructie ‘ja’ te antwoorden wanneer het resultaat van hun worp 2, 3 of 4 is, en ‘nee’ wanneer het resultaat 11 of 12 is. De respondenten worden als het ware geforceerd om een antwoord te geven. Deze geforceerde antwoorden zijn in principe betekenisloos. Wanneer het resultaat van de worp tussen 5 en 10 is, dan wordt de respondent gevraagd de gevoelige vraag naar waarheid met ‘ja’ of ‘nee’ te beantwoorden. Met behulp van elementaire kansberekening kan daarna een populatieschatting voor het gevoelige gedrag worden gemaakt. Ook hier geldt weer dat niemand nog kan achterhalen welke individuen het gevoelig gedrag vertonen, omdat het individuele ‘ja’ of ‘nee’ antwoord, door het inbrengen van het geforceerde antwoord, opnieuw betekenisloos is geworden. Wel kan het voorkomen van het gevoelige gedrag in de totale populatie worden geschat en kunnen verbanden tussen diverse onafhankelijke variabelen en het gevoelige gedrag worden vastgesteld (van der Heijden, van Gils, Bouts en Hox, 2000).

De Geforceerde Response methode is statistisch veel efficiënter dan Warner’s oorspronkelijke design, de steekproefomvang hoeft nu nog maar 2 keer zo groot te zijn om een betrouwbaarheidsinterval te verkrijgen dat gelijk is aan dat van een design dat steunt op directe vragen.

Een nadeel van het gebruik van de Geforceerde Response methode is het feit dat het geforceerde antwoord ‘ja’ gekoppeld is aan de gevoelige vraag. Op het scherm of op het antwoordblad ziet de respondent de volgende gevoelige vraag staan:

– *‘Heeft u wel eens voor meer dan 50 Euro aan materiaal van het werk meegenomen voor eigen gebruik? (bijvoorbeeld kantoomaterialen of diskettes)’*

Als de respondent 2-4 gooit, wordt verwacht dat het antwoord ‘ja’ wordt aangevinkt. Dit ‘ja’ aanvinken, wordt door de respondenten die nog nooit materiaal hebben meegenomen als onprettig en oneerlijk ervaren, waardoor de neiging om mee te werken aan de Geforceerde Response procedure minder kan worden (Boeije en Lensvelt-Mulders, 2002). Als dit een probleem is, kan worden gekozen voor het statistisch even efficiënte, maar psychologisch sterkere ‘Unrelated Question’-design met bekende populatieparameters (Greenberg, Abdul-Ela, Simmons en Horvitz, 1969). De respondent krijgt nu twee vragen voorgelegd. Een gevoelige vraag en een neutrale vraag.

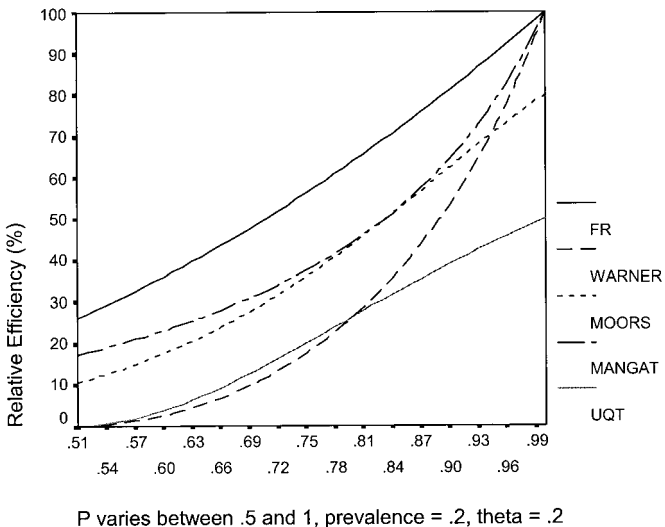
– *Heeft u wel eens voor meer dan 50 Euro aan materiaal van het werk meegenomen voor eigen gebruik? (bijvoorbeeld kantoomaterialen of diskettes)?*

– *Is/was uw moeder jarig tussen 1 januari en 31 maart?*

De dobbelstenen bepalen weer welke vraag beantwoord moet worden. Omdat we het percentage mensen die jarig zijn tussen 1 januari en 31 maart kennen, lijkt deze methode dus op de Geforceerde Response, 25% van de respondenten die de neutrale vraag moeten beantwoorden wordt als het ware geforceerd om ‘ja’ te zeggen. Maar omdat het ‘ja’-antwoord nu gekoppeld is aan de neutrale vraag hebben respondenten daar minder problemen mee. Het Unrelated Question-design kan echter problematisch worden wanneer veel gevoelige vragen moeten worden beantwoord, dit omdat elke neutrale vraag maar éénmaal mag voorkomen, om te voorkomen dat individuen te vaak hetzelfde antwoord moeten geven (altijd ‘ja’ als zij de neutrale vraag moeten beantwoorden). Hier-

door kan bij de respondent het idee ontstaan dat uiteindelijk uit deze antwoordreeksen toch de eigen identiteit kan worden afgeleid. Dit is een serieus nadeel dat heeft geleid tot het mislukken van de RRT conditie bij Justitie onderzoek naar fraude met individuele huursubsidies en de bestrijdingsmiddelenwet (Justitie, 1998).

Het maken van meerdere neutrale vragen over gedrag waarvan het voorkomen in de populatie bekend is, is minder makkelijk dan het op het eerste gezicht lijkt. Iedereen die zo'n vraag kan verzinnen wordt dan ook vriendelijk verzocht hem naar mij op te sturen. Om dit probleem te ondervangen stelt Moor's (1971) voor om een klein deel van de totale steekproef te gebruiken om diverse neutrale gedragingen te schatten. We kunnen hierbij denken aan vragen als 'heeft u een abbonement op de Volkskrant' of 'gaat u vaker dan 2 keer per jaar naar de schouwburg'. De inhoud van deze neutrale vragen kan afhankelijk van de beoogde populatie worden gekozen. Twintig procent van de beoogde respondenten krijgt alleen de vragen over het neutrale gedrag voorgelegd. Wanneer de respondenten random worden getrokken uit de beoogde steekproef kan worden aangenomen dat de zo verkregen schattingen van het neutrale gedrag voor de hele steekproef gelden en dus gebruikt kunnen worden bij het berekenen van de populatie gemiddelden en varianties van het *gevoelige* gedrag. Het nadeel van deze methode is dat het onderzoek weer minder efficiënt wordt, door het verloren gaan van een deel van de steekproef. In figuur 1 wordt de relatieve efficiëntie (in %) van vijf randomized response methoden vergeleken, bij een prevalentie van het gevoelige gedrag in de populatie van 20% en een variabele p , de kans dat de respondent de gevoelige vraag moet beantwoorden. De efficiëntie van het 'Unrelated Question'-design met *bekende* populatie parameters (is /was je moeder jarig tussen 1 januari en 31 maart) voor de neutrale vraag, is gelijk aan die van de Geforceerde Response methode. Het 'Unrelated Question'-design met *onbekende* populatie parameters voor de neutrale vraag is veel minder efficiënt, dit omdat nu ook het neutrale gedrag geschat moet worden.



Figuur 1: Efficiëntie verschillen tussen 5 randomized response methoden, bij een prevalentie van 20% en een variabel gemaakte p (kans dat de respondent de gevoelige vraag moet beantwoorden).

Hoe valide zijn randomized response resultaten?

Als we de resultaten van verschillende Randomized Response Technieken met elkaar vergelijken valt het op dat deze niet eenduidig zijn. Een deel van de onderzoeken wijst uit dat het gebruik van een randomized response techniek betere schattingen van het gevoelige gedrag oplevert, terwijl andere onderzoeken aantonen dat er geen verschil is tussen de resultaten van randomized response onderzoek en de resultaten van condities waarin gebruik wordt gemaakt van directe vragen. Wanneer de resultaten van onderzoeken elkaar tegen spreken wordt het opportuun om een meta-analyse toe te passen. Een meta-analyse is een techniek waarbij de resultaten van verschillende studies statistisch worden gecombineerd om te komen tot algemeen geldende uitspraken. Daarbij wordt natuurlijk rekening gehouden met verschillen in steekproefomvang en varianties tussen studies. Een meta-analyse van de resultaten van 25 jaar vergelijkend RRT onderzoek heeft uitgewezen dat met behulp van randomized response technieken meer valide uitkomsten worden verkregen dan met andere, meer conventionele dataverzamelmethode, zoals schriftelijke en telefonische enquêtes en face-to-face interviews (Lensvelt-Mulders, Hox en van der Heijden, under review). Dat wil zeggen dat er minder sociaal wenselijke antwoorden worden gegeven op de gevoelige vragen. Dit positieve effect is groter naarmate de onderwerpen meer sociaal gevoelig zijn. Zelfs als het onderwerp zeer gevoelig ligt, zoals onderzoek naar bedrijfs- of uitkeringsfraude of onderzoek naar afwijkend seksueel gedrag, worden met behulp van RRT valide schatters van het voorkomen van dit gedrag in de populatie gevonden.

Doordat de resultaten van randomized response onderzoek meer valide zijn, kunnen in potentie de dark numbers worden verkleind. Dark numbers is de naam voor het aantal gedragingen dat zich aan de wetenschappelijke observatie onttrekt, het zogenaamde zwarte circuit. Het is per definitie onmogelijk om cijfers te koppelen aan dark numbers (80 % van de fraudeurs kennen we niet), maar algemeen is het zo dat wanneer de verwachting is dat de dark numbers hoog zijn, het nemen van maatregelen of het maken van beleidsbeslissingen zeer moeilijk wordt. Hoe kleiner het dark number, des te beter het beleid kan worden afgesteld op de situatie. Het gebruik van randomized response technieken leidt tot een betere kennis van het voorkomen van sociaal gevoelig gedrag in de samenleving en een beter inzicht in sociaal gevoelige fenomenen kan bijdragen tot het verbeteren van de besluitvorming in het beleid.

Assumptie van vertrouwen en privacy

De respondent die meewerkt aan randomized response onderzoek weet dat zijn privacy volledig gegarandeerd is. Het gevolg daarvan is dat hij of zij eerder geneigd zal zijn om op een gevoelige vraag naar waarheid te antwoorden. Dit is de belangrijkste assumptie die aan de basis van alle RRT's ligt, doordat de privacy gegarandeerd is, kan en zal de respondent vrijuit praten (Armacost, Hosseini, Morris en Rehbein, 1991; Begin en Boivin, 1980). Veel onderzoeksresultaten worden daarom geïnterpreteerd volgens het adagium 'meer is beter', hogere schatters staan voor meer valide resultaten. Naar de privacyassumptie die aan de basis van het randomized response onderzoek ligt is maar weinig onderzoek gedaan, terwijl uiteindelijk het hele randomized response onderzoek staat of valt met het waar zijn van deze assumptie. In het onderstaande zal ik de resultaten van een paar van deze onderzoeken samenvatten.

– Onderzoek van Landsheer en medewerkers (1999) toont aan dat er een verband

bestaat tussen de mate waarin respondenten begrijpen hoe de RRT in hun voordeel werkt, het vertrouwen dat zij uitspreken in de methode en hun bereidheid om de vragen naar waarheid te beantwoorden. Als respondenten begrijpen hoe de randomized response methode hun privacy beschermt, is er een grotere bereidheid om mee te werken aan het onderzoek. Wanneer het vertrouwen van de respondenten in de randomized response methode enkel gebaseerd is op de introductie van de interviewer en de respondent dus niet de strekking van de techniek begrijpt dan is de bereidheid om mee te werken veel minder groot. Dit werd ook gevonden in een later onderzoek naar de mate waarin respondenten de randomized response techniek begrijpen. Wanneer de respondent in eigen woorden kon uitleggen hoe de methode de privacy beschermt was hij of zij meer bereid de gevoelige vragen te beantwoorden (Lensvelt-Mulders en Boeije, submitted).

- Er zijn studies gedaan naar het effect van variaties in de keuze voor de kans dat de gevoelige vraag moet worden beantwoord (p). De waargenomen veiligheid van de participanten wordt bepaald door de keuze van de kans: de p -waarde. Hoe groter de kans dat de gevoelige vraag moet worden beantwoord, hoe efficiënter het RRT design is, maar hoe kleiner de privacybescherming van de respondenten. De balans tussen efficiëntie en privacy waarborging ligt bij een waarde van p tussen de .75 -.80 (75% tot 80% van de respondenten moet de gevoelige vraag beantwoorden). Hogere p -waarden leiden al snel tot vermindering van het aantal eerlijke antwoorden, zeker wanneer meerdere RRT vragen worden gesteld (Fligner, Poticello en Singh, 1977; Soeken en Macready, 1982).
- Empirisch onderzoek naar het eerlijk en naar waarheid beantwoorden van gevoelige vragen onder randomized response condities toont aan dat 2-10% van de respondenten ‘liegt’, dat wil zeggen zij ontkennen sociaal onaanvaardbaar gedrag waaraan ze zich wel schuldig hebben gemaakt, of zij overdrijven sociaal gewaardeerd gedrag dat ze niet vertonen. Deze resultaten zijn in strijd met de belangrijkste assumptie van de RRT. Dat neemt niet weg dat in deze RRT condities significant minder ‘sociaal wenselijk’ werd gerapporteerd dan in de andere condities (Edgell, Duchan en Himelfarb, 1992).

4. COMPUTER GESTUURD ONDERZOEK

Sinds Warner in 1965 zijn RRT ontwikkelde, is er veel veranderd in de wereld van het grootschalig onderzoek. Een van de nieuwe toepassingen binnen survey onderzoek, waarvan Warner zelfs niet kon dromen, is natuurlijk het gebruik van computertechnieken. Tegenwoordig wordt voor dataverzameling binnen vragenlijstonderzoek gebruik gemaakt van de computer; het gebruik van CATI, CAPI, Access-panels en websurveys neemt steeds meer toe. Het meest bekend en toegepast zijn waarschijnlijk de telefonische enquêtes, waarbij de interviewer vanachter de computer de vragenlijst afneemt. Het wordt echter ook steeds interessanter om respondenten te benaderen voor onderzoek door hen een vragenlijst per diskette (disk-by-mail) of per e-mail toe te sturen en we zien een toenemende belangstelling voor de mogelijkheden van het internet voor het afnemen van vragenlijsten (Crawford, Couper en Lamias, 2001; Schmidt, 1997). Het gebruik van de computer heeft veel voordelen, fouten bij doorverwijzingen en in antwoordcategorieën worden vermeden, de kwaliteit van de data wordt bij goed geïmplementeerd computerondersteund onderzoek beter, en bij grootschalige surveys zijn

de kosten per vragenlijst beduidend lager (De Leeuw, 2001).

Respondenten ervaren het gebruik van CASI (computer assisted self interviewing) als interessant en prettig. Ze vinden dat het gebruik van computer gestuurde vragenlijsten een professionele indruk maakt (Baker en Bradburn, 1992; De Leeuw, Hox en Snijkers, 1995).

CASI en sensitieve onderwerpen.

Het gebruik van complexe en op de respondent toegesneden vragenlijsten heeft ook voordelen wanneer gevoelige onderwerpen worden onderzocht (De Leeuw, Hox, Kef en van Hattum, submitted). Het gebruik van de computer lijkt het gevoel van privacy zelfs te vergroten, waardoor respondenten zichzelf in het antwoordproces meer blootgeven. Doordat er geen interviewer aanwezig is, heeft de respondent meer privacy, wat zich uit in minder sociaal wenselijke antwoorden (Weisbrand en Kiesler, 1996). Dit maakt dat CASI uitermate geschikt is voor de studie van gevoelige onderwerpen.

5. CARR: COMPUTER ASSISTED RANDOMIZED RESPONSE

De belangrijkste aanname bij RRT onderzoek is dat wanneer de respondent het gevoel heeft dat zijn of haar privacy voldoende kan worden gewaarborgd, hij of zij eerder geneigd is zonder terughouding mee te werken aan het onderzoek ook als dit over gevoelige onderwerpen gaat. In bovenstaande paragraaf is beschreven dat de door de respondenten gepercipieerde privacy een van de sterke punten is van computergestuurd onderzoek naar sociaal gevoelige onderwerpen. Daarmee lijkt het voordeel dat het gebruik van randomized response technieken biedt, achterhaald. Immers CASI biedt ook een verhoogd gevoel van privacy protectie en is daarbij ook nog efficiënter. Is er dan nog een noodzaak voor het ontwikkelen van computergestuurd randomized response onderzoek (CARR)? Ik denk van wel, ik denk zelfs dat het gebruik van randomized response technieken in de nabije toekomst noodzakelijker wordt dan ooit. En daar heb ik twee redenen voor.

Op de eerste plaats verliest het gebruik van de computer langzaam maar zeker zijn magie. De computer wordt steeds meer een dagelijks gebruiksvoorwerp, waardoor de respondenten vertrouwer worden met de apparatuur. Ook mensen die niet professioneel met een computer werken, leren steeds beter haar (on)mogelijkheden kennen. Er wordt steeds meer informatie verzameld en vastgelegd met behulp van computers en steeds grotere data bestanden worden aangelegd. In de pers verschijnen vaker negatieve berichten over de manier waarop wordt omgegaan met datafiles, zoals over het verkopen van informatie of het koppelen van gegevens door verschillende organisaties. Hoewel deze berichten meestal niet over onderzoeksgegevens gaan en er ten aanzien van het koppelen van bestanden van verschillende organisaties en de heranalyse van onderzoeksgegevens door derden een duidelijke regelgeving bestaat, kan deze berichtgeving op den duur toch het vertrouwen van respondenten in het gebruik van CASI schaden. Het vertrouwen dat de privacy wordt gewaarborgd door het gebruik van de computer kan dan afnemen. Daarom zal de noodzaak om te zoeken naar onderzoeksmethoden die de privacy werkelijk garanderen in de toekomst eerder toenemen dan afnemen.

De tweede reden om randomized response technieken te gebruiken is het toenemen van mogelijkheden om interviews af te nemen via het internet. Internet leent zich sterk voor onderzoek onder bijvoorbeeld organisaties of speciale groepen respondenten. Omdat

grote groepen individuen tegelijk benaderd kunnen worden, hoeven de voor het randomized response onderzoek noodzakelijke extra grote steekproeven geen bezwaar meer te zijn. Verder worden de toegang tot- en het gebruik van het internet steeds meer gemeengoed, waardoor zeer diverse soorten subpopulaties kunnen worden bereikt. Het nadeel van het gebruik van het internet als bron om data te verzamelen, is dat in principe iemand die via internet een vragenlijst invult altijd te traceren is. Respondenten vernemen via de media dat de politie personen die zich schuldig maken aan bijvoorbeeld het verspreiden van kinderporno op het internet worden getraceerd en opgepakt. Dit kan het psychologische effect hebben dat respondenten zich gaan afvragen of zijzelf ook kunnen worden getraceerd en eventueel worden aangesproken op hun gevoelige gedrag. Dit kan bijvoorbeeld spelen bij personen die in een vragenlijst toegeven dat zij ooit gefraudeerd te hebben met het opgeven van de inkomstenbelasting. Jochem Musch en medewerkers hebben in Duitsland onderzoek gedaan naar de mogelijkheden om belastingontduiking door organisaties te onderzoeken met behulp van vragenlijsten op het internet. Zij gebruikten daarvoor met succes een randomized response design (Musch, Broder en Klauer, 2001).

CARR ontwikkelingen in Nederland

Aan de Universiteit Utrecht doet de capaciteitsgroep Methodenleer en Statistiek op dit moment onderzoek naar de mogelijkheden om randomized response technieken te implementeren in grootschalig computergestuurd surveyonderzoek. Daarbij gaan we uit van de gedachte dat we het beste uit twee werelden kunnen gebruiken om de resultaten van het onderzoek nog verder te optimaliseren. Het randomized response design in een CASI-omgeving kan helpen een optimale datakwaliteit te realiseren.

Voor het Ministerie van Sociale Zaken en Werkgelegenheid is daartoe een studie verricht naar de mogelijkheden om fraude met uitkeringen te onderzoeken met behulp van Computer Assisted Randomized Response (CARR). Als randomized response methode is gekozen voor het eerder beschreven Geforceerde Response design. Uit eerder onderzoek was namelijk gebleken dat de onderzoeksresultaten bij gebruik van de Geforceerde Response methode meer valide waren in vergelijking met designs die op direct vragen steunden (Van der Heijden, van Gils, Bouts en Hox, 1998).

Voor deze studie hebben we 22 personen benaderd die allemaal langer dan een jaar een uitkering hadden volgens de ABW. De respondenten waren tussen 21 en 63 jaar oud, de helft van hen was boven de 45. Twaalf vrouwen en zeven mannen hebben deelgenomen aan het onderzoek. Drie personen hebben afgehaakt om onbekende redenen. Omdat de gevoelige vragen moesten worden beantwoord met behulp van een computer zonder de tussenkomst van een interviewer hebben we ook gechecked hoeveel respondenten al ervaring hadden met het gebruik van een personal computer. Twaalf respondenten gaven aan geen of zeer beperkte ervaring met het gebruik van computers te hebben.

Na een korte introductie van het onderzoek werd aan de respondenten een korte vragenlijst aangeboden die met behulp van een computer moest worden ingevuld. De vragenlijst bestond uit 12 vragen, die betrekking hadden op de regels die het recht op een bijstandsuitkering omgeven. Aan deze twaalf vragen gingen drie oefenvragen vooraf om de randomized response techniek onder de knie te krijgen. Met behulp van twee dobbelstenen die zij zelf moesten gooien, werd bepaald of een respondent geforceerd 'ja' of 'nee' moest afvinken (2, 3, 4, en 11 of 12), of dat hij of zij de vraag naar waarheid met 'ja' of 'nee' moest beantwoorden (5-10). Het invullen van de vragenlijst werd

op videoband opgenomen, terwijl twee onderzoekers in een andere kamer meekeken via een televisiescherm. Hiervoor was door de respondenten tevoren toestemming gegeven. Nadat de vragenlijst was voltooid werd de respondent een interview afgenomen. Dit interview was semi-gestructureerd, de inhoud werd bepaald door een topiclijst in combinatie met vragen afgestemd op de gedragingen van de respondent tijdens het invullen van de vragenlijst. Had een respondent bijvoorbeeld heel erg vaak zitten zuchten, of twee keer met de dobbelstenen gegooid voor een vraag, dan werd dit gedrag in het interview ter sprake gebracht. Het interview werd digitaal opgenomen. De technieken van het meekijken en interviewen zijn ontleend aan het zogenaamde cognitieve laboratorium onderzoek. Cognitief laboratorium onderzoek wordt algemeen gebruikt om vragenlijstonderzoek te evalueren en de kwaliteit daarvan te bewaken en te verbeteren. Hoewel dit instrument gewoonlijk wordt ingezet om de kwaliteit van het vraagproces te onderzoeken, is het ook heel goed bruikbaar voor het evalueren en monitoren van data-verzamelingenprocedures (Snijkers, 2001).

De doelstelling van het onderzoek was om na te gaan in hoeverre deze speciale groep respondenten in staat was om, zonder aanwezige interviewer, zelfstandig een computer-gestuurde vragenlijst met randomized response vragen in te vullen.

Hoewel de voorkennis van de respondenten onderling aanzienlijk uiteenliep en acht respondenten aangaven nog nooit met computers te hebben gewerkt, waren slechts drie respondenten niet in staat om de vragenlijst zelfstandig in te vullen. Alle andere respondenten voltooiden de vragenlijst probleemloos en binnen beperkte tijd. Het gebruik van de computer beschreven de respondenten als leuk, spannend en interessant, iets dat ook in eerder empirisch onderzoek werd gevonden (Baker en Bradburn, 1992). Een aantal respondenten gaf zelf aan dat het gebruik van de computer hun een veilig en prive gevoel gaf, waardoor het makkelijker werd om de vragen naar waarheid in te vullen. Echter twee respondenten gaven gedurende het interview aan dat de computer een onpersoonlijk 'ding' was, een 'machine', en dat zij dergelijke vragen liever hadden beantwoord in een persoonlijk interview.

Privacy was een zeer belangrijk item voor de alle respondenten. Zonder uitzondering werd gevraagd of alles wat in het interview gezegd werd ook in vertrouwen kon worden verteld. Een respondent vond het niet goed dat we video-opnames maakten, maar wel dat we meekeken en het interview opnamen. Voor alle respondenten lag de inhoud van de vragen erg gevoelig. De vragen gingen over de regels die samenhangen met het recht op een uitkering. In feite ging het dus over hun recht op inkomen. Overigens vond geen van de respondenten de vragen aanmatigend of veroordelend.

Bij het bespreken van de randomized response methoden kwamen we tot een paar belangrijke inzichten ten aanzien van de RRT.

Ten eerste hadden we verwacht dat het gebruik van de randomized response techniek zou leiden tot een verzwaring van de cognitieve taken van de respondent. Door de respondenten wordt de techniek inderdaad ervaren als 'twee dingen tegelijk moeten doen'. De door de respondenten meest gebruikte strategie was om de vraag te lezen en het juiste antwoord te bedenken, waarna zij afhankelijk van het aantal ogen dat zij hadden gegooid, beslisten welk antwoord uiteindelijk moest worden aangevinkt, de waarheid of een geforceerd 'ja' dan wel 'nee'. Ondanks deze dubbele taak gaven de respondenten aan dat de randomized response methode hen goed was afgegaan.

Ten tweede bleek dat we een onderscheid moesten maken tussen wat we later herkende als 'liegen' versus 'vals spelen'. Liegen doet de respondent wanneer hij of zij willens

en wetens de vraag niet naar waarheid beantwoord. Dus als er naar waarheid 'ja' moet worden ingevuld terwijl de respondent 'nee' aanvinkt, of andersom. Een respondent speelt vals als hij of zij zich niet houdt aan de randomized response regels. De eerste 10 respondenten speelden regelmatig vals door 'nee' aan te vinken terwijl ze 2, 3, of 4 hadden geworpen en dus 'ja' hadden moeten antwoorden. Volgens hen was dit omdat het oneerlijk aanvoelde om 'ja' in te drukken terwijl het ware antwoord 'nee' was. Zij voelden zich gedwongen tot oneerlijkheid terwijl het doel van een enquête in hun ogen toch het geven van eerlijke antwoorden was. Nadat we, volgend op de oefenvragen, een extra stukje instructie hadden toegevoegd aan de vragenlijst, waarin we duidelijk maakten dat 'eerlijk' in dit bijzondere geval betekende 'werken volgens de regels van de speciale vraagmethode' werd deze vorm van valsspelen nog maar een keer waargenomen.

Het naar waarheid laten invullen van de gevoelige vraag was lastiger. Uiteindelijk konden we drie groepen respondenten onderscheiden. De eerste groep respondenten wist dat zij een regel had overtreden en gaf dat ook toe bij het invullen van de vragenlijst. Om een voorbeeld te geven, op de vraag of men wel een iets had gekregen van vrienden of kennissen en dit niet had opgegeven, antwoordde een van de respondenten dat zij een erfenisje had gehad. Dit had zij niet bij de sociale dienst gemeld, hoewel dat volgens de regels wel had moeten. Zij verklaarde dit door te vermelden dat zij van het geld een aantal noodzakelijke zaken had aangeschaft waaronder een wasmachine. Als zij de erfenis had aangemeld dan had zij een korting op de uitkering tegemoet kunnen zien en in een later stadium moeten aankloppen voor een bijzondere uitkering om de wasmachine alsnog te kunnen kopen. Gezien de problemen van de sociale dienst met hun administratie had zij besloten dat zulke narigheden maar beter konden worden vermeden. Op de vragenlijst vulde zij naar waarheid 'ja' in.

De tweede groep bestond uit respondenten die weliswaar niet naar de regel van de wet hadden gehandeld, maar die dit zelf niet zo voelde. Een van de respondenten had een baan als conductrice afgewezen, omdat zij deze betrekking te gevaarlijk vond, gezien het feit dat ze alleen voor de opvoeding van haar kinderen stond. In de vragenlijst geeft ze toe dat ze passende arbeid heeft geweigerd, ondanks het feit dat zijzelf deze arbeid als niet passend heeft ervaren. Respondenten in deze groep geven in het interview aan alleen in een heel veilige onderzoeksomgeving bereid te zijn deze gevoelige vragen naar waarheid in te vullen.

Ten slotte bleef er een kleine groep van respondenten (4) over die aangaven het onderwerp zo gevoelig te vinden dat zij altijd het meest veilige antwoord zouden geven. Voor hen staat er teveel op het spel, namelijk hun inkomen.

Tenslotte kwamen we tot de conclusie dat de randomized response techniek zich goed leent voor toepassingen in een CASI omgeving, en in een bijzondere populatie.

Het is daarbij belangrijk om heel goed aan de respondenten duidelijk te maken hoe deze speciale vraagtechniek hun privacy optimaal garandeert. Hier moet veel aandacht aan worden gegeven, eventueel kan dit worden duidelijk gemaakt met behulp van een voorbeeld. Tenslotte is het belangrijk om, bij gebruik van de Geforceerde Response methode ook duidelijk te maken dat men eerlijk is door antwoord te geven volgens de regels van het spel, ook als dat vreemd en zelfs oneerlijk aanvoeld. Dit laatste voorkomt dat er een extra bron van fouten wordt toegevoegd, namelijk teveel 'nee' antwoorden als gevolg van vals spelen.

6. CONCLUSIE

Randomized response is een in 1965 ontwikkelde methode voor het onderzoek naar sociaal gevoelige onderwerpen en staat ook nu nog volop in de belangstelling. Doordat de privacy van de respondent gegarandeerd is, wordt de respondent in staat gesteld om eerlijk te antwoorden, ook als de vragen over sociaal onwenselijk of illegaal gedrag gaan. Meta-analyse toont aan dat wanneer een onderzoek over een sociaal gevoelig onderwerp gaat, er gemiddeld meer valide populatie schatters worden gevonden wanneer gebruik wordt gemaakt van randomized response technieken dan wanneer gebruik wordt gemaakt van meer conventionele 'directe vragen'-designs. Dit effect wordt groter als de sociale gevoeligheid van het onderzoek toeneemt. Door het gebruik van nieuwere designs zijn de randomized response technieken efficiënter geworden dan Warner's originele design, waardoor het gebruik van een randomized response techniek aantrekkelijker wordt voor toepassing in large-scale surveys.

Onderzoek naar computer gestuurd interviewen leert dat dit in het algemeen eerlijkere antwoorden oplevert dan het traditionele survey. Het toepassen van de randomized response techniek in een computer gestuurde interview setting verenigt de voordelen van randomized response onderzoek met die van computer gestuurd onderzoek en heeft daardoor de potentie om de meest valide en betrouwbare data op te leveren. Omdat wetenschap en beleid gediend zijn met gegevens van goede kwaliteit is het aan te bevelen om bij onderzoek naar sociaal gevoelige onderwerpen het gebruik van randomized response designs te overwegen.

LITERATUUR

Armacost, R. L., J. C. Hosseini, S. A. Morris en K. A. Rehbein (1991) "An empirical comparison of direct questioning, scenario, and randomised response methods for obtaining sensitive business". *Decision Science* 22: 1073-1087.

Baker, R. P., en N. M. Bradburn (1992) "CAPI: impacts on data quality and survey costs." *Public health conference on records and statistics*.

Begin, G., en M. Boivin (1980). "Comparrison of data gathered on sensitive questions via direct questionnaire, randomized response technique, and a projective method." *Psychological Reports* 47: 743-750.

Boeije, H. en Lensvelt-Mulders, G. (2002) iHonest by chance: A qualitative interview study to clarify respondent's (non)compliance with computer assisted randomized response. i Bulletin Methodologie Sociologique, 75, 24-39.

Boruch, R. F. (1971) "Assuring confidentiality of responses in social research: a note on strategies." *The American Sociologist* 6: 308-311.

Crawford, S. D., M. P. Couper, en M. J. Lamias (2001) "Web surveys: perception of a burden." *Social Science Computer Review* 19: 146-162.

De Leeuw, E., J. Hox, S. Kef, en M. v. Hattum. submitted. "Computer assisted selfinterviewing tailored for special populations; Overcoming the problems of special interviews and sensitive topics".

De Leeuw (1997) Non-response, wat weten we en wat kunnen we er aan doen? (Non-response: theories, facts and actions), In A. E. Bronner, P. Ester, A. J. Olivier, W. F. van Raaij, M. Wendel, en B. Wieringa (Eds.). *Recente ontwikkelingen in het marktonderzoek1997. Jaarboek van de Nederlandse vereniging voor marktonderzoek en information management* (pp. 39-59). Haarlem: de Vrieseborch.

- De Leeuw, E. D. (1999) Vragen naar gevoelige informatie; ene overzicht. *Pedagogische Studien*, 76, 2, 92-103.
- De Leeuw, E. D., J. J. Hox, en G. Snijders (1995) "The effect of computer assisted interviewing on data quality: a review." *Journal of the Market Research Society* 37: 325-344.
- De Leeuw, E. D. (2001). The effect of computer assisted interviewing on data quality: A review of the evidence (CD-rom). In J. Blasius, J. Hox, E. de Leeuw en P. Schmidt (Eds.), *Social science methodology in the new millennium*. opladen, FRG: Leske + Budrich 9edit is recentste CAI overzicht.
- Dillman, D. A. 1978. *Mail and telephone surveys*. Wiley, New York.
- Edgell, S. E., K. L. Duchan, en S. Himmelfarb (1992) "An empirical test of the unrelated question randomized response technique." *Bulletin of the Psychonomic Society* 30: 153-156.
- Ellfers, H., P. G. M. van der Heijden, en M. Hezewijk. opgestuurd. "Explaining regulatory non-compliance. A survey study of rule transgression for two Dutch instrumental laws, applying the randomized response method."
- Farberow, N. (1963) Introduction. in N. Farberow, ed. *Taboo topics*. Atherton Press, New York.
- Fligner, M. A., G. E. Policello, en J. Singh (1977). "A comparison of two randomized response survey methods for the level of respondent protection." *Communications in statistics: theory and methods* A6: 1511-1524.
- Greenberg, B. V., A. A. Abdul-Ela, W. R. Simmons, en D. G. Horvitz (1969) "The unrelated question randomized response model: Theoretical framework". *Journal of the American Statistical Association* 66: 243-250.
- Houtkoop-Steenstra, H., en H. van den Bergh (1994) Responseverhoging in telefonische interviews. *Ontwikkelingen in het Marktonderzoek. Jaarboek 1994, MarktOnderzoeksAssociatie*. De Vrieseborch, Haarlem.
- Houtkoop-Steenstra, H., en H. v. d. Bergh (2000). "Effects of the introductions in large-scale telephone survey interviews". *Sociological Methods and Research* 28: 281-300.
- Hox, J., en E. d. Leeuw (1994) "A comparison of non-response in mail, telephone and face to face surveys". *Quality and Quantity* 28: 329-344.
- Jobe, J. B., en D. J. Mingay (1991) "Cognition and survey measurement: History and overview". *Applied Cognitive Psychology* 5: 175-192.
- Justitie (1998) "Monitoring van beleidsinstrumentele wetgeving." Inspectie voor de rechtshandhaving, Den Haag.
- Landsheer, J. A., P. v. d. Heijden, en G. v. Gils (1999) "Trust and understanding, two psychological aspects of randomized response." *Quality and Quantity* 33: 1-12.
- Lee, R. M. 1993. *Doing research on sensitive topics*. Sage Publications, London.
- Lensvelt-Mulders, G. en H. Boeije, Submitted. Evaluating compliance with the Randomized Response Technique: A qualitative research into the backgrounds of lying and cheating.
- Lensvelt-Mulders, G. J. L. M., J. J. Hox en P. G. M. v. d. Heijden, submitted. "Meta-analysis of randomized response research: 35 year of validation".
- Maddala, G. S. (1983) *Limited dependent and qualitative variables in econometrics*. Cambridge University Press, Cambridge.
- Moors, J. J. A. (1971) "Optimization of the unrelated question randomized response model". *Journal of the American Statistical Association* 66: 627-629.
- Musch, J., A. Broder, en K. C. Klauer (2001) Improving survey research on the World-Wide Web using the randomized response technique in M. Bosnjak, ed. *Dimensions of internet science*. Pabst, Lengerich.
- Peiris, D. R., P. Gregor, en N. Alm (2000) "The effects of simulating human conversation style in a computer-based interview". *Interacting with Computers* 12: 635-650.
- Pondman, L. M. (1999). Overstag?! Reductie van weigeringspercentage in telefonisch onderzoek. In: A. E. Bronner et al. *Recente ontwikkelingen in het marktonderzoek*. (pp 25-34). Jaarboek van de Nederlandse vereniging voor marktonderzoek en informatiemanagement. Haarlem: de Vrieseborch.

- Rasinski, K. A., G. B. Willis, A. K. Baldwin, W. Yeh, en L. Lee (1999) "Methods of data collection, perception of risks and losses, and motivation to give truthful answers to sensitive survey questions". *Applied Cognitive Psychology*: 465-484.
- Scheers, N. J., en M. C. Dayton (1988) "Covariate randomized response models." *Journal of the American Statistical Association* 83: 969-974.
- Schmidt, W. C. (1997) "World Wide Web survey research: benefits, potential problems and solutions." *Behavioral Research Methodes, Instruments and Computers* 29: 274-279.
- Schwarz, N. (1999) "How the questions shape the answers." *American Psychologist* 54: 93-105.
- Singer, E., D. R. Von Thurn, en E. R. Miller (1995) "Confidentiality assurances and responses: A quantitative review of the experimental literature." *Public Opinion Quarterly* 59: 66-77.
- Snijckers, G. (2001) *Cognitive laboratory experiences*. CBS, Heerlen.
- Snijckers, G., en M. Luppens (2000) "The best of two worlds: total design method and new Kontiv design. An operational model to improve respondent co-operation." *Netherlands Official Statistics* 3: 4-10.
- Soeken, K. L., en G. B. Macready (1982) "Respondents' perceived protection when using randomized response". *Psychological Bulletin* 92: 487-498.
- Stoop, I. A. L. (2002) Wie doen niet mee? een analyse van nonresponse naar voorzieningengebruik. in A. Smidts, ed. *Ontwikkelingen in het Marktonderzoek. Jaarboek 2002, MarktOnderzoeksAssociatie*. De Vrieseborch, Haarlem.
- Sudman, S., en N. M. Bradburn (1982) *Asking questions: a practical guide to questionnaire design*. Jossey-Bass, San Fransico.
- Sykes, W., en J. Morton-Williams (1987) "Evaluating Survey Questions". *Journal of Official Statistics* 3: 191-207.
- Van der Heijden, P. G. M., en G. v. Gils (1996) Some logistic regression models for randomized response data. Pages 341-348 in G. Galmatti, ed. *Statistical modelling. Proceedings of the 11th international workshop on statistical modelling.*, Orvieto, Italy.
- Van der Heijden, P. G. M., Van Gils, G., Bouts, J., en Hox, J. J. (1998) "A comparison of randomized response, CASAQ, and direct questioning: eliciting sensitive information in the context of fraud". *Kwantitatieve Methoden* 19: 15-34.
- Van der Heijden, P. G. M., Van Gils, G., Bouts, J., en Hox, J.J. (2000) "A comparison of randomized response, computer-assisted self-interview, and face to face direct questioning." *Sociological Methods and Research* 28: 505-537.
- Van der Heijden, P. G. M., J. J. Hox, en P. Ellfers. in press. "Het meten van regelnaleving." Universiteit Utrecht, Utrecht.
- Van den Hout, A., en P. G. M. van der Heijden (2002a) "Loglinear analysis of randomized response data." *17th International Workshop on Statistical Modelling.*, Chania.
- Van den Hout, A., en P. G. M van der Heijden (2002b) "Randomized response, statistical disclosure control and misclassification: a review." *International Statistical Review*.
- Van Gils, G., P. G. M. v. d. Heijden, en A. Rosebeek (2001a) "Randomized Response: Onderzoek naar regelovertreeding resultaten ABW, WAO en WW ." Ministerie of Social Affairs and Labour, Den Haag.
- Warner, S. L (1965) "Randomized response: A survey technique for eliminating evasive answer bias". *Journal of the American Statistical Association* 60: 63-69.
- Warner, S. L. (1971) "The linear randomized response model". *Journal of the American Statistical Association* 66: 884-888.
- Weisband, S., en S. Kiesler (1996) "Self disclosure on computer forms: meta-analysis and implications". University of Arizona (E, Tucson AR.

5. *Wie het weet die moet dat zeggen*

Het aggregeren van informantgegevens in organisationeel marktonderzoek^{1,2}

G.H. VAN BRUGGEN, G.L. LILIEN en M. KACKER

SAMENVATTING

In marktonderzoek worden gegevens over organisaties vaak verzameld door middel van het ondervragen van zogenaamde informanten. In dit artikel betogen we dat het gebruik van meerdere informanten per organisatie in plaats van één enkele informant de kwaliteit van responsgegevens en daarmee de kwaliteit van onderzoek zal verhogen. Informanten kunnen echter van mening verschillen. Dit leidt tot de vraag hoe opinies geaggregeerd moeten worden. We beschrijven een aantal aggregatiemethoden die gebruik maken van de inschattingen van informanten van de nauwkeurigheid van hun respons en van informatie over de inschattingen van informanten relatief ten opzichte van andere informanten. Onze methoden zijn effectief en relatief gemakkelijk toepasbaar in de praktijk van organisationeel marktonderzoek.

1. INLEIDING

In veel marketing- en marktonderzoekstudies verzamelen onderzoekers gegevens over verschillende variabelen (zoals marketinginspanningen, verkopen en marktaandeel) om vervolgens na te gaan of er een verband tussen deze variabelen bestaat. In marktonderzoek onder organisaties, of organisationeel marktonderzoek, is de organisatie of een organisatieonderdeel het object van onderzoek. Als de juiste gegevens verkregen kunnen worden door gebruik te maken van archieven, bedrijfsadministraties, jaarverslagen etc. dan zullen onderzoekers dit doen. Vaak zullen gegevens echter niet gemakkelijk te verzamelen zijn of zijn bronnen binnen organisaties moeilijk toegankelijk (Kumar, Stern en Anderson 1993). Onderzoeksgegevens zullen dan verzameld moeten worden door gebruik te maken van informanten. Deze informanten geven hun inschatting over de waarde van organisationele variabelen. Informanten rapporteren dus hun eigen subjectieve, retrospectieve inschattingen met betrekking tot eigenschappen van de organi-

¹ Deze bijdrage is gebaseerd op het artikel "Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses" van Gerrit H. van Bruggen, Gary L. Lilien en Manish Kacker dat in November 2002 in het *Journal of Marketing Research* gepubliceerd is. Voor meer achtergronden van de in deze bijdrage beschreven methoden en een uitgebreide beschrijving van de empirische studie verwijzen wij naar dit artikel.

² De auteurs danken redacteur Ale Smidts voor zijn waardevolle commentaar op een eerdere versie van dit artikel.

satie (bijvoorbeeld verkopen en winstcijfers). Zij verschillen hierin van respondenten die informatie over zichzelf als individu verschaffen (bijvoorbeeld hun persoonlijke beoordeling van de kwaliteit van de organisatie).

Onderzoekers die informatie over organisationele variabelen verzamelen door gebruik te maken van informanten worden geconfronteerd met twee belangrijke vragen: 1) moet gebruik gemaakt worden van *één* of van *meer* (multiple) informanten; en indien meer dan één informant gebruikt wordt 2) hoe moeten de gegevens van meerdere informanten dan geaggregeerd worden?

Het is vaak gemakkelijk en bovendien goedkoop om per organisatie slechts één informant om informatie te vragen. Dit gebeurt dan ook vaak in de praktijk. Diverse studies laten echter zien dat het gebruik van meerdere informanten de kwaliteit van de verzamelde gegevens zal vergroten (Seidler 1974; Hogarth 1978; Hill 1982). Diverse onderzoekers bepleiten daarom het gebruik van meerdere informanten in organisationeel marktonderzoek. In deze bijdrage willen we ingaan op de vraag *waarom* het gebruik van gegevens van meerdere informanten de nauwkeurigheid van informatie en van gevonden relaties tussen variabelen vergroot en *hoe* bij gebruik van meerdere informanten de respons van deze informanten geaggregeerd moet worden.

2. HET GEBRUIK VAN INFORMANTEN: ÉÉN OF MEERDERE?

Objectieve organisationele variabelen zoals de verkopen en winst van een bedrijf worden gemeten op het niveau van de organisatie (of business unit). Deze variabelen hebben als kenmerk dat een juiste waarde (true score) voor de organisatie bestaat. De door informanten opgegeven waarde van variabelen zal echter kunnen afwijken van de juiste waarde doordat er sprake is van meetfouten (bijvoorbeeld omdat informanten zich bepaalde informatie niet goed meer herinneren). Deze meetfouten bevatten vaak een systematische - en een random component. Samenvattend:

$$\text{Respons van Informant} = \text{Juiste Waarde} + \text{Meetfout}$$

$$\text{Meetfout} = \text{Systematische Meetfout} + \text{Random Meetfout}$$

De random meetfout ontstaat omdat van informanten gevraagd wordt om als zogenaamde "key informant" inschattingen te maken over de waarde van organisatievariabelen. Het maken van een accurate inschatting van deze waarden is vaak een complexe zaak (Philips 1981) en informanten kunnen hierbij fouten maken. Random meetfouten kennen geen systematisch patroon. De verwachte waarde van de random meetfout is dan ook nul.

De systematische meetfout in de respons van informanten kan zowel individuele oorzaken als organisationele oorzaken hebben. Individuen kunnen systematisch een te hoge of te lage response voor variabelen geven, bijvoorbeeld vanwege sociale wenselijkheid. Een organisationele oorzaak die tot systematische meetfouten kan leiden kan bijvoorbeeld liggen in de afdeling in de organisatie waar de informant werkzaam is. Zo kun-

nen medewerkers van de marketingafdeling de marktgerichtheid van een organisatie systematisch hoger inschatten dan medewerkers van andere afdelingen.

De omvang van de foutencomponent kan substantieel zijn. Zo stelt Philips (1981) dat de informatie van informanten over de waarde van variabelen vaak een variantie kent die voor meer dan 50% is toe te schrijven aan systematische - en random meetfouten.

Onderzoekers zijn vaak geïnteresseerd in relaties tussen variabelen. De relaties die zij vinden zijn die tussen de gerapporteerde responsgegevens van de informanten. Omdat deze respons een foutencomponent bevat zullen de gerapporteerde correlaties niet alleen afhankelijk zijn van de correlaties tussen de werkelijke, juiste waarden van de variabelen, maar ook van de correlaties tussen de foutencomponenten en de omvang van deze foutencomponenten. Om er voor te zorgen dat de gevonden relatie tussen constructen zo goed mogelijk weergeeft wat de relatie tussen de werkelijke waarden van deze variabelen is, is het belangrijk dat de omvang van de foutencomponenten en de correlatie tussen deze foutencomponenten zo klein mogelijk is.

Het gebruik van meerdere informanten en het simpelweg middelen van de respons van deze verschillende informanten tot één enkele organisatiescore leidt tot een verkleining van de random meetfout in organisationele variabelen. De random meetfouten in de respons van individuele informanten worden zo tegen elkaar weggemiddeld. Hoe groter het aantal informanten, hoe kleiner de verwachte random meetfout zal worden. Het *optimale* aantal informanten hangt af van de kosten van het verkrijgen van additionele informanten ten opzichte van de kosten van de overblijvende meetfout. Onderzoek van Ashton en Ashton (1985) op het gebied van voorspellingen laat zien dat een aantal informanten dat tussen twee en vijf ligt effectief is. Libby en Blashfield (1978) stellen dat door het gebruik van drie informanten de meeste winst in nauwkeurigheid kan worden gerealiseerd.

Naast een reductie van de random meetfout is het ook belangrijk om de invloed van de systematische meetfout te minimaliseren. Ook hier geldt dat het gebruik van meerdere informanten een effectief middel is. In tegenstelling tot verwachte waarde van de random meetfout zal de verwachte waarde van de systematische meetfout in de respons van individuele informanten niet gelijk aan nul zijn. Bovendien zal de systematische fout voor de verschillende variabelen, waarvoor een informant informatie verschaft, vergelijkbaar zijn. Een informant die systematisch overschat zal voor alle variabelen een te hoge respons geven. Dit leidt tot een positieve correlatie tussen de foutencomponenten in deze verschillende variabelen. De verwachte waarde van de correlatie tussen systematische meetfouten neemt dus toe als responsgegevens voor variabelen van dezelfde informant afkomstig zijn. Dit zal de gevonden correlatie tussen de variabelen die bestudeerd worden beïnvloeden.

Een simpele manier om de correlatie tussen systematische fouten te reduceren is het gebruiken van verschillende informanten per organisatie voor verschillende variabelen. Zo kan bijvoorbeeld een eerste informant in een organisatie gevraagd worden om een inschatting van de marketinginspanningen terwijl een tweede informant gegevens over marketingresultaten verschaft. De respons van verschillende informanten moet hierbij zo onafhankelijk mogelijk van elkaar zijn. Door het gebruik van meerdere onafhankelijke informanten zal de verwachte waarde van de correlatie tussen systematische foutencomponenten afnemen, waardoor de gevonden correlatie dichterbij de werkelijke of

“true” correlatie ligt. Het gebruik van meerdere informanten maakt het bovendien mogelijk om per construct de meest deskundige informant te selecteren. Ook dit zal mogelijke meetfouten verkleinen. Het is immers voor informanten lastig om over alle variabelen de meest betrouwbare informatie te verstrekken.

Samenvattend stellen we dat in organisationeel marktonderzoek het gebruik van meerdere informanten te prefereren is boven het gebruik van één enkele informant per organisatie. Verschillende informanten kunnen over verschillende variabelen rapporteren. Ze kunnen echter ook over dezelfde variabelen rapporteren. Hierbij zullen ze echter vaak verschillende opinies hebben. Dit leidt tot de vraag hoe met de respons van de verschillende informanten moet worden omgegaan om tot één enkele organisatiewaarde voor variabelen te komen.

3. HET AGGREGEREN VAN INFORMANTENRESPONS

Verschillende informanten die afkomstig zijn uit één en dezelfde organisatie zullen vaak verschillende meningen hebben. Informanten zullen vaak ook verschillen in de nauwkeurigheid van hun respons. Het is daarom belangrijk om na te denken over de vraag hoe de verschillende opinies geaggregeerd moeten worden. Aggregatie is nodig om één enkele organisatiewaarde te verkrijgen. In de literatuur worden zowel behavioraal – als mathematische aggregatiemethoden beschreven.

Volgens de behavioral aggregatiemethode zullen de verschillende informanten die over dezelfde organisatie of over hetzelfde organisatieonderdeel rapporteren, discussiëren over de waarde van een variabele waarover zij van mening verschillen. Zij proberen het verschil van mening weg te werken en het zo eens te worden over één groepswaarde (Ferrel 1985). De behavioral aggregatiemethode biedt zo een directe oplossing voor het aggregatieprobleem. De methode vraagt echter substantiële inspanningen en coördinatie van de informanten. Verder maakt ze het lastig voor informanten om vertrouwelijk en anoniem informatie te geven. Tevens kan de resulterende groepswaarde mogelijk beïnvloed worden door groepsprocessen en machtsverhoudingen. Tenslotte kan het zijn dat de in het overleg meest invloedrijke groepsleden niet de meest accurate zijn (Lar-réché en Moinpour 1983). Al deze factoren kunnen de praktische toepasbaarheid en de nauwkeurigheid van de behavioral aggregatiemethode negatief beïnvloeden.

Omdat de van de informanten gevraagde inspanningen de behavioral aggregatiemethode in de praktijk moeilijk uitvoerbaar maken ontwikkelden Kumar, Stern en Anderson (1993) de zogenaamde hybride consensus-aggregatiemethode. Volgens deze methode wordt voor de variabelen waarover geen grote meningsverschillen tussen de informanten bestaan simpelweg het rekenkundig gemiddelde berekend. Voor variabelen waarvoor grotere verschillen bestaan zullen de leden van een groep door middel van discussie overeenstemming proberen te realiseren. Empirisch onderzoek van Kumar, Stern en Anderson (1993) laat echter zien dat het resultaat van discussies tussen groepsleden vaak dichter ligt bij de mening van informanten met een hiërarchisch gezien hogere positie. Een alternatieve aggregatiemethode die minder last heeft van groepsprocessen is de Delphi methode. Deze methode is echter kostbaar vanwege de hoeveelheid tijd en inspanningen die zij vraagt van de informanten.

Mathematische aggregatiemethoden bieden een alternatief voor de behavioral methoden. Een veelgebruikte mathematische aggregatiemethode is het simpelweg berekenen van het rekenkundig gemiddelde over de verschillende meningen van verschillende informanten. Als informanten echter substantieel van mening verschillen dan zal deze methode zelden tot de meest nauwkeurige groepswaarde leiden (James 1982). Hill (1982) laat zien dat de gemiddelde groepswaarde vaak wel nauwkeuriger is dan de respons van individuele groepsleden maar dat door meer geavanceerde statistische aggregatiemethoden vaak nog nauwkeuriger resultaten verkregen kunnen worden. Zogenaamde Bayesiaanse methoden (Morris 1977) integreren de opinies van verschillende informanten op elegante en theoretische geavanceerde wijze. Ze zijn echter niet altijd even gemakkelijk toepasbaar in praktijksituaties omdat ze veel inspanningen vragen van zowel informanten, die gecompliceerde kansverdelingen moeten inschatten (Clemen en Winkler 1993), als van onderzoekers.

De inspanningen die bij zowel de behavioral als bij de Bayesiaanse aggregatiemethoden van informanten gevraagd wordt hebben ertoe geleid dat ze in de praktijk slechts beperkt worden toegepast. Het berekenen van het rekenkundig gemiddelde is eenvoudig maar zal alleen effectief zijn als er geen systematische foutencomponent in de respons van informanten aanwezig is. Hiervan zal in organisationeel onderzoek zelden sprake zijn.

Als er sprake is van systematische meetfouten in de respons van individuele informanten, dan is het belangrijk om vast te stellen wat de oorzaak van deze fouten is en is het nodig per organisatie(onderdeel) die informant te identificeren waarvan de systematische responsfout het kleinst is. Als met zekerheid vastgesteld kan worden welke informant de kleinste systematische fout in zijn/haar respons kent, dan moet de respons van die informant gebruikt worden als groepswaarde. Als de nauwkeurigheid van de groepsleden niet met zekerheid kan worden vastgesteld, dan moet een gewogen gemiddelde bepaald worden, waarbij de informanten waarvan verwacht wordt dat die meer nauwkeurige informatie verschaffen een groter gewicht krijgen dan de informanten die mogelijk minder nauwkeurig zijn.

Het is nu de vraag hoe we vast kunnen stellen wat de nauwkeurigheid van een individuele informant is en hoe we deze informatie vervolgens kunnen gebruiken in een aggregatieprocedure.

4. METHODEN VOOR DE AGGREGATIE VAN INFORMANTEN-RESPONS

In deze paragraaf introduceren we drie methoden voor het aggregeren van gegevens van informanten. Dit zijn: (1) het ongewogen rekenkundig groepsgemiddelde (dit gemiddelde geldt als referentiewaarde); (2) een gewogen groepsgemiddelde waarbij de gewichten worden afgeleid uit de door de informanten verstrekte gegevens; en (3) een gewogen groepsgemiddelde waarbij de gewichten worden afgeleid uit door de informanten verstrekte informatie met betrekking tot hun eigen vertrouwen in de door hen verstrekte gegevens.

1. Het ongewogen groepsgemiddelde (UNWMEAN)

Volgens onze eerste methode berekenen we een groepswaarde voor een variabele door

het rekenkundig gemiddelde van de individuele scores van groepsleden te bepalen. De waarde van het ongewogen groeps-gemiddelde voor organisatievariabele X (in de empirische toepassing die later zal worden besproken analyseren we acht organisatievariabelen, dus X = 1,..., 8), UNWMEAN_{xi} van groep i, wordt als volgt berekend:

$$\text{UNWMEAN}_{xi} = \frac{\left(\sum_{j=1}^{n_i} X_{ij} \right)}{n_i} \quad (1)$$

Hierbij geldt

X_{ij} = de waarde van variabele X zoals gegeven door informant j in groep i
n_i = aantal informanten in groep i

Stel dat drie informanten rapporteren over het marktaandeel van een bedrijf. Informant 1 schat dit aandeel in op 10%, informant 2 op 20% en informant 3 op 21%. Het ongewogen groeps-gemiddelde voor dit bedrijf is dan 17% ((10+20+21)/3).

2. Het met responsgegevens van informanten gewogen groeps-gemiddelde (WDMEAN)

De tweede methode voor het aggregeren van gegevens van informanten gaat er vanuit dat de *mate van overeenstemming* tussen informanten uit een groep kan worden gebruikt bij het bepalen van een groeps-waarde. De afwijking van een individueel groeps-lid van de opinie van de overige groeps-leden zegt mogelijk iets over de nauwkeurigheid van dit lid. Stel, bijvoorbeeld, dat twee informanten uit een groep van drie leden dezelfde inschattingen voor de waarde van een variabele geven, terwijl een derde groeps-lid een substantieel afwijkende waarde geeft. In een dergelijke situatie wordt de informatie van de twee informanten die het met elkaar eens zijn zwaarder gewogen dan de informatie van de informant die een afwijkende mening heeft. Deze aanpak gaat er dus vanuit dat de juiste waarde in het algemeen dichter bij de inschattingen van de meerderheid in de groep ligt dan bij de inschatting van groeps-leden met een afwijkende mening.

Voor de berekening van WDMEAN_{xi} berekenen we eerst de gewichten voor ieder van de informanten binnen de groep. Dit doen we door eerst DIST_{xij} te berekenen. Deze variabele geeft de absolute afstand tussen informant j's inschatting van de waarde van variabele X en het ongewogen, rekenkundig gemiddelde van groep i waartoe informant j behoort. DIST_{xij} wordt als volgt berekend:

$$\text{DIST}_{xij} = \left| X_{ij} - \text{UNWMEAN}_{xi} \right| \quad (2)$$

Voor het eerder geïntroduceerde getallenvoorbeeld zal de waarde van DIST voor de eerste informant 7 (=|10-17|) bedragen. Voor informant 2 is deze waarde 3 (=|20-17|) en voor de derde informant is deze 4 (=|21-17|).

Het gewicht (WEIGHT_{xij}) dat wordt toegekend aan de inschatting van informant j van de waarde van variabele X is omgekeerd evenredig aan de absolute afstand van deze inschatting tot het ongewogen groeps-gemiddelde voor groep j, ten opzichte van deze afstand voor de andere informanten uit groep i. We berekenen het gewicht van informant j's inschatting voor variabele X (WEIGHT_{xij}) als volgt:

$$\text{WEIGHT}_{xij} = \left(\frac{\left(\sum_{j=1}^{n_i} \text{DIST}_{xij} \right)}{\text{DIST}_{xij}} \right)^{\alpha} \quad (3)$$

In formule (3) introduceren we parameter α die de defaultwaarde 1 kent. Bij deze waarde van 1 is het gewicht van een informant omgekeerd evenredig aan zijn/haar relatieve afstand tot het ongewogen groepsgemiddelde. De waarde van α kan worden verhoogd wanneer we de gewichten van informanten met een inschatting die dicht bij het ongewogen groepsgemiddelde ligt, ten opzichte van informanten met een afwijkende mening, willen verhogen. Als α de waarde 0 heeft dan wegen de inschattingen van de verschillende groepsleden even zwaar (dit is dus gelijk aan het ongewogen groepsgemiddelde).

Uitgaande van de defaultwaarde 1 voor parameter α wordt de waarde van WEIGHT voor de eerste respondent uit het getallenvoorbeeld 2 $(=(7+3+4)/7)$. Voor informant 2 wordt deze waarde 4.67 $(=(7+3+4)/3)$ en voor de derde informant 3.5 $(=(7+3+4)/4)$.

We berekenen nu het op responsgegevens van informanten gebaseerde gewogen groepsgemiddelde (WDMEAN_{xi}) voor variabele X voor groep i. De inschatting van ieder individueel groepslid wordt gewogen op basis van zijn relatieve afstand tot het ongewogen groepsgemiddelde. WDMEAN_{xi} wordt als volgt berekend:

$$\text{WDMEAN}_{xi} = \sum_{j=1}^{n_i} \left(\frac{\text{WEIGHT}_{xij}}{\left(\sum_{j=1}^{n_i} \text{WEIGHT}_{xij} \right)} * X_{ij} \right) \quad (4)$$

Volgens deze aggregatiemethode wordt het gewogen groepsgemiddelde berekend als:

$$(2/(2+4.67+3.5))*10 + (4.67/(2+4.67+3.5))*20 + (3.5/(2+4.67+3.5))*21 = 18.38.$$

Te zien is dat het op deze wijze berekende groepsgemiddelde dicht bij de waarde van de informanten 2 en 3 (die het met elkaar eens waren) ligt dan het ongewogen groepsgemiddelde. De waarde ligt verder af van informant 1 (die een afwijkende mening had).

3. Het met vertrouwen van informanten gewogen groepsgemiddelde (WCMEAN)

De derde aggregatiemethode gebruikt het vertrouwen van informanten in de nauwkeurigheid van hun inschatting van de waarde van variabele X. De inschattingen van informanten met meer vertrouwen worden hierbij zwaarder gewogen dan de inschattingen van informanten die minder zeker zijn. WCMEAN_{xi} , de waarde voor variabele X voor groep i waar informant j's inschatting wordt gewogen met zijn/haar vertrouwen CONF_{xij} in de nauwkeurigheid van deze inschatting, wordt berekend als:

$$WCMEAN_{xi} = \sum_{j=1}^{n_i} \left(\frac{CONF_{xij}^{\alpha}}{\left(\sum_{j=1}^{n_i} CONF_{xij}^{\alpha} \right)} * X_{ij} \right) \quad (5)$$

In formule (5) maken we opnieuw gebruik van parameter α die het mogelijk maakt om de gewichten die worden toegekend aan informanten met minder of meer vertrouwen in hun inschattingen te manipuleren. Parameter α heeft ook hier de defaultwaarde 1. Als deze waarde verhoogd wordt dan wordt het gewicht van informanten met relatief veel vertrouwen in de eigen inschattingen groter. Als de waarde van α gelijk aan 0 wordt gesteld dan leidt dit tot de berekening van het ongewogen groepsgemiddelde.

Ook deze derde aggregatiemethode kunnen we illustreren aan de hand van het eerder gebruikte getalenvoorbeeld. Stel dat informant 1 op een 9-puntsschaal een vertrouwensscore 7 had (in de inschatting 10), informant 2 een score van 8 en informant 3 een score 6. Het met vertrouwen gewogen groepsgemiddelde is dan: $(7/(7+8+6))*10 + (8/(7+8+6))*20 + (6/(7+8+6))*21 = 16.95$.

Na de beschrijving van de drie aggregatiemethoden bespreken we nu een empirische studie waarin de kwaliteit van de aggregatiemethoden wordt geanalyseerd.

5. HET AGGREGEREN VAN “RECALL” GEGEVENS

5.1. Opzet van de studie

Om de kwaliteit van onze aggregatiemethoden te bepalen was het noodzakelijk om gegevens van informanten te verzamelen in een realistische organisatieomgeving. Hierbij moest het mogelijk zijn om de inschattingen van de informanten te vergelijken met een objectieve, correcte (juiste) waarde van de variabele. De MARKSTRAT simulatie (Larréché en Gatignon 1990) is een omgeving die deze mogelijkheden biedt. MARKSTRAT is in het verleden vaak gebruikt voor het bestuderen van besluitvorming in marketing (zie bijvoorbeeld Glazer, Steckel en Winer 1992). In de MARKSTRAT simulatie dienen deelnemers in groepsverband gedurende een aantal tijdsperiodes beslissingen te nemen over zowel strategische - als tactische marketingvraagstukken voor verschillende elkaar beconcurrerende ondernemingen.

De informanten in onze eerste empirische studie waren 67 marketing studenten die in het kader van hun opleiding deelnamen aan een marketingstrategie cursus aan een universiteit in de Verenigde Staten. De studenten vormden twee, drie of vier-persoons groepen die beslissingen moesten nemen voor één van de vijf bedrijven die actief waren in vier MARKSTRAT bedrijfstakken. In totaal opereerden de 67 informanten in 20 groepen. De groepen namen iedere periode beslissingen na bestudering van de resultaten van voorgaande periodes en van marktonderzoeksrapporten. Alle groepen met deelnemers ontvingen dezelfde informatie en hadden dezelfde hoeveelheid tijd om beslissingen te nemen. De beslissingen werden gelijktijdig door alle groepen genomen. Nadat de groepen een aantal ronden aan de MARKSTRAT simulatie hadden deelgeno-

men werd ieder van de deelnemers gevraagd om *individueel* een vragenlijst in te vullen. Onder deze vragen bevonden zich een achttal vragen die de informanten vroegen om de waarden van MARKSTRAT variabelen te reproduceren (recall data). Het ging hier om de waarden van marketing mix variabelen zoals reclame-uitgaven, prijzen en aantallen vertegenwoordigers waarvoor de groepen beslissingen hadden genomen als ook voor het marketingbudget dat de groepen beschikbaar hadden voor toekomstig beleid. Hiernaast werd de informanten gevraagd om per variabele hun *vertrouwen* in hun eigen inschatting te geven op een 9-puntsschaal. Hier betekende een 1 dat de informanten “absoluut niet zeker” waren van hun inschatting terwijl een 9 een “absolute zekerheid” in de nauwkeurigheid van de inschatting betekende. Om de betrokkenheid van de informanten bij de studie te waarborgen, om hun nauwkeurigheid te stimuleren en om samenwerking tussen de leden van een groep te ontmoedigen, ontvingen de informanten die de meest nauwkeurige inschattingen wisten te geven een beloning. Tenslotte is het belangrijk om twee kenmerken van onze onderzoeksopzet te noemen. Ten eerste waren alle deelnemers studenten die geen hiërarchisch of functioneel verschillende posities in hun MARKSTRAT groep innamen. De informanten hadden dus hetzelfde (homogene) perspectief, wat een hiërarchische of functionele bias onwaarschijnlijk maakt. Ten tweede produceert het MARKSTRAT computerprogramma de werkelijke (juiste) waarden voor alle variabelen waar ook de informanten gegevens over verstrekken. Dit maakte het mogelijk om objectief de nauwkeurigheid van de inschattingen van de informanten vast te stellen.

5.2 Resultaten

De aggregatiemethoden, zoals eerder besproken, werden toegepast op de gegevens zoals de informanten in de verschillende MARKSTRAT groepen produceerden. Op deze manier werd voor ieder van de twintig MARKSTRAT groepen een waarde voor de acht variabelen bepaald. Om de nauwkeurigheid van de verschillende methoden te bepalen, berekenden we per methode hoeveel procent de geaggregeerde waarde gemiddeld van de werkelijke waarde afweek (*Mean Absolute Percentage Error (MAPE)*). De MAPE voor groep *i* voor een specifieke aggregatiemethode (gemiddeld over de acht MARKSTRAT variabelen) wordt als volgt berekend:

$$MAPE_i = \left(\sum_{k=1}^8 \left| \frac{\text{Estimated Value of } X_{ki} - \text{Real Value of } X_{ki}}{\text{Real Value of } X_{ki}} \right| * 100\% \right) / 8 \quad (6)$$

In Tabel 1 worden de resultaten voor de verschillende aggregatiemethoden weergegeven.

Tabel 1. Nauwkeurigheid van de drie aggregatiemethoden (standaard afwijking tussen haakjes).

Aggregatie Methode	MAPE
1. ongewogen groeps-gemiddelde	16.30% (10.08)
2. met informanten respons gewogen groeps-gemiddelde	14.20% (10.22)
3. met informanten vertrouwen gewogen groeps-gemiddelde	12.81% (8.37)

De resultaten in Tabel 1 laten zien dat aggregatiemethoden die gebruik maken van gewichten tot een nauwkeuriger bepaling van een groepswaarde leiden ($F=15.45$, $p=.001$). De waarden van MAPE zijn voor methode 2 en 3 kleiner dan voor methode 1. Verder is te zien dat het gebruik van gewichten gebaseerd op het vertrouwen van individuele informanten in hun eigen inschattingen tot nauwkeuriger resultaten leidt dan het gebruik van op de respons gebaseerde gewichten. Het gebruik van op vertrouwen gebaseerde gewichten leidt tot een ruim 20% hogere nauwkeurigheid van de resultaten ($F=20.88$, $p<.001$) vergeleken met het ongewogen groeps-gemiddelde (12.81% vs. 16.30%).

De resultaten in Tabel 1 zijn gebaseerd op het gebruik van een standaardwaarde 1 voor parameter α in formules (3) en (5). Het is nu de vraag of de resultaten van de aggregatiemethoden verbeteren als de waarde van α verandert. Voor de aggregatiemethode 2 (op informanten respons gebaseerd) berekenen we daarom de waarde van α in formule (3) die de waarde van MAPE (Formule 6) minimaliseert. Hiervoor maken we gebruik van de Solver module uit Microsoft Excel. We berekenen een optimale waarde voor α voor ieder van de acht MARKSTRAT variabelen (we noemen dit de variabele-specifieke α). Hiernaast berekenen we ook een optimale α die een gelijke waarde kent voor alle acht MARKSTRAT variabelen (de uniforme α). In Tabel 2 presenteren we de resultaten van deze aanpak.

Tabel 2. Nauwkeurigheid (MAPE) van de aggregatiemethoden bij het gebruik van verschillende α waarden (standaardafwijkingen tussen haakjes).

Aggregatie Methode	$\alpha = 1$ default waarde)	uniforme α (geoptimaliseerd over de acht MARKSTRAT variabelen)	variabele specifieke α (geoptimaliseerd per MARKSTRAT variabele)
met informant respons gewogen groepswaarde	14.20% (10.22)	12.45% (10.76) ($\alpha = 25.70$)	12.35% (10.71) (α range = 2.88–77.00)
met informant vertrouwen gewogen groepswaarde	12.81% (8.37)	7.90% (6.24) ($\alpha = 12.87$)	7.53% (6.07) (α range = 5.46–25.80)

De resultaten in Tabel 2 laten zien dat de nauwkeurigheid van de aggregatiemethode die gebruik maakt van de op informanten respons gebaseerde weging met ongeveer 15% (12.45 vs. 14.20) verbeterd kan worden door optimalisatie van de α waarde ($F=9.11$, $p=.007$). De optimale uniforme α waarde is 25.70, terwijl de optimale waarde voor de variabele specifieke α waarde varieert tussen 2.88 en 77.00. Het verschil tussen de MAPE bij het gebruik van variabele specifieke α waarden en die bij een uniforme α waarde is klein (12.35 vs.12.45). Kennelijk komt de meeste winst in nauwkeurigheid uit het *vergroten van het gewicht van de informanten die overeenstemming vertonen*. Dit gebeurt door een waarde van α te gebruiken die substantieel boven 1 ligt. Hierbij is de nauwkeurigheid van de aggregatiemethode minder gevoelig voor de feitelijke waarde van de α .

De α waarden voor de op informant vertrouwen gebaseerde methoden kunnen op dezelfde manier geoptimaliseerd worden. We berekenen dan een waarde van α in Formule 5 die de waarde van MAPE in Formule 6 minimaliseert. Hierbij bepalen we zowel een optimale uniforme α waarde als een variabele specifieke optimale α waarde. Ook hier laten de resultaten zien dat het vergroten van de gewichten van de informanten met vertrouwen in hun eigen inschattingen tot een substantiële verbetering van de nauwkeurigheid van de aggregatiemethoden leidt. Als we α de waarde 1 geven dan bedraagt de MAPE 12.81. De berekening van een uniforme optimale α waarde leidt tot een daling van de MAPE tot 7.90 terwijl het gebruik van variabele specifieke α waarden leidt tot een MAPE van 7.53. Dit betekent dat door gewogen aggregatie een verbetering in de nauwkeurigheid van de groepswaarden van meer dan 50% verkregen kan worden (7.90 vs. 16.30). Ook hier geldt dat het berekenen van optimale waarden van α per variabele tot nauwelijks betere resultaten leidt dan het gebruik van een uniforme optimale α waarde (7.53 vs. 7.90).

Om onze methoden nu succesvol toe te kunnen passen is het uiteraard noodzakelijk om de waarde van de optimale α te weten. Deze kan bepaald worden door informanten te verzoeken om een aantal vragen te beantwoorden die gerelateerd zijn aan het onderzoeksonderwerp en waarvoor een objectief correct antwoord al bekend is. Deze vragen dienen dezelfde complexiteit te kennen als de uiteindelijke vragen in de vragenlijst. Met behulp van deze informatie kan een optimale waarde van α worden berekend. Deze optimale α waarde kan vervolgens worden gebruikt voor het aggregeren van variabelen waarvoor de objectieve, correcte waarde niet bekend is en waarvoor onderzoekers door middel van de informantgegevens een zo nauwkeurig mogelijke groeps-waarde willen bepalen.

Op basis van de resultaten in Tabel 1 en Tabel 2 kunnen we concluderen dat het gebruiken van een wegingprocedure tot substantieel nauwkeuriger aggregatieresultaten leidt dan wanneer we simpelweg het rekenkundig gemiddelde berekenen. Verder blijkt het gebruik van gewichten die worden gebaseerd op de inschattingen van informanten van hun eigen nauwkeurigheid effectiever te zijn dan het gebruik van gewichten die worden gebaseerd op de respons van informanten. Hierbij is het niet eens noodzakelijk dat informanten voor iedere variabele waarover zij rapporteren aangeven hoeveel vertrouwen zij in de nauwkeurigheid van hun respons hebben. Een beperkter aantal metingen van het "overall" vertrouwen voor een groep vragen leidt ook tot een beperkt verlies in nauwkeurigheid terwijl het beroep dat op informanten gedaan wordt reduceert. Dit laatste leidt mogelijk tot een verhoging van de respons en van de kwaliteit van de verzamelde gegevens.

Tenslotte liet een vervolgstudie zien dat de gebruikte methoden ook voor de aggregatie van voorspellingen van informanten tot goede resultaten leidt. Naast vertrouwenscores gebruikten we hier ook de competentie van informanten als gewichten in onze weging-procedures. Het gebruik van competentiescores als gewichten leidt tot vergelijkbare resultaten zolang het de competentie op een taak betreft die vergelijkbaar is met de taak waarvoor geaggregeerd wordt.

6. CONCLUSIE

In marktonderzoek naar organisatiegegevens zullen informanten een belangrijke bron van informatie zijn. De informatie die door informanten verstrekt wordt zal vaak niet honderd procent nauwkeurig zijn en dus een foutencomponent bevatten. Het gebruik van meerdere informanten (in plaats van één enkele informant) zal daarom de kwaliteit van responsgegevens verhogen en de validiteit van gevonden relaties tussen variabelen marktonderzoek vergroten. Omdat verschillende informanten vaak verschillende meningen zullen hebben over de waarde van organisatievariabelen is het de vraag hoe uit deze verschillende meningen één score voor een organisatie(onderdeel) moet worden afgeleid. De nadruk in dit artikel ligt op de wijze waarop responsgegevens van meerdere informanten geaggregeerd moeten worden om een zo nauwkeurige waarde voor een organisatievariabele te krijgen.

In onze onderzoekscontext leidt het berekenen van een simpel (ongewogen) rekenkundig gemiddelde al tot nauwkeuriger waarden voor organisatievariabelen (vergeleken met de inschattingen van individuele informanten) doordat de invloed van random fouten (error) zoals die aanwezig is in de informatie van individuele informanten afneemt naarmate het aantal informanten toeneemt. Het ongewogen rekenkundig gemiddelde is echter substantieel minder nauwkeurig dan gewogen gemiddelden waarin het vertrouwen van de informant in de eigen informatie of zijn/haar nauwkeurigheid met betrekking tot het verstrekken van gelijksoortige informatie, of responsgegevens als gewichten worden gebruikt. Het gebruik van deze gewichten is effectief omdat ze informatie over systematische fouten in de respons van informanten bevatten.

In tegenstelling tot onderzoek dat laat zien dat het wegen van de respons van individuen niet nodig is (zie bijvoorbeeld Armstrong 1986; Einhorn en Hogarth 1975) tonen onze resultaten dus aan dat weging van de informatie van verschillende informanten wel degelijk effectief is. Een mogelijke verklaring voor de effectiviteit van onze gewichten kan liggen in het type gewichten dat wij gebruiken. In onze aanpak weegt de informatie van meer competente informanten en informanten met meer vertrouwen zwaarder dan informatie van minder competente informanten en informanten met minder vertrouwen. Als gewichten op basis van competentie worden berekend dan is het van belang dat de competentiegewichten op basis van vergelijkbare taken wordt berekend. Competentie en expertise van informanten blijken domein-specifiek te zijn (Ashton en Ashton 1985). Hoewel het voor informanten moeilijk kan zijn om hun eigen nauwkeurigheid in te schatten en hoewel deze inschattingen zelf systematische afwijkingen kunnen kennen (Alba en Hutchinson 2000) blijken zij toch nog steeds nuttige informatie te bevatten. De aggregatiemethoden die wij in dit artikel presenteren hebben het grote voordeel dat zij zowel eenvoudig bruikbaar als effectief zijn. Onze resultaten bevestigen de bevinding van Clemen (1989) dat relatief eenvoudige aggregatiemethoden vaak opvallend effectief zijn als ze vergeleken worden met meer complexe methoden.

Voor marktonderzoekers die onderzoek onder organisaties doen hebben de resultaten van ons onderzoek een aantal *implicaties*. Allereerst is het voor de nauwkeurigheid van de verzamelde gegevens belangrijk om per organisatie niet één maar het drie informanten te ondervragen. Als vervolgens gegevens van meerdere informanten verzameld zijn dan moet niet simpelweg een gewogen gemiddelde bepaald worden om zo een organisatiescore te verkrijgen. Het is beter om een gewogen gemiddelde te berekenen waarbij

of de competentie van informanten of hun vertrouwen in hun eigen inschattingen als gewichten gebruikt wordt. Om onze methode toe te kunnen passen moeten onderzoekers informanten vragen om (1) een inschatting van de kwaliteit van hun respons te geven (bijvoorbeeld op een schaal van 1 tot 9 en (2) een aantal vragen te beantwoorden die gerelateerd zijn aan het onderzoeksonderwerp maar waarvoor een objectief correct antwoord al bekend is. Met behulp van deze laatste informatie kan de competentie van de informant worden vastgesteld. Aan de hand van de antwoorden op deze twee vragen kan een optimale waarde van α worden berekend. Deze optimale α waarde kan vervolgens worden gebruikt voor het aggregeren van variabelen waarvoor de objectieve, correcte waarde niet bekend is en waarvoor onderzoekers door middel van de informantgegevens een zo goed mogelijke waarde willen bepalen.

Al met al denken wij dat wij een aantal effectieve en gemakkelijk toepasbare aggregatiemethoden ontwikkeld hebben die het mogelijk maken om op een correctie manier informatie van meerdere informanten te aggregeren. De gepresenteerde aggregatiemethoden blijken effectief in een gesimuleerde omgeving. Het is belangrijk om ons onderzoek te repliceren in realistische praktijksituaties om zo vast te stellen hoe effectief onze methoden in dergelijke situaties zijn. Verder is het interessant en verdient het aanbeveling de waarde van onze methoden te bepalen in het geval van meer complexe variabelen (zoals de marktgerichtheid of machtspositie van organisaties) dan het soort waarover wij in ons onderzoek gegevens hebben verzameld.

LITERATUUR

- Alba, Joseph W. en J. Wesley Hutchinson, 2000. "Knowledge Calibration: What Consumers Know and What They Think They Know," *Journal of Consumer Research*, 27, September, 123-156.
- Armstrong, J. Scott, 1986. "The Ombudsman: Research on Forecasting: A Quarter-Century Review, 1960-1984," *Interfaces*, 16:1, 89-109.
- Ashton, Alison Hubbard en Robert H. Ashton, 1985. "Aggregating Subjective Forecasts: Some Empirical Results," *Management Science*, 31, 12, 1499-1508.
- Bruggen, Gerrit H. van, Gary L. Lilien en Manish Kacker, 2002. "Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses," *Journal of Marketing Research*, November, Forthcoming.
- Clemen, Robert T., 1989. "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting*, 559-583.
- Clemen, Robert T. en Robert L. Winkler, 1993. "Aggregating Point Estimates: A Flexible Modeling Approach," *Management Science*, 39, 4, 501-515.
- Einhorn, Hillel J. en Robin M. Hogarth, 1975. "Unit Weighting Schemes for Decision Making," *Organizational Behavior and Human Performance*, 13, 171-192.
- Ferrell, William R., 1985. "Combining Individual Judgments," in: G. Wright (ed.), *Behavioural Decision Making*, New York: Plenum Press, 111-145.
- Glazer, Rashi, Joel Steckel en Russell Winer, 1992. "Locally Rational Decision Making: The Distracting Effect of Information on Managerial Performance," *Management Science*, Vol. 38, 212-226.
- Hill, Gayle W., 1982. "Group Versus Individual Performance: Are N+1 Heads Better Than One?," *Psychological Bulletin*, 91, 3, 517-539.
- Hogarth, Robin M., 1978. "A Note on Aggregating Opinions," *Organizational Behavior and Human Performance*, 21, 40-46.

- James, Lawrence R., 1982. "Aggregation Bias in Estimates of Perceptual Agreement," *Journal of Applied Psychology*, Vol. 67, No. 2, 219-229.
- Kumar, Nirmalya, Louis W. Stern en James C. Anderson, 1993. "Conducting Interorganizational Research Using Key Informants," *Academy of Management Journal*, 36, 6, 1633-1651.
- Larréché, Jean Claude en Hubert Gatignon, 1990. *MARKSTRAT2: A Marketing Simulation Game*. Palo Alto, CA: The Scientific Press.
- Larréché, Jean Claude en Reza Moinpour, 1983. "Managerial Judgment in Marketing: The Concept of Expertise," *Journal of Marketing Research*, Vol. XX, 110-121.
- Libby, Robert en Roger K. Blashfield, 1978. "Performance of a Composite as a Function of the Number of Judges," *Organizational Behavior and Human Performance*, 21, 121-129.
- Morris, P.A., 1977. "Combining Expert Judgments: A Bayesian Approach," *Management Science*, 23, 679-693.
- Philips, Lynn W., 1981. "Assessing Measurement Error in Key Informant Reports: A Methodological Note on Organizational Analysis in Marketing," *Journal of Marketing Research*, Vol. XVIII, November, 395-415.
- Seidler, John, 1974. "On Using Informants: A Technique for Collecting Quantitative Data and Controlling for Measurement Error in Organization Analysis," *American Sociological Review*, Vol. 39 (December), 816-831.

6. Innovatie van mobiele dienstverleningen

De intersectie van snelheid en kennis

A. LIEVENS en V. BLAZEVIC

SAMENVATTING

Het turbulente, dynamische karakter van onze hoog-technologische digitale en mobiele markten noodzaakt bedrijven snel op de bal te spelen, waarbij de steeds omvangrijker wordende informatiestroom van onze digitale economie een steeds grotere uitdaging wordt om succesvol te innoveren. Het is exact deze uitdaging – de intersectie van snelheid en kennis – waar onze bijdrage over gaat. We beschrijven een case study onderzoek binnen een Nederlands telecombedrijf, waarbij we stap voor stap de kritieke determinanten van “time-to-market” en “project learning” genereren en samenbrengen binnen één conceptueel raamwerk dat managers van mobiele services kan helpen in het opzetten en opvolgen van succesvolle innovatieprocessen in de toekomst.

1. INLEIDING

Een Japanse manager die ver van huis werkzaam is, houdt zijn mobiele telefoon omhoog voor zijn collega's die hij eerst enkele instructies geeft. Vervolgens begint de groep collega's "happy birthday" te zingen voor de dochter van de manager. Dit wordt opgenomen via de ingebouwde videocamera, waarbij de manager aansluitend het lied e-mailt naar zijn dochter. Dit laatste is geen absurde fantasie, maar een dienstverlening die momenteel wordt aangeboden door DoCoMo, een Japans telecombedrijf¹.

Binnen de markt van de telecommunicatie vertegenwoordigt die van de mobiele telefonie het meest dynamische en snelst groeiende segment (IDC Research 2000). Bijgevolg moeten telecombedrijven, die via hun aangeboden mobiele services toegevoegde waarde wensen te creëren voor hun klanten, zo snel mogelijk een leiderschapspositie opbouwen om lange-termijn succes te verkrijgen. Vooral de snelheid waarmee innovaties in een markt worden geïntroduceerd vormt een essentiële succesfactor binnen de telecomindustrie. Onderzoek heeft inderdaad uitgewezen dat een snelle “time-to-market” van innovaties een duurzaam concurrentieel voordeel creëert (Kessler en Chakrabarti 1999; Bayus 1997). Bovendien zal het bedrijfsresultaat n.a.v. innovatie afhangen van de effectieve verwerking van markt- en technologische informatie door projectleden. Dit laatste impliceert duidelijk dat projectteams kennis opbouwen en bijgevolg leren gedu-

¹ Donal O'Shea and Mary Crowe, “The 3G Mobile Phone: A Manager's Guide”, *Harvard Management Communication Letter*, January 2001.

rende het innovatieproces. Onder meer om die toegevoegde waarde te creëren via nieuwe services is tegenwoordig het efficiënte gebruik van informatie in onze netwerkeconomie een bepalende factor geworden. Snel en effectief innoveren is een cruciaal onderdeel om de concurrentie een stap voor te blijven (Mascitelli 2000; Hurley en Hult 1998). Daarom wensen wij in deze paper snelheid en kennis te bespreken als essentiële componenten om duurzame concurrentiële voordelen te bereiken. Een diepgaande case-study werd opgezet binnen de telecomindustrie met het doel die kritieke factoren te identificeren die snelheid of “time-to-market” en opbouw van kennis (i.e., leren) gedurende het innovatieproces van mobiele services beïnvloeden.

Deze paper is als volgt gestructureerd. Eerst geven we een conceptuele onderbouw met betrekking tot de cruciale rol van kennis en snelheid en besteden aandacht aan het samengaan van diepgang en bereik van informatie binnen de innovatiecontext van mobiele services. Vervolgens beschrijven we kort het onderzoeksontwerp van onze gevalstudie dat ontwikkeld werd binnen een Nederlands telecombedrijf. Na een diepgaande analyse van de gevalstudies volgt een bespreking van de kritieke factoren die leren en “time-to-market” beïnvloeden. Ten slotte formuleren we de belangrijkste conclusies en management implicaties van ons onderzoek.

2. DE INTERSECTIE VAN SNELHEID EN KENNIS

De snelheid van verandering en aanpassing zijn cruciaal geworden nu de dimensies van tijd, locatie en aard van onze zakelijke transacties radicaal veranderd zijn. De directe, onmiddellijke waarneming – of “immediacy” in Engelstalig business-jargon – is in wezen het belangrijkste kenmerk van de digitale economie: handel vindt direct/onmiddellijk plaats, levenscycli van producten worden korter en “first movers” op de markt worden beloond. Alertheid en de capaciteit om snel te reageren op veranderende omstandigheden zijn fundamentele vereisten om te overleven binnen deze omgeving (Tapscott, Lowy en Ticoll 1998). “Time-to-market” vormt bijgevolg een van de belangrijkste drijfveren binnen de markt van de mobiele dienstverleningen. Zoals Paul May (2001, p. 11), stichter van Verista, stelt: *“Mobile commerce is an ostentatiously fast-moving world where ‘Internet time’ looks downright sluggish. The immediacy of mobile phone connection and the disposability of mobile devices combine with an impulsive, lifestyle oriented usage relationship to create a market that expects rapid turnover of concepts, services and brands. Players across the board are actively looking to turn every feature of the mobile landscape into a time-to-market advantage”*. Dus, marktspelers van mobiele services trachten een “first mover” – voordeel op te bouwen door snel nieuwe services aan te bieden. Onderzoek heeft uitgewezen dat dergelijke pioniers een concurrentiële voordeel kunnen uitbouwen inzake merkerkenning. Hieruit resulteert voor deze organisaties op lange termijn een hogere winstgevendheid en marktaandeel (Song *et al.* 2000). Op die manier kunnen organisaties “gedragsvoordelen” verwerven indien de klant weigerachtig staat om over te stappen naar de concurrent omwille van de kosten die ze hierbij ervaren. Deze kosten noemen we “switching costs” en zijn gerelateerd aan de gepercipieerde onzekerheid en informatieklouf die de klant ervaart bij een mogelijke overstap naar een andere mobiele provider. Bovendien heeft onderzoek bijkomend aangetoond dat “time-to-market” een cruciale betekenis heeft binnen industrieën met een hoge technologische intensiteit (Lynn *et al.* 1999; Hultink en Robben 1995). Oorspronkelijk is de telecomindustrie steeds

koploper geweest wat implementatie van toonaangevende technologieën betrof. Daar de concurrentie steeds intenser wordt, moeten bedrijven sneller reageren op de eisen van de eindgebruikers en zich toespitsen op een leiderschapspositie via “time-to-market”. Een vereiste echter om “time-to-market” leiderschap te verwerven zijn efficiënte mogelijkheden/middelen tot informatieverwerking. Informatieverwerking op zich omvat activiteiten die gerelateerd zijn aan de verwerving, verspreiding en interpretatie van informatie binnen een projectteam. Deze teams, die werken rond een concreet innovatieproject, kunnen bijgevolg gezien worden als informatieverwerkende systemen die enerzijds kennis van de organisatie gebruiken en anderzijds bijdragen tot organisatorisch leren. De effectiviteit waarmee projectteams bepaalde leermechanismen hantieren is uiteraard bepalend voor de snelheid en kwaliteit van innovatie. Meyer en Purser (1993) benadrukken dat een verbetering in het leerproces van een organisatie een kern-element vormt in het realiseren van kortere levenscycli van innovatieprojecten. Bijgevolg is “project learning” zeer belangrijk tijdens het ontwikkelingsproces van mobiele service innovaties. De teamleden rond een bepaald innovatieproject creëren een kennisplatform via de uitwisseling van hun respectievelijke expertises en inputs vanuit de externe omgeving. De omvang, complexiteit en snelheid waarmee deze informatie binnen een projectgroep wordt aangeleverd vereist (1) efficiënte informatieverwerking om een snelle doorlooptijd van de innovatie te bewerkstelligen, evenals (2) een effectief gebruik van informatiebronnen om tot succesvolle innovaties te komen.

3. DE NIEUWE INFORMATIE-ECONOMIE: KANSEN VOOR MOBIELE SERVICES

Voordat het Internet haar intrede deed als additioneel informatiekanaal, was het praktisch ondenkbaar dat een bedrijf gepersonaliseerde informatie aan een grote groep consumenten kon leveren. Onze zogenaamde “oude economie” werd gekenmerkt door een trade-off tussen “richness” en “reach” (Evans and Wurster 1997), waarbij “richness” betrekking heeft op de diepgang van de informatie en “reach” het effectieve bereik betreft. Dus, een organisatie kon gestandaardiseerde informatie aanbieden aan een grote groep klanten of deze organisatie kon een beperkt aantal klanten bereiken met gepersonaliseerde/geïndividualiseerde informatie. Het Internet en de daarmee gepaard gaande elektronische communicatiemogelijkheden hebben deze trade-off overbodig gemaakt. Voormelde kenmerken zijn zelfs nog meer uitgesproken voor mobiele services dan voor electronic services. In termen van “richness” kan een provider van mobiele services zeer gedetailleerde informatie verkrijgen van de klant. Een mobiele telefoon is immers een persoonlijk toestel dat zelden gezamenlijk gebruikt wordt met meerdere andere personen. Bijgevolg is “mobiel gedrag” duidelijk gelinkt aan een bepaald individu/persoon. Wanneer de klant bepaalde mobiele services gebruikt dan is zijn locatie gekend. Deze bijkomende informatie kan benut worden door de provider die “on target” gepersonaliseerde services kan aanbieden. Een nauwgezette analyse van “mobiel gedrag” van de klant genereert op haar beurt waardevolle informatie en input voor de ontwikkeling van nieuwe mobiele dienstverleningen. In termen van “reach” bezitten mobiele services ook een uitgesproken voordeel, daar deze het grootste deel van de tijd aan staan. In die zin is de klant meer bereikbaar dan in een elektronische omgeving, waarbij de gebruiker eerst een verbinding moet hebben tot het Internet. In een mobiele service context is de klant vaker bereikbaar, onderweg in een trein, auto of wachtend in

een café. Een deel van die tijd is nu beschikbaar en mobiele service providers zien hun kansen groeien om gepersonaliseerde services aan te bieden.

4. ONDERZOEKSONTWERP CASE-STUDIES

Binnen een in Nederland gevestigd telecombedrijf werden 4 case studies of innovatieprojecten rond mobiele dienstverleningen geselecteerd. Twee criteria werden hiervoor in aanmerking genomen: (1) het niveau van informatieverwerking (hoog versus laag), en (2) “time-to-market” (snel versus traag). Deze cases werden geanalyseerd met het oog op het genereren van de belangrijkste factoren die “project learning” beïnvloeden evenals het inventariseren van de belangrijkste determinanten van “time-to-market”. Zowel desk research als field research werden georganiseerd. Desk research was vooral gebaseerd op interne documentatie verstrekt door het bedrijf. Deze informatie maakte het mogelijk inzicht te verwerven in de strategische markt- en technologische benadering van het bedrijf. Field research werd opgezet via diepte-interviews met teamleden van ieder innovatieproject (Bijvoorbeeld: projectleider, product manager, segment manager, branch manager, advertising manager, training manager, andere projectleden, enz.).

Het eerste project (laag niveau van informatieverwerking en trage “time-to-market”) had betrekking op het ontwikkelen van een basisdienst voor een mobiele service provider: het leveren van “voice and fax mail” services. Een tweede project (laag niveau van informatieverwerking and snelle “time-to-market”) werd ontwikkeld om te verzekeren dat content aangeleverd via SMS up-to-date en aanpasbaar was in functie van specifieke noden van de gebruiker. Een derde project (hoog niveau van informatieverwerking en trage “time-to-market”) was gericht op het concipiëren van een virtueel privaat netwerk waar de leden elkaar kunnen bereiken op een persoonlijk kort nummer op hun professionele telefoon of hun thuis telefoon. Het vierde project (hoog niveau van informatieverwerking en snelle “time-to-market”) bestond uit de ontwikkeling van informatie en “entertainment” services die worden verstrekt via de mobiele telefoon. (voor een uitgebreide omschrijving van de 4 verschillende projecten – zie bijlage)

5. KRITIEKE DETERMINANTEN VAN “PROJECT LEARNING”

Onderzoek van de diverse case studies heeft diverse factoren die “project learning” beïnvloeden aan de oppervlakte gebracht. Hierna beschrijven we deze factoren evenals hun relatie met “project learning”.

5.1. Beslissingsstructuur

Binnen het project omtrent het Virtuele Private Netwerk (VPN) moest het top management beslissen over de hardware die het projectteam diende te hanteren. Deze beslissing en zelfs aankoop gebeurde reeds voor het projectteam was samengesteld. Tijdens de ontwikkeling van VPN bleek dat die hardware zeer complex was en niet strookte met de voorkeur van het projectteam. De teamleden kregen bijgevolg heel wat problemen met betrekking tot de informatieverwerking en het hanteren van de hardware rond dit project wat op zijn beurt een optimale leerervaring van de projectleden in de weg

stond. De autoriteit van de beslissingen leggen bij het top management heeft een fundamentele impact op de effectiviteit van de informatieverwerking door de projectleden. Een aanverwant aspect bij de beslissingsstructuur betreft het proces dat de projectleden moeten doorlopen, waarbij go/no go beslissingen door een stuurcomité worden genomen en dit een voorwaarde is tot de allocatie van resources door het top management. De druk die op die manier wordt gecreëerd om bepaalde beslissingen te nemen kan tevens zeer nefast zijn voor de kennisverwerving van het team. Dit laatste was het geval voor het project rond de informatie- en entertainment services, waarbij het team onder druk werd gezet snel een toeleverancier te selecteren. De deadlines ingegeven vanuit de procesactiviteiten maakten het de projectleden onmogelijk zorgvuldig een geschikte supplier te kiezen. Twee weken nadat een supplier was aangewezen kwam een betere kandidaat op de voorgrond die een veel effectievere bijdrage had kunnen leveren tot de informatieverwerking.

Het project omtrent “voice and fax mail services” werd beïnvloed door de afstand tussen beslissingsbevoegdheid en locatie van de benodigde informatie/expertise om de beslissing te kunnen nemen. Zo zorgde het vertrek van een top manager van een aantal weken voor een immense vertraging aangezien goedkeuring door betreffende manager voor allerhande aspecten vereist was. Onderbreking door de manager met beslissingsbevoegdheid onderbreekt op zijn beurt de nodige informatieverwerkende activiteiten die “project learning” stimuleren. Niettegenstaande het telecombedrijf een relatief flexibele organisatiestructuur had die projecten met veel informatieverwerking en een snelle doorlooptijd aanmoedigde, werd de organisatiestructuur gekenmerkt door een sterk hiërarchische beslissingsstructuur. Deze rigiditeit beperkt de potentiële leerervaringen van betrokken projectleden. Tal van organisaties binnen onze traditionele bedrijven hanteren deze hiërarchische organisatiestructuren en top-down processen. De mobiele services markt echter stelt zeer hoge eisen inzake informatieverwerking en reactiesnelheid. Daarom moet de beslissingsstructuur binnen de telecomindustrie aan deze eisen tegemoet komen tijdens het innovatieproces (Mendelson and Pillai 1999). Communicatie doorheen de diverse hiërarchische lagen van een organisatie zorgt voor onnodige vertragingen en reduceert de initiële kwaliteit van de informatie. Bijgevolg, dient de beslissingsbevoegdheid ook te liggen bij die personen die over de nodige informatie/expertise beschikken. Binnen een dynamische omgeving zoals die van de mobiele telefonie is flexibiliteit en vooral decentralisatie van de beslissingsstructuur aangewezen. Deze decentralisatie impliceert dat projectleden zelf beslissingen moeten kunnen nemen indien zij effectief over de relevante informatie beschikken.

5.2. Informatiebewustzijn

Binnen het telecombedrijf werd informatie omtrent concurrenten en toeleveranciers vooral verkregen via het Internet. Marktinformatie werd verworven via brainstorming sessies en focus groepen. Deze specifieke informatiebronnen voorzagen in heel wat nieuwe waardevolle informatie voor het bedrijf. Teamleden rond de innovatieprojecten werden gestimuleerd om nieuwe informatie met andere teamleden te delen tijdens de project meetings of via verslagen en e-mails. Het bewustzijn omtrent het bestaan van informatie werd tevens actief gecreëerd via het Intranet a.h.v. het rondzenden van elektronische briefings. Deze briefings omvatten meestal de laatste ontwikkelingen op concurrentieel vlak evenals algemene trends binnen de telecom business. Daarenboven konden projectleden specifieke informatie verspreiden en uitwisselen door het gebruik

van speciale “drives” uniek toegankelijk voor projectleden. Met uitzondering van “new content on SMS” maakten alle projecten gebruik van dit kanaal om informatie te verwerken. Binnen het “voice and fax mail” project werden productomschrijvingen en verslagen beperkt elektronisch verspreid. De meeste informatie werd gedistribueerd via hard copy “handouts”. Niettegenstaande de technische mogelijkheden voorhanden waren, werd Intranet beperkt actief gebruikt doorheen het innovatieproces. Deze situatie was dan weer totaal anders voor het project omtrent informatie- en entertainment services, waarbij de drive actief werd gebruikt om informatie te verwerven, het proces beter te plannen en contracten op te volgen. Hier beschouwde de project manager deze drive als een waardevol instrument om “up-to-date” informatie uit te wisselen en stimuleerde dan ook het actieve gebruik ervan. Bijgevolg werd informatie frequenter uitgewisseld en kende dit project een zeer hoge graad van informatieverwerking. Projectleden formuleerden ook duidelijk dat zij over “up-to-date” informatie beschikten, een absolute noodzaak in een dynamische industrie waar bedrijven informatie over de laatste technologische trends, concurrentiële acties en veranderende consumentenvoorkeuren en – levensstijlen op de voet moeten volgen. Diepgaande analyse van de cases toont aan dat informatiebewustzijn een belangrijke antecedent is van “project learning”. Informatiebewustzijn kan gedefinieerd worden als “ *promotion of widespread awareness of new information from the organization’s external environment, including information about markets, new technologies, and changes in customer tastes*” (Mendelson 2000, p. 515) Projectleden moeten continue nieuwe externe informatie verwerven die ze integreren in de uiteindelijke diensteninnovatie. Externe informatieverwerving is echter niet genoeg. Deze informatie moet gedeeld worden met andere projectleden zodat, indien nodig, de informatie zich verder kan verspreiden binnen de organisatie.

5.3. Geheugen van het projectteam

Het “voice and fax mail” project evenals het project rond informatie- en entertainment services werd opgezet toen het bedrijf net was opgericht. Formele procedures en processen waren toen nog niet ontwikkeld. Daarenboven betrof het voor beide projecten een “me-too” service die technologisch zeker niet complex te noemen was. Niettemin waren de teamleden onervaren. Zo had het marketing departement voor het project omtrent informatie- en entertainment services weinig inzicht in de markt. Dit gebrek aan marktkennis vloeyde tevens voort uit de onzekerheid met betrekking tot de fit met de Nederlandse markt.

De twee andere projecten (SMS en VPN) werden opgezet toen bedrijfsprocedures en – processen al meer geformaliseerd waren. In wezen kon het project omtrent “new content on SMS” een spin-off genoemd worden van het initiële project omtrent informatie- en entertainment services. De teamleden waren dus al in zekere mate vertrouwd met bepaalde aspecten van de service evenals de doelgroep. Het VPN-project kreeg te kampen met een ander belangrijk probleem met name dat van turnover van teamleden tijdens het innovatieproces. De effectieve project team manager moest tot 2x toe worden vervangen, waarbij de nieuwe team leader terug de nodige kennis moest opbouwen en dit een vlotte informatieverwerking en beslissingsname in de weg stond. Dit probleem liet het management ook inzien dat de uiteindelijke “lock-in” (i.e., insluiten/vasthouden van) van kennis zeer belangrijk is om zodoende de afhankelijkheid ten aanzien van welbepaalde personen te verminderen en daarbij te verhinderen dat kennis verdwijnt als bepaalde personen het bedrijf verlaten.

Onze projecten illustreren duidelijk het belang van “project team geheugen” als determinant van “project learning”. We kunnen project team geheugen definiëren als de collectieve kennis die ingebed is binnen het innovatieteam. Het geheugen van het innovatieteam omvat gedeelde ideeën, databases, geformaliseerde procedures en routines evenals formele culturele gedragscodes binnen de organisatie (Sinkula, Baker and Noordewier 1997). Deze collectieve, reeds bestaande kennis beïnvloedt alle informatieprocessen, i.e. de creatie, de verspreiding en het delen van nieuwe kennis. Zodoende dient reeds verworven kennis als een bouwsteen tot verdere kennis te worden gezien (Kogut and Zander 1992). Het geheugen dat doorheen vroegere projecten werd opgebouwd verhindert dan ook dat men steeds opnieuw het “warme water” uit vindt. (Drew 1994)

5.4. Vereisten versus middelen tot informatieverwerking

De geanalyseerde case studies illustreren het belang van een duidelijke fit tussen de vereisten inzake informatieverwerking en de mogelijkheden ervan. De vereisten inzake informatieverwerking worden afgeleid uit de druk vanuit de omgeving evenals de taakgerelateerde onzekerheid die samenhangt met de complexiteit van het innovatieproject. De digitale economie is verantwoordelijk voor de creatie van een nog grotere informatiestroom. Information overload is meer dan ooit een probleem. Organisaties beschikken gewoon niet over de nodige informatieverwerkende capaciteit. We hebben reeds vermeld dat voor al onze projecten de uitwisseling van informatie via het Intranet mogelijk was. Toch hebben we moeten vaststellen dat slechts bij 1 project (informatie-entertainment services) Intranet actief werd gebruikt en dit vooral omdat de projectleider het gebruik ervan stimuleerde onder de andere projectleden. Niettegenstaande Intranet de informatieverwerkende capaciteit verhoogt, blijkt dit geen voldoende voorwaarde te zijn. We merken dat organisaties vaak de technologische capaciteiten voorzien, maar de menselijke processen onvoldoende ondersteunen. We moeten nu eenmaal aanvaarden dat we als individuen over beperkte mogelijkheden beschikken inzake cognitieve verwerking (Simon 1979). Bijgevolg dienen we tevens aandacht te besteden aan de humane processen als het om het managen van information overload gaat. Organisaties dienen aandacht te besteden aan hoe hun werknemers vaker kennis kunnen verwerven en delen en ze zodoende aanzetten tot het gebruiken van de technologische middelen. We verwachten dan ook dat projectleden een betere leerervaring zullen doormaken indien er een grote fit bestaat tussen de vereisten inzake informatieverwerking en de uiteindelijke informatieverwerkende capaciteit.

5.5. Aard van de communicatie

Met uitzondering van het project omtrent “voice and fax mail” services begon het innovatieproject voor ieder project met ideeëngeneratie. Projectteams organiseerden brainstorming sessies en creatieve meetings. Na de ideeëngeneratie, dienden teamleden te specificeren in hoeverre ontwikkeling technologisch en economisch haalbaar was. Vervolgens werd de designfase ingezet gevolgd door een testfase. Gedurende al deze activiteiten werd de aard van de communicatie gekenmerkt door een hoge mate van creativiteit. Latere stadia doorheen het ontwikkelingsproces omhelsden in hoofdzaak het voorbereiden en organiseren van de lancering. Ieder betrokken departement werd hierbij geïnstrueerd omtrent zijn taken en over het functioneren van de dienstverlening zelf.

Het project “voice and fax mail” services betrof een imitatie van de concurrentie waarbij later enkel nieuwe additionele services werden ontwikkeld. Hier in dit geval was de communicatie bij aanvang eerder taakgeïntereerd om de geïmiterde technologie te implementeren. Met het oog op de ontwikkeling van de additionele features was brainstorming en inventiviteit vereist. Communicatie werd bijgevolg creatiever om in een later stadium tijdens de voorbereiding van de lancering opnieuw taak-geïntereerd te worden, waarbij vooral aspecten met betrekking tot de organisatie en coördinatie van technische/commerciële aspecten werden besproken. Bijgevolg, stelden we vast dat doorheen het innovatieproces van alle projecten de aard van de communicatie veranderde. Deze vaststelling strookt met deze opgemerkt binnen de innovatieliteratuur (Souder en Moenaert 1992; Blazevic and Lievens 2001) : innovatieve versus coördinerende communicatie. Innovatieve communicatie omvat de creatieve zoektocht naar de oplossing van problemen. Innovatieve communicatie omvat het zoeken naar antwoorden omtrent nieuwe noden/behoefte van de klant. Innovatieve communicatie stimuleert ideeëngeneratie en initieert het ontwikkelingsproces (Souder and Moenaert 1992). Via deze weg reageert het bedrijf om nieuwe informatie vanuit de omgeving met betrekking tot technologische mogelijkheden, veranderende klantennoden of veranderingen in de strategieën van concurrenten. Het projectteam is daadwerkelijk betrokken in de probleemoplossing en de verwerking van nieuwe ideeën. Hieruit volgt dat innovatieve communicatie 1 van de bouwstenen van informatieverwerking en leren binnen het projectteam vormt. Coördinerende communicatie wordt geassocieerd met controle, opdrachten, procedures en feedback tussen staf en ondergeschikten van diverse functionele departementen. Het heeft specifiek te maken met de organisatie van een innovatieproject op zich. Hierbij gaat het over de communicatie met betrekking tot taakverdeling, individuele verantwoordelijkheden en conflicthantering, communicatieaspecten die op hun beurt nodig zijn om leren te faciliteren.

6. KRITIEKE DETERMINANTEN VAN “TIME-TO-MARKET”

Een tweede factor die het succes van nieuwe mobiele services bepaalt is “time-to-market”. Via onze case studies genereerden wij diverse factoren die een bepalende impact op “time-to-market” uitoefenen. Hierna omschrijven we deze factoren evenals hun relatie met “time-to-market”.

6.1. *Project complexiteit*

Het project omtrent de ontwikkeling van een virtueel privaat netwerk (VPN) was een hoogtechnologisch project gekenmerkt door een aanzienlijke complexiteit. Technische problemen waren vaak moeilijk op te lossen en namen heel wat tijd in beslag. Het project “new content on SMS” daartegen werd helemaal niet als complex gepercipieerd aangezien het de ontwikkeling van een gestandaardiseerd “me-too” service betrof. Een andere dimensie binnen die complexiteit had te maken met de omvang van het project. Opnieuw was “new content on SMS” een relatief eenvoudig project, waarbij slechts een beperkt aantal personen en afdelingen betrokken waren. Dit resulteerde in een duidelijke taak- en rolverdeling en vervolgens in een kortere doorlooptijd van het project. Voor het VPN-project en de informatie- en entertainment services waren heel wat afdelingen en personen op hetzelfde ogenblik betrokken. Die samenwerking tussen ver-

schillende teamleden op een complex project gaf aanleiding tot heel wat conflicten. Dit werd nog versterkt door de heterogene samenstelling van het projectteam: de projectleden hadden sterk verschillende achtergronden. Deze spanningen hebben ertoe geleid dat de projectleden alsmaar minder met elkaar gingen communiceren. Daarbij kwam nog de hoge turnover van projectleden voor het VPN-project wat nog meer onzekerheid en ambiguïteit creëerde. De complexiteit van het project wordt tevens beïnvloed door het accuraat gebruik van documentatie. Dit laatste was zeker het geval voor de informatie- en entertainment services waar documentatie gemakkelijk toegankelijk was voor alle teamleden en een overzicht van de projectstadia verspreid werd onder alle projectleden. Tenslotte kunnen culturele verschillen een beduidende impact hebben op project complexiteit. De startfase van het “voice and fax mail” project werd gekenmerkt door een internationale samenwerking, waarbij coördinatie en communicatie vaak moeizaam verliep. Verschillen in gedrag, attitudes hebben dan ook “time-to-market” aanzienlijk vertraagd.

Diverse factoren bepalen de complexiteit van een project. Recent onderzoek heeft aangetoond dat project complexiteit een aanzienlijke impact heeft op de doorlooptijd van innovaties (Clift and Vandenbosch 1999) en bijgevolg een verschillende aanpak inzake informatieverwerking vereisen. Project complexiteit kan gedefinieerd worden als “consisting of many varied interrelated parts possessing two main features: differentiation and interdependency” (Baccarini 1996, p. 201; Williams 1999). Differentiatie wijst op de diverse elementen die in overweging moeten genomen worden: de omvang van het project, de contacten tussen de projectleden, de innovativiteit van het project en de gepercipieerde onzekerheid zijn enkele belangrijke factoren die project complexiteit beïnvloeden. Bedrijven dienen complexiteit te reduceren via 3 C’s daar dit in belangrijke mate “time-to-market” kan versnellen: (1) coördinatie van taken en verantwoordelijkheden, (2) communicatie over voormelde elementen, en (3) controle over het innovatieproces.

6.2. Top management support

In geval van het “voice and fax mail” project ontwikkelde het telecombedrijf een gestandaardiseerde service die als een basisvoorziening voor mobiele operatoren wordt beschouwd bij het betreden van de markt. Zodoende kreeg deze “me-too” service een zeer hoge prioriteit vanwege het top management. Bijgevolg werden voldoende middelen aan dit project toegewezen om de service zo snel mogelijk op de markt te brengen. Eenmaal deze basisdienst gelanceerd was, dienden additionele service features ontwikkeld. Dit kreeg al heel wat minder prioriteit en aandacht vanwege het management, waardoor de ontwikkelingstijd van deze features zeer veel tijd in beslag nam in tegenstelling tot de basisdienst. Ook de andere projecten tonen aan dat de prioriteit die het top management geeft aan een project bepalend is voor een vlotte allocatie van resources en bijgevolg in belangrijke mate “time-to-market” bepaald. Dit is niet onbelangrijk daar het project klimaat en de motivatie van de team leden eronder lijdt indien top management een lage prioriteit aan het project toekent. Bovendien impliceert die lage prioriteit vaak een beperkte informatiedoorstroming van het top management naar de projectleden. Dit laatste was het geval voor de informatie- en entertainment services waar de projectleden door een gebrekkige informatie heel wat vertraging opliepen in het nemen van de nodige beslissingen.

Samengevat kunnen we stellen dat de prioriteit die wordt toegewezen door het top

management een indicatie is van de support dat een projectteam krijgt. Deze support vertaalt zich dan in de middelen die worden toegewezen aan het project evenals de motivatie van de projectleden zelf. Top management op zich blijkt een belangrijke motivator te zijn en stuwende kracht achter “time-to-market”. Projecten met “high priority” en gemotiveerde leden vormen de beste garantie tot een snelle “time-to-market”.

6.3. Informatiemacht van de supplier

De informatiemacht van de supplier heeft gewoonweg alles te maken met de bereidheid van deze supplier om de nodige informatie te verstrekken tijdens het innovatieproces. Voor twee projecten (“voice and fax mail” services en VPN) hadden de projectleden negatieve ervaringen met de bereidheid van de suppliers om informatie door te geven. Vooral de afhankelijkheid van de supplier met betrekking tot technologische aspecten vormde in beide projecten het probleem. Eenvoudige antwoorden op technische vragen lieten op zich wachten. Hierdoor kon de technische afdeling het design van de service niet voltooien en kon het project team niet tot testing overgaan zoals voorzien binnen het planningsschema. Bijgevolg, diende de lancering te worden uitgesteld. Zo was het probleem nog accuter voor “new content on SMS” waarbij de supplier zonder enige reden pertinent weigerde om nog informatie te verstrekken. Zelfs na bemiddeling van de overheid was de supplier niet bereid informatie te verschaffen. Het technische departement is zodoende in een “trial en error”-straatje terechtgekomen. Gelukkig werd deze tegenslag gecompenseerd door efficiënte outsourcing. Diverse taken, waaronder het aanleveren van de “content” voor de gebruiker, gebeurde door externe bedrijven. De analyse van de cases leert ons dat bedrijven voorzichtig moeten zijn in de keuze van hun partners en voldoende maatregelen moeten nemen om overdreven afhankelijkheid van de supplier te vermijden. De wereld van de mobiele services is dynamisch, maar dient vooral gesitueerd binnen onze digitale netwerkeconomie waarbij bedrijven netwerken van suppliers, distributeurs, providers en klanten dienen op te zetten. Deze zogenaamde “Electronic Business Communities” vormen samenwerkingsplatformen die een basisvereiste vormen voor innovatie in onze mobiele wereld (Tapscott, Lowy en Ticoll 1998). De uitdaging voor het management blijft uiteraard deze netwerken te sturen zodat snelle service innovatie mogelijk blijft.

6.4. Technologische synergie

Het project omtrent VPN maakte gebruik van hardware technologieën die waren aangekocht door het management voordat het projectteam was samengesteld. Deze technologie stemde helemaal niet overeen met de noden van het projectteam. Dit stond in schril contrast met het project omtrent informatie- en entertainment services, waar de gebruikte technologie ontwikkeld was door het moederbedrijf en zeer goed aansloot bij de technologienoden van het projectteam. Aangezien het project “new content on SMS” een “spin-off” was van voormeld informatie- en entertainment” services was de synergie inzake technologieën wel duidelijk aanwezig. De aanwezige projectleden konden duidelijk beroep doen op hun ervaring opgedaan tijdens de ontwikkeling van de informatie- en entertainment services. De mate waarin technologische synergie aanwezig is, bepaald wel degelijk de ontwikkelingstijd van de innovatie. Technologische synergie wordt gedefinieerd als de fit tussen de noden van het innovatieproject en de bestaande technologie binnen het bedrijf (Lievens et.al. 1999; Cooper 1979; Cooper en

Kleinschmidt 1987). Technologische synergieën bevorderen “time-to-market” aangezien de technische ontwikkeling minder complex en tijdrovend is. Rekening houdend met de hoog-technologische intensiteit van de innovaties binnen de telecomindustrie dient deze factor te worden benut.

7. CONCLUSIE EN MANAGEMENT IMPLICATIES

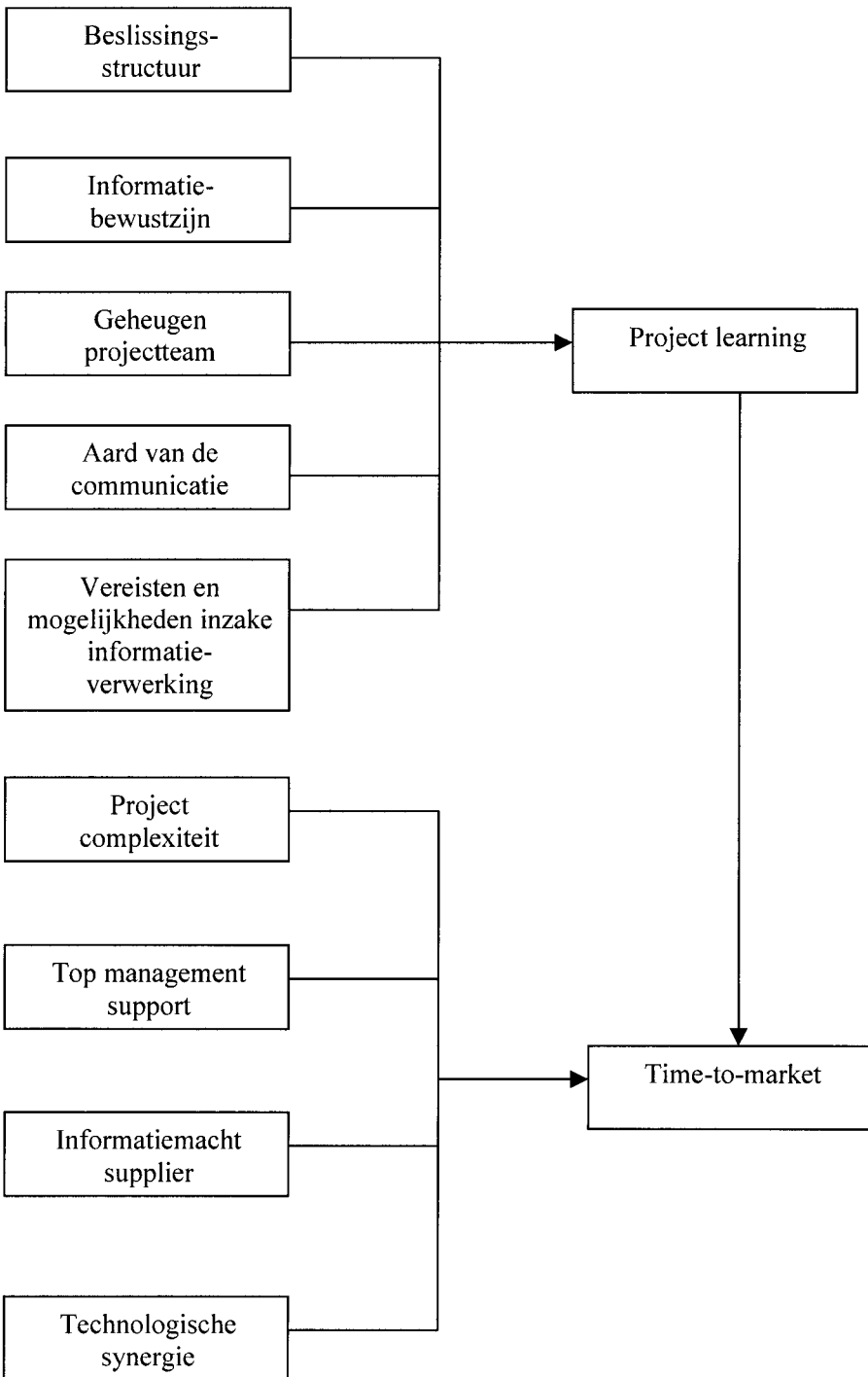
De opkomst van het Internet heeft onze wereld drastisch veranderd. Naast de toenemende digitalisering van onze economie kunnen we stellen dat de wereld tevens mobiel wordt. De penetratie van de mobiele telefonie groeit enorm en ongeveer 500 miljoen mensen op deze wereld kopen mobiele telefoons. Daarenboven, zal de “3G” technologie die enorme opportuniteiten creëert voor de snelle transfer van data via mobiele netwerken de vraag naar mobiele services verder doen toenemen. Een snelle ontwikkeling van mobiele services is het antwoord om tegemoet te komen aan deze dynamische omgeving.

We hebben een gevalstudie-onderzoek (case-study) opgezet binnen de telecomindustrie met het oog op het identificeren van die factoren die bijdragen tot succesvolle dienstinnovaties. Twee essentiële succesfactoren kwamen uiteindelijk uit de case studies naar voor: “time-to-market” en “project learning” (Zie Figuur 1).

In overeenstemming met de visie dat organisaties informatieverwerkende systemen zijn, hebben we het project team opgevat als een informatieverwerkend systeem dat kennis binnen de organisatie ontwikkeld. Organisaties moeten konstant leren om bij te benen met de technologische trends, concurrentiële ontwikkelingen en veranderende klantennoden (Lievens, de Ruyter en Lemmink 1999). De “snelle” vertaling van deze technologische - en marktinformatie in succesvolle services is van vitaal belang binnen de snel evoluerende mobiele service omgeving. Bijgevolg, is de intersectie van snelheid en kennis een kritiek punt in het onderzoeken van product- of dienstinnovatie binnen organisaties.

“Project learning” omvat informatieverwerkende activiteiten die een kennisbasis opbouwen. Om dit te bewerkstelligen moet het management een evenwicht in innovatieve en coördinerende communicatie voorzien. Daarbij dienen de capaciteiten inzake informatieverwerking een voldoende fit te vertonen met de vereisten inzake informatieverwerking. Onze digitale economie creëert hier alvast een volledig nieuw communicatieplatform, waarbij mobiele services en betere communicatiesystemen (Intranet, Extranet) heel wat opportuniteiten creëren inzake informatieverwerkende capaciteit. Niettemin, hebben onze case studies aangetoond dat meer communicatiekanalen (onder de vorm van e-mail, Intranet) niet noodzakelijk een efficiëntere informatieverwerking opleveren. Afgezien van de informatieverwerkende capaciteit die wordt geleverd via de technologie, blijft de gepercipieerde information overload een belangrijk probleem. Management dient dan ook aandacht te besteden aan het begeleiden van de humane processen om extra communicatiekanalen (i.e., informatieverwerkende capaciteit) te hanteren.

De organisatorische context blijft een belangrijke determinant van “project learning”. Een ondersteunend, harmonieus, participatief klimaat vormen de beste condities voor de effectieve uitwisseling van kennis. Managers dienen een organisatorisch klimaat te creëren dat leergedrag aanmoedigt en faciliteert. Leren is een sociaal proces (Blazevic en Lievens 2001) dat werknemers met elkaar in contact brengt. Participatie en flexibiliteit



Figuur 1 : Conceptueel raamwerk met betrekking tot de kritieke determinanten van “project learning” en “time-to market”.

teit tijdens het nemen van beslissingen stimuleert leren en is cruciaal voor de motivatie van het personeel. Via participatie is de zelfrealisatie en tevredenheid hoger. De uiteindelijk motivatie om deel te nemen aan informatieverwerkende activiteiten is veel groter en draagt bij tot het leren van de projectleden.

Het geheugen van het projectteam is een andere bepalende factor van “projectleren” daar team leden hun expertise, know-how en reeds verworven kennis uitwisselen. In hun aanwervingspolitiek dienen managers aan de ene kant voldoende diversiteit in te bouwen in hun project teams zodat leren wordt gestimuleerd, maar aan de andere kant dienen zij meer aandacht te besteden aan de “lock-in” van kennis. Te vaak worden project teams voor een dilemma geplaatst als een spilfiguur (Bijvoorbeeld projectleider) plots het team verlaat en een deel van de waardevolle kennis meeneemt zonder dat deze reeds formeel werd vastgelegd.

Het innovatiemanagement moet het informatiebewustzijn voldoende stimuleren via inter-departementele interfaces en het uitwisselen, verspreiden en delen van intern en extern gegenereerde informatie. De nodige training, ondersteuning, maar ook incentives die worden gekoppeld aan leergedrag kunnen deel uitmaken van een ondersteunend klimaat inzake leren.

Management dient, gezien de turbulente, dynamische context waarbinnen mobiele services worden ontwikkeld, “time-to-market” als een strategische prioriteit beschouwen. Een belangrijke voorwaarde is de support van het top management en de prioriteit die aan het innovatieproject wordt gegeven. Deze support bepaalt uiteindelijk de allocatie van resources en de toewijding/motivatie van het hele project team.

Zogenaamde “electronic business platforms” of netwerken van suppliers, providers, distributeurs en klanten zouden aan het management een antwoord kunnen bieden voor het oplossen van de informatie-assymetrieën die typisch ontstaan bij overdreven afhankelijkheid van welbepaalde suppliers. Ten slotte moet het innovatiemanagement de complexiteit van het project managen en technologische synergieën trachten te bewerkstelligen om zo “Time-to-market” gunstig te beïnvloeden.

LITERATUUR

- Baccarini D. 1996. The Concept of Project Complexity – A Review. *International Journal of Project Management* **14**: 201-204.
- Blazevic V., Lievens A. 2001. Learning during the New Financial Service Innovation Process: Antecedents and Performance Effects. *Journal of Business Research* forthcoming.
- Clift TB, Vandenbosch MB. 1999. Project Complexity and Efforts to Reduce Product Development Cycle Time. *Journal of Business Research* **45**: 187-198.
- Cooper RG. 1979. The Dimensions of Industrial New Product Success and Failure. *Journal of Marketing* **43**: 93-103.
- Drew, S.: *Business Re-engineering in Financial Services*, Pitman Publishing, London, 1994.
- Evans PB, Wurster TS. 1997. Strategy and the New Economics of Information. *Harvard Business Review*: 71-82.
- Evans PB, Wurster TS. 1999. Getting Real about Virtual Commerce. *Harvard Business Review*: 71-82.
- Hultink EJ, Robben, HSJ. 1995. Measuring New Product Success: The Difference that Time Perspective Makes. *Journal of Product Innovation Management* **12**: 392-405.
- Hurley RF, Hult GTH. 1998. Innovation, Market Orientation, and Organizational Learning: An Integration and Empirical Examination. *Journal of Marketing* **62**: 42-54.
- Kessler EH, Chakrabarti AK. 1996. Innovation Speed: A Conceptual Model of Context, Antece-

- dents, and Outcomes. *Academy of Management Review* **21**: 1143-1191.
- Kogut B, Zander U. 1992. Knowledge of the Firm, Combinative Capabilities, and The Replication of Technology. *Organization Science* **3**: 383-397.
- Lievens A, Moenaert RK, S'Jegers R. 1999. Linking Communication to Innovation Success in the Financial Services Industry: A Case Study Analysis. *International Journal of Service Industry Management* **10**: 23-47.
- Lievens A., de Ruyter K. and Lemmink J. 1999. Learning during new banking service development: a communication network approach to marketing departments. *Journal of Service Research* **2**: 145-163.
- Lynn GS, Skov RB, Abel KD. 1999. Practices that Support Team Learning and Their Impact on Speed to Market and New Product Success. *Journal of Product Innovation Management* **16**: 439-454.
- Mascitelli R. 2000. From Experience: Harnessing Tacit Knowledge to Achieve Breakthrough Innovation. *Journal of Product Innovation Management* **17**: 179-193.
- May P. 2001. *Mobile Commerce*. Cambridge University Press: New York.
- Mendelson H, 2000. Organizational Architecture and Success in the Information Technology Industry. *Management Science* **46**: 513-529.
- Mendelson H, Pillai RR. 1999. Information Age Organizations, Dynamics and Performance. *Journal of Economic Behavior and Organization* **38**: 253-281.
- Meyer C, Purser RE. 1993. Six Steps to Becoming a Fast Cycle-Time Competitor. *Research Technology Management*. 41-48.
- O'Shea D, Crowe M. 2001. The 3G Mobile Phone: A Manager's Guide. *Harvard Management Communication Letter*.
- Simon HA. 1979. Rational Decision Making in Business Organizations. *American Economic Review* **69**: 493-513.
- Sinkula JM, Baker WE, Noordewier, TA. 1997. Framework for Market-based Organizational Learning: Linking Values, Knowledge and Behavior. *Academy of Management Science* **25**: 305-318.
- Song MX, Di Benedetto CA, Song, LZ. 2000. Pioneering Advantage in New Service Development: A Multi-Country Study of Managerial Implications. *Journal of Product Innovation Management* **17**: 378-392.
- Souder WE, Moenaert RK. 1992. Integrating Marketing and R&D Project Personnel Within Innovation Projects: An Information Uncertainty Model. *Journal of Management Studies* **29**: 485-511.
- Tapscott D, Lowy A, Ticoll D. 1998. *Blueprint to the Digital Economy: Creating wealth in the era of E-business*, McGraw-Hill: New York.
- Williams TM. 1999. The Need for New Paradigms for Complex Projects. *International Journal of Project Management* **17**: 269-273.
- Woodruff D. Wireless Services. Retrieved June 25th, 2001 from the World Wide Web: <http://www.hoovers.com/industry/snapshot/profile/0,3519,49,00.html>

Bijlage: Beschrijving van de innovatieprojecten

Time-to-market

Traag

Snel

		Project I	Project II
<i>Laag</i>	Korte omschrijving	<ul style="list-style-type: none"> •Project I betreft de ontwikkeling van voice en fax mail services. Dit omvat “postpaid” voice mail, “prepaid” voice mail en “post-paid” business voice mail. 	<ul style="list-style-type: none"> •Deze service is een spin-off van project IV en werd ontwikkeld om “new content via SMS” aan te bieden. Het wordt in eerste instantie aangeboden voor “postpaid” klanten, waarbij later “prepaid” users zullen volgen. Gebruikers kunnen de laatste informatie inwinnen over topics zoals verkeer, weerbericht, beursnieuws, enz.
	Strategische doelstelling	<ul style="list-style-type: none"> •De service moest eenvoudig en gebruiksvriendelijk zijn om gebruik door de klant aan te moedigen, loyaliteit te verzekeren en een goed imago voor de organisatie te ontwikkelen. De doelstelling voor de business voice mail was professionele gebruikers als doelgroep aan te trekken. 	<ul style="list-style-type: none"> •De doelstelling bestond erin om de inhoud, verzonden naar mobiele telefoons, up-to-date en aanpasbaar te maken aan de specifieke voorkeuren van de klant.
	Levering/verstrekken van de service	<ul style="list-style-type: none"> •De service was gericht naar “postpaid” packages en in een beperkte vorm aangeboden aan “prepaid” users. Voor het business package werden fax mail services toegevoegd om een volledig set van mailbox features te ondersteunen. 	<ul style="list-style-type: none"> •De inhoud kan verkregen worden op het Internet via het versturen van een SMS of via de mobiele telefoon.
<i>Hoog</i>	Korte omschrijving	<p style="text-align: center;">Project III</p> <ul style="list-style-type: none"> • Deze dienst betreft een virtueel privaat netwerk (VPN) die in belangrijke mate waarde toevoegt aan het interne netwerk van de organisatie. Leden kunnen elkaar bereiken op een persoonlijk kort nummer op hun professionele telefoon of hun thuis telefoon. 	<p style="text-align: center;">Project IV</p> <ul style="list-style-type: none"> •Project IV omvat het ontwikkelen van een platform voor het aanbieden van informatie- en entertainment services. Informatie services omvat inhoud over topics zoals verkeersinformatie en financiële informatie. Entertainment omvat o.m. humor en horoscopen.
	Strategische doelstelling	<ul style="list-style-type: none"> • De doelstelling was om het marktaandeel te verhogen in de business markt, meer business traffic te capteren van corporate networks, loyaliteit van de klant verhogen en het realiseren van een concurrentieel voordeel. 	<ul style="list-style-type: none"> •De doelstelling bestond erin een “me-too” product te lanceren, waarbij de snelheid van marktlantering van belang was met betrekking tot de aanwezigheid in de markt van “value added services”.
	Levering/verstrekken van de service	<ul style="list-style-type: none"> • De service wordt op vraag van de klant aangeboden, daar het VPN op maat wordt gemaakt voor het vragende bedrijf. 	<ul style="list-style-type: none"> •De service wordt verstrekt via SMS, waarbij de gebruiker specifieke informatie kan vragen of via het Internet waar de klant kan intekenen voor het verkrijgen van bepaalde informatie.

7. Veel keus en toch weinig moeite

*De invloed van assortimentsvariëteit op slagingskans en keuzemoeilijkheid**

E. VAN HERPEN

*When it comes to product variety, shoppers tend to send out mixed signals.
They like supermarkets to offer a wide variety of product.
At the same time, they think that stores carry too many items.*
Rafferty, 1993

SAMENVATTING

Voordat consumenten een product kiezen vormen ze verwachtingen over de slagingskans en de keuzemoeilijkheid die ze gaan ondervinden in een winkel. Detaillisten denken over het algemeen dat het verhogen van variëteit de slagingskans vergroot, maar ook de keuzemoeilijkheid. In dit artikel worden verschillende componenten van variëteit onderscheiden: het aantal producten in het assortiment, de spreiding in attribuutniveaus en de samenhang tussen attributen. Er wordt aangetoond dat het effect van variëteit op de verwachting van keuzemoeilijkheid niet hetzelfde is voor elk van deze componenten. Het is mogelijk om de variëteit van een winkelassortiment te laten stijgen zonder dat consumenten afgeschrikt worden vanwege een hogere verwachte moeilijkheid om te kiezen. Gevolgen voor detaillisten en marktonderzoekers worden besproken.

1. INLEIDING

De laatste jaren is er een explosieve stijging te zien van het aantal producten dat in winkelassortimenten wordt aangeboden. In vergelijking met 13 gerechten op het menu in de jaren 70, heeft MacDonaldis bijvoorbeeld zijn assortiment met meer dan 200% laten stijgen tot 43 gerechten eind jaren 90 (Cox & Alm 1998). Omdat het Internet geen ruimtebeperkingen heeft, kunnen assortimenten daar extreem toenemen. Zo biedt JCPenney, een warenhuis uit de VS, maar liefst 200,000 verschillende producten online aan (Estienny 2000). Het omgaan met deze grote assortimenten is een uitdaging voor detaillisten. Als gevolg hiervan bestaat er een toenemende belangstelling voor *category management* en voor assortimentsonderzoek.

De vraag is of consumenten het wenselijk vinden dat de omvang en variëteit van assortimenten zo is toegenomen. In het begin van de jaren '40, toen supermarkten veel min-

¹ Dit artikel is gebaseerd op het proefschrift "Perceptions and Evaluations of Assortment Variety" (2001) van Erica van Herpen; promotor: Rik Pieters (Tilburg University). De auteur wil Rik Pieters bedanken voor zijn substantiële bijdragen aan dit onderzoek.

der producten aanboden dan nu, klaagden consumenten al over de vele schappen met producten (Converse & Spencer 1942). We kunnen ons daarom afvragen of de enorme variëteit van vandaag de dag het niet ontzettend moeilijk maakt voor consumenten om een keuze te maken. Dit artikel laat zien welke invloed assortimentsvariëteit heeft op de verwachtingen van consumenten over slagingskans en keuzemoeilijkheid. In tegenstelling tot de algemene aanname dat meer variëteit betekent dat keuzes moeilijker worden (Kahn & McAlister 1997) laat dit onderzoek zien dat de keuzemoeilijkheid niet per se toeneemt bij hogere variëteit. Dit gebeurt door een onderscheid te maken in verschillende componenten van variëteit, en de effecten van deze componenten te bekijken. Het is dus mogelijk voor een detaillist om de variëteit van het assortiment te laten toenemen zonder dat consumenten de indruk hebben dat het moeilijker wordt om te kiezen.

2. SLAGINGSKANS VERSUS KEUZEMOEILIKHEID

Er zijn verschillende redenen voor een consument om assortimenten met veel variëteit te prefereren boven assortimenten met weinig variëteit (Hoch, Bradlow & Wansink 1999, Kahn 1998). In assortimenten met veel variëteit is het waarschijnlijker dat een favoriet product ook daadwerkelijk aanwezig is, kan een consument op verschillende momenten andere producten kiezen, en de productkeuze uitstellen totdat hij/zij in de winkel staat. Bovendien biedt variëteit de gelegenheid om over een productcategorie te leren, en deze te verkennen. Consumenten die winkelen leuk vinden kunnen waarschijnlijk meer afleiding en stimulatie vinden in gevarieerde assortimenten. Verder weten we uit eerder onderzoek dat consumenten variëteit prefereren, zelfs wanneer dit betekent dat minder gewaardeerde producten worden opgenomen in een set (Ratner, Kahn & Kahneman 1999).

Over het algemeen genomen is een hoge variëteit dus aantrekkelijk omdat het de kans om te slagen, en een geschikt product te vinden, vergroot. Desalniettemin kan een hoge variëteit ook negatieve consequenties hebben. Te veel variëteit kan de consument verwarren, en de moeite die het kost om te kiezen vergroten (Iyengar & Lepper 2000). De grotere complexiteit van het assortiment leidt ertoe dat meer inspanning, zowel in termen van moeite als benodigde tijd, nodig is om alle keuzemogelijkheden in detail te beoordelen. Bovendien kan meer variëteit leiden tot meer spijt achteraf, of tot ontevredenheid over het gekozen product (Iyengar & Lepper 2000). Tabel 1 geeft een overzicht van de kosten en baten van assortimentsvariëteit. Hoewel er andere kosten en baten bestaan, zoals uit de tabel blijkt, zal dit artikel zich richten op de slagingskans (aanwezigheid van een geschikt product) en de cognitieve keuzemoeilijkheid van consumenten.

Tabel 1. De baten en kosten van assortimentsvariëteit voor consumenten.

Baten van een hoge assortimentsvariëteit	Kosten van een hoge assortimentsvariëteit
– Aanwezigheid van een geschikt product is meer waarschijnlijk	– Cognitieve moeite
– Flexibiliteit in keuzes	– in het onderscheiden van producten
– Leren over een productcategorie	– in het nemen van de beslissing
– Afleiding en ‘fun’	– Tijd
	– Geanticipeerde spijt

Een veel gebruikt framework om productkeuze te onderzoeken is het ‘accuracy-effort framework’ (Klein & Yadav 1989; Payne, Bettman & Johnson 1988; Ratchford 1982). De aanname van dit framework is dat consumenten een afweging maken tussen de nauwkeurigheid van een beslissing en de moeite die het kost om tot deze beslissing te komen. Het maximaliseren van de nauwkeurigheid van de keuze en het minimaliseren van de cognitieve moeite van de keuze worden beschouwd als zogenaamde meta doelstellingen, die de belangrijkste motieven van het keuzegedrag omvatten (Bettman, Luce & Payne 1998).

Om nauwkeurigheid en moeite te bepalen, is in voorgaand onderzoek gebruik gemaakt van computersimulaties (Payne, Bettman & Johnson 1988), is de beslissingstijd bepaald (Hendrick, Mills & Kiesler 1968), of is de consument na het nemen van een beslissing gevraagd naar zijn/haar perceptie van nauwkeurigheid en moeite (Klein & Yadav 1989). Voorgaand onderzoek richt zich derhalve op de ervaring van nauwkeurigheid en moeite tijdens of na het beslissingsproces. Er is echter niet onderzocht wat de verwachting van de consument is, vóórdat de beslissing genomen wordt. Juist deze eerste indruk en verwachtingen van de consument zijn belangrijk voor het verdere verloop van het beslissingsproces. Of consumenten een winkel al dan niet inlopen zal afhangen van hun verwachtingen over de slagingskans en de keuzemoeilijkheid die met het assortiment van de winkel gepaard gaan.

Nauwkeurigheid en moeite worden vaak gepresenteerd als een noodzakelijke trade-off die gemaakt moet worden (Beach and Mitchell 1987; Payne, Bettman & Johnson 1988). Consumenten dienen een compromis te sluiten tussen de druk voor meer accurate beslissingen en de weerstand tegen het inzetten van extra inspanning. Binnen de detailhandel worden de positieve en negatieve kanten van variëteit ook vaak als tegen-gestelde krachten gezien. Een verhoging van de complexiteit in het assortiment en van de resulterende verwarring bij de consument worden gezien als onontkoombare gevolgen van een verhoging van de variëteit in een assortiment (Kahn & McAlister 1997). Maar is het wel zo dat een verhoging van de variëteit automatisch leidt tot een grotere keuzemoeilijkheid?

3. DE VARIËTEIT VAN EEN ASSORTIMENT

Om de gevolgen van variëteit beter in kaart te kunnen brengen, zal deze sectie eerst het begrip ‘variëteit’ nader bekijken. Sommige onderzoekers stellen variëteit gelijk aan de omvang van een assortiment: een groot assortiment is meer gevarieerd dan een klein assortiment (Chiang & Wilcox 1997). Hierbij wordt echter geen rekening gehouden met de verschillen tussen producten (Kahn & Lehmann 1991). Een assortiment met 10 blauwe stropdassen wordt dan als even gevarieerd aangeduid als een assortiment met 5 blauwe en 5 rode stropdassen. Daarom hebben Hoch, Bradlow en Wansink (1999) het begrip variëteit ruimer gedefinieerd, en zijn zij uitgegaan van de mate van overeenkomst tussen de producten in het assortiment. Neem als voorbeeld het assortiment met drie stropdassen van Tabel 2. Stropdas 1 en stropdas 2 verschillen nauwelijks van elkaar; alleen hun motief is anders. Stropdas 1 en stropdas 3 verschillen daarentegen zowel in kleur, materiaal als motief. De totale variëteit in een assortiment kan nu bepaald worden door de verschillen in paren van producten bij elkaar te nemen. Deze benadering gaat ervan uit dat de consument alle producten één voor één met elkaar vergelijkt. Dit is wellicht een minder realistische aanname wanneer consumenten gecon-

fronteerd worden met grote en complexe assortimenten. Stelt u zich voor dat u een kledingzaak binnen loopt met heel veel stropdassen. Eén van de eerste dingen die op zullen vallen, zijn de kleuren van de stropdassen, en misschien ook het materiaal en motief. Door rond te kijken kunt u zich een indruk vormen van de variëteit in deze attributen, zonder dat u elke individuele stropdas met elke andere stropdas vergelijkt.

Tabel 2. Samenstelling van een hypothetisch assortiment van stropdassen.

Attributen	Stropdas 1	Stropdas 2	Stropdas 3
Kleur	Blauw	Blauw	Groen
Materiaal	Katoen	Katoen	Zijde
Motief	Strepen	Stippen	Stippen

Er is dan ook een andere manier om de variëteit van een assortiment te benaderen. In plaats van naar producten te kijken, kunnen attributen met elkaar vergeleken worden (Bettman, Luce & Payne 1998). Er kunnen twee variëteitcomponenten onderscheiden worden op basis van de attributen: de *spreiding* in attribuutniveaus en de *samenhang* tussen attributen. Spreiding in attribuutniveaus heeft te maken met de relatieve frequentie waarin de verschillende attribuutniveaus voorkomen in een assortiment. Het assortiment van tabel 2 heeft twee blauwe en één groene stropdas. Dit is meer gevarieerd dan een assortiment met drie blauwe stropdassen. In het assortiment is de samenhang tussen kleur en materiaal sterk: alle blauwe stropdassen zijn van katoen, en alle groen stropdassen van zijde. Wanneer in het assortiment ook een blauwe zijden stropdas zou voorkomen, zou de variëteit toegenomen zijn. Assortimenten zijn meer gevarieerd wanneer de spreiding in attribuutniveaus hoog is (een zo gelijk mogelijke verdeling van de producten over de attribuutniveaus) en wanneer de samenhang tussen attributen zwak is (geen systematische verbanden tussen de attributen).

Voorgaand onderzoek heeft het belang van attributen in de evaluatie van assortimenten benadrukt. Wanneer consumenten geconfronteerd worden met grote en gevarieerde assortimenten, zullen ze over het algemeen de attribuutinformatie benadrukken en behulpzaam vinden (Bettman, Luce & Payne 1998; Huffman & Kahn 1998), en reageren op veranderingen in de beschikbaarheid van attribuutniveaus (Boatwright & Nunes 2001). De componenten van variëteit (assortimentsgrootte, spreiding van attribuutniveaus en samenhang tussen attributen), geven dan ook een goede voorspelling van de variëteitperceptie van consumenten (Van Herpen & Pieters 2002).

4. DE GEVOLGEN VAN VARIËTEIT

4.1. Invloed van variëteit op slagingskans

Over het algemeen is de kans om een geschikt product te vinden hoger in grotere assortimenten (Koelemeijer & Oppewal 1999). Grotere assortimenten geven immers meer vrijheid om te kiezen. Een vergelijkbaar effect zal waarschijnlijk optreden voor de overige variëteitcomponenten (spreiding van attribuutniveaus en samenhang tussen attributen). Wanneer de producten in een assortiment meer van elkaar verschillen, stijgt de kans dat er een geschikt product aanwezig is. Daarom kan er gesteld worden:

- H1: Een verhoging van de assortimentsvariëteit, door middel van een grotere assortimentsomvang en/of meer spreiding van attribuutniveaus en/of een zwakkere samenhang tussen attributen, leidt tot een grotere slagingskans.

4.2. Invloed van variëteit op keuzemoeilijkheid

Een vergroting van de omvang van een assortiment zorgt ervoor dat de keuzemoeilijkheid groter wordt (Malhotra 1982). Grotere assortimenten zijn meer complex, aangezien meer producten geëvalueerd moeten worden. Hoewel de effecten van de variëteit-componenten niet eerder als zodanig zijn onderzocht, is er wel onderzoek gedaan naar een gerelateerd begrip – de gelijkenis tussen producten. Wanneer de spreiding van attribuutniveaus hoog is, en de samenhang tussen attributen zwak, lijken de producten weinig op elkaar.

Een hoge gelijkenis tussen producten kan zowel positieve als negatieve gevolgen voor de keuzemoeilijkheid hebben. Aan de ene kant zal het moeilijker worden om producten van elkaar te onderscheiden, en de verschillen tussen producten te achterhalen (Kahn & McAlister 1997; Stone & Schkade 1991). Aan de andere kant zullen de afwegingen (trade-offs) die gemaakt moeten worden kleiner zijn en zal een verkeerde keuze minder gevolgen hebben. Keuzes worden moeilijker wanneer elk product zowel voor- als nadelen heeft (Shugan 1980; Chatterjee & Heath 1996). Bij een hoge gelijkenis tussen producten zullen deze relatieve voor- en nadelen kleiner zijn.

Wanneer producten in een assortiment erg op elkaar lijken, kunnen consumenten verwachten dat het moeilijk gaat worden om de producten te onderscheiden en hun keuze te maken. Maar ze kunnen ook verwachten dat de afwegingen die ze moeten gaan maken minder groot zullen zijn. Deze effecten zouden tegen elkaar weg kunnen vallen. Dit leidt tot de volgende hypothesen:

- H2a: Een verhoging van de assortimentsvariëteit, door middel van een grotere assortimentsomvang, leidt tot een grotere keuzemoeilijkheid.
- H2b: Een verhoging van de assortimentsvariëteit, door middel van meer spreiding van attribuutniveaus en/of een zwakkere samenhang tussen attributen, leidt niet tot een grotere keuzemoeilijkheid.


5. OPZET VAN HET EXPERIMENT

5.1. Stimuli en design

Om de hypothesen te toetsen is een experiment opgezet waarbij de respondenten verschillende assortimenten met vaatwassers te zien kregen. Het experiment werd via de computer afgenomen, met behulp van het programma Authorware (Kellogg & Bhatnagar 2002). De vaatwassers werden in drie kolommen weergegeven, gegroepeerd op merknaam, en met een beschrijving van hun attributen (zie Tabel 3 voor een beschrijving van de attributen en Figuur 1 voor een voorbeeld). De productcategorie van vaatwassers was gekozen, omdat het een belangrijke aankoop is, met een redelijk hoog niveau van complexiteit. De respondenten waren 62 studenten aan de Universiteit van Tilburg, die ieder 12 assortimenten evalueerden. Het experiment had een 2 (assortimentsgrootte) x 3 (spreiding van attribuutniveaus) x 2 (samenhang tussen attributen)

within-subjects design. De assortimenten bestonden uit 12 of 18 producten. De spreiding van attribuutniveaus was gemanipuleerd in 3 niveaus: laag (sommige niveaus waren afwezig), middelmatig (ongelijke verdeling over alle niveaus), en hoog (gelijke verdeling over alle niveaus). De samenhang tussen de attributen was zo hoog of zo laag mogelijk, zonder dat er replica's in het assortiment voorkwamen. Tabel 3 geeft de details van het experiment.

Brand:	Bauknecht	Zanussi	A.E.G.	/
Number of programs:	6	6	4	
Average cycle time:	91 min.	91 min.	86 min.	
Average water use:	12 liter	12 liter	16 liter	
Average energy use:	1,05 kWh	1,05 kWh	0,9 kWh	∅
Brand:	Bauknecht	Zanussi	A.E.G.	/
Number of programs:	6	6	4	∅
Average cycle time:	91 min.	91 min.	86 min.	/
Average water use:	12 liter	12 liter	16 liter	∅
Average energy use:	0,9 kWh	0,9 kWh	1,05 kWh	/
Brand:	Bauknecht	Zanussi	A.E.G.	∅
Number of programs:	6	4	4	r
Average cycle time:	91 min.	86 min.	86 min.	∅
Average water use:	16 liter	16 liter	12 liter	∅
Average energy use:	1,05 kWh	1,05 kWh	0,9 kWh	/
Brand:	Bauknecht	Zanussi	A.E.G.	/
Number of programs:	6	4	4	S
Average cycle time:	86 min.	86 min.	91 min.	/
Average water use:	12 liter	16 liter	16 liter	∅
Average energy use:	1,05 kWh	0,9 kWh	0,9 kWh	r
				∅



*Dishwasher
Department*

Figuur 1. Voorbeeld van een vaatwasser assortiment uit het experiment (klein, lage spreiding, hoge samenhang).

5.2. Procedure en maten

Aan de respondenten werd verteld dat het doel van het experiment was om internet winkels te onderzoeken, en ze werden verzocht om vragen te beantwoorden over de assortimenten vaatwassers in 12 verschillende winkels. Deze assortimenten werden in willekeurige volgorde aangeboden. Tevens kregen de respondenten een klein informatiepakketje met achtergrondinformatie over vaatwassers, om ze bekend te maken met de attributen die in de studie gebruikt worden. Deze informatie kwam uit een artikel van de Consumentengids, aangevuld met een tabel met de attributen en attribuutniveaus die in de studie gebruikt werden. Voor de vragen werd een 10-puntsschaal gebruikt, met twee items per construct. De verwachte keuzemoelijkheid werd gemeten door 'Kiezen tussen de vaatwassers van deze winkel is ... [heel erg moeilijk – heel erg gemakkelijk]' en 'Als me gevraagd zou worden een keuze te maken in deze winkel, dan moet ik daar ... [veel over nadenken – weinig over nadenken]' met een Cronbach α (gemiddeld over de assortimenten) van .88. De perceptie van assortimentsvariëteit werd gemeten door 'Dit assortiment vaatwassers biedt ... [zeer weinig variëteit – zeer veel variëteit]' en 'Het assortiment vaatwassers in deze winkel is ... [helemaal niet

Tabel 3. Gebruikte attributen van vaatwassers.

Attribuut	Attribuutniveaus	Spreiding van de attribuutniveaus		
		Laag	Medium	Hoog
Merk	AEG; Bauknecht; Zanussi; Siemens; Bosch; Whirlpool	3 niveaus in gelijke proporties	3 niveaus komen 1 keer voor, anderen in gelijke proporties	6 niveaus in gelijke proporties
Energieverbruik (kWh)	0.9; 1.05; 1.25	1 niveau afwezig	1 niveau komt slechts 2 keer voor	3 niveaus in gelijke proporties
Waterverbruik (liter)	10; 12; 16	1 niveau afwezig	1 niveau komt slechts 2 keer voor	3 niveaus in gelijke proporties
Aantal programma's	4; 5; 6	1 niveau afwezig	1 niveau komt slechts 2 keer voor	3 niveaus in gelijke proporties
Tijdsduur van het programma (min)	120; 91; 86.	1 niveau afwezig	1 niveau komt slechts 2 keer 2 keer voor	3 niveaus in gelijke proportie

afwisselend – heel erg afwisselend]' met een gemiddelde Cronbach α van .82. Tenslotte werd de verwachte kans op slagen gemeten door 'De kans dat dit assortiment een geschikte vaatwasser heeft is ... [zeer klein – zeer groot]' en 'Hoe waarschijnlijk is het dat je in dit assortiment een goede vaatwasser kunt vinden ... [helemaal niet waarschijnlijk – zeer waarschijnlijk]' met een gemiddelde Cronbach α van .90. De vragen verschenen in bovenstaande volgorde op het scherm. Gemiddeld duurde het experiment 20 minuten, en de respondenten kregen ongeveer 4 Euro als beloning.

6. RESULTATEN

6.1. Beschrijving van de data

De evaluaties van elk van de assortimenten worden gegeven in Tabel 4. De data zijn geanalyseerd met behulp van variantieanalyses (ANOVAs) met herhaalde metingen, en Tabel 5 geeft hiervan de resultaten. Aangezien de interactie effecten niet significant zijn, zullen alleen de hoofdeffecten besproken worden. Voor *variëteitpercepties* zijn significante hoofdeffecten gevonden voor assortimentsgrootte ($F = 160.2$; $p < .001$), spreiding van de attribuutniveaus ($F = 133.0$; $p < .001$) en samenhang tussen de attributen ($F = 23.8$; $p < .001$). Alle effecten zijn in de verwachte richting: consumenten percipiëren meer variëteit wanneer een assortiment groot is (gemiddelde variëteit $GV = 4.96$) dan wanneer het klein is ($GV = 3.97$); wanneer spreiding in attribuutniveaus hoog is ($GV = 5.23$) versus medium ($GV = 4.69$) versus laag ($GV = 3.49$); en consumenten percipiëren meer variëteit wanneer de samenhang tussen attributen zwak is ($GV = 4.70$) dan wanneer de samenhang tussen attributen sterk is ($GV = 4.24$).

Tabel 4. Evaluaties van de vaatwasser assortimenten ($n = 62$).

No.	Variëteit			Assortiment evaluaties	
	Grootte	Spreading van attribuut niveaus	Samenhang tussen attributen	Gemiddelde verwachte slagingskans (sd) ¹	Gemiddelde verwachte keuzemoeilijkheid (sd) ¹
1	12	laag	hoog	4.43 (1.09)	6.77 (1.50)
2	12	laag	laag	4.52 (1.33)	6.78 (1.55)
3	12	medium	hoog	4.72 (1.14)	7.07 (1.33)
4	12	medium	laag	5.09 (1.23)	7.26 (1.35)
5	12	hoog	hoog	4.90 (1.31)	6.92 (1.33)
6	12	hoog	laag	5.44 (1.07)	7.23 (1.44)
7	18	laag	hoog	4.80 (1.33)	7.63 (1.53)
8	18	laag	laag	5.05 (1.36)	7.44 (1.48)
9	18	medium	hoog	5.26 (1.07)	7.61 (1.45)
10	18	medium	laag	5.49 (1.08)	7.76 (1.38)
11	18	hoog	hoog	5.62 (1.19)	7.91 (1.57)
12	18	hoog	laag	6.07 (0.96)	7.88 (1.63)

¹ Slagingskans en keuzemoeilijkheid gemeten op een 10-puntsschaal, met 10 = grote kans om te slagen, respectievelijk, 10 = moeilijke keuze.

Tabel 5. Resultaten van de ANOVAs.

	Verwachte slagingskans			Verwachte keuzemoeilijkheid		
	<i>F</i> -waarde	<i>df</i>	<i>p</i> -waarde	<i>F</i> -waarde	<i>df</i>	<i>p</i> -waarde
Assortimentsomvang	36.2	1	<.001	36.5	1	<.001
Spreading van attribuutniveaus	26.1	2	<.001	3.0	2	.053
laag vs. rest ¹	27.8	1	<.001	4.3	1	.042
medium vs. hoog ¹	21.2	1	<.001	.3	1	.604
Samenhang tussen attributen	20.8	1	<.001	.6	1	.443

¹ Post hoc testen zijn gebruikt om assortimenten met een lage spreading af te zetten tegen het gemiddelde van de overige assortimenten, en om assortimenten met medium spreading af te zetten tegen assortimenten met hoge spreading van attribuutniveaus.

6.2. Slagingskans

Ook voor de geanticiperde slagingskans zijn de verwachte hoofdeffecten gevonden. Consumenten verwachten eerder te kunnen slagen wanneer het assortiment groot is (gemiddelde slagingskans $GS = 5.38$) dan wanneer het klein is ($GS = 4.85$; $F = 36.2$; $p < .001$), wanneer de attribuutniveaus in het assortiment gelijkelijk over de producten verdeeld zijn ($GS = 5.51$) dan wanneer de attribuutniveaus ongelijk verdeeld zijn ($GS = 5.14$ voor medium spreading, en 4.70 voor lage spreading; $F = 26.2$; $p < .001$), en wanneer de samenhang tussen attributen in het assortiment zwak is ($GS = 5.28$) dan wanneer de samenhang sterk is ($GS = 4.95$; $F = 20.8$; $p < .001$). Deze resultaten ondersteunen hypothese 1.

6.3. Keuzemoeilijkheid

Hypothese 2a wordt ondersteund: de verwachte keuzemoeilijkheid is groter in grote assortimenten (gemiddelde keuzemoeilijkheid $GM = 7.70$) dan in kleine assortimenten ($GM = 7.01$; $F = 36.5$; $p < .001$). Het algehele effect van de spreiding in attribuutniveaus benadert significantie ($F = 3.0$; $p = .053$). Nadere analyses laten zien dat assortimenten waarin enkele attribuutniveaus niet voorkomen, een significant lagere verwachte keuzemoeilijkheid hebben ($GM = 7.15$) dan de overige assortimenten ($GM = 7.43$ voor medium spreiding en 7.49 voor hoge spreiding; $F = 4.3$; $p = .042$). Wanneer alle attribuutniveaus voorkomen, heeft de spreiding van de niveaus echter geen significant effect ($F = 0.3$; $p = .604$). De samenhang tussen de attributen heeft geen significant effect op de verwachte keuzemoeilijkheid ($F = .6$; $p = .443$). Hypothese 2b wordt dus ten dele ondersteund door de data.

6.4. Discussie

De resultaten voor de verwachte slagingskans ondersteunen hypothese 1, en laten zien dat meer gevarieerde assortimenten een hogere gepercipieerde kans van slagen bieden. De verwachte keuzemoeilijkheid is echter niet in alle gevallen hoger wanneer een assortiment meer variëteit biedt. Consumenten verwachten dat keuzes moeilijker zijn wanneer er meer producten en/of meer attribuutniveaus in het assortiment aanwezig zijn. Een meer evenwichtige verdeling van de attribuutniveaus, of een zwakkere samenhang tussen attributen, doen de verwachte keuzemoeilijkheid echter *niet* stijgen, hoewel de waargenomen variëteit en de verwachte slagingskans dan wel toenemen. Het blijkt dus mogelijk te zijn om meer variëteit te bieden in een assortiment zonder dat het moeilijker lijkt om te kiezen.

7. CONCLUSIES

Dit artikel onderzocht het effect van assortimentsvariëteit op de verwachtingen die consumenten hebben over de slagingskans en keuzemoeilijkheid in een winkel. Door de variëteit van een assortiment te verhogen kunnen detaillisten de perceptie van de slagingskans voor hun klanten vergroten. Detaillisten vrezen echter dat de keuzemoeilijkheid hiermee ook toeneemt, en dat dit klanten afschrikt. De studie heeft aangetoond dat keuzemoeilijkheid weliswaar toeneemt wanneer nieuwe producten of attribuutniveaus in het assortiment worden geïntroduceerd, maar niet wanneer de spreiding van de attribuutniveaus hoger is en wanneer de samenhang tussen de attributen zwakker is. Het is dus mogelijk om de verwachte slagingskans van een assortiment te laten toenemen, zonder dat de verwachte keuzemoeilijkheid stijgt.

7.1. Beperkingen

Dit onderzoek heeft enkele beperkingen. In de eerste plaats bekijkt het een situatie waarin consumenten geen duidelijke voorkeur voor een bepaald product hebben. Wanneer consumenten een favoriet product hebben, zullen ze misschien minder gevoelig zijn voor veranderingen in de variëteitcomponenten zolang dit favoriete product aanwezig blijft (Broniarczyk, Hoyer & McAlister 1998). Door een productcategorie te

nemen waar respondenten relatief weinig directe ervaring en kennis over hebben, werd variëteit in dit onderzoek benadrukt en apart onderzocht zonder storende effecten van productpreferentie. Wanneer marktonderzoekers juist geïnteresseerd zijn in productcategorieën waar sterke voorkeuren voor producten aanwezig zijn, wordt aangeraden om naast de variëteitcomponenten ook mogelijke heuristieken van variëteit (zoals schapruimte en aanwezigheid van een favoriet product) te onderzoeken.

In de tweede plaats heeft variëteit verschillende kosten en baten voor de consument, naast slagingskans en keuzemoeilijkheid. Winkelen is bijvoorbeeld een vrijetijdsbesteding geworden voor veel mensen, en dit kan gevolgen hebben voor het soort assortimenten waarin zij geïnteresseerd zijn. Gevarieerde assortimenten zouden in dit geval meer stimulatie en afleiding kunnen bieden. Voor consumenten die winkelen als vrijetijdsbesteding ('fun shoppers') is keuzemoeilijkheid waarschijnlijk van minder belang dan voor consumenten die snel even een product willen kopen ('run shoppers'). Dergelijke individuele en/of situationele verschillen zijn in dit onderzoek niet aan bod gekomen.

Ten derde, de effecten van factoren als presentatiewijze, verkooppersoneel, en schapruimte zijn buiten beschouwing gebleven. Deze factoren kunnen echter de keuzemoeilijkheid beïnvloeden. Zo kan het gemakkelijker lijken om te kiezen, wanneer een verkoper de attribuutniveaus aan de koper voorlegt ("u kunt kiezen uit deze vier kleuren sofa's en uit deze vier materialen"), dan wanneer alle verschillende producten worden voorgelegd (Huffman & Kahn 1998).

7.2. Implicaties voor marktonderzoekers en detaillisten

Er zijn diverse implicaties van het onderzoek, zowel voor marktonderzoekers als voor detaillisten. Het onderscheid van variëteit in componenten, die een verschillend effect hebben op keuzemoeilijkheid, geeft aan dat variëteit een meerdimensionaal concept is. Marktonderzoekers wordt dan ook aangeraden om de componenten van variëteit uit te splitsen, en de effecten van deze componenten apart te bekijken. Een gedetailleerde analyse van een assortiment kan aangeven waarom het, bijvoorbeeld, een lage perceptie van variëteit heeft. Het zou kunnen zijn dat sommige attribuutniveaus slechts zelden voorkomen. Het is ook mogelijk dat bepaalde combinaties van attributen helemaal niet voorkomen, wat een indicatie kan geven voor de introductie van succesvolle nieuwe producten.

In dit onderzoek zijn assortimenten aangeboden via het computerscherm. Marktonderzoekers zouden ook kunnen overwegen om bestaande winkelassortimenten te nemen, om winkelschappen na te bouwen in een laboratorium, of om de winkelomgeving via *virtual reality* technieken na te bootsen. Aan elk van deze mogelijkheden zijn voor- en nadelen verbonden. Hoewel echte winkels de meest realistische omgeving bieden voor participanten, is het erg tijdrovend en kostbaar om assortimenten in een winkel te veranderen voor een onderzoek. Vergelijking tussen winkels met verschillende assortimenten kan een oplossing bieden, maar data over vergelijkbare winkels is in de praktijk niet altijd aanwezig. Het nabootsen van assortimenten in een laboratorium of met behulp van de computer kan dan een goed alternatief zijn. Met een dergelijke opzet kunnen verschillende versturende factoren (andere klanten, hoeveelheid voorraad van producten, en dergelijke) onder controle gehouden worden, en kan gericht gekeken worden naar de causale invloed van een verandering van assortimentsopbouw. Deze experimentele onderzoeksmethode zal met name goede inzichten geven in de directio-

nele effecten van een verandering (bijvoorbeeld of de keuzemoeilijkheid zal stijgen of dalen bij een bepaalde aanpassing van het assortiment). Voor het doen van specifieke voorspellingen over bijvoorbeeld het aantal consumenten dat een assortiment prefereert zal het realisme van de winkelomgeving belangrijker worden. Hier kunnen in de toekomst de vernieuwde *virtual reality* technieken nuttig blijken te zijn.

Assortimentsonderzoek is niet alleen relevant voor detaillisten. Ook producenten bieden vaak een set van producten aan klanten aan. Inzicht in de percepties en evaluaties van deze productsets kan ook voor hen erg interessant zijn. Daarnaast zullen zowel detaillisten als producenten geïnteresseerd zijn in de productkeuzes die consumenten maken. Aangezien onderzoek heeft aangetoond dat deze keuzes beïnvloed worden door de andere producten die in een assortiment worden aangeboden (Drolet, Simonson & Tversky 2000; Prelec, Wernerfelt & Zettelmeyer 1997) kan juist het opnemen van productkeuzes binnen assortimentsonderzoek nieuw inzicht opleveren.

Dit artikel geeft richtlijnen voor detaillisten die veel variëteit willen aanbieden, zonder het erg moeilijk te maken voor hun klanten om een product te kiezen. Door hun assortimenten goed te managen, en spreiding van attribuutniveaus hoog te houden en samenhangen tussen attributen zwak, kunnen zij dit bereiken. Ook binnen een beperkte schapruimte kan de variëteit van een assortiment op deze manier verhoogd worden, terwijl de keuzemoeilijkheid hanteerbaar blijft.

Bij extreem grote internet assortimenten, zoals de 200.000 producten van JCPenney die in de inleiding genoemd worden, zal keuzemoeilijkheid hoog zijn. Het zou dan niet verwonderlijk mogen zijn dat het verwijderen van producten uit het assortiment kan leiden tot een stijging van de verkoop. Zo laten Boatwright en Nunes (2001) zien dat de eliminatie van bijna de helft van de producten van een internet supermarkt leidt tot een gemiddelde verkoop stijging van 11% in de verschillende productassortimenten. Dit toont aan dat een voortdurende verhoging van het aantal producten in een assortiment de verkopen kan benadelen.

De componenten van variëteit kunnen tevens gekoppeld worden aan de positionering van winkels. Wanneer de spreiding over attribuutniveaus hoog is, en de samenhang tussen attributen zwak, biedt een winkel veel variëteit. Deze '*variety stores*' hebben een breed assortiment, met zeer verschillende producten. Zij bieden voor elk wat wils. Daarentegen kenmerken *speciaalzaken* zich door een lage spreiding over attribuutniveaus en een zwakke samenhang tussen attributen. Deze winkels hebben een beperkte groep producten, maar binnen deze groep zijn allerlei verschillende producten beschikbaar. Een totaal andere assortimentvorm is het '*scrambled assortment*', waarbij de winkel een hele verscheidenheid aan producten heeft, maar in alle verschillende groepen een beperkte keus. Hierbij is de spreiding over attribuutniveaus hoog, en de samenhang tussen attributen sterk. In zo'n assortiment is het voor de consument moeilijk te bepalen welke producten precies aanwezig zullen zijn. Tenslotte bestaan er winkels met een *beperkt assortiment*. Deze winkels hebben een beperkte groep producten, en binnen deze groep enkele sterk op elkaar lijkende producten. Dit kan een winkel zijn die zich richt op een niche in de markt. Afhankelijk van de positionering die door een winkel gekozen is, zal de assortimentstrategie en de hoeveelheid variëteit die wordt aangeboden anders zijn. Een inventarisatie van de variëteitcomponenten kan aangeven in hoeverre deze strategische positionering uitgewerkt is in de assortimentsopbouw en welke verbeteringen mogelijk zijn.

7.3. Toekomstig onderzoek

Toekomstig onderzoek zou kunnen nagaan hoe de voorkeur van consumenten voor gevarieerde assortimenten afhangt van individuele kenmerken. Zo kunnen experts, vanwege hun grote kennis in een productcategorie, misschien beter omgaan met variëteit dan leken. De perceptie van keuzemoeilijkheid zal dan voor een gegeven assortiment lager zijn bij experts dan bij leken. Bovendien zullen leken hun oordeel over variëteit wellicht vooral baseren op assortimentsomvang, terwijl experts ook de verschillen tussen de producten zien, en de spreiding over attribuutniveaus en samenhang tussen attributen meenemen in hun oordeel over variëteit.

Het effect van de ordening van producten op variëteitperceptie en assortimentsevaluatie, is een andere vraag voor toekomstig onderzoek. Eerder onderzoek heeft aangetoond dat geordende assortimenten anders gewaardeerd worden dan ongeordende assortimenten (Hoch, Bradlow & Wansink 1999). Het effect van verschillende manieren van ordening is echter niet bekend. Winkels kunnen hun assortimenten ordenen op basis van merk (soep), productvariant (vloeibaar wasmiddel), grootte (televisies), of gelegenheid (wenskaarten). Binnen deze eerste ordening kunnen de producten weer verder geordend worden. Assortimenten kunnen verschillen in de mate van ordening, het belang van de attributen waarop geordend wordt, en de manier van ordening (in rijen, kolommen, of blokken). Deze variabelen zouden de variëteitperceptie van consumenten kunnen beïnvloeden.

Een andere mogelijke richting voor toekomstig onderzoek is het verschil tussen de verwachtingen en ervaringen van consumenten. De meeste mensen hebben wel eens een ervaring waarbij ze tevoren verwachtten om gemakkelijk een keuze te kunnen maken, maar de daadwerkelijke keuze moeilijker blijkt dan gedacht. Inzicht in de redenen waarom verwachtingen en ervaringen niet met elkaar overeenkomen, en manieren om dit te managen, kan een interessant onderzoeksgebied zijn voor marktonderzoekers die detaillisten willen helpen met hun assortimentssamenstelling.

LITERATUUR

- Beach, Lee Roy, and Terence R. Mitchell, 1978, "A Contingency Model for the Selection of Decision Strategies," *Academy of Management Review*, 3, 439-449.
- Bettman, James R., Mary Frances Luce, and John W. Payne, 1998, "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25 (3), 187-213.
- Boatwright, Peter, and Joseph C. Nunes, 2001, "Reducing Assortment: An Attribute-Based Approach," *Journal of Marketing*, 65 (3), 50-63.
- Broniarczyk, Susan M., Wayne D. Hoyer, and Leigh McAlister, 1998, "Consumers' Perceptions of the Assortment Offered in a Grocery Category: The Impact of Item Reduction," *Journal of Marketing Research*, 35 (2), 166-176.
- Chatterjee, Subimal, and Timothy B. Heath, 1996, "Conflict and Loss Aversion in Multiattribute Choice: The Effects of Trade-Off Size and Reference Dependence on Decision Difficulty," *Organizational Behavior and Human Decision Processes*, 67 (2), 144-155.
- Chiang, Jeongwen, and Ronald T. Wilcox, 1997, "A Cross-Category Analysis of Shelf-Space Allocation, Product Variety, and Retail Margins," *Marketing Letters*, 8, 183-191.
- Converse, P.D., and Charles Spencer, 1942, "What Housewives Think of the Super-Market," *Journal of Marketing*, 6 (4), 371-374.
- Cox, W. Michael, and Richard Alm, 1998, "The Right Stuff; America's Move to Mass Customization," *Federal Reserve Bank of Dallas 1998 Annual Report*, 3-26.

- Drolet, Aimee, Itamar Simonson, and Amos Tversky, 2000, "Indifference Curves that Travel with the Choice Set," *Marketing Letters*, 11 (3), 199-209.
- Estienny, Corinth, 2000, "Online with the Front Line," *Discount Merchandiser*, 40 (1), 37-40.
- Hendrick, Clyde, Judson Mills, and Charles A. Kiesler, 1968, "Decision Time as a Function of the Number and Complexity of Equally Attractive Alternatives," *Journal of Personality and Social Psychology*, 8 (3), 313-318.
- Hoch, Stephen J., Eric T. Bradlow, and Brian Wansink, 1999, "The Variety of an Assortment," *Marketing Science*, 18 (4), 527-546.
- Huffman, Cynthia, and Barbara E. Kahn, 1998, "Variety for Sale: Mass Customization or Mass Confusion?" *Journal of Retailing*, 74 (4), 491-513.
- Iyengar, Sheena S., and Mark R. Lepper, 2000, "When Choice is Demotivating: Can One Desire Too Much of a Good Thing," *Journal of Personality and Social Psychology*, 79 (6), 995-1006.
- Kahn, Barbara E., 1998, "Marketing in the 21st Century; Dynamic Relationships With Customers: High-Variety Strategies," *Journal of the Academy of Marketing Science*, 26 (Winter), 45-53.
- Kahn, Barbara E., and Donald R. Lehmann, 1991, "Modeling Choice Among Assortments," *Journal of Retailing*, 67 (3), 274-299.
- Kahn, Barbara E., and Leigh McAlister, 1997, *Grocery Revolution, The New Focus on the Consumer*, Reading: Addison-Wesley Educational Publishers, Inc.
- Kellogg, Orson, and Veera Bhatnagar, 2002, *Authorware 6, Training from the Source*, Berkeley: Macromedia Press.
- Klein, Noreen M., and Manjit S. Yadav, 1989, "Context Effects on Effort and Accuracy in Choice: An Enquiry into Adaptive Decision Making," *Journal of Consumer Research*, 15 (4), 411-421.
- Koелеmeijer, Kitty, and Harmen Oppewal, 1999, "Assessing the Effects of Assortment and Ambience: A Choice Experimental Approach," *Journal of Retailing*, 75 (3), 319-345.
- Malhotra, Naresh K., 1982, "Information Load and Consumer Decision Making," *Journal of Consumer Research*, 8 (4), 419-430.
- Payne, John W., James R. Bettman, and Eric J. Johnson, 1988, "Adaptive Strategy Selection in Decision Making," *Journal of Experimental Psychology*, 14 (3), 534-552.
- Prelec, Drazen, Birger Wernerfelt, and Florian Zettelmeyer, 1997, "The Role of Inference in Context Effects: Inferring What You Want from What Is Available," *Journal of Consumer Research*, 24 (1), 118-125.
- Rafferty, Dan, 1993, "Trim the Dead Wood," *Progressive Grocer*, 72 (September), 42-43.
- Ratchford, Brian T., 1982, "Cost-Benefit Models for Explaining Consumer Choice and Information Seeking Behavior," *Management Science*, 28 (2), 197-212.
- Ratner, Rebecca K., Barbara E. Kahn, and Daniel Kahneman, 1999, "Choosing Less-Preferred Experiences for the Sake of Variety," *Journal of Consumer Research*, 26 (1), 1-15.
- Shugan, Steven M., 1980, "The Cost of Thinking," *Journal of Consumer Research*, 7 (2), 99-111.
- Stone, Dan N., and David A. Schkade, 1991, "Numeric and Linguistic Information Representation in Multiattribute Choice," *Organizational Behavior and Human Decision Processes*, 49 (1), 41-59.
- Van Herpen, Erica, 2001, *Perceptions and Evaluations of Assortment Variety*, CentER Dissertation Series, Tilburg University.
- Van Herpen, Erica, and Rik Pieters, 2002, "The Variety of An Assortment: An Extension to the Attribute-Based Approach," *Marketing Science*, 21 (3), 331-341.

8. *De toepassing van het Mokkenschaalmodel voor acquisitiepatroonanalyse in de financiële dienstverlening*¹

L.J. PAAS

SAMENVATTING

Inzicht in de combinaties en volgordes waarin consumenten producten aanschaffen kan belangrijk zijn voor relatiemarketing. Deze marketingbenadering is namelijk gebaseerd op de verkoop van meerdere producten in plaats van een enkel product van dezelfde afnemer. De onderstaande studie is gericht op het vinden van een geschikte techniek voor het bundelen van productbezit en voor het bepalen van de volgordes waarin consumenten producten in deze bundels aanschaffen. We demonstreren de toepasbaarheid van Mokkenschaalanalyse voor deze doeleinden. De toepassing betreft twee soorten gegevens: indicaties voor productbezit in crosssectie enquêtegegevens en in crosssectie transactiegegevens. De genoemde toepassingen geven consistente resultaten en komen overeen met de manier waarop consumenten de producten daadwerkelijk afnemen. Verder demonstreren we dat het Sequentiële Logit Beslismodel bruikbaar is om de factoren die van invloed zijn op het bezit van verschillende producten in een acquisitievorgorde te achterhalen. De beschreven toepassing is in de financiële dienstverlening, maar het is aannemelijk dat de benadering ook toepasbaar is in andere sectoren.

1. INLEIDING

Onderzoek naar consumentengedrag is meestal gericht op het verklaren van één enkel type gedrag. Echter, de huidige studie is gericht op de combinaties en volgordes waarin verschillende consumentengedragingen voorkomen. We geven aan dat dit soort combinaties en volgordes relevant zijn voor, bijvoorbeeld, een bedrijf dat wil inspelen op behoeften gedurende de hele klantlevenscyclus. Hiervoor is inzicht nodig in combinaties en in ontwikkelingen van consumentenbehoeften over de tijd. Het benodigde onderzoek is uitvoerbaar met een benadering die we acquisitiepatroonanalyse (APA) noemen.

APA is bruikbaar voor het definiëren van de productaankopen van consumenten op een relatief algemeen niveau; de definiëring is niet op het niveau van een enkel product, zoals een inboedelverzekering of een auto, maar op een niveau van categorieën, zoals verzekerings- of mobiliteitsproducten. Producten die consumenten vaak in combinatie

¹ Dit hoofdstuk is gebaseerd op het proefschrift “Acquisition pattern analysis with Mokken scales: Applications in the financial services market” (2002) van L.J. Paas, promotores: A.A.A. Kuijlen en W.F. van Raaij.

bezitten, worden aan dezelfde bundel toegewezen en beschouwen we als complementair. Deze bundels heten ‘karakteristieke verzamelingen’ (McFall, 1969). APA richt zich, echter, niet alleen op het bundelen van producten maar ook op het vinden van hiërarchische volgordes waarin producten worden aangeschaft (acquisitievorgordes), ofwel (LaPlaca et al., 1985, p. 126):

“Consumers think of their household purchases in terms of sets of similar items rather than individual products. For example, consumer durables could be grouped into a “comfort set” (e.g., electric blanket, air conditioner, automatic washer, etc.), a “clean set”, a “cook set”, an “entertainment set”, etc.” Acquisition patterns could then be identified across these sets of consumer durables... Kasulis et al. (1979) postulated that common patterns of consumer durable acquisitions resulted from a common utility structure among the population in question. The utility structure provided a rational basis for prioritizing an evoked set of possible consumer durable alternatives which were in competition for the limited resources available to purchase the goods.”

APA is niet alleen toepasbaar op duurzame producten maar ook op financiële producten (Kamakura et al., 1991; Soutar en Cornish-Ward, 1997). De relevantie van deze benadering voor relatiemarketing in de financiële dienstverlening beschrijven we aan de hand van een voorbeeld. Stel dat consumenten, die een beleggingsfonds kopen, meestal al een spaarrekening hebben. Een logische volgende stap voor spaarrekeningbezitters is dan de afname van een beleggingsfonds. Onder deze omstandigheden dient een financiële dienstverlener haar spaarrekeningbezitters een beleggingsfonds aan te bieden. Dit is belangrijke informatie, want als de consument het beleggingsfonds bij de concurrent afneemt, dan verliest financiële dienstverlener X een deel van de mogelijke toekomstige omzet.

Ondanks het potentieel van APA zijn er methodologische vragen die we moeten beantwoorden. Het is, bijvoorbeeld, niet bekend welke techniek geschikt is voor de toepassing van APA. Ook weten we niet welke soorten data over productbezit bruikbaar zijn. Zo analyseren financiële dienstverleners vaak transactiegegevens in plaats van enquêtegegevens. Transactiegegevens zijn gebaseerd op activiteiten van klanten bij de dienstverlener en bevatten vaak ook indicaties voor productbezit. Een probleem is dat er weinig bekend is over de toepasbaarheid van APA op transactiegegevens.

Nog een probleem is dat niet alleen de combinaties waarin consumenten producten bezitten en volgordes van acquisitie interessant zijn voor relatiemarketing, ook verbanden tussen dit soort informatie en andere gegevens, zoals de leeftijd en het inkomen van de consument, kunnen een rol spelen. Als bekend is dat vooral spaarrekeningbezitters in de leeftijd van 30 tot en met 50 jaar met een bovenmodaal inkomen geneigd zijn om een beleggingsfonds af te nemen, dan is het aan te raden om de verkoop van beleggingsfondsen vooral op deze categorie spaarrekeningbezitters te richten (indien zij het laatstgenoemde product nog niet bezitten). Er is echter weinig bekend over de geschiktheid van verschillende technieken voor het uitvoeren van onderzoek naar de relatie tussen acquisitievorgordes en andere variabelen. In aansluiting op het bovenstaande hebben we drie doelstellingen in dit hoofdstuk:

1. Het omschrijven van de toepassing van APA in de financiële dienstverleningsmarkt. Deze omschrijving is input voor de twee onderstaande methodologische doelstellingen.
2. Het selecteren van een geschikte techniek voor de uitvoering van APA en het controleren of deze techniek toepasbaar is op zowel enquêtegegevens als transactiegegevens.

3. Ten behoeve van de praktische waarde van APA bepalen we een techniek voor het analyseren van verbanden tussen de uitkomsten van APA en andere gegevens over de consument.

In aansluiting op het eerste punt stellen we in paragraaf 2 de theorie over APA aan de orde en beschrijven we deze in termen van de financiële markt. Voor de tweede doelstelling bespreken we in paragraaf 3 een methodologie voor de uitvoering van APA. Het zal blijken dat zowel karakteristieke verzamelingen als acquisitievorgordes zijn te herleiden uit data van één punt in tijd. Vanwege het tekort aan longitudinale gegevens, is dit een belangrijk voordeel. In paragraaf 4 introduceren we de data die is geanalyseerd ten behoeve van de onderstaande rapportage en wordt ook de analyseopzet omschreven. Daarna passen we in paragraaf 5 APA toe op enquêtegegevens en in paragraaf 6 op transactiegegevens uit een marketingdatabase van een Nederlandse financiële dienstverlener. Voor de derde doelstelling introduceren we in paragraaf 7 een benadering die een koppeling ondersteunt met andere gegevens zoals de leeftijd en het inkomen van de consument. Het hoofdstuk wordt afgesloten met een discussie (paragraaf 8).

2. THEORIE

Hieronder beschrijven we de toepassing van APA op het bezit van financiële producten. We bespreken eerst in paragraaf 2 de theorie waarop APA is gebaseerd en daarna beschrijven we de twee basisbegrippen, zijnde: karakteristieke verzamelingen en acquisitiepatronen. Om te beginnen, de veronderstelling die ten grondslag aan APA ligt, is dat productbezit samenhangt met financiële behoeften van consumenten; attributen van financiële producten maken het voor consumenten mogelijk om aan saillante financiële behoeften te voldoen. De consument koopt het product, wanneer bevrediging van de relevante behoefte meer waarde voor haar/hem heeft dan het geld en de tijd die nodig zijn voor de aankoop en het bezit van het product (Clawson en Vinson, 1978; Gutman en Vinson, 1979). Een probleem is echter, dat consumentengedrag niet alleen wordt beïnvloed door latente behoeften, maar ook door, bijvoorbeeld, beperkingen in het productaanbod, de bekendheid van de consument met het aanbod op de markt en de cognitieve capaciteiten van het individu (Engel et al., 1986).

Maar niet alleen benaderingen die zijn gebaseerd op productbehoeften hebben beperkingen. Kuijlen (1993) bespreekt de theoretische en methodologische problemen die een rol spelen bij de toepassing van attitudemodellen voor het verklaren en voorspellen van consumentengedrag (belangrijke modellen zijn geïntroduceerd door, bijvoorbeeld, Azjen en Fishbein (1980), Engel et al. (1986) en Kotler (1971). Nu kunnen deze modellen een verklaring geven voor het gedrag van sommige individuen in bepaalde situaties, maar het is onbekend voor welke individuen en in welke situaties de modellen opgaan (Verhallen en Pieters, 1984). Kuijlen (1993) geeft verder aan dat attitudemodellen vooral in de financiële dienstverlening moeilijk toepasbaar zijn, omdat productaankopen vaak niet samenhangen met consumentenbehoeften. Dit komt, bijvoorbeeld, door de rol van tussenpersonen, die zich vaak richten op hogere commissies in plaats van klantbehoeften, en de lage betrokkenheid van consumenten bij financiële producten. Kuijlen (1993) concludeert dat consumentenonderzoek in de financiële dienstverleningssector niet alleen over latente psychologische concepten moet gaan, maar ook over aankoopgedrag.

In aansluiting op het bovenstaande is deze studie gericht op een techniek die is gebaseerd op productaankopen, acquisitiepatroonanalyse (APA). Maar voor een optimale benutting van APA moet deze benadering verder worden gedefinieerd. Eerder is in McFall (1969) aangegeven dat consumenten producten in zogenaamde karakteristieke verzamelingen indelen, zoals de huishoudelijke verzameling, de keukenverzameling en de verzameling van audio- en video apparatuur. Er werd verondersteld dat consumenten die al één of meer producten uit een dergelijke verzameling bezitten een verhoogde kans hebben op de aankoop van een ander product uit dezelfde karakteristieke verzameling. Echter, andere onderzoekers gaven aan dat APA bruikbaar is voor het verrichten van onderzoek naar de volgordes waarin consumenten producten aanschaffen (bijvoorbeeld, Kamakura et al., 1991; Kasulis et al., 1979; Soutar and Cornish-Ward, 1997). De derde interpretatie van APA verenigt deze twee visies op APA. LaPlaca et al. (1985) en Paas (1998) suggereerden dat APA bruikbaar is voor onderzoek naar zowel karakteristieke verzamelingen als voor onderzoek naar acquisitiepatronen. In de onderstaande discussie sluiten wij ons aan bij deze laatstgenoemde interpretatie van APA. We voegen verder toe dat de volgende twee stappen in APA sequentieel genomen dienen te worden:

1. *Onderzoek naar karakteristieke verzameling*: het indelen van producten die vaak gecombineerd in het bezit zijn bij dezelfde consumenten.
2. *Onderzoek naar acquisitiepatronen*: zoeken naar volgordes waarin consumenten producten in elke karakteristieke verzameling aanschaffen.

Beide stappen zijn relevant voor relatiemarketing, het doeleinde waarvoor APA inzetbaar is. Onderzoek naar karakteristieke verzamelingen is interessant, omdat de resultaten van deze stap van APA inzicht geven in de complementariteit van verschillende producten. Hiermee weet de marketeer dus welke pakketten of combinaties van producten kunnen worden aangeboden aan de consument (Poiesz en Van Raaij, 2002). Maar ook inzicht in de volgorde waarin consumenten producten aanschaffen is van belang. We weten dan welk product de meest logische extensie vormt op het pakket van producten dat de consument momenteel bezit. Let wel, volgordes zijn vooral interessant in zoverre dat deze opgaan voor producten in dezelfde karakteristieke verzameling. Voor producten in dezelfde karakteristieke verzameling geldt namelijk dat de aankoop van één product uit de verzameling impliceert dat er een grotere kans is dat andere producten uit de verzameling ook worden aangeschaft. Het belang van het laatste is aan de hand van een fictief voorbeeld uit te leggen. Stel dat 90% van de consumenten product A bezit en dat 50% van de bezitters van product A ook product B hebben. Verder heeft 80% van de consumenten die product A niet bezitten, product B wel. In dit geval wordt product A meestal voor B aangeschaft, maar is het niet aan te raden om product B alleen aan bezitters van product A aan te bieden. Aangezien er relatief veel product B bezitters zijn te vinden onder consumenten die product A niet bezitten, is het onwaarschijnlijk dat product B een vervolgstap vormt op product A.

Op dit punt merken we op dat, ondanks de centrale rol van vertoond gedrag (productaankopen), APA niet als een “revealed preference approach” kan worden beschouwd. APA heeft weliswaar deze gerichtheid op vertoond gedrag gemeen met de “revealed preference approach”, maar de laatstgenoemde benadering zou bij de toepassing op aankoopgedrag gebaseerd zijn op de veronderstelling dat een behoefte direct herleidbaar is uit de aankoop van een enkel product. Bij de toepassing van APA veronderstel-

len we daarentegen dat er een behoefte is op een generiek niveau, welke de aankoop van meerdere producten beïnvloedt. Een generieke behoefte beschouwen we als een klasse van specifieke behoeften die saillant kunnen zijn ten gevolge van hetzelfde algemene probleem. Zo zijn, bijvoorbeeld, de specifieke behoeften voor een auto, een fiets of een abonnement voor de trein of bus, alle relevant voor de generieke behoefte om een transportprobleem op te lossen. APA verschilt van de ‘revealed preference approach’ in de zin dat er rekening wordt gehouden met de mogelijkheid dat er producten kunnen worden gekocht die niet samenhangen met een generieke behoefte van de consument. Deze ‘mismatch’ is dan te wijten aan, bijvoorbeeld, cognitieve beperkingen van de koper, beperkingen in het aanbod of onbekendheid met het totale aanbod binnen de financiële markt. APA veronderstelt dus, in tegenstelling tot de ‘revealed preference approach’, een probabilistisch verband tussen een latente generieke behoefte van consumenten en productaankopen.

In APA definiëren we welke producten samenhangen met een zelfde generieke behoefte. Dit geschiedt met behulp van zogenaamde schaaltechnieken, zoals het model van Guttman, de Mokkenschaal of de Raschschaal. Deze technieken stammen uit de discipline van ‘Item Response Theory (IRT)’ en zijn gebaseerd op de assumptie dat er een verzameling stimuli (items) is, die aan een verzameling personen wordt geïntroduceerd (subjecten). Schaaltechnieken maken het mogelijk om te bepalen in hoeverre de reacties van subjecten op de items worden bepaald door één latente eigenschap (Lord and Novick, 1968; geciteerd in Molenaar, 1982, blz. 7):

“Any theory of latent trait supposes that an individual’s behaviour can be accounted for, to a substantial degree, by defining certain human characteristics called *traits*, quantitatively estimating the individual’s standing on each of these traits, and then using the numerical values obtained to predict or explain performance in relevant situations (...) Much of psychological theory is based on trait orientation, but nowhere is there any necessary implication that traits exist in any physical or physiological sense. It is sufficient that a person behaves as if she/he were in possession of an amount of each of a number of relevant traits and that she/he behaves as if these amounts substantially determined her/his behaviour.”

Indien we het wel of niet bezitten van een product registreren als een binair item, dan zijn schaaltechnieken ook toepasbaar op het bezit van financiële producten. Wanneer een subject de waarde ‘1’ heeft op een dergelijk item, dan bezit deze persoon het betreffende product, een waarde ‘0’ betekent dat zij/hij het product niet bezit. In een data set met n subjecten en k producten krijgen we dus een bezitsmatrix van n maal k nullen en énen. In deze matrix representeert rij n het totale productbezit van subject n , in kolom k is aangegeven wie product k bezit.

Als we beschikking hebben over deze soort data, dan zijn schaaltechnieken dus bruikbaar ter ondersteuning van een definiëring van klantbehoeften op het niveau van een generieke behoefte; de latente eigenschap die het productbezit verklaart, kan worden geïnterpreteerd in termen van een generieke behoefte. Maar een kanttekening is op zijn plaats. De definiëring van karakteristieke verzamelingen (de verzameling producten die vanwege dezelfde generieke behoefte wordt aangeschaft) moet niet alleen gebaseerd zijn op de resultaten van schaalanalyse (McFall, 1969, p. 51):

“Most of the established sets will overlap since they will all be sub-sets of the universal set – the total market offering at a given place and time (...) A product should be included in a characteristic set when its attributes are similar to those of the set in question.”

Vergelijkbare overwegingen spelen een rol in IRT. Mokken (1971) bespreekt een verzameling van twaalf items die op één schaal indeelbaar zijn of die kunnen worden ingedeeld op vier schalen, welke samenhangen met vier correlerende latente variabelen. Net als McFall (1969) geeft Mokken (1971) aan dat de analist de modelkeuze dient te baseren op inhoudelijke overwegingen. In termen van APA betekent dit dat we rekening moeten houden met de kenmerken van financiële producten, naast de resultaten van schaaltechnieken, bij het formuleren van karakteristieke verzamelingen.

Uit de inleiding op dit hoofdstuk blijkt al dat na het vinden van karakteristieke verzamelingen, APA bruikbaar is voor de bepaling van volgordes waarin consumenten producten in een karakteristieke verzameling aanschaffen. In termen van de toepassing van APA op bezitspatronen van financiële producten stelden Kamakura et al. (1991, blz. 331) het volgende op basis van acquisitievorgordes die zijn gevonden in de VS:

“The combination of financial-need priorities and resource availability at different stages of the household’s life-cycle influences the sequence in which different financial services are acquired (...) A household is expected to satisfy its basic objectives (such as liquidity, cash reserve and insurance) before allocating funds for satisfying higher-order investment objectives for capital appreciation, inflation, hedging and speculation (...) Thus, the relationship between products may be viewed as ‘complementary over time’ (coverage of basic objectives/products enhances the probability that higher objectives/products will be attended to)”.

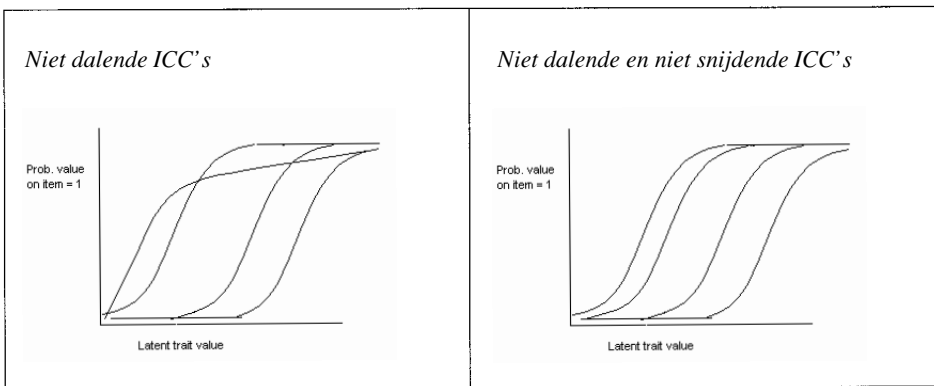
Het gebruik van schaaltechniek voor onderzoek naar karakteristieke verzamelingen en acquisitievorgordes wordt verder toegelicht in de volgende paragraaf.

3. EEN METHODOLOGIE VOOR APA

Een schaaltechniek voor APA dient onderzoek naar zowel karakteristieke verzamelingen als naar acquisitievorgordes te ondersteunen. Mathematisch gezien voldoet Mokkenschaaanalyse aan deze eisen (Paas en Molenaar, 2002). Maar empirisch onderzoek is nodig om na te gaan of de resultaten in echte data ook voldoen. De volgende Mokkenschaa-modellen zijn relevant voor APA (Paas, 1998): (1) Het model voor Monotone Homogeniteit (MH-model) en (2) Het model voor dubbele monotonie (DM-model). Met het MH-model kan een onderzoeker toetsen of subjecten aan de hand van een bepaalde set producten hiërarchisch kunnen worden geordend. Het DM-model is bestemd om te toetsen of items, zoals producten, hiërarchisch zijn te ordenen. Hieronder wordt aangegeven dat het ordenen van subjecten belangrijk is in termen van karakteristieke verzamelingen; ordenen van items heeft betrekking op acquisitiepatronen. We beschrijven eerst kort de basisbegrippen van deze modellen en daarna bespreken we de toepassing voor APA verder. We vermijden technische uitwijdingen, voor meer informatie over Mokkenschaaanalyse verwijzen we naar Mokken (1971) en Sijtsma en Molenaar (2002).

Voor de beschrijving van Mokkenschaaanalyse komen we terug op het eerder gedefinieerde begrip “latente variabele”. Een uitgangspunt voor de toepassing van Mokkenschaaanalyse ten behoeve van APA is dat bezit van producten sterk wordt beïnvloed door de positie van producten op een latente variabele en de waarde die consumenten hebben op deze variabele. Deze posities van producten en consumenten zijn niet direct te meten, maar leiden we af uit de bezitspatronen (de $n \times k$ tabel waarin met nullen en énen is aangegeven welk product elke consument bezit). Nu is de kans dat ieder subject

een bepaald product bezit gerepresenteerd door een zogenaamde “Item Characteristic Curve (ICC)”. Een ICC representeert de relatie tussen de waardes die subjecten hebben op een veronderstelde latente variabele en de antwoorden die zij geven op gemeten items. Deze ICC’s kunnen verschillende vormen aannemen en de vorm bepaalt welke hiërarchische veronderstellingen mogelijk zijn. Zo kunnen de ICC’s niet-dalend zijn, zoals weergegeven in figuur 1.1. In dit geval betekent een hogere positie op een latente variabele een grotere kans dat de positieve antwoordcategorie op het item wordt ingevuld door het subject. Het MH-model veronderstelt dit soort ICC’s en ondersteunt een ordening van subjecten op de latente variabele. ICC’s kunnen naast niet-dalend ook nog niet-snijdend zijn, zie figuur 1.2. Dit sluit aan op de vereisten van het DM-model en betekent dat de ordening van items altijd hetzelfde is, ongeacht de waarde die het subject heeft op de latente variabele. Het laatste verantwoordt de ordening van producten op een latente variabele (Mokken, 1971; Sijtsma en Molenaar, 2002). Verder in deze paragraaf zullen we zien dat deze twee vormen van ordenen van belang zijn voor APA, maar eerst bespreken we de toetsen voor Monotone Homogeniteit en Dubbele Monotonie.



Figuur 1. Voorbeelden van ICC's.

Mokken (1971) toonde aan dat de mate waarin datapatronen voldoen aan de assumpties van het MH-model is vast te stellen aan de hand van de Loevinger coëfficiënt voor homogeniteit. Verschillende varianten van deze coëfficiënt zijn nodig om de consistentie van datapatronen met het MH-model vast te stellen. Er is een test waarmee we de monotone homogeniteit van alle producten in de verzameling kunnen testen (H), van elk afzonderlijk product (H_i) en van twee producten onderling (H_{ij}). We beschouwen een verzameling items als monotoon homogeen, indien aan de volgende condities is voldaan: (1) $H \geq 0,30$; (2) alle $H_i \geq 0,30$; (3) alle $H_{ij} \geq 0,30$; (4) H significant van 0 afwijkt en (5) alle H_i significant van 0 afwijken.

Of bezitspatronen voldoen aan de assumpties van het DM-model is te toetsen met behulp van zogenaamde P-matrices. De $P(+,+)$ -matrix bevat, voor alle mogelijke combinaties van twee producten, het aandeel van de subjecten dat beide producten in bezit heeft. Bijvoorbeeld, element $p(i,j)$ geeft het aandeel subjecten dat zowel product i als product j bezit. In de $P(+,+)$ -matrix zijn de producten geordend naar oplopend bezit, producten die weinig subjecten hebben staan links en boven in de matrix, producten die relatief vaak bezitten staan rechts en onder. In de $P(+,+)$ -matrix dient er voor

het voldoen aan de assumpties van het DM-model sprake te zijn van olopende of gelijkblijvende aandelen in rijen en kolommen, dalingen zijn niet toegestaan. De P(-,-)-matrix ziet er hetzelfde uit, maar presenteert de aandelen subjecten die beide producten niet bezitten. In deze matrix mogen de aandelen in rijen en kolommen juist niet olopen, maar moeten zij dalen of gelijk blijven. In Tabel 1 representeren de vetgedrukte elementen schendingen in de P-matrices (Tabel 1 is een hypothetisch voorbeeld uit Mokken (1971)).

Tabel 1. Fictief voorbeeld van P-Matrices.

Items →		P(1,1) - matrix				
		1	2	3	4	5
↓	1	-	0.12	0.18	0.16	0.18
	2	0.12	-	0.20	0.22	0.25
	3	0.18	0.20	-	0.26	0.38
	4	0.16	0.22	0.26	-	0.42
	5	0.18	0.25	0.38	0.42	-
	p_i	0.20	0.30	0.40	0.50	0.60

Items →		P(0,0)-matrix				
		1	2	3	4	5
↓	1	-	0.62	0.58	0.46	0.38
	2	0.62	-	0.50	0.42	0.35
	3	0.58	0.50	-	0.36	0.38
	4	0.46	0.42	0.36	-	0.32
	5	0.38	0.35	0.38	0.32	-
	p_i	0.20	0.30	0.40	0.50	0.60

In termen van APA hebben het MH- en het DM-model belangrijke inhoudelijke betekenissen (Paas, 1998). Indien patronen in een bezitsmatrix voldoen aan de assumpties van het MH-model, dan behoren deze producten tot dezelfde karakteristieke verzameling. Er is dan één generieke behoefte van invloed op de aankoop van producten en door complementariteit bestaat er een positieve samenhang tussen het bezit van producten. Indien bezitspatronen naast de assumpties van het MH-model ook voldoen aan de assumpties van het DM-model, dan worden producten meestal gekocht in een volgorde die is gebaseerd op bezitspercentages; producten die in het bezit zijn van een relatief groot aandeel van de subjecten worden doorgaans afgenomen voor producten die relatief weinig subjecten hebben. Stel, bijvoorbeeld, dat de bezitspatronen van de volgende drie producten voldoen aan de assumptie van het MH-model en het DM-model: betaalrekening, creditcard en lening. Stel verder dat 90% van de consumenten een betaalrekening heeft, 50% heeft een creditcard en 10% een lening. In dit hypothetische voorbeeld bezitten consumenten deze producten vaak in combinatie en schaffen consumenten doorgaans eerst een betaalrekening aan, daarna een creditcard en als laatste een lening. Verder verhoogt de aanschaf van een betaalrekening de kans dat een creditcard wordt aangeschaft en consumenten die al een creditcard bezitten hebben op hun beurt weer een verhoogde kans op afname van de lening.

In de volgende paragraaf volgt de beschrijving van een studie aan de hand waarvan we onderzoeken of bezitspatronen van producten in echte data ook Mokka-schalen kunnen

vormen. Toetsing is op verschillende soorten data om de algemene toepasbaarheid van Mokka-schaalanalyse voor APA te onderzoeken. Ten behoeve van de praktische relevantie bespreken we ook methoden voor onderzoek naar verbanden tussen de resultaten van APA en de waardes die subjecten hebben op andere variabelen.

4. METHODE

4.1. Data

In de inleiding gaven we al aan dat APA met behulp van het Mokka-schaalmodel is toegepast op de bezitsmatrices van verschillende soorten financiële diensten in enquêtebestanden en op transactiegegevens. In paragraaf 5 passen we APA toe op enquêtegegevens. Dit zijn de meest gebruikte data voor APA. In paragraaf 6 passen we Mokka-schaalanalyse toe op een bezitsmatrix van financiële producten uit een marketing-database van een financiële dienstverlener (transactiegegevens). Deze analyse is relevant, omdat we hiermee algemenere toepasbaarheid van APA aantonen. Dat wil zeggen, we gaan na of APA niet alleen toepasbaar is op enquêtegegevens, zoals in het verleden is gebeurd, maar brengen ook de toepasbaarheid op transactiegegevens in kaart. Verder zijn de resultaten bruikbaar voor de validatie van de in paragraaf 5 gerapporteerde uitkomsten. Indien enquêtegegevens en transactiegegevens een consistent beeld geven van karakteristieke verzamelingen en acquisitiepatronen, is er meer zekerheid over de gevonden patronen. Daarnaast worden de transactiegegevens gebruikt voor onderzoek naar verbanden tussen acquisitiepatronen en andere gegevens. De resultaten van dit onderzoek zijn gerapporteerd in paragraaf 7. Hieronder bespreken we eerst het belang van analyse op transactiegegevens.

Er is een aantal voordelen verbonden aan de toepassing van APA op transactiegegevens. Ten eerste zijn transactiegegevens een neerslag van het feitelijke gedrag uit het verleden en daarom niet beïnvloed door factoren zoals de sociale wenselijkheid van antwoorden, geheugeneffecten, de weigering om bepaalde informatie door te spelen aan de enquêteur, etc. Dergelijke factoren zijn wel van invloed op enquêtegegevens. In de tweede plaats zijn transactiegegevens goedkoper. Voor grotere financiële instellingen is het bijna ondenkbaar om alle klanten te interviewen over productbezit.

Toch is er een nadeel verbonden aan het gebruik van transactiegegevens; er is sprake van inzicht in het bezit van producten bij maar één bedrijf. Een financiële dienstverlener X kan maar in beperkte mate achterhalen (via interviews of overboekingen), welke producten een cliënt bij de concurrent bezit. Dit kan de resultaten van APA beïnvloeden. Zo kunnen de resultaten van APA op transactiegegevens een gevolg zijn van marketingactiviteiten uit het verleden. Bijvoorbeeld, als in het verleden aan klanten bij financiële dienstverlener X alleen product B wordt aangeboden indien zij product A al bezitten, dan wordt product B bij financiële dienstverlener X doorgaans na product A afgenomen. Hier zou er dus sprake zijn van een 'self-fulfilling prophecy'.

Om dit probleem te ondervangen, zouden we transactiegegevens kunnen verrijken met enquêtegegevens. We vragen dan een representatieve steekproef klanten van financiële dienstverlener X, om aan te geven welke producten zij bij concurrenten van X bezitten. Alle klanten interviewen is waarschijnlijk te duur, maar via imputatietechnieken en voorspellingsmodellen kunnen we cliënten, die niet zijn geïnterviewd, scoren op de kans dat zij verschillende producten bij de concurrent bezitten. Deze informatie zou

kunnen dienen als input voor APA. Echter, voor de hieronder besproken analyse zijn dergelijke enquêtegegevens niet beschikbaar. Wel gebruiken we enquêtegegevens ter validatie van karakteristieke verzamelingen en acquisitievorgordes die zijn afgeleid uit transactiegegevens. Indien toepassing van APA op beide soorten data dezelfde resultaten geeft, dan hebben we waarschijnlijk te maken met valide en betrouwbare uitkomsten. Dus de enquêtegegevens zijn bruikbaar naast de beschikbare transactiegegevens.

De hieronder geanalyseerde enquêtegegevens zijn in 1994 verzameld door Telepanel van Universiteit van Amsterdam (thans is het Telepanel ondergebracht bij Centerdata aan de Universiteit van Tilburg). We gebruikten een steekproef van 743 subjecten die representatief is voor de consumenten in Nederland met een jaarinkomen van minstens fl. 100.000,- (45.000 Euro). Deze subjecten gaven aan of ze de volgende vier producten bezitten: (1) Spaarrekening; (2) Koopsompolis; (3) Beleggingsfonds; (4) Aandelen. De bezitsmatrix is als Tabel 2 weergegeven.

Tabel 2. Bezitsmatrix Telepanel data.

Spaarrekening	Koopsompolis	Beleggingsfonds	Aandelen	Aantal subjecten	Proportie van de subjecten
0	0	0	0	53	7,13%
1	0	0	0	187	25,17%
1	1	0	0	101	13,59%
1	1	1	0	69	9,29%
1	1	1	1	118	15,88%
0	1	0	0	27	3,70%
0	0	1	0	7	0,94%
0	0	0	1	4	0,54%
1	0	1	0	70	9,42%
1	0	0	1	23	3,10%
0	1	1	0	8	1,08%
0	1	0	1	1	0,13%
0	0	1	1	7	0,94%
1	1	0	1	28	3,77%
1	0	1	1	33	4,44%
0	1	1	1	7	0,94%
Totaal				743	100%

Verder is er is een representatieve steekproef geanalyseerd van 6490 subjecten uit de marketingdatabase van een Nederlandse financiële dienstverlener (financiële dienstverlener X). We hebben informatie over het bezit van de volgende producten: (1) Betaalrekening; (2) Spaarrekening; (3) Beleggingsfonds; (4) Aandelen. Deze gegevens worden gebruikt ter validatie van de resultaten van de toepassing van APA op de hierboven beschreven enquêtegegevens.

Het bestand met de transactiegegevens is ook bruikbaar om aan te tonen hoe het verband tussen de resultaten van APA en andere gegevens is te onderzoeken. Tabel 3 presenteert de beschikbare variabelen voor het onderzoek naar de geneigdheid van subjecten om een volgende stap te nemen in de genoemde acquisitievorgorde. We gebruiken in de rest van de tekst de afkortingen in Tabel 3. Van belang bij de bespreking van de analyses is dat inkomen is weergegeven met twee variabelen. De eerste variabele, 'Income_Amount', representeert het bedrag aan inkomen dat het subject laat storten op

haar/zijn betaalrekening bij financiële dienstverlener X. Deze variabele neemt de waarde '0' aan als er geen inkomen wordt overgemaakt. De tweede variabele, 'Income_Transfer', neemt ook de waarde '0' aan indien de klant geen inkomen laat overmaken naar een betaalrekening bij financiële dienstverlener X, maar is '1' indien dit wel gebeurt. Dit gebruik van dummy-variabelen maakt de omgang met ontbrekende waarden (missings) mogelijk, zoals het geval is voor inkomen als dit niet wordt overgemaakt naar financiële dienstverlener X. De gevonden effecten hebben daarnaast ook een inhoudelijke betekenis. Bijvoorbeeld, als alleen Income_Transfer significant effect heeft, in een bepaalde fase van de acquisitievolverde, dan is alleen het wel of niet overboeken van inkomen naar financiële dienstverlener X relevant. Wanneer daarnaast Income_Amount effect heeft, dan is de hoogte van het overgemaakte inkomen ook van belang.

Tabel 3. Variabelen voor het SLD-model.

Afkorting	Beschrijving
Income_Transfer	Indicatie of er inkomen wordt overgemaakt naar financiële dienstverlener X
Income_Amount	Hoeveelheid overgemaakt inkomen
Years_Client	Aantal jaren dat iemand klant is bij financiële dienstverlener X
Age	Leeftijd van het subject in jaren

4.2. Analyses

Met behulp van het programma MSP (Sijtsma en Molenaar, 2002) is Mekkenschaal-analyse toegepast op de data in Tabel 2. Indien de producten geen schaal vormen wordt het item met de grootste schendingen verwijderd. Dat kan dus het item met de laagste waarde op de itemgerelateerde coëfficiënt voor monotone homogeniteit (H_1) zijn of het item dat de grootste schendingen veroorzaakt in de P-Matrices. Daarna wordt getoetst of de resterende items wel een schaal vormen. Indien dit weer niet het geval is wordt weer het slechtste schalende item verwijderd, totdat de overgebleven producten een schaal vormen volgens de criteria die in paragraaf 3 zijn genoemd.

Naast informatie over de karakteristieke verzamelingen en acquisitievolverdes is inzicht in het verband tussen acquisitievolverdes en andere variabelen ook bruikbaar voor marketingdoeleinden. Wanneer dit verband wordt onderzocht, gaat het meestal om de samenhang tussen: (1) Het totale aantal producten dat consumenten uit een acquisitievolverde hebben afgenomen en (2) De waarden die consumenten hebben op andere variabelen (Kamakura et al., 1991; Soutar en Cornish-Ward, 1997). Deze benadering kan voor sommige doelen te onnauwkeurig zijn, omdat verschillende factoren in de verschillende fases van de acquisitievolverde een (andere) rol kunnen spelen. Om dit te verklaren introduceren we een hypothetische acquisitievolverde bestaande uit producten A, B en C. Het is mogelijk dat vooral risicomijdende klanten die product A bezitten geneigd zijn om de volgende stap te zetten in de acquisitievolverde, afname van product B. Echter, het is mogelijk dat in de stap van product B naar C, de bereidheid om risico te nemen geen rol speelt, maar juist leeftijd en de hoogte van het inkomen relevant zijn. Een model dat rekening houdt met het belang van verschillende eigenschappen in verschillende fases van de acquisitievolverde is het 'Sequential Logit Decision Model' (SLD-model) (Madalla, 1983). Dit laatste wordt verder toegelicht in paragraaf 7, waarbij we het SLD-model ook gebruikten om een koppeling te leggen tussen de output van APA en de gegevens in Tabel 3.

5. ANALYSERESULTATEN ENQUETEGEREGEVENS

In deze paragraaf worden de resultaten van de analyses op de enquêtegegevens gerapporteerd, met daarin informatie over het bezit van de volgende producten: spaarrekening, koopsompolis, beleggingsfonds en aandelen. Vooraf formuleerden we twee mogelijke uitkomsten van de analyse met MSP:

- H_0 : De bezitspatronen van de vier genoemde producten vormen geen Mokkenschaal.
 H_1 : De bezitspatronen van de vier genoemde producten vormen wel een Mokkenschaal.

De output van de analyse toont aan dat de data voldoen aan de assumpties van het MH-model, zie Tabel 4. De schaalwaarde (H) is groter dan 0.30. Verder zijn alle itemcoëfficiënten voor monotone homogeniteit (H_i) groter dan de minimumwaarde 0.30 en valt op te merken dat alle $H_{ij} \geq 0,30$. H en alle H_i 's wijken significant af van 0 (alle p -waarden in Tabel 4 zijn kleiner dan 0,001), hetgeen ook een vereiste is volgens de criteria in paragraaf 3. Tot zo ver zijn de resultaten in overeenstemming met hypothese H_1 . De data voldoen echter niet aan de assumpties van het DM-model, de P -matrices geven aan dat deze assumpties worden geschonden door de koopsompolis, zie de vetgedrukte elementen in Tabel 5. Zoals in paragraaf 3 is aangegeven, betekent dit een schending van de assumpties van het DM-model en dient de koopsompolis te worden verwijderd uit de set om na te gaan of de andere producten wel leiden tot bezitspatronen welke in overeenstemming zijn met de assumpties van het DM-model.

Verwijdering van de koopsompolis leidt tot bezitspatronen die voldoen aan de assumpties van zowel het MH-model als die van het DM-model. We concluderen daarom dat de vier bovengenoemde producten tot dezelfde karakteristieke verzameling behoren (MH-model) en dat de spaarrekening, beleggingsfondsen en aandelen in een vaste volgorde worden aangeschaft (DM-model).

Tabel 4. Monotone Homogeniteit Telepanel-gegevens.

Product	Percentage bezit elk product	H_i	Significantie
Aandelen	30%	0,48	$p < 0,001$
Beleggingsfonds	43%	0,41	$p < 0,001$
Koopsompolis	48%	0,32	$p < 0,001$
Spaarrekening	85%	0,34	$p < 0,001$
		$H = 0,39$	$p < 0,001$

De inhoudelijke interpretatie van de generieke behoefte die ten grondslag ligt aan de gevonden karakteristieke verzameling is in termen van wat Katona (1975) 'kunnen en willen' zou noemen. De producten in de karakteristieke verzameling worden alleen afgenomen indien consumenten hiertoe de financiële mogelijkheden hebben (kunnen). Dit betekent dat het geld niet nodig is voor andere verplichtingen zoals een hypotheek, de afbetaling op een auto of aan dagelijkse uitgaven. Consumenten moeten de producten ook willen. Dit hangt waarschijnlijk samen met de attitude ten aanzien van sparen. Bij bezit van beleggingsfondsen en aandelen is er waarschijnlijk ook sprake van bereidheid om financiële risico's te accepteren.

Tabel 5. P-matrices voor de Telepanel-gegevens.

		P(+,+) - matrix			
		1	2	3	4
Aandelen	(1)	-	0,2221	0,2073	0,2719
Beleggingsfonds	(2)	0,2221	-	0,2719	0,3903
Koopsompolis	(3)	0,2073	0,2719	-	0,4253
Spaarrekening	(4)	0,2719	0,3903	0,4253	-
P_i		0,30	0,43	0,48	0,85

		P(-,-) - matrix			
		1	2	3	4
Aandelen	(1)	-	0,4953	0,4266	0,1279
Beleggingsfonds	(2)	0,4953	-	0,3594	0,1144
Koopsompolis	(3)	0,4266	0,3594	-	0,0956
Spaarrekening	(4)	0,1279	0,1144	0,0956	-
P_i		0,30	0,43	0,48	0,85

Voor de gevonden acquisitievolverde (eerst spaarrekening, dan beleggingsfonds gevolgd door aandelen) en de verwerping van de koopsompolis uit deze volgorde is een verklaring te formuleren in termen van de inmiddels herziene belastingwetgeving in Nederland. In 1994 konden gezinnen vastgestelde rentebedragen van spaarrekeningen belastingvrij ontvangen; fl. 1.000,- (\pm 450 Euro) per volwassene in het huishouden en fl. 500,- (\pm 225 Euro) per kind. Dezelfde belastingvrije bedragen waren in 1994 van toepassing op het dividend dat men ontving uit beleggingsfondsen. Consumenten boekten daardoor eerst geld naar een spaarrekening, totdat ze belasting moesten betalen over de rente. Het bedrag dat leidde tot belasting werd vervolgens overgemaakt naar beleggingsfondsen, totdat ook daar de belastingvrije voet werd bereikt. Daarna gingen consumenten over tot de afname van aandelen.

De belastingwetgeving biedt ook een verklaring voor de verwerping van de koopsompolis uit de acquisitievolverde. Dit product leidde in 1994 niet tot belastingvrije rente of dividenden, maar het geld dat werd ingelegd op een koopsompolis was wel belastingaftrekbaar. Het gestorte geld werd van het inkomen afgetrokken en dit leidde tot de betaling van minder belasting. Het geld dat werd ingelegd op de koopsompolis werd met een waardeverhoging uitgekeerd na een vooraf besproken periode, meestal na de pensionering van de consument. Belasting werd betaald over het uitgekeerde bedrag, met als voordeel dat het belastingpercentage na de pensionering meestal relatief laag is. Belastingvoordelen op de koopsompolis hingen dus nauw samen met de hoogte van het inkomen en hooguit indirect met het opgebouwde spaargeld van consumenten. Misschien werden de spaarrekening, het beleggingsfonds en aandelen daardoor wel in een vaste volgorde afgenomen en de koopsompolis niet. We merken wel op dat het hier gaat om een post-hoc verklaring en dat er geen gegevens in de Telepanel-data beschikbaar zijn ter controle van de geformuleerde veronderstellingen. Tevens is de belastingwetgeving omtrent de koopsompolis inmiddels veranderd, hetgeen ook gevolgen kan hebben voor in termen van de uitkomsten van APA.

6. ANALYSERESULTATEN TRANSACTIEGEGEVENS

Ter controle van de resultaten in paragraaf 6 passen we APA toe op de beschikbare transactiegegevens, met daarin informatie over bezit van de volgende vier producten: betaalrekeningen, spaarrekeningen, beleggingsfondsen en aandelen. De resultaten uit paragraaf 6 worden ondersteund indien het acquisitiepatroon dat uit transactiegegevens wordt afgeleid consistent is met het in paragraaf 5 gerapporteerde patroon. Verder is het de bedoeling dat het gevonden acquisitiepatroon consistent is met de daadwerkelijke data waarop de klanten de producten afnamen.

Toepassing van het eerdergenoemde programma MSP toont aan dat de bezitspatronen van de producten in de transactiegegevens voldoen aan de assumpties van zowel het MH-model als het DM-model, zie Tabel 6. Ook kunnen melden dat alle $H_{ij} \leq 0,30$ en dat er geen schendingen zijn in de P-matrices. Dus volgens de uitkomsten van APA op de transactiegegevens worden de spaarrekening, beleggingsfondsen en aandelen in de genoemde volgorde afgenomen, zoals ook bleek uit de analyse van de enquêtegegevens. Verder zien we dat de afname van betaalrekeningen meestal vooraf gaat aan de afname van de drie eerder genoemde producten.

Zoals eerder aangegeven impliceert het voldoen van de data aan de assumpties van het MH- en DM-model dat de geanalyseerde producten tot dezelfde karakteristieke verzameling behoren en dat producten, die in het bezit zijn van een groter deel van de consumenten, doorgaans worden aangeschaft voor producten die minder consumenten hebben. We kunnen de uitkomsten valideren aan de hand van de werkelijke data waarop klanten producten afnamen (transactiedata zijn ook beschikbaar in het geanalyseerde bestand). Het blijkt dat 95% van de klanten van financiële dienstverlener X die zowel de betaalrekening als de spaarrekening bezitten, de betaalrekening afnamen voorafgaand aan de spaarrekening. Verder nam 83% van de klanten die zowel de spaarrekening als het beleggingsfonds bezitten, de spaarrekening af voor het beleggingsfonds. Tot slot nam 86% van de klanten die bij financiële dienstverlener X zowel een beleggingsfonds als een aandelenfonds hebben, het beleggingsfonds eerst af. Deze consistentie betekent dat onderzoekers die alleen bezitspatronen, en geen transactiedata, tot hun beschikking hebben, bezitspatronen kunnen gebruiken om te onderzoeken in welke volgorde klanten diensten meestal afnemen. Gegeven de consistentie met de resultaten in paragraaf 5 kunnen we aangeven dat zowel enquêtegegevens als transactiegegevens geschikt zijn voor de toepassing van APA.

Tabel 6. Monotone Homogeniteit Transactiegegevens.

Product	Percentage bezit elk product	H_i	Significantie
Betaalrekening	82%	0,79	$p < 0,001$
Spaarrekening	43%	0,72	$p < 0,001$
Beleggingsfondsen	6%	0,68	$p < 0,001$
Aandelen	2%	0,75	$p < 0,001$
Algemene gegevens		$H = 0,72$	$p < 0,001$

7. KOPPELING VAN ACQUISITIEVOLGORDES MET ANDERE GEGEVENS

In deze paragraaf bespreken we een techniek waarmee het verband tussen de output van APA en andere variabelen is te onderzoeken. Deze techniek heet het “Sequential Logit Decision Model (SLD-model)” (Maddala (1983) en wordt toegepast op de acquisitievorgorde die we in paragraaf 6 rapporteerden. Voor de bouw van het SLD-model voerden we de volgende vier logit-analyses uit in SPSS, één voor elke stap in de acquisitievorgorde:

1. De eerste analyse richtte zich op het onderscheid tussen subjecten die geen betaalrekening hebben bij financiële dienstverlener X en subjecten die dit product wel hebben. Let wel, bezit van de betaalrekening bij de concurrent is niet af te leiden uit de transactiegegevens van financiële dienstverlener X en wordt dus niet in het SLD-model meegenomen. Dit geldt ook voor bezit van de producten die corresponderen met de drie andere fases van de acquisitievorgorde.
2. Voor de tweede analyse gebruikten we alleen de gegevens van betaalrekeningbezitters. Deze analyse richtte zich op het onderscheid tussen betaalrekeningbezitters die ook een spaarrekening hebben bij financiële dienstverlener X en betaalrekeningbezitters die geen spaarrekening hebben bij financiële dienstverlener X.
3. De derde analyse werd toegepast op spaarrekeningbezitters. Hier ging het om het maken van onderscheid naar bezit van beleggingsfondsen bij financiële dienstverlener X.
4. De vierde en laatste analyse richtte zich op het onderscheid tussen beleggingsfondsbezitters die aandelen hebben bij financiële dienstverlener X en beleggingsfondsbezitters die geen aandelen hebben bij X.

Alleen significante effecten (op het 0,05-niveau) zijn gehandhaafd. Niet-significante effecten zijn stapsgewijs uit de modellen verwijderd. In elke stap werd het effect waarbij de kans op toeval het grootste was uit het model verwijderd en voerden we de analyse opnieuw uit met de overgebleven variabelen. Deze benadering lijkt op ‘backward stepwise regression’ in SPSS. De overgebleven significante effecten zijn gerepresenteerd in Tabel 7. Deze Tabel presenteert ook een indicatie voor de kwaliteit van elk model: de mate waarin een cut-off onderscheid maakt tussen klanten die al de volgende stap in de acquisitievorgorde hebben genomen en klanten die dit (nog) niet hebben gedaan.

Tabel 7. SLD-model.

	Stap 1: Niets → CA	Stap 2: CA → SA	Stap 3: SA → IT	Stap 4: IT → S
<i>Effect:</i>				
Constant	1,8458	-0,8804	2,7355	-0,6108
Income_Transfer	1,7475	0,3927	- 0,3861	
Income_Amount	0,0007	0,0001	0,0002	
Years_Client		0,0280	0,0207	-0,0112
Age				
<i>Kwaliteit:</i>				
Hoge-score groep	99,3%	62,5%	17,4%	27,5%
Lage-score groep	57,4%	36,4%	7,5%	22,6%
Legend	CA=Betaalrek.	SA=Spaarrek.	IT=Beleggingsfondsen	S=Aandelen

De betekenis van het model demonstreren we aan de hand van de tweede stap. Overboeken van inkomen heeft positief effect op de kans dat de betaalrekeningbezitter ook een spaarrekening heeft. Verder zien we dat naarmate er meer inkomen wordt overgeboekt en naarmate de betaalrekeningbezitter langer klant is bij financiële dienstverlener X, er een grotere kans is op het bezit van spaarrekening bij financiële dienstverlener X. Het model heeft een sterk onderscheidend vermogen. We zien dat 62,5% van de betaalrekeningbezitters, met eigenschappen die samenhangen met bezit van de spaarrekening, het laatstgenoemde product ook daadwerkelijk bezitten. Dit geldt maar voor 36,4% van de betaalrekeningbezitters die de genoemde positieve eigenschappen niet hebben. Van belang voor marketingtoepassingen is dat consumenten met eigenschappen die samenhangen met het bezit van een product, maar het product niet hebben, het meest geneigd zijn om het product van interesse af te nemen (Breur en Paas, 2000). Dit betekent dat 37,5% van de betaalrekeningbezitters in de Hoge-score groep, die geen spaarrekening bezit, in aanmerking komt voor een aanbod om het laatstgenoemde product af te nemen. De gegevens die betrekking hebben op de andere stappen in Tabel 6 zijn op een vergelijkbare wijze te interpreteren. Ten behoeve van bondigheid worden deze punten niet verder besproken. Wel merken we op dat sommige variabelen in verschillende stappen van de acquisitievorgorde voorkomen of tegenovergestelde effecten hebben, zie bijvoorbeeld het effect van *Years_Client* in stappen 3 en 4. Dit verschijnsel is een rechtvaardiging voor het onderzoeken van de effecten per stap.

8. DISCUSSIE

In aansluiting op de drie doelstellingen die in de inleiding van dit hoofdstuk zijn geformuleerd hebben we aangetoond dat APA aan de hand van Mokka-schaal bruikbaar is om producten te bundelen in karakteristieke verzamelingen en inzicht kan geven in de volgorde waarin consumenten producten afnemen (acquisitievorgordes). Ook vonden we dat APA vrij algemeen toepasbaar is, zowel op enquêtegegevens als op transactiegegevens. Deze toepassing op verschillende soorten data leidde tot consistente resultaten. Verder bespraken we een techniek waarmee onderzoekers het verband tussen uitkomsten van APA en andere variabelen kunnen onderzoeken.

Op dit punt dienen we op te merken dat Mokka-schaalanalyse naar aanleiding van de analyses, die in deze paper zijn gerapporteerd, ook is toegepast op andere enquêtegegevens. In Paas (2001) is er een acquisitievorgorde gerapporteerd, die volgens enquêtegegevens uit 17 Europese landen redelijk stabiel is. Het blijkt dat consumenten de volgende zes producten meestal in de genoemde volgorde afnemen: (1) Betaalrekening; (2) Cheques; (3) Pas voor betaalautomaten; (4) Creditcard waarop afrekening meteen na de uitgave plaatsvindt; (5) Creditcard met uitgestelde afrekening; (6) Creditcard met maandelijkse aflossing.

Wat betreft de implicaties kunnen we aangeven dat APA aansluit op de meest recente ontwikkelingen in marketing. Volgens Poiesz en Van Raaij (2002) kunnen bedrijven zich in de toekomst vooral onderscheiden door weloverwogen pakketten van producten aan te bieden die inspelen op combinaties van klantbehoeften. Inzicht in karakteristieke verzamelingen kan een richtlijn geven voor de samenstelling van dit soort pakketten. Verder geven acquisitievorgordes aan in welke volgorde faciliteiten uit pakketten het beste zijn aan te bieden aan consumenten. Een koppeling tussen acquisitievorgordes en andere klantgegevens, met het SLD-model, is bruikbaar om te bepalen welke cliënten

in aanmerking komen voor het volgende product of een volgende faciliteit in het pakket. Echter, vervolgonderzoek is nodig. Toekomstige studies kunnen gericht zijn op combinaties en volgordes waarin consumenten producten in diverse domeinen afnemen. Ook kan er worden onderzocht of er verbanden bestaan tussen verschillende karakteristieke verzamelingen en acquisitiepatronen van uiteenlopende domeinen. Verder zou onderzoek naar de tijd die ligt tussen de afname van producten relevant zijn. Dergelijke tijdsgebonden aspecten zijn te onderzoeken met longitudinale data.

LITERATUUR

- Ajzen, I., Fishbein, M., 1980. *Understanding Attitudes and Predicting Social Behaviour*. Englewood Cliffs: Prentice-Halls.
- Breur, T. and Paas, L.J., 2000. Building predictive models for direct mail: A framework for choosing training- and test data. *Journal of Database Marketing* 8, 9-16.
- Clawson, C.I., Vinson, D.E., 1978. Human values: A historical and interdisciplinary analysis. *Advances in Consumer Research* 5, 396-402.
- Engel, J.F., Blackwell, R.D., Miniard, P.W., 1986. *Consumer Behaviour*. Chicago: Dryden Press.
- Gutman, J., Vinson, D.E., 1979. Value structures and consumer behaviour. *Advances in Consumer Research* 6, 47-57.
- Kamakura, W.A., Ramaswami, S.N., Srivastava, R.K., 1991. Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing* 8, 329-349.
- Kasulis, J.L., Lusch, R.F., Stafford, E.F., 1979. Consumer acquisition patterns for durable goods. *Journal of Consumer Research* 6, 47-57.
- Katona, G., 1975. *Psychological Economics*. New York: Elsevier.
- Kotler, P., 1971. *Marketing Decision-Making: A Model Building Approach*. New York: Holt, Rinehart and Winston.
- Kuijlen, A.A.A., 1993. *De Scenariobenadering: Een Onderzoek naar Complexe Consumentenbeslissingen met behulp van Computergestuurd Enquêteren (Proefschrift)*. Tilburg: KUB.
- LaPlaca, P.J., Punj, G., Randazzo, N., 1985. A multicultural model of consumer durable acquisition patterns. *Advances in Consumer Research* 12, 125-129.
- Lord, F.M., Novick, M.R., 1968. *Statistical Theories of Mental Test Scores*. Reading: Addison Wesley.
- Madalla, G.S., 1983. *Limited-Dependent and Qualitative Data in Econometrics*. Cambridge: Cambridge University Press.
- McFall, J., 1969. Priority patterns and consumer behaviour. *Journal of Marketing* 33, 50-55.
- Mokken, R.J., 1971. *A Theory and Procedure of Scale Analysis*. Reading: Mouton.
- Molenaar, I.W., 1982. Mensen die het beter meten. *Kwantitatieve Methoden* 3, 3-29.
- Paas, L.J., 1998. Mokken scaling characteristic sets and acquisition patterns of durable- and financial products. *Journal of Economic Psychology* 19, 353-376.
- Paas, L.J., 2001. Acquisition patterns of products facilitating financial transactions: A cross-national investigation. *International Journal of Bank Marketing* 19, 266-275.
- Paas, L.J., Molenaar, I.W., 2002. An evaluation of methodologies for acquisition pattern analysis. Submitted for publication.
- Poiesz, Th.B.C., Van Raaij, W.F., 2002. *Synergistische Marketing: Een Visie op Oorzaken en Gevolgen van Veranderd Consumentengedrag*. Amsterdam: Prentice Hall.
- Sijtsma, K., Molenaar, I.W., 2002. *Introduction to Nonparametric Item Response Theory*. Thousand Oaks: Sage Publications.
- Soutar, G.N., Cornish-Ward, S., 1997. Ownership patterns for durable goods and financial assets: A Rasch analysis. *Applied Economics* 29, 903-911.
- Verhallen, Th.M.M., Pieters, R.G., 1984. Attitude theory and behavioural costs. *Journal of Economic Psychology* 5, 223-249.

9. *Het jongste vooruitzicht*

De toekomst volgens jongeren in Nederland

H. VINKEN, P. ESTER en L. VAN DUN

SAMENVATTING

Een heldere en beloftevolle toekomst is niet langer vanzelfsprekend. De toekomst is niet langer voorspelbaar. Dat is de boodschap van vele contemporaine jeugdstudies. Het omgaan met deze onzekerheden zal een belangrijkere rol gaan spelen in de toekomstoriëntaties van jongeren. Het geeft jongeren de kans om zelf vorm en betekenis te geven aan hun leven. Dat is bedreigend, maar tevens kansrijk. In veel jeugdstudies wordt echter geen aandacht geschonken aan wat jongeren zelf van hun toekomst verwachten of hoe zij zich op die toekomst voorbereiden. In dit artikel, dat is gebaseerd op een reeks recente toekomstonderzoeken, geven we daarvan een eerste proeve van inzicht.

Na een korte theoretische introductie, bespreken we in het empirische deel de toekomstverwachtingen van jongeren op een drietal voor hen belangrijke domeinen: opleiding, werk en inkomen en de combinatie werk en privé. Jongeren zien voor zichzelf een mooie toekomst naderen die ze optimistisch tegemoet treden. Ze onderkennen wel dat dat een toekomst is die veel (meer) druk met zich brengt, vooral omdat ze op *alle* terreinen willen uitblinken. Ze vertrouwen daarbij op een sterke inzet van onder andere de overheid en sociale partners, maar ze zijn onzeker of die dat ook zullen doen. Het jongste vooruitzicht is een vorm van optimisme in onzekerheid.

1. DE BELOFTE VAN ONZEKERHEID

Een toekomst met een vaste baan, met een werkzaam leven bij een en dezelfde baas of zelfs in een en hetzelfde beroep, met een vast pensioen, een toekomstig intiem bestaan met die ene levensgezel, maar ook een toekomst met heldere tijdsvakken van groei (leren), inspanning (werken en opvoeden) en ontspanning (vrijetijd besteden, met pensioen zijn), zowel over de gehele levensloop als over de dag bezien, lijkt verleden tijd. Collectieve rollen, ritmes en rituelen eroderen, zo laat ook Nederlands tijdsbestedingsonderzoek zien (Van den Broek & Breedveld, 2000). Hetzelfde lijkt op te gaan voor rotsvaste en breedgedragen waarden en normen of voor ruim gedeelde en beklijvende culturele iconen, helden of symbolen die voor grote groepen mensen het leven nu en in de toekomst richting kunnen geven (Ester et al., 1993; Vinken et al., 1997; Diepstraten et al., 1999a). Levensbeschouwingen en ideologieën die glorieuze toekomst beloven met religieuze of met rode horizonten, geven evenmin nog enig houvast.

Met het verdwijnen van vaste sociale en culturele bakens verliest de toekomst haar voorspelbaarheid. Garanties zijn niet meer te geven. Voor jongeren, die evident nog veel toekomst voor zich hebben, is het wegvallen van een vanzelfsprekende toekomst niet zonder consequenties. In het toekomstperspectief van jongeren zal onzekerheid een

belangrijkere plaats krijgen, zo is de verwachting (Dieleman et al., 1999), en in het handelen gericht op de toekomst zullen het ontwikkelen en onderhouden van vaardigheden om adequaat op onzekerheden te reageren prominenter worden. De toekomst ligt open en vele jongeren hebben de kans daar zelf vorm en betekenis aan te geven. Veranderingsgezindheid, oog voor alternatieven, de vaardigheid uit de alternatieven de juiste keuze te kunnen maken, het vermogen beslissingen te nemen en sterke schouders om de consequenties hiervan zelf te dragen lijken de condities om een onzekere toekomst beloftevol te maken. Onder deze condities verandert het wegvallen van voorgeschreven en gebaande paden naar de toekomst in een kans om als individu meer zeggenschap en greep te krijgen op het verloop van de eigen toekomst en deze in te richten naar eigen preferenties.

Deze condities worden daarmee ook nieuwe eisen, eisen die sterk ingebed zijn in het primaat dat in jeugdstudies gegeven wordt aan begrippen die met het woord 'zelf' beginnen. Zelfsturing, zelfregie, zelfcontrole, zelfredzaamheid, zelfontplooiing, zelfkennis en uiteraard zelfstandigheid zijn enkele van deze termen.¹ De populariteit van die termen is tot grote hoogte gestegen, of het oog nu gericht wordt op het onderwijs, ouder-kindrelaties, werksituaties of op marketing en reclame. Het primaat van het 'zelf' vergroot met andere woorden de afstand tussen jongeren en hun opvoeders (ouders, leraren, professionele opvoeders en andere publieke gezagsdragers) en de pedagogische onzekerheid van de laatsten. In dit verband is eveneens de grote belangstelling voor het 'monitoren' van wie jongeren nu zijn en waar zij met hun leven heen willen te plaatsen. Onderzoekers, wetenschappelijke jongerenexpertteams en jongerenmonitoringsinstanties richten hun zoeklichten tot op heden voornamelijk op wie de jongeren van vandaag zijn en verzamelen systematisch inzicht over de condities waaronder jongeren nu hun identiteit verwerven, hoe zij aan de samenleving participeren, met welke middelen zij welke vruchten van die samenleving plukken of hoe hun waarden en normen eruit zien. Inzicht in wat jongeren zelf van hun toekomst verwachten, hoe zij zich op die toekomst voorbereiden en hoe zij met onzekerheden omgaan ontbreekt merkwaardigerwijs in het groeiende kennisarsenaal over jongeren. Het belang van dit inzicht is groot gegeven het vertoog over de toekomst als een open en individueel regisseerbaar project. Een eerste proeve van dit inzicht hebben wij onlangs gegeven met de publicatie 'Aan ons de toekomst' (Vinken et al., 2002) over een serie toekomstonderzoeken die eind jaren negentig op ons initiatief in Nederland onder verschillende generaties zijn uitgevoerd. Dit artikel vat enkele resultaten samen van dit onderzoek. In de volgende paragraaf beschrijven we ratio en achtergronden van dit onderzoek. Vervolgens gaan we in paragraaf 3 in op enkele sturende basisbegrippen die zich met name uit psychologische en sociologisch-pedagogische studies over toekomstoriëntaties van jongeren laten afleiden. Paragraaf 4 rapporteert de empirische inzichten voor die kerndomeinen die jongeren vooraan in hun toekomstoriëntatie hebben staan: onderwijs, arbeid en inkomen, en de combinatie van werk en privé. De analyses worden voorafgegaan door een blik op de kernelementen van het toekomstperspectief van jongeren en op de maatschappelijke probleemagenda die in dit perspectief dominant is. Door ook een vergelijking met dertigers te maken wordt de eigenheid van de toekomstoriëntaties van jongeren (hier met name twintigers) in deze paragraaf sterk benadrukt. Paragraaf 5 concludeert.

2. ‘AAN ONS DE TOEKOMST’

De Interdepartementale Commissie Jeugdonderzoek (CJO) stelt vast dat er weinig bekend is, ook niet uit de vele survey-onderzoeken die er tegenwoordig in Nederland onder jongeren gedaan worden, over het toekomstperspectief van jongeren (Dieleman et al., 1999; zie ook Heemskerk et al., 1999). Dit ontbreken van kennis wordt als problematisch ervaren juist nu van jongeren met betrekking tot de inrichting van hun toekomst een grote mate van zelfcontrole en eigen verantwoordelijkheid wordt gevraagd.

De CJO heeft IRIC, Institute for Research on Intercultural Cooperation, gevraagd om in samenwerking met Globus, Institute for Globalization and Sustainable Development en OSA, Organisatie voor Strategisch Arbeidsmarktonderzoek – alle gevestigd op de campus van de Universiteit van Tilburg², met hun recente onderzoeken rondom toekomstperspectieven van jonge Nederlanders na te gaan hoe jongeren zich op hun toekomst voorbereiden, hoe ze tegen hun eigen toekomst aankijken en tegen de toekomst van een aantal belangrijke maatschappelijke trends en instituties (zoals onderwijs, betaalde arbeid, zorgarbeid). Veel populaire vertogen stellen dat jongeren ‘bij de dag’ leven en nauwelijks op een vol ontwikkeld besef van de toekomst te betrappen zijn. Andere, haast even gretige observaties verhalen van een ‘teveel’ aan toekomstbesef bij jongeren, hetgeen zich veelal laat vertalen in een sterke fixatie van jongeren op een strakke planning van de eigen carrière en de persoonlijke financiële toekomst. De vraag is of jongeren (anders dan de volwassenen van nu, ofwel de jongeren van vroeger tijden) een ander besef van toekomst hebben, zowel in kwantitatieve als in kwalitatieve zin. Kijken ze juist korter of verder vooruit (in jaren gerekend), zijn ze bezig met ‘de’ toekomst en vullen ze die toekomst in met betekenissen die op ‘hoop’ dan wel op ‘vrees’ duiden? Ook klinkt er zorg om de generatie van de toekomst door: welke opdrachten stellen jongeren zich, welke keuzes maken zij, welke onzekerheden en problemen ervaren ze en welke waarden en normen hanteren ze hierbij?

Hoe jongeren de toekomst zien en hoe vooral zij op de toekomst anticiperen hebben IRIC, Globus en OSA in 1999 in twee grootschalige toekomstonderzoeken onderzocht. In deze twee onderzoeken, in het voor- en najaar van 1999, hebben jongeren (en ouderen) via het zogenaamde CentERpanel een reeks van vragen beantwoord.³

Het *eerste onderzoek* is uitgevoerd in opdracht van de Geassocieerde Persdiensten (GPD). De verwachtingen, angsten en dromen over de ééntwintigste eeuw en de mate waarin voorziene veranderingen hen ook persoonlijk zullen raken, komen in dit onderzoek onder andere aan bod. Het onderzoek behandelt de beelden van het verleden, het heden en de toekomst van zo’n 230 jongeren tussen de 16 en 28 jaar. We hebben uit dit onderzoek gegevens over dezelfde beelden van oudere Nederlanders.⁴ In het vervolg zullen we geen vergelijking maken met alle Nederlanders maar alleen met het leeftijdscohort direct boven deze jongeren: de dertigers (een kleine 350 personen tussen de 29 en 39 jaar). We zullen hier niet verder uitweiden over de achtergronden van aanpak en methoden van het onderzoek.⁵ Opgemerkt zij dat we zoveel mogelijk kwalitatief (verkregen uit open vragen) en kwantitatief onderzoeksmateriaal (uit gesloten vragen) hebben gecombineerd.

Het *tweede onderzoek* concentreert zich op de breed gevoelde druk van taakcombinatie binnen en tussen de domeinen van arbeid, zorg en vrijetijd. De maatschappelijke dwang en persoonlijke drang naar individuele excellentie binnen deze domeinen roept meer algemeen de vraag op welke verwachtingen jonge (en oudere) Nederlanders koes-

teren als het gaat om arbeid, zorg en vrijetijd in het verdere verloop van de éérentwintigste eeuw. Het *OSA Toekomst van de Arbeid Survey* is de eerste, meer omvangrijke studie in ons land waarin toekomstverwachtingen rond de thematiek van arbeid, zorg en vrijetijd in de éérentwintigste eeuw worden verkend. Leidraad is de vraag welke ontwikkelingen jongeren de komende vijfentwintig jaar, tussen nu en 2025, voor waarschijnlijk houden, de wenselijkheid van deze ontwikkelingen en de mate waarin deze ontwikkelingen de eigen leefsituatie zullen raken.⁶ We presenteren hier de antwoorden van 350 jongeren (16-28 jaar) en vergelijken die met de antwoorden van zo'n 400 dertigers (29-39 jaar).⁷

3. EXTENSIE, AFFECTIE, CONTROLE EN CULTUUR

Het voorgaande maakt duidelijk dat er een grote maatschappelijke behoefte is de toekomstverwachtingen van jongeren te kennen. De wetenschappelijke kennis over wat jongeren van de toekomst verwachten en hoe zij daarop inspelen is veelal of sociaal-psychologisch van karakter of sociologisch-pedagogisch met een sterke hang naar cultuuranalyse. Onderzoek in de sociaal-psychologische traditie heeft empirisch een beperkte reikwijdte want is doorgaans gebaseerd op kleine en selectieve jongerenpopulaties. We zullen aangeven welke begrippen er in onderzoek naar jongeren en toekomst in deze traditie bestaan en zo ook wat we in ons onderzoek met enkele begrippen bedoelen. Centraal in de sociologisch-pedagogisch gefundeerde analyses over culturele veranderingen staan geleidelijke of meer plotselinge veranderingen in de omgeving van jongeren die er voor zorgen dat zij op een andere manier anticiperen op hun toekomst dan volwassenen. We schetsen enkele voorbeelden van deze gedachtegang, waarbij we ook hier aantekenen dat het empirisch materiaal in deze traditie nog tamelijk beperkt is.

Aan het begin van de jaren zeventig worden in Duitsland de eerste empirische onderzoeken verricht naar toekomstperspectieven (Lamm & Trommsdorff, 1974; zie ook Lamm et al., 1976; Schmidt et al., 1978; Trommsdorff et al., 1979). Deze (en vele latere) onderzoeken bouwen voort op een rationele benadering van keuzegedrag. Toekomstoriëntaties zijn in deze benadering sterk gerelateerd aan het zich doelen stellen, het plannen en besluiten vormen en het doelgericht handelen. Trommsdorff (zie ook Trommsdorff, 1994, 2000) en de haren gaan ervan uit dat iemand in lijn met zijn of haar toekomstoriëntaties bepaalde activiteiten plant, rekening houdend met verwachte veranderingen in de eigen omgeving. Jongeren kunnen in hun nabije toekomst meer dan enige andere leeftijdsgroep rekenen op belangrijke veranderingen in de transitie van adolescentie naar volwassenheid. Jongeren zitten in de levensfase met de sterkste verdichting maar ook de hoogste urgentie van toekomstkeuzes. De gemaakte keuzes en de veranderingen die hierop volgen, beïnvloeden op hun beurt weer de toekomstoriëntaties. De feitelijk veranderde omstandigheden (bij voorbeeld voor het eerst de socialisatie als werknemer ondergaan) kunnen betekenen dat toekomstoriëntaties verbreed of verdiept worden of zelfs een andere affectieve betekenis krijgen.

Belangrijk bij de wisselwerking tussen toekomstperspectieven en (geanticipeerde en feitelijke) keuzes is het controlemechanisme (zie m.n. Trommsdorff, 1994). Behalve het verschil in interne en externe 'locus of control' (de overtuiging dat je zelf veel of zelf juist weinig invloed op je omstandigheden hebt) wordt het onderscheid tussen primaire en secundaire controle gemaakt. Primair is de controle-overtuiging dat je zelf de

omgeving kunt beïnvloeden en deze ook poogt te veranderen om zo je eigen doelen te verwezenlijken. Secundaire controle is de overtuiging dat de omgeving de grenzen stelt en dit wordt gepaard aan het mechanisme je eigen doelen bij te stellen naar gelang de omstandigheden.

Eerste empirische inzichten laten zien dat jongeren van een primair controlemechanisme uitgaan (Flammer, 1988; Baltes & Baltes, 1990; Trommsdorff, 1994, 2000). Jongeren structureren hun toekomstgericht gedrag zo dat ze vooraleerst pogen de omgeving aan hun doelen aan te passen. Later in de levensloop komt het secundair controlemodel in beeld. Ervaringen met gepleegde investeringen in de verandering van de omgeving (de winst- en verliesrekening) en een toenemend aantal leeftijdsgebonden beperkingen (werk, gezin, fysieke omstandigheden) reduceren de aantrekkelijkheid van nieuwe, soortgelijke investeringen. Aanpassing van doelen aan de ervaren werkelijkheid, het secundaire controlemodel, ligt voor oudere leeftijdsgroepen voor de hand (soms begeleid door gerichte, selectieve investeringen). Een enkele bron laat zien dat primaire controle ook een dominante strategie is in individualistische (Westerse) culturen (Essau & Trommsdorff, 1993; vgl. Hofstede, 2001). In deze culturen leggen mensen meer nadruk op primaire controle dan mensen in groepsgeoriënteerde, collectivistische culturen die er eerder naar neigen een secundaire strategie te volgen. Het verschil zit vooral in het navolgen van eigen doelen ongeacht of juist rekeninghoudend met de ander.

Het merendeel van het sociaal-psychologisch onderzoek richt zich op wat de basiselementen van toekomstoriëntaties zijn. Toekomstoriëntaties bevatten verschillende, cognitieve, motivationele en affectieve elementen: van de lengte van het tijdsperspectief of voorstellingsvermogen, ook wel de extensie van het toekomstbeeld genoemd, tot en met een optimistisch of pessimistisch affect. In onderzoek ligt het zwaartepunt vaak op één aspect: de dichtheid ('density') van de toekomstoriëntatie ofwel het aantal gebeurtenissen en omstandigheden waarover een individu zich respectievelijk geen en wel zorgen maakt ('hopes' en 'fears').⁸ Consequent in de top drie staan in beide gevallen onder jongeren: scholing, werk en intieme relaties. Succes of falen op school, het wel of juist geen geschikte baan vinden, en het treffen of mislopen van de ideale vrouw of man zijn de belangrijkste 'hopes' en 'fears'. Meisjes noemen meer gebeurtenissen, zowel meer 'hopes' als meer 'fears', over intieme relaties. Jongens doen dat meer over werk.

Toekomstoriëntaties hebben een sterke band met de normatieve levensloop en met identiteitsformatie. Wat men van en voor zichzelf verwacht komt nauw overeen met de collectief geformuleerde ontwikkelingsstaken bij de overgang van adolescentie naar volwassenheid en met verwachtingen over wat de 'normale' volwassen rollen van mannen en vrouwen en van lager en hoger opgeleiden zijn. We leren uit het onderzoek dat we ons bij vraagstukken over de oriëntatie op de eigen toekomst van jongeren niet alleen moeten richten op 'hopes' en 'fears' op de kerndomeinen (school, werk, gezin), maar ook op de extensie (of men vooruit kijkt en hoe ver men dan vooruit kijkt) en de affectie (de mate van optimisme). We laten nu zien dat er nog een aspect bij komt: het maatschappelijke.

In de sociaal-psychologische traditie ligt de nadruk op het individu en op fundamentele processen die zich bij een individu in de interactie met zijn of haar directe sociale omgeving afspelen. De maatschappelijke en de culturele factor, alsmede de effecten van processen van sociale en culturele veranderingen op toekomstbeelden hebben daarin een weinig prominente plaats.

Een in Nederland invloedrijk vertoog over jongeren en hun keuzegedrag dat wel oog heeft voor deze veranderingen is dat van het Leidse sociologen- en pedagogenteam rondom Du Bois-Reymond (o.a. Du Bois-Reymond et al., 1990, 1994, 2000). Twee onderling verwante veranderingsprocessen spelen in dit vertoog een belangrijke rol: de opkomst van de onderhandelingshuishouding en de verschuiving van de standaardbiografie naar de keuzebiografie. De opkomst van de onderhandelingshuishouding heeft alles te maken met het evenwichtiger worden van de machtsbalans tussen ouders (en andere opvoeders) en kinderen.⁹ Anders dan in voorgaande bevelshuishoudingen, geïdentificeerd als het opvoedingsmodel van mensen die voorafgaand aan de jaren zestig van de twintigste eeuw jong waren, staan in de onderhandelingshuishoudingen begeleiden in plaats van leiden, uitleggen in plaats van voorschrijven, overleggen in plaats van aanvaarden en gehoorzamen centraal. De sleutelbegrippen en de sleutelvaardigheden zijn reflectie en communicatie.

Daarnaast valt de Leidse wetenschappers een transitie op van strak gefixeerde standaardbiografieën naar zogenoemde individualistische keuzebiografieën. De centrale elementen in deze nieuwe biografieën zijn een toegenomen variatie in het aantal, in de timing en in de volgorde van levensgebeurtenissen en statuspassages die jongeren meemaken.¹⁰ De vanzelfsprekendheid van te maken keuzes is met de grotere variatie afgenomen. Jongeren pogen keuzes uit te stellen of pogen het een en wisselen dan weer naar het ander (Du Bois-Reymond et al., 2000: 224-225).

De Nederlandse kijk op jongeren en hun toekomst heeft nauwe banden met een Duits jeugdsociologisch perspectief waarin sterk op de individualisering van de jeugdfase wordt ingespeeld.¹¹ Ook hierin wordt verondersteld dat de levensloop van jongeren individualiseert en de-standaardiseert. De ultieme consequentie is dat het hele idee van patronen in levensfasen passé is. Er is geen ideaal-typisch patroon meer dat voorspellen kan op welke leeftijd, of in welke levensfase, jongeren (en dus ook ouderen) ervaringen hebben op het gebied van relaties, scholing, werk, huisvesting et cetera. Du Bois-Reymond et al. (2000: 211) spreken al van de jokoïsering van de jeugdige levensloop. Jongeren bereiken de volwassenenstatus op sommige terreinen, terwijl ze op andere nog in een jong stadium verkeren. Ouderen laten op gezette tijden hun vastigheid op het ene terrein voor wat het is en nemen weer jeugdige beginnersrollen op andere terreinen op. Dan weer uitzendkracht en dan weer gedegen vaste-werknemer, dan in de status van partner of vader, dan weer in die van vrijgezel, een tijdje werken, een tijdje studeren, sabbatsverlof tussendoor en/of volop maatschappelijk actief blijven na 65 jaar. De status van volwassenheid en die van jeugd is niet meer eenduidig en zo is de transitie van de ene naar de andere fase dat evenmin. De toekomstkeuzes die jongeren maken, staan niet meer in relatie tot een helder beeld van de volwassenenstatus. Deze redeneringen zijn niet meer, maar ook niet minder, dan interessante theoretische noties.

Dit is anders bij een recente versie van de Duitse Shell Jugendstudie, *Jugend 2000*. Het zwaartepunt van *Jugend 2000* is de oriëntatie van jongeren op de toekomst, zowel in persoonlijk als maatschappelijk opzicht (Fischer et al., 2000). Theoretisch startpunt is het idee dat processen als flexibilisering en globalisering tot een grotere onzekerheid over de toekomst hebben geleid en vastigheid en eeuwigheidswaarde, of het nu om werk of om privé gaat, tot achterhaalde concepten hebben gemaakt. Uit het empirisch deel van *Jugend 2000* blijkt dat dit geenszins het geval is, ook niet als wordt vergeleken met vroegere jaren.¹²

Een groot aantal onderzoeken naar het veranderingsproces van individualisering geeft voorts voldoende evidentie dat vaste ankerpunten in termen van waarden en normen, of

die nu over levensbeschouwing, de rollen van mannen en vrouwen of over de juiste smaak in de vrijetijdsbesteding gaan, aan invloed op de keuzes van mensen hebben ingeboet (o.a. Ester et al., 1993, 1994). De keuze uit het palet van waarden en normen is minder aan instituties als kerk, familie of sociale groep gebonden en meer een zaak van individuele keuze geworden. De keuzevrijheid van mensen om hun leven naar eigen inzichten in te richten lijkt daarmee toegenomen. Voorgescreven trajecten waarmee de jeugdstatus voor de volwassenenstatus wordt ingewisseld, zijn eveneens minder vanzelfsprekend geworden.

4. TRENDS, THEMA'S EN TRANSITIES VAN DE TOEKOMST

In ons onderzoek hebben we voorrang gegeven aan de toekomstbeleving en toekomstverwachtingen zoals die bij jongeren zelf op een breed aantal terreinen leven. Centrale nadruk hebben de kwantitatieve (extensie) en kwalitatieve aspecten (affectie) van de toekomstvisie van jongeren. Waar mogelijk hebben we aandacht geschonken aan een persoonlijk (de eigen toekomst) en een sociaal (de toekomst van anderen) of maatschappelijk perspectief (de toekomst van de samenleving). In dit artikel komen de voor jongeren primair belangrijke toekomstdomeinen aan de orde: opleiding, werk en gezin (met name in relatie tot werk).¹³

We zijn vooraleerst benieuwd naar de eerste generieke verschillen in toekomstperspectieven. In een reeks vragen is ingegaan op de extensie en affectie van het toekomstperspectief. We bezien onder andere hoe twee cohorten, twintigers en dertigers, en de verschillende sekse- en opleidingsgroepen de toekomst van het eigen persoonlijke leven en dat van Nederland en de wereld in kwantitatieve en kwalitatieve zin beoordelen. Om ook te weten of jongeren meer dan andere cohorten 'bij de dag' leven, wordt eerst nagegaan of ze nog wel een samenhang waarnemen tussen wat er in hun levensloop vroeger, nu en in de toekomst plaatsvindt. Vervolgens analyseren we de visies van beide cohorten op een aantal uitspraken die laten zien hoe planmatig zij met de toekomst omgaan. Tabellen 1 tot en met 3 vatten samen.

De extensie van de toekomstvisie, zowel betrokken op het persoonlijk leven als op Nederland en de wereld als zodanig, is gelijk onder de cohorten twintigers en dertigers. Beide cohorten kijken in persoonlijke en maatschappelijke zin even ver vooruit (gemiddeld tot ongeveer de komende 10 jaar).

Wel zijn twintigers significant optimistischer over de toekomst van het eigen leven en dat van Nederland en de wereld. Dit fenomeen zullen we nog vele malen terugzien. Jongeren, twintigers, hechten net zoveel aan de samenhang tussen verleden, heden en toekomst als dertigers. Een tijdperceptie waarbij men berust dat alles toeval, chaos en onsamenhangend is, heeft geen draagvlak onder jongeren. Voor de grote meerderheid hangen levensdomeinen onderling als in de loop van het leven samen.¹⁴

Of jongeren zich wel bezighouden met de toekomst zijn we ook nagegaan. In tabel 2 staan de resultaten.

Rond de helft van de twintigers en van de dertigers kiest voor het antwoord 'ik hou me niet met de toekomst bezig': een opmerkelijk hoog aandeel. Toch nog ongeveer tweevijfde van beide cohorten gaat wel bewust met de toekomst om en zegt dingen te plannen voor zijn toekomst. Slechts een tiende wil er echt liever niet over nadenken. Dat betekent al met al toch, als we de eerste en laatste categorie optellen, dat de meerder-

Tabel 1. Vooruitkijken in persoonlijk en maatschappelijk opzicht en de samenhang verleden-heden-toekomst.

	Cohort		Sekse		Opleiding				R ²
	<29	29-39	man	vrouw	laag	midden	hoog	allen	
- hoe ver vooruit eigen leven? ^{ae}	2,94	2,77	2,91	2,77	2,79	2,90	2,73	2,83	.02
- de toekomst van eigen leven wordt...? ^b	2,73	2,48	2,59	2,57	2,63	2,51	2,67	2,58	.06
- hoe ver vooruit NL/wereld? ^{ce}	3,26	3,12	3,32	3,04	2,76	3,23	3,39	3,17	.05
- de toekomst van NL/wereld wordt...? ^b	2,49	2,15	2,33	2,25	2,43	2,23	2,29	2,29	.05
- samenhang tussen verleden-heden-toekomst? ^d	2,48	2,48	2,32	2,64	2,34	2,51	2,55	2,48	.05

Bron: CentERdata Millennium Enquête

Toelichting: a: gemiddelden van 1 = het komende jaar, 2 = de komende 2 tot 5 jaar, 3 = de komende 10 jaar, 4 = de periode van de rest van mijn leven; b: gemiddelden van 1 = slechter, 2 = idem, 3 = beter; c: gemiddelden van 1 = het komende jaar, 2 = de komende 2 tot 5 jaar, 3 = de komende 10 jaar, 4 = de periode van de rest van mijn leven, 5 = de periode tot één of twee generaties na mij, 6 = de periode over honderd jaar en later; d: gemiddelden van 1 = nauwelijks samenhang; het leven van mensen hangt van toeval aan elkaar, 2 = elk levensdomein heeft zijn eigen samenhang tussen heden, verleden en toekomst; tussen de levensdomeinen is er nauwelijks verband, 3 = er is samenhang; alle levensfasen zijn duidelijk aan elkaar geschakeld; e: de exacte formulering van de vraag luidt 'Als ik aan de toekomst van mijn eigen leven/Nederland en de wereld denk, dan denk ik aan gebeurtenissen in ...'; vet = F-waarde significant bij $p \leq 0.05$ getoetst met ANOVA (MCA); R² = verklaarde variantie van alle variabelen te zamen.

Tabel 2. Hoe gaat u met de toekomst van uw eigen leven om?

%	Cohort		Sekse		Opleiding			
	<29	29-39	man	vrouw	laag	midden	hoog	allen
- ik plan nu dingen voor mijn toekomst	38	40	40	38	38	37	47	39
- ik hou me niet met mijn toekomst bezig, de toekomst komt vanzelf	53	50	46	56	47	55	46	51
- ik wil er liever niet over nadenken	9	10	14	5	16	8	6	9

Bron: CentERdata Millennium Enquête

Toelichting: Hier worden de resultaten van enkelvoudige kruistabellen gepresenteerd, de verschillen zijn getoetst met logistische regressie, met dertigers, vrouwen en hoger opgeleiden als referentiecategorie; vet = significant bij $p \leq 0.05$.

heid van zowel twintigers als dertigers zich niet bewust met de toekomst wil bezighouden. Het zijn met name lager opgeleiden en mannen die er liever niet over nadenken en vrouwen die aangeven zich er expliciet niet mee bezig te willen houden.

Rest de vraag of jongeren in het algemeen anders om gaan met de toekomst, bij voorbeeld weinig geneigd zijn hun leven uit te stippelen of planmatig te anticiperen op de toekomst. De uitspraken over dit type omgangsstrategieën, uiteengelegd op een breed scala van terreinen, in tabel 3 geven uitkomst.

In vergelijking tot dertigers geven twintigers significant vaker aan dat de toekomst onvoorspelbaar is en ze liever 'bij de dag' leven, liever geen bindingen met verenigingen en clubs aangaan om zo tijd te sparen, minder neigen om elke dag te plannen, liever zo ongebonden mogelijk, met zo weinig mogelijk verplichtingen, door het leven gaan, hun sociale contacten minder graag ver van te voren plannen en het significant minder van belang vinden nu al geld opzij te leggen voor hun oude dag.

Al met al een stevige vingerwijzing dat jongeren hun toekomst weinig planmatig en zo open en vrij mogelijk willen benaderen. Daar staat tegenover dat zij het, anders dan dertigers, wel prettiger vinden om over de toekomst na te denken. Als ze dit doen, zijn ze evident optimistischer over de toekomst van hun eigen leven en van de samenleving. Ze vertalen hun optimisme, zo blijkt op andere plaatsen in het onderzoek, naar andere Nederlanders: van anderen denken ze ook dat zij de toekomst zonnig inzien.

Hoe vertaalt zich deze optimistische toekomstvisie als we een aantal maatschappelijke vraagstukken nader in beeld brengen? Voorafgaand aan de domeinspecifieke toekomstvisies gaan we na hoe meer in het algemeen gezien jonge cohorten maatschappelijke verschijnselen en problemen agenderen. Op die wijze ontstaat een meer concreet beeld over waar jongeren in hun maatschappelijke agenda van de toekomst de nadruk leggen. Beide cohorten is een breed palet van maatschappelijke problemen voorgelegd met de vraag of deze problemen in deze eeuw zullen toenemen of afnemen in Nederland. Hiermee kunnen we in kaart brengen of de twee cohorten verschillen in maatschappelijk probleembesef en waar op het maatschappelijk vlak hun 'hopes' en 'fears' liggen. Tabel 4 doet verslag.

Onder twintigers is er een opmerkelijke eensgezindheid over wat er in Nederland aan maatschappelijke verschijnselen toe- en afnemen. Afnemen volgens hen, als we letten op de top-drie van de fenomenen waarop jongeren een neer- of opwaartse trend waarnemen, de achterstand van vrouwen, het computer-analfabetisme en de betaalbaarheid van sociale voorzieningen. Een toename van communicatiemogelijkheden, genetische manipulatie en werkdruk wordt door jongeren voorspeld.

Voor iets meer dan de helft van de voorgelegde maatschappelijke problemen geldt, dat de toekomstverwachtingen van twintigers en dertigers significant verschillen. Het jongste cohort is *systematisch* minder pessimistisch over deze problemen. We treffen dit grotere optimisme aan bij verwachtingen rond werkloosheid, armoede, achterstand van vrouwen, oorlogsdreiging, genetische manipulatie, computer-analfabetisme, criminaliteit, spanningen tussen etnische groepen, asielzoekers, files en discriminatie. Twintigers blijken eveneens optimistischer met betrekking tot oudedagsvoorzieningen, economische groei, en gezondheidsvoorzieningen.

Tabel 3. Het leven met een vooruitziende blik.

	Cohort		Sekse		Opleiding			R ²	
	<29	29-39	man	vrouw	laag	midden	hoog		
- de toekomst is onvoorspelbaar, ik leef liever bij de dag	2,44	2,71	2,76	2,46	2,38	2,59	2,80	2,60	.06
- ik ben zo min mogelijk lid van clubs of verenigingen, zodat ik meer tijd overhoud	3,14	3,35	3,19	3,33	3,28	3,28	3,22	3,26	.01
- ik vind dat iemand elke morgen zijn dag moet plannen	3,74	3,44	3,45	3,66	3,31	3,63	3,60	3,56	.05
- ik ga het liefst ongebonden, met zo weinig mogelijk verplichtingen door het leven	2,74	3,14	2,90	3,04	2,83	3,04	2,94	2,98	.04
- ik vind het prettig over de toekomst na te denken	2,55	2,71	2,57	2,71	2,56	2,71	2,57	2,65	.02
- ik vind het belangrijk geen tijd te verspillen	2,85	2,75	2,79	2,78	2,51	2,78	2,98	2,78	.03
- ik vind het belangrijk nu een deel van mijn inkomen in verzekeringen te stoppen	2,88	2,82	2,72	2,95	2,80	2,75	3,02	2,84	.03
- het verdere verloop van mijn leven heb ik redelijk goed uitgestippeld	3,17	3,21	3,17	3,22	2,97	3,11	3,52	3,19	.05
- ik vind het belangrijk nu te sparen voor 'moeilijke' tijden	2,48	2,48	2,46	2,50	2,38	2,38	2,75	2,48	.03
- ik vind het belangrijk zo vroeg mogelijk toekomstplannen te maken	3,25	3,14	3,09	3,26	2,96	3,16	3,37	3,18	.04
- het liefst plan ik de tijd die ik aan sociale contacten besteed ruim van tevoren	3,53	3,28	3,29	3,44	3,25	3,35	3,50	3,37	.03
- ik vind het belangrijk nu geld opzij te leggen voor mijn oude dag	2,95	2,55	2,79	2,61	2,89	2,51	2,90	2,70	.07

Bron: *CentERdata Millennium Enquête*

Toelichting: Gemiddelden van 1 = zeer mee eens tot 5 = zeer mee oneens; vet = F-waarde significant bij $p \leq 0.05$ getoetst met ANOVA (MCA); R² = verklaarde variantie van alle variabelen tezamen.

Tabel 4. *Maatschappelijke problemen in de toekomst van Nederland.*

	Cohort		Sekse		Opleiding			R ²	
	<29	29-39	man	vrouw	laag	midden	hoog		
- werkloosheid	1,59	1,91	1,90	1,70	1,99	1,73	1,75	1,79	.06
- armoede	1,65	2,14	1,91	2,00	1,99	1,97	1,89	1,96	.09
- achterstand vrouwen	1,24	1,35	1,36	1,27	1,56	1,28	1,16	1,31	.08
- oudedagsvoorziening	1,83	1,49	1,57	1,65	1,68	1,69	1,41	1,61	.08
- spanningen godsdiensten	2,02	2,13	2,13	2,05	2,20	2,08	2,03	2,09	.02
- communicatiemogelijkheden	2,93	2,90	2,91	2,91	2,83	2,93	2,95	2,91	.02
- werkdruk	2,65	2,73	2,67	2,72	2,83	2,71	2,58	2,70	.02
- economische groei	2,50	2,37	2,39	2,44	2,60	2,40	2,30	2,42	.04
- oorlog	1,75	2,03	1,85	1,99	2,08	1,92	1,79	1,92	.08
- solidariteit	1,98	1,85	1,84	1,96	1,96	1,96	1,75	1,90	.03
- gezondheidsvoorzieningen	2,33	1,98	2,16	2,05	2,21	2,12	2,01	2,11	.06
- genetische manipulatie	2,61	2,73	2,70	2,67	2,69	2,64	2,77	2,69	.02
- milieuvervuiling	2,03	2,07	1,89	2,22	2,20	2,04	1,98	2,05	.04
- computer-analfabetisme	1,26	1,46	1,43	1,35	1,56	1,39	1,26	1,39	.04
- betaalbaarheid sociale voorzieningen	1,48	1,44	1,51	1,40	1,59	1,47	1,33	1,45	.03
- criminaliteit	2,48	2,65	2,61	2,57	2,73	2,57	2,49	2,59	.04
- spanningen tussen etnische groepen	2,29	2,43	2,41	2,35	2,41	2,40	2,30	2,38	.02
- asielzoekers	2,59	2,71	2,69	2,64	2,87	2,64	2,53	2,66	.05
- files	2,61	2,80	2,69	2,77	2,72	2,77	2,67	2,73	.04
- discriminatie	2,13	2,38	2,30	2,28	2,33	2,32	2,19	2,29	.03
- bevolkingsgroei	2,53	2,41	2,55	2,38	2,66	2,43	2,32	2,46	.05

Bron: *CentERdata Millennium Enquête*

Toelichting: Gemiddelden op een schaal van 1 = neemt af tot 3 = neemt toe; vet = F-waarde significant bij $p \leq 0.05$ getoetst met ANOVA (MCA); R² = verklaarde variantie van alle variabelen te zamen.

Hoop en vrees in meer maatschappelijke zin vertaalt zich niet zelden naar uiteenlopende visies op de verhouding tussen specifieke groepen in de samenleving. Hiermee raken we aan het thema van de sociale cohesie. Sociale cohesie staat vanouds centraal op de politieke en beleidsagenda, zeker in tijden van economische laagconjunctuur. Het concentreert zich op de vraag of de maatschappelijke samenhang in Nederland onder druk komt te staan en maatschappelijke groepen en geledingen verder uiteengroeien. Hoe denken twintigers en dertigers hier over, zijn er verschillen en zo ja blijkt dan ook nu weer dat het jongste cohort wat minder pessimistisch oordeelt over het probleem van de sociale cohesie? Tabel 5 geeft het antwoord.

We zien feitelijk slechts één verschijnsel waarbij gemiddeld genomen twintigers verschillen zien afnemen: die tussen mannen en vrouwen. Toenemende verschillen waargenomen door jongeren zijn er niet eenduidig vast te stellen. Onder de dertigers is het beeld niet veel anders. In de meerderheid van de gevallen blijken twintigers en dertigers geen andersluidend oordeel te hebben over door hen verwachte ontwikkelingen in

Table 5. Verschillen tussen groepen Nederlanders.

	Cohort		Sekse		Opleiding				R ²
	<29	29-39	man	vrouw	laag	midden	hoog	allen	
- buitenlanders in Nederland en Nederlanders	1,87	2,07	2,15	1,85	2,04	2,03	1,88	1,99	.05
- mensen met verschillende geloven	2,03	2,14	2,14	2,05	2,12	2,16	1,95	2,09	.02
- mensen in verschillende regio's of provincies	1,76	1,79	1,83	1,73	1,95	1,78	1,64	1,78	.03
- arme en rijke mensen	2,29	2,60	2,49	2,48	2,43	2,50	2,50	2,48	.06
- werkenden en niet-werkenden	2,26	2,30	2,33	2,24	2,35	2,25	2,31	2,29	.01
- mensen met en zonder vaste baan	2,06	2,31	2,28	2,15	2,42	2,20	2,05	2,21	.06
- hoger en lager opgeleide mensen	2,36	2,50	2,38	2,49	2,40	2,47	2,41	2,44	.02
- ouders en kinderen	1,96	2,02	2,09	1,92	2,14	1,97	1,93	2,00	.03
- mannen en vrouwen	1,53	1,67	1,62	1,62	1,94	1,57	1,45	1,62	.09
- generaties onderling	2,03	1,96	2,02	1,95	2,17	1,92	1,95	1,98	.03
- mensen die wel en niet met computers kunnen omgaan	1,98	1,93	2,02	1,89	2,07	1,89	1,96	1,95	.01
- mensen met veel en weinig vrije tijd	2,22	2,30	2,31	2,24	2,27	2,26	2,28	2,27	.01

Bron: CentERdata Millennium Enquête

Toelichting: Gemiddelden op een schaal van 1 = neemt af tot 3 = neemt toe; vet = F-waarde significant bij $p \leq 0.05$ getoetst met ANOVA (MCA); R² = verklaarde variantie van alle variabelen te zamen.

maatschappelijk onderscheid tussen een reeks van sociale groeperingen. Wanneer de verschillen evenwel significant zijn, blijken twintigers in relatie tot dertigers wederom wat optimistischer. Dit geldt met name voor de problematiek van allochtonen, de verschillen tussen arme en rijke mensen, het onderscheid tussen personen met en zonder een vaste baan, tussen hoger en lager opgeleiden en tussen mannen en vrouwen.

Kortom, het vergelijken van toekomstverwachtingen van twintigers en dertigers leidt tot een tweeledige conclusie: voor vele maatschappelijke vraagstukken treffen we aan dat beide generaties in gelijke mate pessimistisch dan wel optimistisch gestemd zijn en indien zich intergenerationale verschillen voordoen dan zijn dertigers systematisch pessimistischer dan twintigers.

Eén van de eerste specifieke kerndomeinen waarover jongeren 'hopes' en 'fears' hebben is onderwijs. Onderwijs is het cruciale levensdomein voor elke jonge generatie. Toekomstchansen worden goeddeels in dit domein bepaald en onderwijs speelt daarom ook een grote rol in de hoop en vrees die jongeren over de toekomst uiten. Meer zelfstandig keuzes maken en verantwoordelijkheden dragen worden gepaard aan meer communiceren, uitleggen en verantwoorden. Anticipatie en reflectie zijn kernkwaliteiten geworden, vooral in een samenleving waarin de verdeling van posities voorname-

lijk langs het behaalde opleidingsniveau verloopt en waarin kennis en vaardigheden onmisbare factoren zijn. In een dergelijke samenleving is keuzevrijheid ook keuzedwang. Geen keuze maken mag niet en zelf keuzes maken mag alleen als deze uitgelegd en verantwoord kunnen worden. Slechte keuzes maken is desastreus in een kennis-samenleving. Binnen deze culturele kaders, met zijn geboden en verboden, moeten jongeren opereren, graag of niet.

Jongeren doen dat graag. Een toenemend aantal jongeren volgt onderwijs, de leeftijd waarop ze van school komen loopt op, meer jongeren uit lagere sociale milieus en vooral meisjes hebben hun aanvankelijke onderwijsachterstand ingelopen. De omstandigheden echter waaronder de jongste generatie onderwijs volgt lijken minder gunstig

Tabel 6. Richting en wenselijkheid van rol scholing en kennis (selectie).

	Cohort		Sekse		Opleiding			R ²	
	<29	29-39	man	vrouw	laag	midden	hoog		
- mate waarin het hoger onderwijs op de eisen van de arbeidsmarkt gericht is ^a	2,24	2,39	2,44	2,18	2,45	2,22	2,33	2,32	.05
neemt toe en is ^b	2,24	2,42	2,50	2,20	2,41	2,45	1,92	2,34	.07
- zich blijvend moeten bijscholen op het werk ^a	2,23	2,04	2,09	2,16	2,22	2,09	2,03	2,12	.04
neemt toe en is ^b	2,32	2,24	2,26	2,28	2,30	2,27	2,25	2,27	.00
- combineren van het volgen van opleidingen met werk ^a	2,31	2,16	2,26	2,18	2,22	2,25	2,16	2,22	.02
neemt toe en is ^b	2,25	2,27	2,31	2,22	2,39	2,18	2,24	2,26	.02
- belang van computers voor het werk ^a	1,72	1,62	1,73	1,61	1,79	1,58	1,67	1,67	.04
neemt toe en is ^b	2,32	2,47	2,42	2,39	2,45	2,35	2,51	2,41	.01
- aantal mensen dat, als ze ergens gaan werken, eisen stelt aan de mogelijkheden voor scholing op het werk ^a	2,30	2,31	2,38	2,23	2,59	2,17	2,16	2,31	.09
neemt toe en is ^b	2,06	2,09	2,05	2,10	2,08	2,10	2,03	2,08	.00
- kans om er als werknemer een periode tussen-uit te gaan (zgn. sabbatsverlof) ^a	2,48	2,40	2,53	2,32	2,65	2,36	2,26	2,43	.06
neemt toe en is ^b	2,06	2,13	2,26	1,99	1,83	2,28	1,96	2,11	.09
- snelheid waarmee kennis en vaardigheden van werknemers verouderen ^a	2,32	2,20	2,23	2,27	2,30	2,27	2,10	2,25	.02
neemt toe en is ^b	3,51	3,80	3,65	3,72	3,43	3,79	3,75	3,69	.08

Bron: OSA Toekomst van de Arbeid Survey

Toelichting: Gemiddelden lopen van a: 1 = neemt sterk toe tot 5 = neemt sterk af; b: 1 = zeer wenselijke tot 5 = zeer onwenselijke ontwikkeling; vet = F-waarde significant bij $p \leq 0.05$ getoetst met ANOVA (MCA); R² = verklaarde variantie van alle variabelen te zamen.

dan die van andere naoorlogse generaties. De ruime vrijheden de schoolloopbaan naar believen in te richten, zoals met name de jaren-zestiggeneratie die had, zijn beknot. De prestatie- en werkdruk zijn toegenomen. Het eisenpakket dat vanuit de samenleving richting school gestuurd wordt, wordt haast dagelijks verzwaard. Ook jongeren onderkennen dit, zoals tabel 6 aangeeft.

Naar de toekomst toe bezien, onderkennen twintigers, precies als dertigers, onder andere dat het onderwijs zich in toenemende mate op de eisen vanuit de samenleving, bij voorbeeld vanuit de arbeidsmarkt, oriënteert. De jonge cohorten (twintigers en dertigers) beschouwen een dergelijke oriëntatie overigens niet negatief, maar zien dat eerder als een wenselijke ontwikkeling. Dat geldt in het algemeen ook voor de grotere betekenis die zij aan scholing en vooral ook aan kennis en vaardigheden op het gebied van informatie- en communicatietechnologie toekennen. In de toekomstvisie van de jongere cohorten nemen een leven lang leren, de combinatie leren met werken en ook ICT op het werk een grote vlucht. Daar staat wel wat tegenover. Het is waarschijnlijk en wenselijk dat men eisen stelt over scholing als men ergens gaat werken en dat men de kans krijgt sabbatsverlof op te nemen. Waarschijnlijk en onwenselijk acht men de versnelde veroudering van kennis en vaardigheden van werknemers. Realiseren jongeren zich volgens tabel 4 dat werkdruk een reëel fenomeen is, nu voegen ze er aan toe dat de druk om kennis en scholing permanent te onderhouden en te blijven innoveren onontkoombaar is. Desalniettemin lijkt de jongste generatie klaar te zijn voor een toekomst waarin permanent scholen en werken in met name kennisintensieve banen steeds nauwer verweven zijn.

De volgende bron van hoop en vrees is voor jongeren werk, betaalde arbeid wel te verstaan. Positief zijn twintigers, zoals is af te lezen uit tabel 7, toch vooral over de eigen beroepskansen en carrièremogelijkheden, de toenemende eigen regie in het werk om zelf tot afspraken met werkgevers over arbeidsvoorwaarden te komen en de versterkende nadruk op een goed loon en salaris in de toekomst. Prominent lijkt dus de eigen ruimte om materieel vooruitgang te boeken.

Negatief zijn jongeren ook en wel over het groeiend aantal buitenlanders dat in Nederland komt werken, het verdwijnen van de ambachtelijke baan en de baan voor lageropgeleiden, meer wedijver en competitie tussen werknemers, de afnemende loyaliteit van werknemers, een toekomst waarin een scala van aspecten van werkdruk en stress op het werk toeneemt¹⁵, de privatisering van de sociale zekerheid, het langer doorwerken van ouderen en het afkalven van de AOW. Het is een combinatie van vertrouwen in meer eigen regie, juist in materieel opzicht, en van vrees voor het verdwijnen van oude, collectief geregelde zekerheden en voor de negatieve gevolgen van het als individu de touwtjes in handen moeten nemen. Zoals we al eerder aangaven gaat hoop op één terrein gepaard aan vrees op datzelfde terrein. In dit geval wordt eigen regie zowel verwelkomd als gevreesd.

Indien twintigers en dertigers verschillen in hun oordeel over toekomstontwikkelingen binnen het domein van arbeid, inkomen en sociale zekerheid dan is de jongste groep echter wel systematisch meer optimistisch of minder pessimistisch. Twintigers beoordelen tal van ontwikkelingen steeds iets zonniger dan dertigers. Deze bevinding loopt als een rode draad door onze analyses. Over trends op de arbeidsmarkt zijn beide groe-

Tabel 7. Richting en wenselijkheid van veranderingen in de arbeidsmarkt (selectie).

	Cohort		Sekse		Opleiding			R ²	
	<29	29-39	man	vrouw	laag	midden	hoog		
A.									
- eigen beroepskansen en carrièremogelijkheden	2,20	2,49	2,27	2,45	2,57	2,25	2,27	2,36	.05
B.									
- het maken van individuele afspraken tussen een werknemer en zijn/haar werkgever over inkomen en andere arbeidsvoorwaarden ^a	2,34	2,31	2,25	2,39	2,46	2,34	2,07	2,32	.04
<i>neemt toe en is^b</i>	2,45	2,68	2,63	2,54	2,61	2,53	2,68	2,59	.02
- nadruk op goed salaris ^a	2,23	2,55	2,49	2,34	2,57	2,37	2,27	2,41	.05
<i>neemt toe en is^b</i>	2,25	2,33	2,40	2,21	2,22	2,27	2,47	2,29	.02
- het aantal buitenlanders (al dan niet Europeaan) dat in Nederland werkt ^a	2,32	2,18	2,17	2,32	2,37	2,17	2,21	2,24	.04
<i>neemt toe en is^b</i>	3,28	3,36	3,44	3,22	3,43	3,38	2,99	3,33	.05
- het aantal ambachten ^a	3,91	3,87	3,90	3,87	3,75	3,97	3,86	3,89	.02
<i>neemt af en is^b</i>	3,88	4,07	3,95	4,02	4,01	3,98	3,99	3,99	.02
- de hoeveelheid werk voor laagopgeleiden ^a	3,48	3,46	3,52	3,42	3,29	3,57	3,51	3,47	.02
<i>neemt af en is^b</i>	3,81	4,03	3,83	4,03	3,83	4,03	3,85	3,94	.06
- wedijver en competitie tussen werknemers op het werk ^a	2,45	2,27	2,33	2,36	2,50	2,28	2,29	2,35	.03
<i>neemt toe en is^b</i>	3,69	4,04	3,92	3,89	3,90	3,90	3,93	3,90	.04
- de trouw van werknemers aan de bedrijven/organisaties waar ze voor werken ^a	3,76	3,82	3,83	3,76	3,64	3,79	4,10	3,80	.05
<i>neemt af en is^b</i>	3,62	3,54	3,62	3,53	3,57	3,62	3,47	3,57	.01
- het gemak waarmee werknemers ontslagen kunnen worden ^a	2,72	2,41	2,38	2,70	2,49	2,59	2,50	2,54	.06
<i>neemt toe en is^b</i>	4,07	4,13	4,01	4,22	4,06	4,24	3,81	4,10	.05
- het tempo van het werk zelf kunnen regelen ^a	3,01	3,47	3,38	3,17	3,35	3,32	2,97	3,27	.08
<i>neemt af en is^b</i>	3,97	3,90	3,87	3,99	3,82	4,00	3,86	3,93	.02
- last van stress door het werk hebben ^a	2,27	1,88	2,21	1,91	1,95	2,11	2,10	2,05	.10
<i>neemt toe en is^b</i>	4,30	4,30	4,35	4,26	4,25	4,28	4,46	4,30	.02
- privatisering van de sociale zekerheid? ^a	2,40	2,24	2,34	2,28	2,65	2,19	2,05	2,31	.09
<i>neemt toe en is^b</i>	3,89	4,03	3,99	3,99	4,09	3,95	3,98	3,99	.01

Tabel 7. Vervolg.

	Cohort		Sekse		Opleiding			R ²	
	<29	29-39	man	vrouw	laag	midden	hoog		
- het aantal ouderen dat verplicht wordt om langer door te werken om de verzorgingsstaat in stand te houden ^a	2,43	2,18	2,23	2,33	2,69	2,08	2,20	2,28	.15
<i>neemt toe en is^b</i>	3,92	3,88	3,75	4,03	3,86	3,98	3,69	3,90	.04
- de kans dat de AOW wordt afgeschaft ^a	2,40	2,23	2,23	2,37	2,50	2,21	2,24	2,30	.05
<i>neemt toe en is^b</i>	4,12	4,38	4,29	4,26	4,32	4,26	4,27	4,28	.02
- de hoogte van de AOW voor toekomstige generaties ^a	3,70	3,80	3,71	3,80	3,68	3,72	4,00	3,76	.03
<i>neemt af en is^b</i>	4,17	4,25	4,22	4,22	4,27	4,22	4,17	4,22	.01

Bron: OSA Toekomst van de Arbeid Survey

Toelichting: A. Gemiddelden lopen van 1 = met zeer veel vertrouwen tot 5 = met zeer weinig vertrouwen, B. Gemiddelden lopen van a: 1 = neemt sterk toe tot 5 = neemt sterk af en b: 1 = een zeer wenselijke ontwikkeling tot 5 = een zeer onwenselijke ontwikkeling; vet = F-waarde significant bij $p \leq 0.05$ getoetst met ANOVA (MCA); R² = verklaarde variantie van alle variabelen te zamen.

pen niet geheel eensgezind. Twintigers zijn over het algemeen zeer optimistisch over de eigen beroepskansen en hun toekomstige arbeidsmarktpositie en relatief minder pessimistisch over de keerzijden van een zich doorzettende flexibilisering, de onderlinge wedijver en competitie op het werk, toekomstige arbeidsverhoudingen en -contracten (loyaliteit aan het bedrijf, CAO's, werken in loondienst, versoepeling van het ontslagrecht). In absolute zin, echter, zijn ook twintigers bezorgd over deze keerzijden. Dit geldt zeker ook voor ontwikkelingen die men voor waarschijnlijk houdt binnen het domein van de sociale zekerheid: privatisering van de sociale zekerheid en het knagen aan de wortels van de AOW.

Op tal van punten laat de groep twintigers evenzeer pessimistische geluiden horen, zij het verhoudingsgewijs minder dan dertigers. Het lijkt erop dat de groep dertigers wat gevoeliger is voor de negatieve aspecten van trends binnen het domein van arbeid, inkomen en sociale zekerheid. De groep twintigers lijkt in dit opzicht wat meer onbevangingen. Onbevangingen is een goede zaak, maar heeft ook een schaduwzijde. Bij tegenvallende ontwikkelingen is men mentaal minder geëquipeerd want slechter voorbereid. Immers de huidige twintigers hebben tijdens hun formatieve periode louter tijden van economische voorspoed meegemaakt en beschikken dus niet over een op eigen ervaringen gebaseerde psychologische bagage om ook slechte tijden het hoofd te bieden. Dat blijkt ook uit een ander soort toekomstuitrusting: zij lopen significant achter bij dertigers als het gaat om het treffen van financiële toekomstvoorzieningen.

De maatschappelijke dynamiek van "goede tijden" en "slechte tijden" die zo bepalend is voor individuele weerbaarheid, is goedgevoelend aan de jongste groep Nederlanders voor-

bijgegaan. Jeugdwerkeloosheid, bijvoorbeeld, is een fenomeen dat ze slechts kennen van horen zeggen. De achterkant van het jeugdig optimisme is daarmee de vraag hoe jongeren zullen reageren op tegenvallende maatschappelijke ontwikkelingen en hoe het in dit opzicht gesteld is met hun individuele vaardigheden.

Het derde domein waarover jongeren droom- en doembeelden hebben is dat van de intieme relaties. In ons onderzoek hebben we dit thema vooral benaderd vanuit twee mogelijk strijdige idealen, idealen die bij het combineren van een privéleven en een arbeidzaam bestaan een dominante rol spelen. Veranderde idealen over wat dé werkne-mer en hét gezinslid van de toekomst is, zijn sterk bepalend voor hoe jongeren met keuzes op dit terrein in het traject naar volwassenheid omgaan. Een grote toewijding aan taken in de betaalde arbeid en een sterke betrokkenheid bij wat zich in privéleven aan verantwoordelijkheden aandient, zijn aan het begin van het nieuwe millennium de overheersende idealen. Een centraal onderdeel in het beleidsdebat van vandaag is het scheppen van voorwaarden om mensen in de gelegenheid te stellen om hun aspiraties over het combineren van beide idealen te realiseren. Het zijn leden van de jongste cohorten die de ruimte die geboden wordt en de grenzen die gesteld worden om deze aspiraties te verwezenlijken het sterkst voelen. Dit geeft een grote urgentie aan inzicht in hoe zij ontwikkelingen op deze terreinen beoordelen en waarderen.

De verwachtingen rondom werken en zorgen zijn hooggespannen onder jongeren. Ze blaken van optimisme over betaald zorgverlof, goede regelingen op het werk om thuis de nodige zorg te verlenen, kinderopvang bij bedrijven tot en met een eerlijker verdeling van topbanen over de seksen. Alle zaken die jongeren zeker niet zelf als eersten onder regie hebben, maar vooral in de handen van sociale partners en de overheid liggen. Tabel 8 rapporteert.

Gemiddeld genomen denken jongeren dat het werken en zorgen in de toekomst makkelijker wordt en opnieuw blijkt dit vertrouwen sterk gebaseerd op uitbreiding van de nodige regelingen en voorzieningen. Zijn de eigen voorspelde gedragskeuzes in lijn met dit gematigd optimisme? Uit het onderzoek komt naar voren (niet in de tabel) dat bijna de helft van de jongeren later een parttime baan wil naast kinderen. Toch nog een kwart werkt liever fulltime naast ouderschap. Andere modellen van het combineren van werk en zorg voor kinderen hebben nauwelijks aanhang onder jongeren. Voorts menen jongeren vooral zelf meer tijd aan betaalde arbeid te gaan besteden in de toekomst. Meer tijd aan de zorg voor kinderen of voor anderen of meer tijd aan ouderschapsverlof (hoewel jongeren dat meer zeggen te willen dan dertigers) is niet sterk in het toekomstbeeld van jongeren verankerd. Het optimisme over de combineerbaarheid van arbeid en zorg wordt niet vertaald in eigen gedragskeuzes die opvallend meer ruimte bieden gemakkelijker aan beide taken toe te komen.

Wat wel opvallend is, is de overeenkomst tussen de inschattingen van twintigers en dertigers over de combineerbaarheidskwestie. Betaald zorgverlof, kinderopvang en meer sekselijkheid bij taken in de betaalde en onbetaalde arbeid stellen zij allen in het vooruitzicht. Nog opvallender is de traditionele waardering van deze trends door de jongste cohorten. Twintigers achten fulltime werken bij het hebben van kinderen, mannen die thuis kunnen zijn, kinderopvang en sekselijkheid minder wenselijk dan dertigers. Deze laatste kiezen vaker voor parttime werken naast ouderschap, zij het ook voor fulltime werken naast ouderschap. De gedachte dat jongere cohorten meer voor een afwisselingsmodel (werken-ouderschap-werken) kiezen, blijkt in onze gegevens

Tabel 8. Richting en wenselijkheid van trends in arbeid en zorg.

	Cohort		Sekse		Opleiding			R ²	
	<29	29-39	man	vrouw	laag	midden	hoog		
- betaald zorgverlof voor mannen en vrouwen ^a	2,30	2,28	2,34	2,24	2,35	2,29	2,20	2,29	.02
<i>neemt toe en is</i> ^b	2,15	2,01	2,22	1,94	2,27	1,96	2,07	2,07	.07
- tijd voor het gezin ^a	2,95	3,01	2,86	3,10	2,99	3,01	2,92	2,99	.02
<i>neemt toe en is</i> ^b	1,94	1,83	1,98	1,79	2,02	1,80	1,91	1,88	.06
- combinatie fulltime werken en kinderen ^a	2,46	2,45	2,29	2,59	2,35	2,47	2,56	2,45	.04
<i>neemt toe en is</i> ^b	3,60	3,82	3,70	3,73	3,64	3,75	3,81	3,72	.02
- werkgevers geven mannen gelegenheid voor tijd thuis ^a	2,43	2,38	2,46	2,35	2,54	2,38	2,22	2,40	.03
<i>neemt toe en is</i> ^b	2,03	1,82	2,04	1,80	1,96	1,89	1,87	1,91	.08
- vrouwen kiezen gezin boven carrière ^a	2,96	3,32	3,18	3,15	2,95	3,24	3,34	3,17	.08
<i>neemt toe en is</i> ^b	3,75	3,56	3,75	3,54	3,74	3,67	3,34	3,63	.06
- kinderopvang bij bedrijven ^a	2,10	2,08	2,14	2,05	2,22	2,02	2,05	2,09	.03
<i>neemt toe en is</i> ^b	2,24	2,09	2,23	2,11	2,24	2,12	2,12	2,16	.02
- aantal uren dat men buitenshuis wil werken ^a	3,07	3,35	3,28	3,19	3,06	3,26	3,45	3,23	.07
<i>neemt af en is</i> ^b	2,56	2,50	2,64	2,40	2,36	2,57	2,55	2,52	.03
- seksegelijkheid tijd aan zorg voor huishouden ^a	2,43	2,41	2,39	2,44	2,33	2,50	2,34	2,42	.02
<i>neemt toe en is</i> ^b	2,28	2,00	2,34	1,93	2,24	2,08	2,01	2,13	.13
- seksegelijkheid tijd aan zorg voor kinderen ^a	2,42	2,35	2,38	2,38	2,30	2,48	2,25	2,38	.02
<i>neemt toe en is</i> ^b	2,22	1,96	2,10	2,06	2,20	2,06	1,92	2,08	.06
- seksegelijkheid buitenshuis werken ^a	2,14	2,23	2,20	2,18	2,28	2,15	2,14	2,19	.01
<i>neemt toe en is</i> ^b	2,58	2,69	2,75	2,55	2,73	2,63	2,53	2,64	.02
- seksegelijkheid verdeling topbanen ^a	2,30	2,41	2,45	2,29	2,52	2,28	2,29	2,36	.04
<i>neemt toe en is</i> ^b	2,08	1,96	2,15	1,93	2,17	1,97	1,92	2,02	.06

Bron: OSA Toekomst van de Arbeid Survey

Toelichting: Gemiddelden lopen van a: 1 = neemt sterk toe tot 5 = neemt sterk af; b: 1 = een zeer wenselijke ontwikkeling tot 5 = een zeer onwenselijke ontwikkeling; vet = F-waarde significant bij $p \leq 0.05$ getoetst met ANOVA (MCA); R² = verklaarde variantie van alle variabelen te zamen.

geen steun te vinden. Het veelal als ideaal afgeschilderde combinatiemodel (parttime werken en ouderschap) is, zoals aangegeven, voor iets minder dan de helft van de twintigers een zinvolle toekomstoptie. Bij dertigers (die deze keuze vaak al gerealiseerd hebben) is dat precies de helft. Uit onze open vragen blijkt dat meerderheden van beide cohorten voorts menen dat het combineren van arbeid en zorg (voor kinderen) in de toekomst gemakkelijker wordt. Redenen zijn bij beide groepen de hoop op meer kinderopvang, verlofregelingen, flexibilisering en een betere afstemming van werktijden, hoop die een sterk beroep doet op een intensieve beleidsinzet op dit terrein.

Gezien het optimisme en de hoop op betere tijden, zoals die met name in de domeinen werk en privé spelen (hoop die overigens sterk inzet op verruiming van regelingen en voorzieningen) springt het onbestemde gevoel van tijdsdruk tamelijk in het oog. Mogelijk dat de vroege betrokkenheid en binding van twintigers met volwassen rollen, ook afgemeten aan hun zorgen over de keerzijden van flexibilisering en privatisering, hun hooggespannen verwachtingen over de toekomst van het combineren van werk en privé en hun (elders in ons onderzoek gebleken) sterk ontwikkeld waardebesef dat zij als individu onafhankelijk van collectiviteiten (als kerk, overheid of vakbond) hun eigen keuzes kunnen maken, dit onbestemde gevoel tijd tekort te komen veroorzaken. Nog geen dertiger, maar zich volop bewust van een toekomst met een toenemende druk om op een breed aantal terreinen zelf te moeten presteren. Wat voorheen pas ging spelen rond de veertig, zo zou men kunnen speculeren, speelt zich nu voor de dertig af: de *quartlife crisis* neemt de plaats in van de *midlife crisis*.

5. OPTIMISME IN ONZEKERE TIJDEN

Het eerste dat opvalt aan het toekomstperspectief van jongeren is dat het op het algemene niveau onbestemd en op het niveau van bepaalde levensdomeinen juist heel helder is. In algemene zin zeggen ze tien jaar vooruit te kijken, zowel wat betreft ontwikkelingen in hun persoonlijke leven als die in de samenleving. Meerderheden zeggen echter zich niet met 'de' toekomst bezig te houden, liever bij de dag te leven en vooral hun dagelijkse en sociale leven niet bijzonder strak vooruit te plannen. Jongeren vinden het prettig over de toekomst na te denken en ze verwachten ook een mooie, betere toekomst, maar op dit algemene niveau lijkt de toekomst ze als thema niet erg aan te spreken.

Dat verandert als het perspectief naar specifieke ontwikkelingen in hun toekomstige leven en in dat van de samenleving wordt verlegd. Jongeren blaken dan van optimisme. Ze menen en wensen met name dat de achterstand van vrouwen in maatschappelijk opzicht wordt ingelopen, er meer seksegelijkheid in het werk gaat ontstaan en de verschillen tussen mannen en vrouwen in dit opzicht steeds kleiner worden. Ook technologie in het werk en in de communicatieve en medische hoek gaat aan belang winnen en ook hierover is optimisme het sleutelwoord. De toekomstoriëntatie richt zich vooral in positieve zin (hoop) op seksegelijkheid en technologie en in negatieve zin (vrees) op de stijgende werkdruk, de druk op sociale voorzieningen en de druk op de samenleving als zodanig door spanningen tussen etnische groepen, criminaliteit en asielzoekers.

Druk is een woord dat we bij diverse voorspellingen tegenkomen, zo ook in die ten aanzien van het onderwijs. Hier gaat het vooral om tijdsdruk, de als negatief gewaardeerde pressie om binnen korte tijd een studie af te ronden. Jongeren lijken terdege te voelen dat ze in dit domein keuzes moeten maken, keuzes die ze moeten kunnen uitleggen en die van groot belang zijn voor het latere leven. Overigens waarderen ze een sterkere arbeidsmarktorientatie van het onderwijs en onderkennen ze het belang permanent aan de eigen kennis en vaardigheden te blijven schaven, ook na de formele schoolperiode. Richting werkgevers eisen zij dan wel ruimte, bij voorbeeld via het sabbatsverlof.

Kijkend naar het terrein van de arbeid zien we het woord druk prominent terugkeren. Werkdruk en werkstress zijn sterk gevreesde maar als onvermijdelijk voorspelde toekomstbeelden. Jongeren zijn sterk optimistisch getint over hun eigen beroepskansen en

ook over de mate waarin zij zelf met werkgevers afspraken kunnen maken over inkomen en arbeidsvoorwaarden. Het geloof in eigen regie is groot. Tegelijk vrezen zij de komst van buitenlanders op de arbeidsmarkt, de teloorgang van het ambacht en van het werk voor laagopgeleiden, de toename in wedijver en competitie en een afname in de loyaliteit van werknemers. Ook de privatisering van de sociale zekerheid, waarbij zij zelf hun toekomst moeten verzekeren, wordt zowel als onvermijdelijk als ook onwenselijk gezien, net als het langer doorwerken van toekomstige ouderen en de afbrokkeling van de AOW. Al met al wensen jongeren hun toekomst op de arbeidsmarkt in eigen hand te nemen, maar vrezen zij gelijktijdig de gevolgen van deze zelfideologie. Het uit beeld raken van collectieve voorzieningen en meer druk op het individu om in wedijver, permanent en bezien over de levensloop langer te presteren zijn de aanstaande onwelkome toekomstige voor jongeren.

Hoe moet dat als de aspiraties in werk gekoppeld worden aan de ambities thuis en in de vrijetijd? Ook dan stellen jongeren duidelijke voorwaarden. Ze zijn zeer optimistisch over de toekomstige combinatiemogelijkheden van arbeid en zorg, maar rekenen daarbij op sabbatsverlof, zorgverlof, goede werkgevers die mannen ruimte geven thuis te zijn, kinderopvang bij bedrijven, en ook aan de top een eerlijker man-vrouwverhouding. Het eigen voorspelde keuzegedrag is niet zonder meer op dit vertrouwen afgestemd. Parttime werken naast ouderschap is de keuze van vele jongeren, maar toch vooral van vrouwen. Fulltime werken naast ouderschap is nog steeds de keuze van jonge mannen. Ook voorspellen jongeren juist meer tijd voor betaalde arbeid vrij te maken. Een dergelijke voorspelling doen zij niet voor zorg thuis en het nemen van zorgverlof. Zonder een sterke uitbreiding van voorzieningen om werk en zorg te combineren wordt het combineren van arbeid en zorg met een dergelijk voorspeld gedragsrepertoire een ware duivelskunst.

Hebben twintigers een eigen en onderscheiden toekomstvisie of hoe verhoudt zich het voorgaande profiel ten opzichte van dat van dertigers?

Bij de algemene maatschappelijke trends blijkt dat maar mondjesmaat het geval. Als zich echter verschillen voordoen dan geldt in nagenoeg alle gevallen dat twintigers minder pessimistische toekomstverwachtingen hebben dan dertigers. In het bijzonder kan vastgesteld worden dat de jongste leeftijdsgroep optimistischer is over een reeks van maatschappelijke problemen in Nederland. Ze hebben weliswaar oog voor deze problemen, maar hun verwachtingen zijn wat minder somber gekleurd. Beide leeftijdsgroepen oordelen tamelijk gelijkkluidend over het vraagstuk van de sociale cohesie in de toekomstige Nederlandse samenleving.

Ook zijn kwaliteiten als reflectie en anticipatie met betrekking tot het domein onderwijs opmerkelijk goed ontwikkeld bij jongeren. Dat wil echter niet zeggen dat er sterke verschillen zijn met het hun voorgaande cohort. Beide cohorten nemen de sterke oriëntatie van onderwijs op de eisen van de arbeidsmarkt waar, zij het dat vooral twintigers dit primair als wenselijke ontwikkeling zien. Het toenemend belang van kennis en ICT-vaardigheden wordt door beide als wenselijk onthaald, net als het een leven lang leren en de combinatie van leren en werken. Ook het meer eisen stellen voor mogelijkheden tot scholing en sabbatsverlof zodra men gaat werken, zien beide cohorten als waarschijnlijk en wenselijk. Leden van de jongste generatie zijn, net als die voor hen, klaar voor een toekomst van permanent leren en werken in kennisintensieve banen.

Het grotere optimisme – of geringere pessimisme – van jongeren keert ook terug in het domein van arbeid, inkomen en sociale zekerheid. De jongste groep ziet weliswaar ook

de schaduwzijden van de trend naar verdergaande flexibilisering van de arbeid, naar hogere werkdruk, afnemende arbeidsvoldoening, toename van wedijver en concurrentie op de werkvloer, verzakelijking van de arbeidsverhoudingen, onregelmatige werktijden e.d., maar toont zich ook hier wat minder bezorgd. Beide groepen voorspellen ICT en telewerken een grote toekomst. De jongste groep is optimistischer over de eigen beroepsloopbaan en inkomenszekerheid. Beide cohorten zijn bezorgd om privatiseringstendensen en verminderde solidariteit binnen het domein van de sociale zekerheid, ook hier zijn twintigers echter wat minder pessimistisch.

Jongeren zien, samenvattend, voor zichzelf een mooie toekomst naderen die ze zeker optimistisch tegemoet treden. Ze onderkennen wel dat dat een toekomst wordt met meer druk, zowel als het om het onderhouden van kennis en vaardigheden gaat, om het werk, om de combinatie van werk en zorg of om de ruimte om aan zichzelf toe te komen en leuke dingen te kunnen doen. Zij lijken ondanks hun rotsvaste geloof in eigen kunnen te rekenen op een sterke inzet van anderen, van sociale partners en van de overheid, met name om het opzetten of het instandhouden van goede verlofregelingen, opvangmogelijkheden, sociale (toekomst)voorzieningen en ruimer beschikbare faciliteiten in de 24-uurssamenleving.¹⁶ Jongeren zijn meer dan dertigers optimistisch over de toekomst, maar ze hebben net als dertigers een scherp oog voor onzekerheden. Het jongste vooruitzicht is een vorm van optimisme in onzekerheid. Ze lijken rekening te houden met een stijgende druk om op tal van domeinen te presteren, maar zijn onzeker of andere partijen nodig voor die prestatie dat ook zullen doen.

NOTEN

1. Zie ook ZSE (2002) over controverses rondom het failliet van 'Fremdsozialisation' enerzijds en 'Selbstsozialisation' als nieuw richtinggevend begrip in de jeugdstudies anderzijds. In een sociologisch kader valt de gestegen populariteit op van verwante theorieën rondom individualisering, reflexiviteit en detraditionalisering (m.n. Heelas et al., 1996; Beck et al., 1994) en de kritiek op dit type theorievorming, zeker ook vanuit empirische hoek, onder verwijzing naar het bestaan van sterke sociale verschillen terzake (b.v. reflexiviteit als kenmerk van betergesitueerden) en de aanhoudend traditionele en sterk uniforme keuzen die groepen in de maatschappij in hun levensloop gezamenlijk maken (m.n. Derks, 2000; Elchardus, 1999; Stevens & Elchardus, 2001).
2. Zie voor meer informatie over deze instituten de websites <http://www.uvt.nl/iric>, <http://www.uvt.nl/globus> en <http://www.uvt.nl/osa>.
3. Zie voor nadere informatie over het CentERpanel Blankert et al., 1998.
4. Het onderzoek, de *CentERdata Millennium Enquête*, is gehouden onder bijna 1400 Nederlanders in mei 1999 en is representatief voor Nederlanders van alle leeftijden (ouder dan 16 jaar), geslachten, opleidings- en inkomensniveaus.
5. Zie hiervoor Diepstraten et al., 1999b.
6. Zie voor een verdere uitwerking van deze drieslag tussen gepercipieerde maatschappelijke ontwikkelingen, de mate van wenselijkheid van deze ontwikkelingen en de verwachte invloed van deze ontwikkelingen op het eigen bestaan Diepstraten et al. 1999b.
7. Voor dit onderzoek is een steekproef getrokken van bijna 1600 respondenten die in de week-ends van 22 oktober, 29 oktober en 26 november 1999 het OSA Toekomst van de Arbeid Survey kregen voorgelegd. De vragenlijst vormt opnieuw een combinatie van open en gesloten vragen, een combinatie die, optimale informatiemogelijkheden biedt. De steekproef (N=1574) is representatief voor de Nederlandse bevolking naar kernindicatoren als leeftijd,

- opleidingsniveau en inkomen. Voor sekse is dit minder het geval, het aantal mannen is oververtegenwoordigd waardoor de data zijn herwogen. Ook naar sekse is het onderzoek nu representatief.
8. Zie o.a. Israelashvili, 1997; Kasurinen, 1997; Malmberg & Norrgård, 1999; Malmberg & Trempela, 1997; Nurmi, 1992, 1994; Pulkinen & Rönkä, 1992; Seginer & Halabi-Kheir, 1998.
 9. Het begrippenpaar bevels- en onderhandelingshuishouding stamt uit een invloedrijk artikel van De Swaan (1982).
 10. Zie ook Vinken, 1997, voor een overzicht en empirische uitwerking.
 11. Te beginnen bij de thesen van Fuchs, 1983.
 12. De proportie jongeren dat zowel persoonlijk als maatschappelijk veeleer optimistisch en hoopvol (dan neutraal of somber/pessimistisch) is, is in 1999, na een aanvankelijk scherpe daling tussen 1991 en 1996, weer sterk toegenomen. De helft van de jongeren is in 1999 persoonlijk optimist. Ronduit pessimistisch is een tiende deel van de jongeren (bijna constant sinds het midden van de jaren tachtig). De maatschappelijk optimisten waren alleen in 1981 in de minderheid ten opzichte van maatschappelijk pessimisten. In 1999 is dan weer het optimismeniveau bereikt van 1991. De gemiddelde reikwijdte van het toekomstperspectief (hoe ver er gemeten in jaren vooruit gekeken wordt) is de laatste twintig jaar nauwelijks toe- of afgenomen: een derde van de jongeren kijkt tot één jaar vooruit. Hoe ouder jongeren worden, hoe verder ze vooruit kijken. Twee derde heeft in 1999 een scherp beeld van de eigen toekomst en ruim drie kwart denkt zelf (zeker) invloed op de eigen toekomst te kunnen uitoefenen. Jugend 2000 biedt geen houvast om te concluderen dat jongeren onder invloed van aanhoudende werkloosheid (vergeleken met Nederland in Duitsland tamelijk fors) of onder invloed van flexibilisering, globalisering of andere metaforen voor maatschappelijke hyperinnovatie het een niet meer van het ander kunnen onderscheiden. Integendeel, jongeren verwerpen het idee dat ze niet meer kunnen rekenen op een levenlang één beroep uitoefenen, dat ze zich steeds moeten aanpassen aan verandering, dat ze zich voorbereiden op korte bindingen, permanente omschakelingen, omdenken en opnieuw van vooraf aan beginnen leren en leren beginnen. Ze bereiden zich optimistisch, met een helder toekomstbeeld en tamelijk vastbesloten voor om hun beroepsmogelijkheden en kansen op een bevredigend privébestaan uit te baten. Ook uit de kwalitatieve data, verkregen uit reeksen open individuele en groepsinterviews, blijkt dat jongeren zich in hun toekomstplannen afkeren van sombere cultuurpessimistische toekomst (Fuchs-Heinritz, 2000: 394-395).
 13. Voor inzicht in meer domeinen en meer algemene beelden over maatschappelijke trends, politieke en culturele ontwikkelingen verwijzen we naar 'Aan ons de toekomst' (Vinken et al., 2002).
 14. Zie ook Diepstraten et al., 1999b: 64 en verder voor een dergelijk betoog.
 15. Hier zijn slechts twee aspecten vermeld. Zie Vinken et al., 2002 voor meer cijfers en detailinformatie rondom werkdruk en werkstress.
 16. Zie Vinken et al., 2002 en vooral ten aanzien van de 24-uurseconomie ook Ester & Vinken, 2001.

LITERATUUR

- Baltes, P. B. & M. M. Baltes (1990). *Successful aging: perspectives from the behavioral sciences*. Cambridge: Cambridge University Press.
- Beck, U., A. Giddens & S. Lash (1994). *Reflexive modernization*. Politics, tradition and aesthetics in the modern social order. Stanford, CA: Stanford University Press.
- Blankert, W., M.A. Conijn, A.F.M. Felix, D. Sikkel & C.M. Vis (1998). *Data quality at a Telepanel: A research program*. Tilburg: Tilburg University: CENTERDATA.
- Bois-Reymond, M. du & M. Oechsle (Hrsg.) (1990). *Neue Jugendbiographie? Zum Strukturwandel der Jugendphase*. Opladen: Leske + Budrich.

- Bois-Reymond, M. du, E. Peters & J. Ravesloot (1994). *Keuzeprocessen van jongeren. Een longitudinale studie naar veranderingen in de jeugdfase en de rol van de ouders*. Den Haag: VUGA.
- Bois-Reymond, M. du, W. Plug, Y te Poel & J. Ravesloot (2000). *En dan maar vaker kijken...?* Onderwijs- en werktrajecten van jongeren: een longitudinaal onderzoek. Paper gepresenteerd tijdens de Onderwijsresearchdagen 2000 (24-26 mei) te Leiden.
- Broek, A. van den & K. Breedveld (2000). Rollen, ritmes en routines. *Sociale Wetenschappen* 49, 2: 9-26.
- Derks, A. (2000). *Individualisme zonder verhaal*. Brussel: VUB Press.
- Dieleman, A., J. van der Lans, T.F.M. ter Bogt (1999). *Het heft in eigen handen*. Over zelfsturing en sociale betrokkenheid van jongeren. Assen: Van Gorcum.
- Diepstraten, I., P. Ester & H. Vinken (1999a). *Mijn generatie*. Zelfbeelden, jeugdervaringen en lotgevallen van generaties in de twintigste eeuw. Tilburg: Syntax Publishers (2e druk).
- Diepstraten, I., P. Ester & H. Vinken (1999b). *Een vooruitziende blik*. Verwachtingen van Nederlanders voor de ééntenwintigste eeuw en hun herinneringen aan de twintigste eeuw. De CentER-data Millennium Enquête. Tilburg: Globus/Katholieke Universiteit Brabant.
- Elchardus, M. (red.) (1999). *Zonder maskers. Een actueel portret van jongeren en hun leraren*. Gent: Globe.
- Essau, C. & G. Trommsdorff (1993). Kontrollorientierung von Jugendlichen im Kulturvergleich. *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie* 13: 311-325.
- Ester, P. & L. Halman (red.) (1994). *De cultuur van de verzorgingsstaat*. Een sociologisch onderzoek naar waardenoriëntaties in Nederland. Tilburg: Tilburg University Press.
- Ester, P., L. Halman & R. de Moor (red.) (1993). *The individualizing society*. Tilburg: Tilburg University Press.
- Ester, P. & H. Vinken (2000). *Van later zorg*. Verwachtingen van Nederlanders over arbeid, zorg en vrijetijd in de 21ste eeuw. Het OSA TOEKOMST VAN DE ARBEID SURVEY. Tilburg: Organisatie voor Strategisch Arbeidsmarktonderzoek.
- Ester, P. & H. Vinken (2001). *Een dubbel vooruitzicht*. Doembeelden en droombeelden van arbeid, zorg en vrije tijd in de 21e eeuw. Bussum: Uitgeverij Coutinho.
- Fischer, A., et al. (2000). *Jugend 2000*. 2 Bände. Leverkusen: Leske + Budrich.
- Flammer, A. (1988). *Entwicklungstheorien: psychologische Theorien der menschlichen Entwicklung*. Bern: Huber.
- Fuchs, W. (1983). Jugendliche Statuspassage oder individualisierte Jugendbiographie? *Soziale Welt* 34: 341-371.
- Fuchs-Heinritz, W. (2000). Lebensentwürfe: Eindrücke aus dem qualitativen Material. In A. Fischer et al., *Jugend 2000*. 2 Bände. Leverkusen: Leske + Budrich (Zweiter Band: 394-395).
- Heelas, P., S. Lash & P. Morris (1996). *Detraditionalization*. Critical reflections on authority and identity. Cambridge, MA (etc.): Blackwell Publishers.
- Heemskerk, J.A.R., et al. (1999). *Vernieuwen & investeren*. Den Haag: Ministerie van Volksgezondheid, Welzijn en Sport, Directie Jeugd beleid, CJO.
- Hofstede, G. (2001). *Culture's Consequences*. Comparing values, behaviors, institutions, organisations across nations. Second Edition. Thousand Oaks, CA (etc.): Sage Publications.
- Israelashvili, M. (1997). School adjustment, school membership and adolescents future expectations. *Journal of Adolescence*, 20, 5: 525-535.
- Kasurinen, H. (1997). Adolescents' ideas about their future in a changing society. In H. Helve & J. Bynner (eds.), *Youth and life management: research perspectives*. Helsinki: Helsinki University Press.
- Lamm, H., R. W. Schmidt & G. Trommsdorff (1976). Sex and social class as determinants of future orientation (time perspective) in adolescents. *Journal of Personality and Social Psychology* 34, 3: 317-326.
- Lamm, H. & G. Trommsdorff (1974). Group influences on probability judgments concerning social and political change. *Psychological Reports* 35: 987-996.
- Malmberg, L. E. & S. Norrgård (1999). Adolescents' ideas of normative life-span development

- and personal goals. *Journal of Adolescence* 22, 1: 33-47.
- Malmberg, L. E. & J. Trempele (1997). Anticipated transition to adulthood: the effect of educational track, gender, and self-evaluation on Finnish and Polish adolescents' future orientation. *Journal of Youth and Adolescence* 26, 5: 517-537.
- Nurmi, J. E. (1992). Age differences in adult life goals, concerns and their temporal extension: a life course approach to future-oriented motivation. *International Journal of Behavioral Development* 15, 4: 487-508.
- Nurmi, J. E. (1994). The development of future orientation in life-span context. In Z. Zaleski (ed.), *Psychology of future orientation*. Lublin, Poland: Towarzystwo Naukowe KUL.
- Pulkkinen, L. & A. Rönkä (1992). Personal control over development, identity formation, and future orientation as components of life orientation: a developmental approach. University of Jyväskylä, Finland.
- Schmidt, R.W., H. Lamm & G. Trommsdorff (1978). Social class and sex as determinants of future orientation (time perspective) in adolescents. *European Journal of Social Psychology*, 8: 71-90.
- Seginer, R. & H. Halabi-Kheir (1998). Adolescent passage to adulthood: Future orientation in the context of culture, age, and gender. *International Journal of Intercultural Relations* 22, 3: 309-328.
- Stevens, F. & M. Elchardus (2001). *Speelplaats als cultureel centrum*. De beleving van de leefwereld van jongeren (conceptversie). Brussel: Onderzoeksgroep TOR.
- Swaan, A. de (1982). Uitgaansbeperking en uitgaansangst: over de verschuiving van de bevelshuishouding naar onderhandelingshuishouding. In A. de Swaan, *De mens is de mens een zorg*. Amsterdam: Meulenhoff.
- Trommsdorff, G. (1994). Future time perspective and control orientation: social conditions and consequences. In Z. Zaleski (ed.), *Psychology of future orientation*. Lublin, Poland: TN KUL.
- Trommsdorff, G. (2000). Subjective experience of social change in individual development. In J. Bynner & R. K. Silbereisen (eds.), *Adversity and challenge in life in the new Germany and England*. London: MacMillan Press.
- Trommsdorff, G., H. Lamm & R. W. Schmidt (1979). A longitudinal study of adolescents' future orientation (time perspective). *Journal of Youth and Adolescence*, 8, 2: 131-147.
- Vinken, H. (1997). *Political values and youth centrism*. Theoretical and empirical perspectives on the political value distinctiveness of Dutch youth centrists. Tilburg: Tilburg University Press.
- Vinken, H., P. Ester & C. de Bont (1997). Westerse jongeren als culturele seismografen: een trendstudie. In P. Ester, J. Geurts & M. Vermeulen (red.), *De makers van de toekomst*. Over nut en noodzaak van toekomstverkenningen voor beleidsonderzoek. Tilburg: Tilburg University Press.
- Vinken, H., P. Ester, H. Dekkers & L. van Dun (2002). *Aan ons de toekomst*. Toekomstverwachtingen van jongeren in Nederland. Assen: Van Gorcum.
- ZSE (2002). *Zeitschrift für Soziologie der Erziehung und Sozialisation*. Schwerpunkt: Selbstsozialisation. Weinheim: Juventa Verlag (22. Jg., 2. Vj. 2002).

10. Waarden in Nederland

Segmentatie van doelgroepen

E. HESSING en A.M.H. REULING

SAMENVATTING

Oprachtgevers voor marktonderzoek – zowel uit het commerciële als het overheidsdomein – hebben steeds (meer) behoefte aan een goede segmentatie van hun markt. Echter, klassieke segmentaties, gebaseerd op sociodemografische en –economische achtergrondkenmerken, voldoen nauwelijks meer omdat het verband tussen zulke kenmerken en de centrale prikkel (de reden om te segmenteren) is verdwenen.

Het hier gepresenteerde WIN-modelTM is een nieuwe, op waardenhiërarchieën gebaseerde segmentatie die met werkelijk zeer veel variabelen een relatie onderhoudt en daarmee dus veel verschillende opdrachtgevers een goed inzicht kan verschaffen in (de segmentering van) hun markten en de doelgroepen die voor hen interessant zijn.

1. INLEIDING

Lange tijd waren socio-demografische kenmerken voor marktonderzoekers dé variabelen bij uitstek om segmentaties op te baseren: bepaalde producten, diensten of beleidvoornemens spreken jongere, hoogopgeleide mannen meer aan dan oudere, laagopgeleide vrouwen; in de noordelijke regio is men gemiddeld zuiniger met energie dan in de Randstad, etc, etc. Tegenwoordig komen dergelijke analyses onvoldoende tegemoet aan de behoefte van opdrachtgevers.

Terwijl de huidige marktwerking vraagt om een steeds beter inzicht in specifieke doelgroepen, blijken mensen steeds minder gemakkelijk in eenvoudige hokjes onder te brengen. Gedrag, ook consumptiegedrag, vloeit voort uit een diversiteit aan opvattingen, leefstijlen, behoeften en mogelijkheden, die dwars door alle maatschappelijke geleidingen vorm krijgt. Er is bijvoorbeeld allang geen sprake meer van één jeugdcultuur of van één consumptiestijl die past bij alle mensen ouder dan 50 jaar. De voortgaande individualisering van de Westerse mens nodigt uit tot eclecticisme en mensen kiezen dan ook uit allerlei stijlen en tradities wat hen het beste bevalt. Segmenteren is daardoor een delicate aangelegenheid geworden, en dat terwijl de behoefte aan een goede segmentatie eerder groter dan kleiner geworden is.

In dit artikel presenteren wij NIPO's WIN-modelTM als een nieuwe, krachtige segmentatie, gebaseerd op *Waarden In Nederland*.

2. HET WIN-MODEL™

2.1. Segmenteren

Een goede segmentatie voldoet volgens Gankema en Wedel (1992) aan zeven criteria:

- *identificeerbaarheid* (het moet duidelijk zijn tot welk segment iemand behoort);
- *toegankelijkheid* (de groepen moeten bereikbaar zijn);
- *voldoende omvang* (er mogen geen te kleine groepjes overblijven),
- *heterogeniteit* (de segmenten moeten duidelijk van elkaar verschillen),
- *stabiliteit* (de segmenten mogen niet voortdurend veranderen),
- *homogene respons* (segmentleden moeten min of meer hetzelfde reageren op centrale prikkels zoals reclame, voorlichting, e.d.); en
- *bewerkbaarheid* (er moeten aanknopingspunten zijn voor (marketing-)strategie).

Het segmenteren zelf gebeurt op basis van een beperkt aantal kenmerken, de segmentatiebasis. Dergelijke bases kunnen algemeen of (product)specifiek zijn en direct waarneembaar of afgeleid zijn. Als direct waarneembare, algemene bases kunnen dienen: culturele variabelen (nationaliteit, religie, etc.), geografische (regio, urbanisatiegraad), demografische (leeftijd, leefvorm) en/of socio-economische kenmerken (inkomen, opleiding, welstand). Zulke segmenten zijn doorgaans goed identificeerbaar, toegankelijk en stabiel. Het probleem is echter dat de relatie met de afhankelijke variabele (het te beïnvloeden gedrag, de afname van diensten, het consumptiegedrag, e.d.) meestal zwak is. Anders gezegd, de respons is niet homogeen (genoeg).

Worden de segmenten bepaald met behulp van psychografische variabelen zoals persoonlijkheidskenmerken, leefstijlen e.d. dan spreken we van afgeleide, algemene bases. Het WIN-model™ is zo'n algemene, niet aan specifieke omstandigheden, diensten of producten gerelateerde segmentatie van de Nederlandse bevolking, die ontwikkeld werd op basis van waarden-hiërarchieën.

Waarden spelen zo'n centrale rol in de cognitieve en emotionele structuur van mensen, dat zij belangrijke determinanten vormen van houdingen, leefstijl, leefdoelen en (consumenten)gedrag. Hoewel de relatie tussen waarden en gedrag vaak niet direct zichtbaar is, blijkt een (gedeeld) waardepatroon wel een goede voorspeller voor (overeenkomsten in) gedrag. Bijkomend voordeel – zeker vanuit de optiek van de markt-onderzoeker – is dat waarden stabiel zijn over langere perioden.

2.2. De rol van waarden

In de literatuur over waarden wordt meestal uitgegaan van de vijf kenmerken van Schwartz en Bilsky (1987). Volgens hen zijn waarden:

- a) opvattingen,
- b) die betrekking hebben op gedrag of op een wenselijke situatie;
- c) die zich niet beperken tot een of enkele gedragsdomein(en);
- d) die richting geven aan gedrag of de evaluatie van gedrag;
- e) en die zich te rangordenen naar belangrijkheid.

Waarden hebben betrekking op wat iemand voor zichzelf belangrijk vindt in het leven. Dat is niet voor iedereen gelijk. De een zal bepaalde waarden belangrijker vinden dan de ander, hetgeen impliceert dat niet iedereen dezelfde waardenhiërarchie hanteert én dat mensen te typeren zijn aan de hand van hun waardenhiërarchie.

Voor het WIN-model™ is gebruik gemaakt van de zogenaamde Rokeach Value Survey

(Rokeach 1973). Deze zijn vaak gebruikt in onderzoek (Schwartz en Bilsky 1987, 1990, Kamakura en Mazon 1991), soms in een enigszins gewijzigde vorm, maar veelal zonder ter discussie te stellen of deze waarden nog wel gelden in een tijd, waarin zich multiculturele samenlevingen ontwikkelen temidden van een steeds verder globaliserende economie.

Oppenhuisen (2000) nam dit vraagstuk serieus onder de loep en inventariseerde de huidige waarden van Nederlanders. In 20 kwalitatieve diepte-interviews werden met behulp van fotocollages – die fasen uit het leven verbeeldde – in eerste instantie 1372 waarden verzameld. Deze zijn op grond van strenge criteria samengevoegd tot 160 fundamentele waarden en voorgelegd aan een panel van 2000 respondenten. Met behulp van Principale Componenten Analyse en Clusteranalyses komt Oppenhuisen uiteindelijk uit op zes factoren, die alle een (verschillend) element van ‘vrijheid’ tegenover ‘binding’ in zich hebben:

Factor 1: relaties – prestatie;

Factor 2: sociaal georiënteerd – individueel georiënteerd;

Factor 3: waarden van oudsher – waarden van nu;

Factor 4: zekerheid – uitdaging;

Factor 5: gezinsleven – vrijheid;

Factor 6: conformistisch zijn – je eigen weg gaan.

Wij zullen echter laten zien dat de twee dimensies, die ten grondslag liggen aan het WIN-model™ een waardenruimte definiëren, waarin de dimensies van Oppenhuisen gemakkelijk zijn terug te vinden. M.a.w., Rokeach’ waarden mogen dan ‘gedateerd’ lijken, er valt nog steeds uitstekend mee te werken. Of zoals Schwartz en Bilsky betoogden: deze waarden zijn universeel.

2.3. Instrumentele waarden en eindwaarden

De Rokeach-waarden worden verdeeld in twee waardenreeksen, de zogenaamde ‘eindwaarden’ en ‘instrumentele waarden’. Instrumentele waarden hebben betrekking op gewenst gedrag en uitstraling, zoals, ‘ruimdenkend’, ‘eerlijk’ ‘betrouwbaar’, terwijl eindwaarden verwijzen naar wat men wil bereiken in het leven (‘geluk’, ‘comfort’, ‘liefde’). Voor het bereiken van de wenselijke manier van leven is de manier van optreden instrumenteel.

Respondenten van NIPO panels¹ krijgen (thuis op hun beeldscherm) als eerste de 18 eindwaarden te zien met de opdracht ze alle 18 te ordenen naar belangrijkheid voor hun eigen leven. Daarna herhaalt de procedure zich met de 18 instrumentele waarden. De belangrijkste waarden krijgen zo de laagste score, de minst belangrijke de hoogste.

‘Geluk’, ‘zekerheid voor het gezin’ en ‘vrede’ zijn de (eind)waarden die voor de Nederlandse burger anno 2002 in het algemeen bovenaan staan. ‘Vergeving’, ‘prestatie’ en ‘een stimulerend leven’ komen onder aan de ladder. Van de reeks instrumentele waarden waardeert men ‘eerlijk’, ‘betrouwbaar’ en ‘behulpzaam’ het meest, terwijl men ‘ambitieuw’, ‘intellectueel’ of ‘logisch’ het minst belangrijk vindt (tabel 1. Zie verder Hessing en Reuling, 2002).

2.4. De waardenruimte

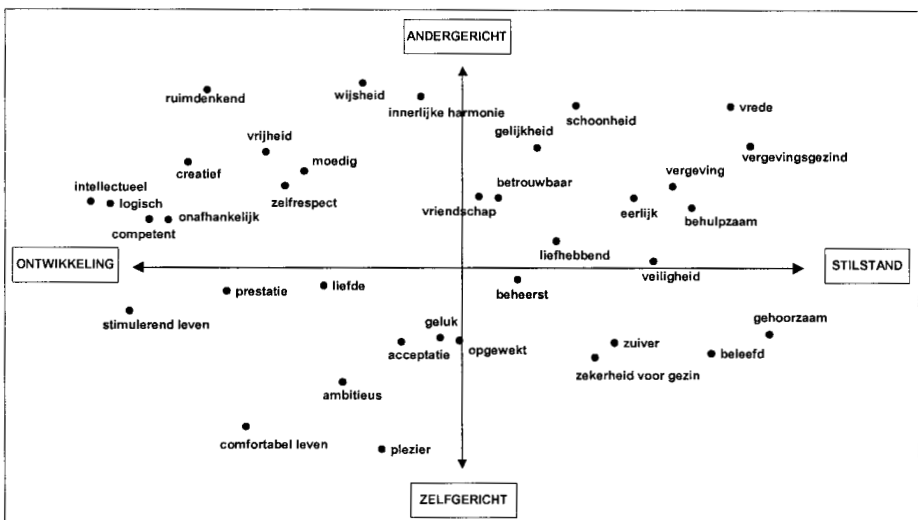
Met behulp van een niet-lineaire, non-metrische, Principale Componenten Analyse is de positie van de waarden in een 2-dimensionele ruimte vastgelegd (Figuur 1).

Tabel 1. Gemiddelde rangscores van de waarden in de totale steekproef (n=1500).

Eindwaarden	Gemiddelde rangscore	Instrumentele waarden	Gemiddelde rangscore
Geluk	4,9	Eerlijk	3,8
Zekerheid voor gezin	5,2	Betrouwbaar	4,6
Wereld in vrede	7,0	Behulpzaam	7,8
Zelfrespect	7,1	Opgewekt	7,9
Vrijheid	7,4	Liefhebbend	8,0
Vriendschap	7,5	Onafhankelijk	8,8
Innerlijke harmonie	8,5	Ruimdenkend	8,8
Plezier	8,6	Beleefd	8,9
Gelijkheid	9,3	Vergevingsgezind	9,1
Wijsheid	9,9	Beheerst	9,3
Acceptatie	10,7	Moedig	10,4
Volwassen liefde	10,8	Creatief	11,2
Schoonheid	11,1	Zuiver	11,6
Nationale veiligheid	11,1	Gehoorzaam	11,8
Comfortabel leven	11,8	Competent	11,8
Stimulerend leven	13,0	Ambitius	12,2
Gevoel van prestatie	13,0	Intellectueel	12,3
Vergeving	14,3	Logisch	12,8

Bron: NIPO, 2002

De assen van de ruimte zijn te interpreteren als: ‘ontwikkeling – stilstand’ (horizontaal) en ‘zelfgericht – op anderen gericht’ (verticaal). De positie van de respondenten in de waardenruimte werd bepaald door hun objectscores als input voor een clusteranalyse te gebruiken. Na enig experimenteren bleek de 8-clusteroplossing het beste resultaat te geven.



Bron: NIPO, 2002

Figuur 1. De Nederlandse waardenruimte.

Een vergelijking van de positie die elke waarde binnen de segmenten inneemt (Kruskal-Wallis toets), toont aan dat alleen de positie van ‘vriendschap’ niet significant verschilt, terwijl ‘betrouwbaar’ slechts zeer gering differentieert. Alle andere waarden hebben duidelijk verschillende posities in de waardenhiërarchie. De rangordening van de waarden per cluster maakt zichtbaar welke waarden binnen een cluster belangrijk en welke minder belangrijk worden gevonden (tabel 2).

Tabel 2. Gemidd. rangscore per cluster (lichtgrijs: belangrijk; donkergrijs: onbelangrijk).

Clusternummer	1	2	3	4	5	6	7	8
<i>Eindwaarden</i>								
Acceptatie	11,7	10,3	9,7	10,7	9,3	10,0	11,9	12,2
Comfortabel leven	14,9	11,2	8,2	11,9	6,6	9,1	13,8	14,6
Gelijkheid	7,4	8,9	11,1	9,0	12,7	10,7	8,8	8,1
Geluk	5,8	4,0	3,9	4,7	3,5	5,5	6,4	6,1
Innerlijke harmonie	7,5	9,5	11,8	8,3	11,0	8,5	5,1	5,9
Liefde	13,1	10,6	9,0	10,8	8,7	9,7	11,8	12,0
Nationale veiligheid	8,6	9,9	10,5	11,9	13,2	13,8	13,4	10,3
Plezier	11,9	6,8	4,6	8,2	5,5	7,6	12,2	12,3
Prestatie	14,7	13,9	12,2	13,4	11,0	10,4	11,8	13,9
Schoonheid	8,6	11,0	13,6	11,4	14,0	12,8	9,1	9,0
Stimulerend leven	15,3	14,0	10,8	13,0	10,0	10,4	12,0	14,7
Vergeving	10,6	14,9	16,4	14,7	15,9	16,2	15,4	12,4
Vrede	3,8	6,4	9,6	6,5	11,8	10,4	6,7	4,2
Vriendschap	7,3	7,8	7,7	7,7	7,0	7,3	7,7	7,3
Vrijheid	8,6	8,2	8,8	7,2	7,1	5,6	5,1	7,0
Wijsheid	9,3	12,4	12,7	9,6	11,5	8,6	5,9	7,0
Zekerheid voor gezin	3,8	2,8	3,4	5,5	5,4	8,6	8,6	6,9
Zelfrespect	8,0	8,5	7,1	6,6	6,9	5,7	5,4	7,1
<i>Instrumentele waarden</i>								
Ambitieuus	13,3	12,3	9,7	12,9	9,3	11,2	13,7	14,2
Beheerst	8,1	9,1	9,7	9,1	9,9	10,5	10,3	9,4
Behulpzaam	5,3	6,8	8,4	8,3	9,9	10,7	9,4	5,9
Beleefd	7,0	6,4	7,2	9,0	10,5	11,7	13,8	10,2
Betrouwbaar	4,1	4,3	4,2	4,7	5,0	5,7	4,8	4,4
Competent	13,9	13,7	12,8	11,4	10,2	9,0	9,2	10,9
Creatief	13,1	12,7	11,8	11,4	10,5	8,3	7,2	10,8
Eerlijk	2,6	2,8	3,9	3,5	4,9	6,3	4,7	3,7
Gehoorzaam	9,5	9,3	10,1	12,5	13,1	15,2	16,3	12,6
Intellectueel	14,3	14,9	13,8	11,7	10,6	8,5	8,6	11,4
Liefhebbend	7,6	6,6	7,1	8,0	8,2	10,2	9,5	8,6
Logisch	14,4	14,7	13,8	12,5	11,3	9,4	9,8	12,6
Moedig	11,5	11,7	10,8	10,4	9,4	8,3	9,2	9,6
Onafhankelijk	11,3	9,8	8,7	9,0	6,9	5,1	6,4	9,3
Opgewekt	8,5	6,9	6,0	7,6	6,7	8,7	10,5	9,9
Ruimdenkend	10,5	11,3	11,0	8,6	8,6	6,0	3,6	6,5
Vergevingsgezind	5,6	8,3	11,8	8,9	13,6	12,2	9,0	7,6
Zuiver	10,4	9,7	10,2	11,6	12,4	14,2	14,9	13,2

Uit tabel 2 kunnen we bijvoorbeeld lezen dat segment 5 ‘comfortabel leven’ het belangrijkste vindt van alle clusters (laagste gemiddelde rangnummer 6.6) en dat segment 1 dat juist niet vindt (gemiddeld 14.9).

Door nu uit tabel 3 zowel de belangrijkste als minst belangrijke waarden per cluster te selecteren, wordt duidelijk hoe de segmenten van elkaar verschillen (tabel 3).

Tabel 3. Verschillen tussen clusters.

Cluster	Relatief belangrijk	Relatief onbelangrijk
1	gelijkheid, veiligheid, schoonheid, vergeving, vrede, beheerst, behulpzaam, eerlijk, vergevingsgezind	comfortabel leven, liefde, prestatie, stimulerend leven, competent, creatief, onafhankelijk
2	zekerheid voor gezin, beleefd, gehoorzaam, liefhebbend, zuiver	zelfrespect, intellectueel, logisch, moedig, ruimdenkend
3	plezier, opgewekt	innerlijke harmonie, vergeving, vrijheid, wijsheid
4	scoort als middengroep, geen opvallende waarden	scoort als middengroep, geen opvallende waarden
5	geaccepteerd worden, comfortabel leven, geluk, liefde, stimulerend leven, ambitieus	gelijkheid, schoonheid, vrede, vergevingsgezind
6	prestatie, competent, intellectueel, logisch, moedig, onafhankelijk	veiligheid, zekerheid voor gezin, beheerst, behulpzaam, eerlijk, liefhebbend
7	innerlijke harmonie, vrijheid, wijsheid, zelfrespect, creatief, ruimdenkend	Geluk, zekerheid voor gezin, beleefd, gehoorzaam, opgewekt, zuiver
8	nooit als eerste. Als goede 2-e: gelijkheid, innerlijke harmonie, schoonheid	geaccepteerd worden, plezier, ambitie

Bron: NIPO, 2002

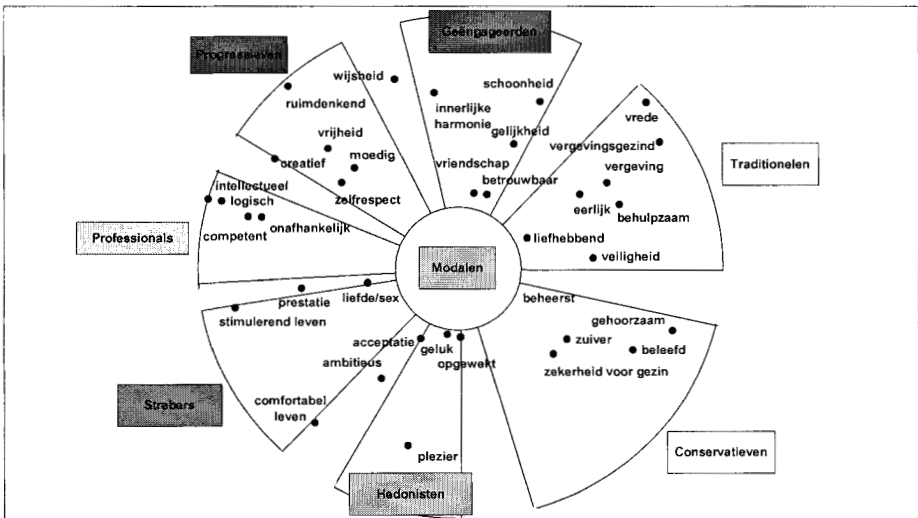
3. DE WAARDENORIENTATIES VAN DE SEGMENTEN

De gevonden clusters blijken alle een geheel eigen waardenoriëntatie te hebben. Ze vormen 8 segmenten van de Nederlandse samenleving, waarvoor we (vooruitlopend op hierna volgende bevindingen) de volgende namen hebben gekozen:

1: Traditionelen, 2: Conservatieven, 3: Hedonisten, 4: Modalen, 5: Ambitieuze materialisten, 6: Professionals, 7: Progressieven en 8: Geëngageerden.

De 8 clusters omvatten respectievelijk 15, 16, 12, 19, 7, 9, 9 en 13 procent van de Nederlandse bevolking.

In de waardenruimte zijn de segmenten (ongeveer) geplaatst in figuur 2.



Bron: NIPO, 2002

Figuur 2. Het waardeptraan van de segmenten.

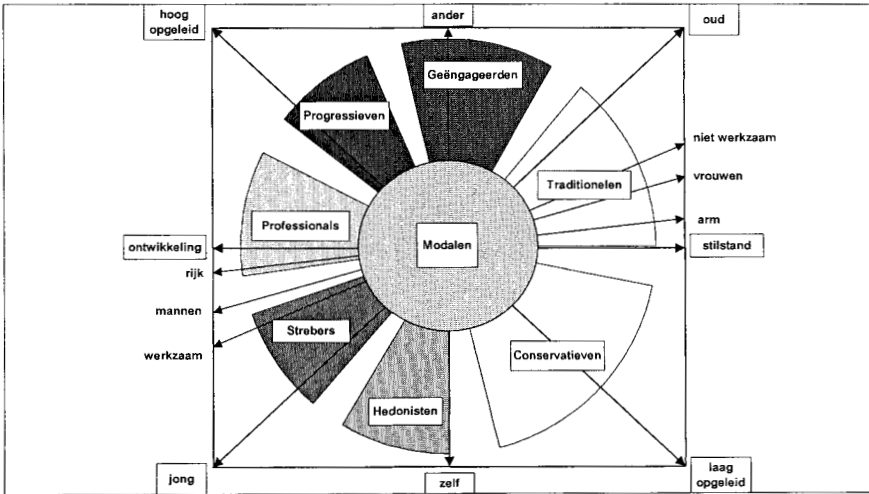
3.1. Verwantschap en tegenstelling in de waardenruimte

De horizontale en verticale as vertegenwoordigen 2 dimensies die ten grondslag liggen aan de waarden van Rokeach. De respondenten groeperen zich als wolken rond de waarden die zij het belangrijkste vinden. Opmerkelijk is dat de segmenten niet alleen qua waarden van elkaar verschillen, maar ook op achtergrondkenmerken als leeftijd, geslacht, opleiding, al dan niet werkzaam zijn. Bovendien blijkt de volgorde van bijvoorbeeld het jongste naar het oudste segment (zie tabel 4 van boven naar beneden) grafisch weer te geven als een lijn door het model. Deze loopt van de Ambitieuze materialisten naar de Traditionelen (van linksonder naar rechtsboven).

Tabel 4. Gemiddelde leeftijd en opleiding van de segmenten.

	Gemiddelde leeftijd		% met havo- opleiding of hoger
Ambitieuze materialisten	38	Conservatieven	15.3
Hedonisten	41	Traditionelen	15.4
Professionals	42	Hedonisten	16.2
Modalen	44	Modalen	27.1
Conservatieven	45	Ambitieuze materialisten	36.0
Progressieven	50	Geëngageerden	45.8
Geëngageerden	52	Professionals	64.7
Traditionelen	53	Progressieven	73.2

Hierna zijn enkele van deze achtergrondlijnen in figuur 3 aangegeven.

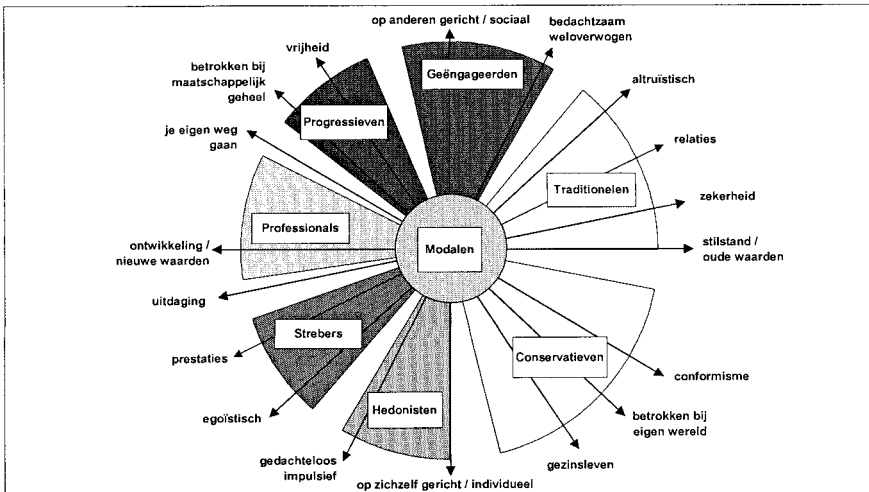


Bron: NIPO, 2002

Figuur 3. De segmenten en enkele achtergrondkenmerken.

Leg in gedachten een lijn haaks op de leeftijdslijn, beginnend linksonder. De leeftijdslijn volgend, passeert de haakse lijn de Ambitieuze materialisten, de Hedonisten en de Professionals, om daarna via de Modalen naar de Conservatieven, Progressieven en Geëngageerden te gaan en bij de Traditionelen te eindigen. De volgorde van ‘passeren’ correspondeert met de volgorde van de gemiddelde leeftijd. De lijn toont derhalve de relatie tussen leeftijd en de segmenten.

De kracht van het WIN-model™ schuilt daarin dat zulke relaties voor werkelijk zeer veel variabelen gevonden werden. Enkele voorbeelden zijn weergegeven in figuur 4.



Bron: NIPO, 2002

Figuur 4. Enkele schalen in het model, waaronder de dimensies van Oppenhuizen.

In de figuur zijn ook zes dimensies weergegeven die vergelijkbaar zijn met de dimensies die Openhuisen vond².

De implicatie van dit lijnenfenomeen is dat segmenten die dicht bij elkaar liggen, zoals Ambitieuze materialisten en Hedonisten, een zekere verwantschap met elkaar vertonen, terwijl tegenover elkaar liggende segmenten, zoals die van Conservatieven en Progressieven, op veel punten in hoge mate tegengesteld zullen zijn.

3.2. Betekenis van de positie in de waardenruimte.

De verwantschappen en tegenstellingen tussen de acht segmenten – met de Modalen als neutrale middengroep – vloeien voort uit de positie die de waarden innemen in de waardenruimte. Schwartz en Bilsky hebben de clustering van waarden in de tweedimensionale waardenruimte opgevat als waardendomeinen. De positie van de domeinen ten opzichte van elkaar zegt volgens hen iets over de ‘psychologische afstand’ tussen de waarden. Domeinen die aan elkaar grenzen en daarmee elkaars waarden enigszins delen, vertonen meer overeenkomst met elkaar dan domeinen die verder van elkaar af of tegenover elkaar liggen. De waarden die het ene segment belangrijk vindt hebben dus een tegengestelde lading (en doel) ten opzichte van de waarden die het tegenover liggende segment belangrijk vindt.

De overgang van het ene waardepatroon naar het andere (of van het ene segment naar het andere) is een geleidelijke. De geleidelijkheid loopt van het ene uiterste (bijvoorbeeld altruïsme) naar het andere uiterste (egoïsme). Zo zijn Hedonisten het meest impulsief van alle segmenten en Geëngageerden juist het meest bedachtzaam. Na de Hedonisten zijn Ambitieuze materialisten het meest impulsief, daarna Conservatieven en Professionelen. In het midden, bij de Modalen, klapt de as om naar bedachtzaamheid: Traditionelen zijn wat meer bedachtzaam, Progressieven nog wat meer en Geëngageerden tenslotte het meest. Zij zijn dus ook het minst impulsief van allen (zie figuur 4).

De lijnen lopen overigens niet altijd precies recht. Progressieven bijvoorbeeld zijn vooruitstrevend en op anderen gericht. Zij vormen verreweg het meest ‘linkse’ segment en stemmen vooral op politieke partijen als Groen Links, SP en D66. Professionals zijn ook vooruitstrevend, zij het duidelijk minder sociaal en meer zelfgericht. Wat zij met Progressieven delen – en waarom ze in de waardenruimte naast elkaar liggen – is hun gerichtheid op ontwikkeling en vooruitgang. Professionals plaatsen zichzelf op een links-rechtsschaal echter juist het meest rechts van alle segmenten en stemmen vooral op de VVD. Zouden we een politieke links-rechts lijn willen trekken, dan zal deze bepaald niet recht blijken te zijn.

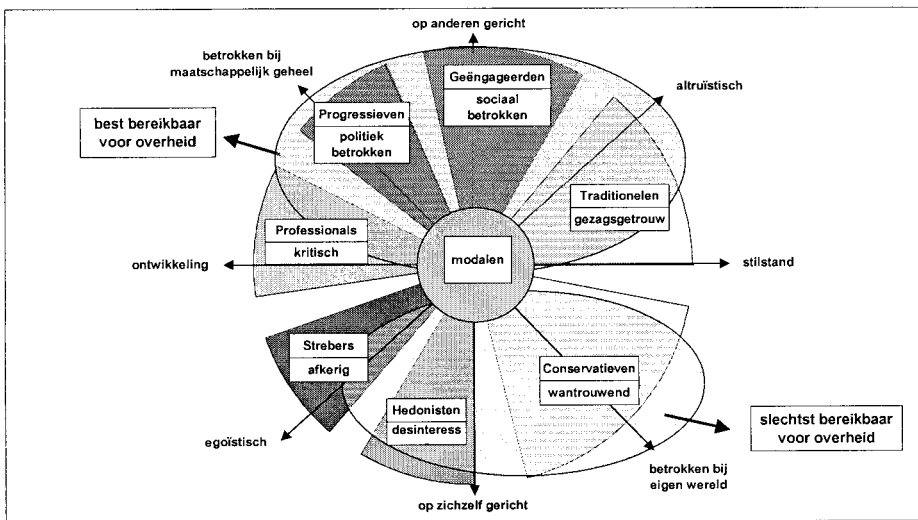
4. TOEPASSINGEN VAN HET WIN-MODEL™

De toepassingsmogelijkheden van het WIN-model™ zijn legio. Groot voordeel is dat a) de segmentatiebasis nauwelijks uitbreiding vergt van de metingen en dat b) de segmentatie is samengebond in één variabele die net als alle andere achtergrondvariabelen gebruikt kan worden om snel inzicht te krijgen in de (markt)data. Zo blijkt een Audi TT het meest gewaardeerd te worden door Professionals, spreekt een nieuw tijdschrift

over psychologie Progressieven en Geëngageerden meer aan dan Hedonisten en Conservatieven en moet men voor de verkoop van bijzondere *gadgets* bij Ambitieuze materialisten aankloppen. Traditionelen zullen zuinig zijn met energie, terwijl Ambitieuze materialisten veruit de hoogste energierekening hebben. Conservatieven hebben moeite met de nieuwe belastingformulieren en Progressieven kunnen haarfijn aangeven hoe deze er beter uit hadden kunnen zien.

Bij het bepalen van specifieke doelgroepen biedt het WIN-model™ eveneens vele mogelijkheden. Van de segmenten zijn uitgebreide profielen beschreven, die met elk onderzoek verder zijn aan te vullen (zie Hessing en Reuling, 2002). Er is een scala aan achtergrondkenmerken, houdingen, wensen, kennis, opvattingen, gevoelens, af- en -voorkeuren, vele gedragingen zoals mediagedrag, politiek stemgedrag en milieugedrag, (consumptie)-patronen en vrijetijdsbestedingen van de segmenten bekend. Ook is bekend hoe segmenten het best te bereiken en aan te spreken zijn, met welke type boodschap, in welke stijl en via welke media. Het model kan dan ook uitstekend gebruikt worden voor allerlei onderzoek van de (semi)overheid: voorbereiding, ondersteuning en evaluatie van campagnes, beleidsevaluaties, het meten van draagvlak voor beleid, etc.

Daarnaast zijn er mogelijkheden om het WIN-model™ in de commerciële sfeer toe te passen. maar anders dan bij bijvoorbeeld het NIPO/TNS-model ‘NeedScope’, dat specifiek gericht is op het vaststellen van de behoeften in de markt en domeinspecifiek blootlegt welke merken en producten bij de zes onderscheiden behoeftensegmenten passen, zal het WIN-model™ eerder te gebruiken zijn als een algemene oriëntatie op doelgroepen.

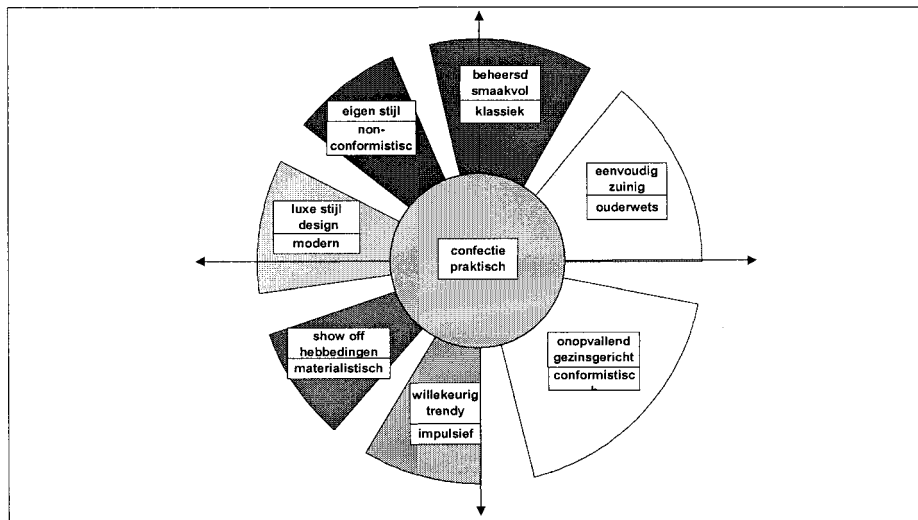


Bron: NIPO, 2002

Figuur 5. Benaderbaarheid van de segmenten door de overheid.

Figuur 6 geeft een indruk van de stijl en smaak van de segmenten. Door het panel domeinspecifieke vragen voor te leggen kunnen dergelijke algemene termen uiteraard nog veel verder uitgediept worden.

Mogelijk kan het model ook domeinspecifieke informatie opleveren en als een aanvulling op NeedScope worden gebruikt of zelfs als een minder prijzig alternatief. Hier is echter nog onvoldoende onderzoek naar verricht.



Bron: NIPO, 2002

Figuur 6. Stijlen (van bijvoorbeeld kleding) van de segmenten.

De segmentvariabele kan tevens effectief dienst doen als selectiecriteria bij het trekken van steekproeven uit de grote panels. Is er voor een onderzoek bijvoorbeeld behoefte aan een steekproef met mensen die geïnteresseerd zijn in een technologisch innovatief product, dan zullen deze bij uitstek voorkomen in de linker segmenten en niet in de rechter. Zoekt men een groep die graag leest, dan is deze te vinden in de bovenste segmenten en weinig in de onderste. Voor nieuwe trends onder jongeren kan men terecht bij de segmenten linksonder en kunnen de segmenten rechtsboven buiten beschouwing worden gelaten.

Het WIN-model™ is nog steeds in ontwikkeling en zal met elk nieuw onderzoek verder worden uitgebreid. De ervaringen tot nu toe hebben in elk geval aangetoond dat deze segmentatie in zeer veel gevallen tegemoet komt aan de behoefte van opdrachtgevers aan een krachtige, veelzeggende segmentatie.

NOTEN

1. Voor de ontwikkeling van het WIN-modelTM is gebruik gemaakt van de data uit de Milieu Gedrags Monitor, een longitudinaal panel-onderzoek in opdracht van het ministerie van VROM. Dit onderzoek is 10 jaar lang afgenomen in NIPO's Telepanel (ca 1500 huishoudens). Sinds kort wordt gebruik gemaakt van NIPO's CAPI@Home, een veel omvangrijker Internet-panel met ca 30.000 huishoudens.
2. Wij legden de afgelopen vier jaar onze panel-respondenten diverse gevalideerde schalen voor, zoals de Attitude-scale van Fishbein en Ajzen (1975), de Egoism-scale van Weigel et al. (1999), De Ways-to-Live-scale van Morris (1973), de Low-Self-Control-scale van Grasmick et al. (1993) en diverse schalen uit politicologische onderzoeken.

LITERATUUR

- Fishbein, Martin and Ajzen, Icek (1975). *Belief, Attitude, Intention and Behavior: An introduction to Theory and Research*. Reading MA: Addison-Wesley.
- Gankema, H.G.J. en Wedel, M. (1992). Marktsegmentatieonderzoek in theorie en praktijk. In: *Recente Ontwikkelingen in het Marktonderzoek, Jaarboek 1992-'93*, 207-222, De Vrieseborch-Haarlem.
- Grasmick, Harold, G., Tittle, Charles R., Bursik, Robert J., & Arneklev, Bruce J. (1993). Testing the Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime. *Journal of Research in Crime and Delinquency* 30, 5-29.
- Hessing-Couvret, Ellen en Reuling, Albert M.H. (2002). *Waarden In Nederland, een segmentatie van de Nederlandse bevolking*. Op: NIPO.nl via dienstenpakket bij producten.
- Morris, Charles (1956). *Varieties of Human Value*. Chicago: University of Chicago Press.
- Oppenhuisen, Joke (2000). *Een schaap in de bus? Een onderzoek naar waarden van de Nederlander*. Proefschrift, Universiteit van Amsterdam.
- Oppenhuisen, Joke en Sikkel, Dirk (2001). *Waarden van de Nederlander*. Artikel voor het congres 'Eetmaal van de Communicatiewetenschap' 22 en 23 februari 2001, Universiteit van Amsterdam.
- Rokeach, Milton (1973). *The Nature of Human Values*. New York, Free Press.
- Schwartz, Shalom H., Bilsky, Wolfgang (1987). Toward A Universal Psychological Structure of Human Values. *Journal of Personality and Social Psychology*, 53: 3, 550-562.
- Schwartz, Shalom H., Bilsky, Wolfgang (1990). Toward a Theory of the Universal Content and Structure of Values: Extensions and Cross-Cultural Replications. *Journal of Personality and Social Psychology*, 58: 5, 878-891.
- Weigel, Russell H., Hessing, Dick J. & Elffers, Henk (1999). Egoism: concept, measurement and implications for deviance. *Psychology, Crime and Law*, 5, 349-378.

11. *Korte termijn effecten van merknaamveranderingen*

J. AKKERMANS en K.J. ALSEM*

SAMENVATTING

Vaak wordt gesteld dat het veranderen van de naam van een merk het merk veel schade berokkent. In dit artikel wordt onderzocht in welke omstandigheden dat wel of niet waar is. De studie beschrijft de resultaten van een experimentele studie naar de korte termijn effecten van merknaamveranderingen op vier componenten van merkwaarde: merkbekendheid, merkassociaties, gepercipieerde kwaliteit en merktrouw. Tevens is onderzocht of er verschillen zijn in effecten bij intrinsiek (instrumenteel) en extrinsiek (abstract) gepositioneerde merken en wat de invloeden zijn van categorie-betrokkenheid, merkdifferentiatie en naamspceptie. Extrinsieke merken blijken gevoeliger voor merknaamveranderingen dan intrinsiek gepositioneerde merken. Merken met weinig onderscheid zijn ook gevoeliger voor naamsvveranderingen dan meer gedifferentieerde merken. De perceptie met betrekking tot de naamsvverandering (passendheid en grootte van de wijziging) is van invloed op de effecten van naamsvwijzigingen op merkwaarde. Al met al doen zich niet in alle situaties sterke dalingen in merkwaarde voor en dienen organisaties hun communicatie ten tijde van merknaamsvveranderingen zorgvuldig in te zetten.

1. MERKEN EN NAAMSVVERANDERINGEN

Het bouwen van sterke merken is een actueel thema in de marketingliteratuur. Sterke merken kennen een hoge merkbekendheid en een duidelijke positionering. Daarnaast genieten ze een uitgesproken voorkeur bij een groep consumenten en een hoge mate van trouw. Sterke merken zijn in staat om sterke posities in te nemen bij de detailhandel en zijn daarnaast beter bestand tegen bedreigingen uit de externe omgeving dan de minder sterke merken. Ook kunnen 'onder de paraplu' van sterke merken succesvolle productintroducties van nieuwe variëteiten of van nieuwe producten plaatsvinden. Tevens wordt het besef dat merken een bepaalde waarde vertegenwoordigen en als investeringen moeten worden opgevat steeds sterker. Kortom, het waarderen van merken wordt steeds belangrijker (Leefflang, 1996). Om een zo hoog mogelijke waardering voor een merk te creëren, dienen de merkenbouwers een uitgebreid en langdurig proces te doorlopen, waarbij het van belang is dat niet alleen de fysieke eigenschappen van het product min of meer constant zijn, maar waarbij het merk tevens een constant beeld naar buiten toe uitstraalt. Een vergelijking met menselijke relaties is hier op zijn plaats: als je iemand langdurig kent leer je zijn gewoontes en gedrag kennen en zul je uiteinde-

* De auteurs danken Prof. Dr. Janny Hoekstra die van dit artikel de redacteur was.

lijk een idee hebben over iemand zijn karakter. Past dat karakter bij je eigen preferenties, dan kan een relatie ontstaan. Gaat die persoon zich van de ene op de andere dag plotseling totaal anders gedragen, dan raak je verward en zal de relatie een knauw krijgen. Zo is het ook met merken. Aanbieders van merken dienen hun merken door consistente en geïntegreerde marketingcommunicatie maar ook door goede, betrouwbare prestaties, een karakter mee te geven dat bij de doelgroep past. Zo kan een relatie ontstaan.

Des te verwonderlijker is het dat steeds meer bedrijven ervoor kiezen om de merknaam te wijzigen. Het laten verdwijnen van een (vaak winstgevende) merknaam en deze vervangen door een nieuwe naam lijkt een recent fenomeen in marketingland. Zo heet Smith Chips tegenwoordig Lay's, Melkunie is inmiddels Campina, PTT Post wordt TPG Post, Libertel gaat op in Vodafone, Jif is Cif geworden, en in een verder verleden is Raider veranderd in Twix en heette Nissan Datsun.

Redenen voor naamswijzigingen zijn:

- *Internationalisering*. Voor internationaal opererende merken is het efficiënter om één merknaam te hebben. Dan kan bijvoorbeeld worden volstaan met één commercial in alle landen.
- *Fusies*. Als merken of bedrijven fuseren gebeurt het vaak dat een nieuwe naam wordt gekozen. Zo is uit de fusie van het Nieuwsblad van het Noorden, Groninger Dagblad en de Drentse Courant het Dagblad van het Noorden ontstaan.
- *Gewenste imagowijzigingen*. Als merken een naam hebben met een ongewenst imago, kan het simpeler zijn het imago te veranderen met een nieuwe naam dan met de bestaande naam. Zo heeft Foster Parents Plan na alle commotie over dubieus bestede middelen in 2001 haar naam veranderd in Plan International.

Overigens heeft de discussie over nieuwe namen bij corporate merken (namen van organisaties) ook een sterke interne component: het is gebruikelijk dat elke partij zoveel mogelijk van zichzelf wil terugzien. Dat leidt soms tot combinaties van namen (ABN AMRO) wat niet altijd gunstige gevolgen heeft voor de duidelijkheid van de positionering van de nieuwe organisatie. Soms ook worden nieuwe namen gekozen (Hoogovens plus British Steel wordt Corus), maar lukt het niet met communicatie de nieuwe naam een goede lading te geven, waardoor het wellicht beter was geweest een link naar de oude naam te behouden.

Aan de aanbodzijde zijn dus genoeg redenen te bedenken om namen te veranderen. Maar wat voor effecten heeft een merknaamverandering op de vraagzijde van de markt: de consumenten? Vinden consumenten het verwarrend als namen wijzigen? En zo ja, leidt de verandering tot andere associaties en een geringere merktrouw? Of valt het wel wat mee met deze effecten? Bij effecten van merknaamveranderingen moet een onderscheid gemaakt worden in:

- Korte termijn effecten: effecten gedurende een periode van bijvoorbeeld enkele weken of maanden (afhankelijk van het soort product);
- Lange termijn effecten: effecten die na de overgangperiode nog blijven bestaan. Voor merken zullen de lange termijn effecten uiteindelijk het belangrijkste zijn.

In de literatuur wordt aan geen van beide effecten veel aandacht besteed. Wel wordt

vaak opgemerkt dat veranderingen in merknamen gevaarlijk en kostbaar zijn (Keller, 1998; Franzen, Bouwman, 2001), maar empirisch onderzoek naar effecten van merknaamveranderingen ontbreekt vrijwel geheel. Dit in tegenstelling tot bijvoorbeeld effecten van het introduceren van merkextensions op factoren als merkimage en merktrouw (Aaker, Keller, 1990; Reddy, Holak, Bhat, 1994; Lane, 200; Bottomley, Holden, 2001). Voorzover er al van enig onderzoek sprake is, betreft het korte termijn effecten. Studies naar lange termijneffecten van naamswijzigingen zijn ons onbekend. Het uitvoeren van dergelijke studies is dan ook bijzonder lastig: ten eerste vereist dit kwantitatieve, consistente metingen gedurende een lange periode, en ten tweede is het zelfs dan nog lastig het geïsoleerde effect van merknaamswijzigingen te onderzoeken. Ook in dit artikel beperken we ons tot het onderzoeken *van de korte termijn effecten op de waarde van een merk*. De opzet van de studie vertoont enige gelijkenis met de eerder genoemde studies naar effecten van merkextensions op percepties van consumenten. Ook bij die studies werden consumenten geconfronteerd met (hypothetische) situaties waarna met behulp van vragenlijsten inzicht werd verkregen in korte termijn effecten van extensions. Het nieuwe van deze studie is dat effecten van merknamen worden onderzocht en dat niet één, maar verschillende dimensies van merkwaarde worden geanalyseerd. Hiermee geven we gevolg aan de opmerkingen van Keller (2002) en Ambler (2000) die stellen dat onderzoek naar merkwaarde bij voorkeur niet beperkt moet blijven tot onderzoek van één kengetal, omdat de toestand waarin een merk verkeert nu eenmaal niet in één getal is uit te drukken. Zo bestudeert een piloot in een vliegtuig ook diverse meetinstrumenten en is zelfs de toestand van het meest besproken item ter wereld niet in één getal te vangen: het weer.

Het vervolg van dit artikel is als volgt opgebouwd. In paragraaf 2 beschrijven we het conceptuele model dat uitgangspunt vormt voor de analyses. Paragraaf 3 bevat de opzet van de empirische studie. In paragraaf 4 beschrijven we de resultaten waarna paragraaf 5 is gewijd aan conclusies en aanbevelingen.

2. CONCEPTUEEL MODEL

Eerst dient het begrip merkwaarde van een meetbare inhoud te worden voorzien, vervolgens geven we aan welke beïnvloedende factoren we in de studie meenemen.

2.1. Merkwaardebegrippen

De laatste jaren is er een toenemende aandacht in de marketing ontstaan voor merkwaardering. Door het meten van de merkwaarde bestaat er voor ondernemingen de mogelijkheid en de wenselijkheid om de consequenties van het merkbeleid te evalueren, waaronder bijvoorbeeld de invloed van een merknaamverandering. In de marketingliteratuur wordt merkwaarde aangeduid met het begrip 'brand equity'. Dit begrip kent verschillende betekenissen. Feldwick (1996) onderscheidt drie verschillende beschrijvingen van het begrip 'brand equity'.

1. In eerste instantie kan het begrip 'brand equity' worden gebruikt als de financiële waarde van een merk. In dit geval wordt brand equity gezien als de totale waarde van een merk wanneer het wordt verkocht of wanneer het op de balans wordt geplaatst.

2. Ook vertaalt het begrip ‘brand equity’ zich als een maatstaf voor de mate waarin consumenten verbonden zijn met het merk. Om deze vorm van brand equity te concretiseren worden een viertal benaderingen onderscheiden:
 - a. Benadering op basis van prijsgevoeligheid. Sterke merken (merken met een hoge brand equity) kunnen een hogere prijs vragen en/of zijn minder prijsgevoelig dan concurrerende merken.
 - b. Benadering op basis van aankoopgedrag. Consumenten met een hoge ‘share of customer’ binnen een productcategorie worden beschouwd als emotioneel gebonden aan een merk en zijn minder snel geneigd zijn een substituut-product te accepteren.
 - c. Benadering op basis van attitude. Deze benadering geeft de voorkeur aan van een consument voor een merk; de verbintenis van een consument met een merk, zonder factoren als prijs en distributie hierin mee te nemen. Het is de affectieve benadering van een merk.
 - d. Benadering op basis van merkbekendheid. Wanneer een merk als eerste in het geheugen van de consument opkomt bij het noemen van de productcategorie, is het merk als het ware ‘eigenaar’ van deze categorie en heeft hiermee een groot voordeel ten opzichte van zijn concurrenten. Deze benadering vertelt echter weinig over de mate waarin een consument verbonden is met een merk.
3. Tenslotte kan het begrip ‘brand equity’ vertaald worden als beschrijving van een merk op basis van de associaties die een consument heeft met een merk. Merkimago kan ook wel in dit verband worden gebruikt. In tegenstelling tot de tweede beschrijving van brand equity, is dit de meer cognitieve benadering van een merk. Er is, volgens Feldwick, niet één operationele beschrijving van het begrip ‘brand equity’. Op verschillende momenten kunnen verschillende benaderingen zich doen gelden (Feldwick, 1996).

2.2. Merkwaardemodel en beïnvloedende factoren

De waarde van een merk kan dus gezien worden als een financiële maatstaf van een merk, als een maatstaf op basis van de waardering van de consument en, tenslotte, als een beschrijving van een merk op basis van associaties. Om de effecten van een merknaamverandering op de merkwaarde te meten, zal gebruik worden gemaakt van een combinatie van de tweede en derde beschrijving van Feldwick van het begrip ‘brand equity’. Alleen deze beschrijvingen geven de waarde die het merk vertegenwoordigt in het hoofd van de consument weer. De financiële benadering doet dit niet.

In het merkwaardemodel van Aaker (1991) worden de tweede en derde beschrijving van Feldwick gecombineerd. Dit model bestaat uit een vijftal elementen: merkbekendheid, merkassociaties, gepercipieerde kwaliteit, merktrouw en overige bronnen, zoals patenten. De laatstgenoemde component laat zich lastig meten en wordt daarom in deze studie niet meegenomen. De overige vier componenten zijn ook terug te vinden bij Keller (2002) en vormen feitelijk ook de basis van de in de praktijk veel gebruikte merkwaarde monitor van de Young en Rubicam Group (Keller, 2002, p. 509 e.v.). Op elk van de vier componenten gaan we nu in.

Merkbekendheid

Het aanleren van merknamen is erg belangrijk, gezien het feit dat merkwaarde in het hoofd van de consument gekoppeld dient te worden aan een bepaalde naam (McCrac-

ken & Macklin, 1998). De naamgeving voor een merk is dus erg belangrijk voor het creëren van een hoge spontane of geholpen naamsbekendheid en hiermee voor het creëren van een hoge merkwaarde. De (logische) verwachting is dat door het merk van naam te veranderen (*ceteris paribus*) de merkbekendheid zal dalen. Doel van de studie op dit punt is inzicht te krijgen in de mate waarin dit effect zich voordoet.

Merkassociaties

Merkassociaties zijn mentale verbintenissen in het geheugen van de consument tussen een merknaam en de symbolische betekenissen die een consument aan dit merk verbindt. Deze betekenissen worden gevormd op basis van directe of indirecte ervaringen met het merk. Door het vormen van een associatief netwerk rondom het product, krijgt het merk dus een bepaalde betekenis voor de consument (Franzen & Bouwman, 2001, p. 144-162). Keller (2002) stelt dat merkassociaties sterk, uniek en favoriet moeten zijn. Als een merknaam wordt veranderd, zullen naar verwachting de associaties minder sterk worden. Of een bedrijf graag wil dat bestaande associaties worden overgedragen op de nieuwe merknaam, is afhankelijk van het doel dat het bedrijf voor ogen heeft met de merknaamverandering.

Gepercipieerde kwaliteit

Het gaat hier om de perceptie van de consument met betrekking tot de totale kwaliteit van een product in vergelijking met relevante alternatieven. Gepercipieerde kwaliteit is dus een subjectieve beoordeling, gebaseerd op alle mogelijke ervaringen en confrontaties van consumenten met het merk. De merknaam wordt vaak gezien als een indicator voor de kwaliteit van het product (Biswas, Sherrell, 1993). Een merknaamverandering kan de gepercipieerde kwaliteit doen afnemen, maar ook doen toenemen. De verwachting in de studie is dat wanneer de merknaam van een product wordt veranderd, de gepercipieerde kwaliteit zal wijzigen.

Merktrouw

Merktrouw neemt een prominente plaats in bij de vaststelling van merkwaarde. Een hoge mate van merktrouw is vaak de kern van de waarde van een merk. In de literatuur worden vele manieren beschreven om merktrouw te meten. In deze studie wordt merktrouw gemeten op drie manieren:

- De *share of customer*: het aandeel dat het merk heeft in de categorie-aankopen van een consument, ofwel het 'marktaandeel' van het merk bij een klant. Een merk met het hoogste share of customer wil zeggen dat het merk het hoofdmerk van een consument is in de betreffende productcategorie.
- Een hoge share of customer is geen garantie voor toekomstige aankopen. Daarom dient volgens Baldinger en Rubinson (1996) tevens *de attitude van de klant ten opzichte van het merk* te worden meegenomen. Zij stellen dat hoe positiever de attitude van de klant ten opzichte van het merk is, hoe meer kan worden aangenomen dat klanten trouw blijven of worden aan het merk.
- Tenslotte dient de klant ook een hoge mate van *betrokkenheid met het merk* te hebben, wil er een garantie zijn voor merktrouw in de toekomst (v. Melick, 1999). Wanneer klanten een sterke binding hebben met een merk, zijn zij niet snel geneigd van merk te wisselen en bieden daarom garantie voor toekomstige omzet.

In het empirisch onderzoek wordt voor elk van deze variabelen nagegaan of en zo ja in welke mate deze merktrouw componenten verminderen als gevolg van merknaamveranderingen.

Samenvattend is de te onderzoeken vraag derhalve welke invloed merknaamwijzigingen hebben op de (korte termijn) merkbekendheid, merkassociaties, gepercipieerde kwaliteit en merktrouw. De vraag is of zich daadwerkelijk effecten voordoen en zo ja of deze effecten sterk zijn. Hiermee wordt inzicht verkregen in de mate waarin merknaamveranderingen al of niet schadelijke effecten hebben op merkwaarde. Dit is een relevante vraag omdat veelal wordt gesteld dat deze effecten sterk zijn.

Op de te onderzoeken relatie zijn een aantal factoren van invloed die eveneens in de studie worden opgenomen. Deze interveniërende variabelen zijn:

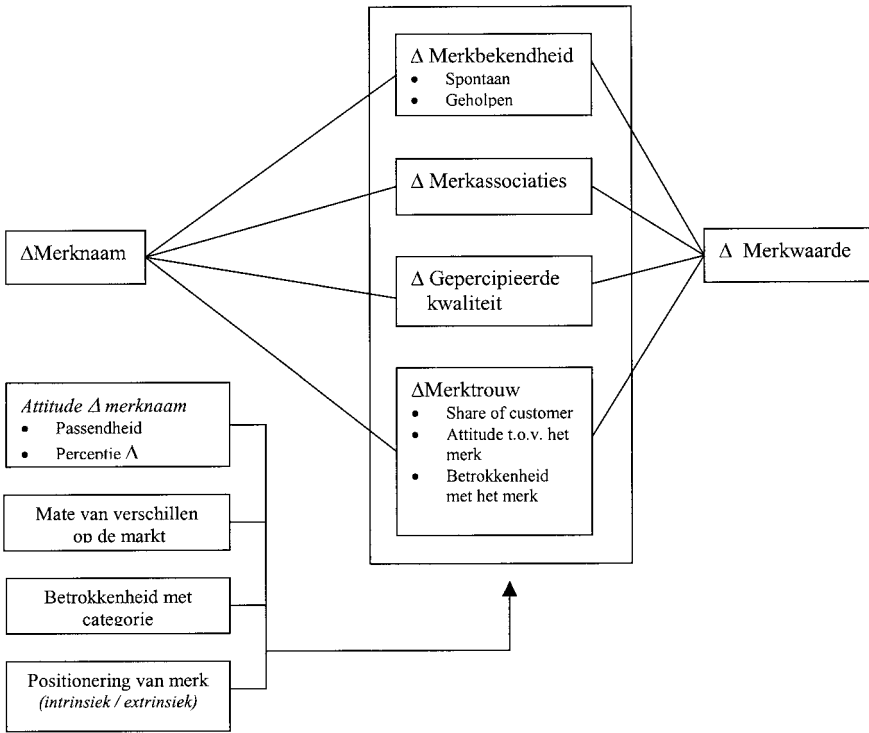
- *De perceptie van de merknaamverandering.* Hierbij onderscheiden we twee componenten: de gepercipieerde *passendheid* van de nieuwe merknaam bij het merkimago en de gepercipieerde *grootte van de merknaamverandering*. De hypothesen zijn dat als consumenten de nieuwe naam goed vinden passen bij het (oude) merkimago er minder invloed is op merkwaarde, en dat grote merknaamveranderingen een sterkere invloed zullen hebben dan kleine naamswijzigingen.
- *De (gepercipieerde) verschillen van de merken op de markt.* Dit is een maatstaf voor de mate van concurrentie op de markt. De hypothese is dat wanneer de gepercipieerde verschillen van merken hoger is, een merknaamverandering van een merk een minder sterk negatief effect zal hebben op de merkwaarde.
- *Betrokkenheid met de categorie.* Consumenten die een grote mate van betrokkenheid met de productcategorie hebben, bijvoorbeeld omdat ze een hoge gebruiksfrequentie hebben (hoog productgebruik), zijn beter op de hoogte van de eigenschappen van merken. De hypothese is dat deze consumenten een verandering in de merknaam minder associëren met merkwaardedalingen dan consumenten met een lage betrokkenheid.
- *Positionering van het merk.* Merken kunnen intrinsiek of extrinsiek worden gepositioneerd. Intrinsiek positionering is op basis van de fysieke eigenschappen van het product. Extrinsieke positionering is positionering op niet-fysieke eigenschappen, zoals het imago. Verwacht wordt dat bij merken met een extrinsieke positionering de rol van de merknaam bij de associaties van het merk een sterkere rol speelt dan bij intrinsiek gepositioneerde merken en dat dus merknaamveranderingen bij extrinsiek gepositioneerde merken een sterkere invloed op merkwaarde zullen hebben dan bij intrinsiek gepositioneerde merken.

Het conceptuele model uit deze studie is samengevat in figuur 1, waarin Δ staat voor verandering.

3. OPZET EMPIRISCHE STUDIE

3.1. Methode van onderzoek

Voor de toetsing van de hypothesen is een experiment gehouden bij circa 200 studenten. De respondenten kregen een vragenlijst met vragen over de merkwaarde van een bestaand intrinsiek en extrinsiek merk (nulmeting). Daarna werd verteld dat de namen van de merken werden veranderd en werd de nieuwe naam gemeld. Van het intrinsieke merk werd gezegd dat de reden was dat alleen merken uit de gezondheidszorg nog op –ium mochten eindigen. Het extrinsieke merk had besloten tot een merknaamverandering gezien de buitenlandse ambities van het merk. Na het lezen van dit bericht kregen



Figuur 1. De invloed van een merknaamverandering.

de respondenten weer een aantal stellingen voorgelegd om zo de merkwaarde na de naamswijziging te kunnen meten. Het voordeel van deze opzet is dat het geïsoleerde effect van een merknaamswijziging wordt gemeten. Er is wel voor gekozen enige argumenten voor de naamswijziging aan te geven. Dit is derhalve een beperkte vorm van communicatie rondom de naamswijziging. Dit is gedaan om de situatie meer overeenkomstig de praktijk te doen zijn, waarin doorgaans ook naamswijzigingen worden ondersteund met communicatie.

Elke respondent kreeg een merknaamverandering van een intrinsiek en een extrinsiek gepositioneerd merk voorgelegd. Voor beide merken zijn kleine en grote naamswijzigingen bedacht. In totaal zijn daarmee vier situaties gecreëerd: intrinsiek, grote en kleine verandering, en extrinsiek, grote en kleine verandering. Teneinde de respondenten niet te vermoeien, kreeg elke respondent twee van deze situaties voorgelegd: een merk met een grote en het andere merk met een kleine naamswijziging. Per situatie bedroeg het aantal respondenten 92.

3.2. Merknamen

Gekozen is voor twee snellopende consumentenproducten (fast moving consumer goods). Bij dergelijke producten zijn in de praktijk het snelst resultaten te meten van merknaamveranderingen. Het intrinsiek gepositioneerde merk is tandpastamerk Zenidium. Dit merk beroept zich in haar uitingen vooral op een technologisch en weten-

schappelijk verantwoorde samenstelling. Het extrinsiek gepositioneerde merk is Hertog Jan bier. Bier is bij uitstek een categorie die op basis van imago wordt gepositioneerd. Beide merken zijn niet marktleider, en hebben een redelijke bekendheid.

De nieuwe merknamen (tabel 1) zijn ontwikkeld door merknaambureau Globrands. Gestreefd is nieuwe namen te ontwikkelen in dezelfde stijl als de oude naam. Zendium is een abstracte naam met een technisch imago. De nieuwe naam met een kleine verandering is Zendial, de nieuwe naam met een grote verandering is Xerial. De merknaam Hertog Jan is een ‘anekdotische’ merknaam met een middeleeuws, ambachtelijk karakter. De nieuwe merknamen voor Hertog Jan kennen tevens dezelfde stijl. De nieuwe naam met kleine verandering is Prince Jan en de nieuwe naam met grote verandering is Compagne Biere.

Tabel 1. Gebruikte merknamen in de empirische studie.

	Intrinsiek merk	Extrinsiek merk
Huidige naam	Zendium	Hertog Jan
Kleine verandering	Zendial	Prince Jan
Grote verandering	Xerial	Compagne Biere

De ene groep van 92 respondenten kregen de namen Zendial en Compagne Biere voor-gelegd, de andere groep de namen Prince Jan en Xerial.

3.3. Meetmethode

Elk van de te meten variabelen (zowel de merkwaarde componenten als de achtergrondvariabelen) is gemeten met meerdere stellingen. De gebruikte stellingen zijn deels gebaseerd op Yoo et al. (2000) en zijn deels zelf ontwikkeld. Voor het toetsen van de interne consistente betrouwbaarheid van de gebruikte schalen is gebruik gemaakt van de Cronbach Alfa. Op grond van deze analyses zijn enkele stellingen uit het bestand verwijderd. Bijlage A geeft weer welke stellingen zijn overgebleven (als voorbeeld is bier genomen) en welke Cronbach Alfa uiteindelijk resteerde.

Teneinde de hypothesen te toetsen zijn de componenten van merkwaarde bepaald als gemiddelden van de desbetreffende stellingen. Met behulp van gepaarde t-toetsen is bepaald of de gemiddelden (bijvoorbeeld merkbekendheid voor en na de merknaamverandering) significant zijn veranderd. Regressie-analyse is gebruikt om relaties te toetsen tussen een afhankelijke variabele (bijvoorbeeld merkassociatieveranderingen) en onafhankelijke variabelen (zoals betrokkenheid bij de categorie). Ten slotte is met een factoranalyse onderzocht of de vier componenten van merkwaarde kunnen worden samengevat in één overkoepelende factor ‘merkwaarde’.

4. RESULTATEN

In tabel 2 is weergegeven in hoeverre de verschillende factoren van merkwaarde na de naamsverandering significant zijn veranderd. Tevens is dit aangegeven voor de totale merkwaarde voor en na de verandering. Deze analyse is uitgevoerd, nadat uit de factor-

analyse naar voren is gekomen dat de vier verschillende factoren kunnen worden samengevoegd in één totaalfactor (de factorscores van de componenten merkwaarde voor en na verandering komen sterk overeen; de principale componentenanalyse komt tot één factor). In de tabel zijn de effecten van een kleine en grote naamsverandering samen genomen.

Tabel 2. Veranderingen in componenten van merkwaarde voor en na naamsverandering (=significant op .05 niveau).*

	Vershil gemiddelden	T-waarde
Merkbekendheid		
• Zendium	-0,1098	-1,041
• Hertog Jan	-0,7604	-7,769*
Merkassociaties		
• Zendium	0,1772	3,847*
• Hertog Jan	0,2308	3,753*
Geperc. Kwaliteit		
• Zendium	0,1739	2,349*
• Hertog Jan	-0,3949	-4,361*
Merktrouw		
• Zendium	-0,1139	-2,293*
• Hertog Jan	-0,2947	-5,507*
Merkwaarde		
• Zendium	0,0309	0,667
• Hertog Jan	-0,3880	-7,034*

Overzicht bevindingen

Uit tabel 2 kan het volgende worden geconcludeerd. Na de naamsverandering:

- is alleen bij het extrinsiek gepositioneerde merk (Hertog Jan) de merkbekendheid gedaald.
- zijn de diverse merkassociaties bij beide merken sterker geworden.
- is de kwaliteitsperceptie bij het intrinsiek gepositioneerde merk gestegen en bij het extrinsiek gepositioneerde merk gedaald.
- is de merktrouw bij beide merken significant verminderd.
- is de totale gepercipieerde merkwaarde van het intrinsiek gepositioneerde merk al met al niet veranderd, maar van het extrinsiek gepositioneerde merk wel (daling).

Het wordt derhalve bevestigd dat het intrinsieke merk ‘minder last’ heeft van naamsveranderingen dan het extrinsieke merk. Dit valt te verklaren uit het feit dat bij merken waarbij het abstracte imago belangrijk is, de naam een sterke drager van dat imago is. Wat ook wordt bevestigd is dat de merkbekendheid en de merktrouw bij een naamsverandering dalen. Opmerkelijk daarbij is dat bij het intrinsiek gepositioneerde merk de daling in de merkbekendheid niet significant is. Kennelijk zijn de respondenten van mening dat zij ondanks de naamsverandering de eigenschappen van het product nog steeds goed kennen.

Bij de gepercipieerde kwaliteit is opvallend dat deze bij het intrinsiek gepositioneerde merk is gestegen. Kennelijk wordt bij het op eigenschappen gepositioneerde merk de

naamsverandering automatisch verbonden aan een kwaliteitsverbetering. Bij het extrinsiek gepositioneerde merk ligt dit juist omgekeerd.

Merkassociaties

De merkassociaties zijn bij beide merken in waardering toegenomen. Hierbij dient echter te worden opgemerkt dat het hier een totaalscore betreft. Wanneer gekeken wordt naar de verandering per afzonderlijke associatie blijkt alleen de associatie ‘vernieuwend’ bij het extrinsiek gepositioneerde merk niet significant te veranderen. Alle andere associaties veranderen wel significant. Er kan niet geconcludeerd worden of een stijging of een daling van de waarde van een afzonderlijke associatie negatief dan wel positief is. Dit is afhankelijk van het doel dat een bedrijf voor ogen heeft met een merknaamverandering.

In dit verband is het interessant om te kijken naar een van de verklarende variabelen, te weten *de mate van verandering in de merknaam*. Naar alle waarschijnlijkheid wil een bedrijf in het geval van een kleine merknaamverandering zoveel mogelijk vasthouden aan de bestaande associaties en is een sterke verandering in de totale associatiewaarde niet gewenst. In het geval van een grote naamsverandering wil het bedrijf naar alle waarschijnlijkheid de bestaande associaties zoveel mogelijk loslaten en nieuwe associaties creëren rondom het merk. Een grote naamsverandering zal dan waarschijnlijk een grote verandering teweeg brengen in de bestaande associaties.

Het verschil in effect van een kleine versus een grote merknaamverandering op veranderingen in associaties is gemeten aan de hand van gepaarde t-toetsen voor de vier verschillende merknamen die zijn gebruikt in het onderzoek. De resultaten van deze t-toetsen zijn opgenomen in tabel 3.

Tabel 3. Verandering associaties kleine versus grote merknaamverandering; *=significant op .05 niveau.

	Verskil in associatiescore	Correlatie
Zendium		
• Klein	-0,0971	0,006
• Groot	0,2109	0,620*
Hertog Jan		
• Klein	0,0355	0,519
• Groot	0,1725	0,461*

Uit de tabel komt naar voren dat de totaalwaarde van de associaties bij een kleine merknaamverandering nauwelijks verandert. De verandering is bij beide merken niet als significant aan te merken. Door de kleine merknaamverandering van ‘Zendium’ in ‘Zendial’ en ‘Hertog Jan’ in ‘Prince Jan’, zijn de respondenten geneigd het merk en daarmee de nieuwe merknaam ongeveer in dezelfde mate op de verschillende associaties te waarderen als voor de naamsverandering. De grote naamsverandering van ‘Zendium’ in ‘Xerial’ en ‘Hertog Jan’ in ‘Compagne Biere’ brengt echter veel meer teweeg bij de totale waardering op de verschillende associaties. Door de grote naamsverandering verandert de totaalscore significant ten opzichte van de totaalscore voor de verandering. Respondenten associëren de nieuwe merknaam en dus het merk verschillend

door een grote naamsverandering. Deze scores komen overeen met de verwachtingen omtrent de veranderingen in associaties.

Bij beide merken treedt op gelijke wijze een verandering op in de bestaande associaties bij een kleine of een grote naamsverandering. Bij een grote naamsverandering veranderen de bestaande associaties significant en bij een kleine naamsverandering blijven de bestaande associaties ongeveer gelijk.

Effecten van achtergrondvariabelen

Ten slotte wordt in tabel aangegeven in hoeverre verbanden zijn gevonden tussen de verschillende componenten van merkwaarde na naamswijziging enerzijds en de volgende achtergrondvariabelen anderszijds: attitude ten opzichte van de nieuwe merknaam (combinatie van passendheid en gepercipieerde verandering van merknaam), betrokkenheid met categorie, en de mate van verschil van de merken in de categorie. In de studie zijn uitsluitend enkelvoudige relaties onderzocht en zijn geen interacties tussen de achtergrondvariabelen meegenomen omdat deze effecten op grond van de inhoud van de achtergrondvariabelen niet worden verwacht.

Tabel 4. Verband tussen achtergrondvariabelen en merkwaarde na naamswijziging.

	R / Bèta	R2	Regr. Coëf.	Constante
Attitude naam				
• Bekendheid	0,475	0,226	0,454*	1,423
• Associaties	0,204	0,042	0,148*	2,645
• Gepercipieerde kwaliteit	0,371	0,138	0,309*	2,386
• Trouw	0,163	0,027	0,151*	1,375
• Merkwaarde	0,473	0,224	0,266*	1,956
Betrokkenheid Categorie				
• Bekendheid	0,035	0,001	0,025	2,525
• Associaties	0,148	0,022	0,082*	2,791
• Gepercipieerde kwaliteit	0,029	0,001	-0,018	3,237
• Trouw	0,128	0,016	0,090	1,503
• Merkwaarde	0,105	0,011	0,045	2,517
Mate van verschil				
• Bekendheid	0,261	0,068	0,185**	2,075
• Associaties	0,144	0,021	0,080	2,846
• Gepercipieerde kwaliteit	0,318	0,101	0,161**	2,891
• Trouw	0,325	0,106	0,226**	1,070
• Merkwaarde	0,221	0,049	0,093**	2,313

* = significant op .05 niveau ; ** significant op .05 niveau voor intrinsieke merk; niet significant bij extrinsieke merk

Uit tabel 4 kunnen de volgende conclusies worden getrokken.

De *attitude ten opzichte van de merknaam* blijkt samen te hangen met de effecten van een naamswijziging op de merkwaarde. Een meer positieve attitude ten opzichte van de merknaamverandering, zal zorgen voor hogere waarden van de verschillende factoren van merkwaarde na een merknaamverandering. Hoe positiever de attitude ten opzichte van de merknaamverandering, hoe minder de merkwaarde dus zal dalen op korte termijn. Dit geldt voor beide merken in de studie.

De *mate van betrokkenheid met de categorie* is alleen bepalend voor de mate waarin bestaande associaties worden overgedragen op de nieuwe merknaam. De mate van betrokkenheid blijkt positief gerelateerd te zijn aan de mate waarin bestaande associaties op korte termijn veranderen na een merknaamverandering. Hoe hoger de betrokkenheid met de categorie, hoe groter de verandering in de bestaande associaties is. Bij consumenten die in hoge mate betrokken zijn met een productcategorie, zullen de associaties rondom een merk in sterke mate veranderen na een merknaamverandering. Dit verband is bij beide merken gevonden.

De *mate van verschil met andere merken* blijkt alleen een causaal verband te vertonen met de merkwaarde na een naamsverandering bij het intrinsiek geïmagineerde merk. Bij het extrinsiek geïmagineerde merk zijn geen significante verbanden gevonden. Hoe meer het merk afwijkt van de andere merken in de markt, hoe hoger de merkwaarde na naamswijziging en dus hoe minder sterk de waarde van het merk daalt na een merknaamverandering.

5. CONCLUSIES EN AANBEVELINGEN

De in dit artikel beschreven analyses leiden tot een aantal conclusies betreffende effecten van merknaamveranderingen. Er is geen onderzoek verricht naar mogelijke effecten van begeleidende communicatie rondom de merknaamverandering. Niettemin schetsen we in het navolgende enkele mogelijke communicatieve richtlijnen om de gewenste effecten van merknaamveranderingen te bewerkstelligen.

Conclusies

Uit dit onderzoek naar de korte termijn effecten van merknaamveranderingen blijkt dat niet in alle gevallen merknaamveranderingen leiden tot nadelige effecten op de diverse factoren die de merkwaarde bepalen. Zo blijkt alleen bij het extrinsiek geïmagineerde merk de naamswijziging tot een significante daling in de merkwaarde te leiden. Bij het intrinsiek geïmagineerde merk gebeurt dit in dit onderzoek niet. De oorzaak hiervan kan liggen in de bevinding dat bij het intrinsiek geïmagineerde merk de naamswijziging automatisch werd geassocieerd met kwaliteitsverbeteringen: als de naam verandert zal het product ook wel zijn verbeterd. Bij het meer op imago geïmagineerde merk gebeurt het omgekeerde: de gepercipieerde kwaliteit daalt na een naamswijziging.

Aanbevelingen

Een eerste aanbeveling is derhalve dat producenten van merken waarbij abstracte eigenschappen een grote rol spelen, veel voorzichtiger met naamswijzigingen moeten omspringen dan producenten van merken waarbij fysieke eigenschappen belangrijker zijn (zoals wasmiddelen, luiers enzovoort).

Andere mogelijke aanbevelingen voor aanbieders van merken die overwegen de merknaam te wijzigen zijn de volgende.

1. Aanbevelingen voor intrinsiek en extrinsiek geïmagineerde merken

- Wanneer een producent graag wil dat de consumenten totaal nieuwe associaties vormen rondom een merk, dient hij te kiezen voor een grote merknaamverandering met een merknaam die zo min mogelijk overeenkomsten vertoont met de

oude merknaam. Wanneer de producent echter wil dat een consument na een merknaamverandering zoveel mogelijk vast blijft houden aan de bestaande associaties, dient hij te kiezen voor een kleine verandering van de merknaam.

- Voor een minder sterke daling van de mate van merktrouw na een merknaamverandering, dient een bedrijf zich in zijn communicatie-uitingen voornamelijk te richten op de minder merktrouwe klanten, zodat zij minder snel geneigd zijn om over te stappen op een ander merk. Daarnaast dient het bedrijf zich te richten op de klanten die na de verandering zijn overgestapt, om deze terug te winnen. Tevens verdient het aanbeveling om de merktrouwe klanten te belonen voor hun hoge mate van trouw voor en na de merknaamverandering.

2. Aanbevelingen voor intrinsiek gepositioneerde merken

- Voor een minder sterke daling van de naamsbekendheid dienen producenten de consument duidelijk te maken dat, ondanks de nieuwe merknaam, het product verder niet veranderd is.
- Door een verandering in de naam van het merk, zijn consumenten van mening dat dit een positieve ontwikkeling is voor een intrinsiek gepositioneerd merk en dat de kwaliteit dus ook wel verbeterd zal zijn. De onderneming dient dit echter wel waar te maken. Wanneer de kwaliteit helemaal niet zo goed is als consumenten door de naamsverandering verwachten, zullen zij alsnog overstappen op een ander merk dat de verwachtingen wel kan waarmaken.
- De merkwaarde van een intrinsiek gepositioneerd merk daalt minder sterk, wanneer het merk duidelijk afwijkt van de concurrentie. Bedrijven dienen dan ook te allen tijde duidelijk de unieke voordelen, die het merk biedt, te communiceren. Voor intrinsieke merken geldt in dit geval dat de uniekheid van de producteigenschappen duidelijk naar voren moet komen.

3. Aanbevelingen voor extrinsiek gepositioneerde merken

- Voor een minder sterke daling van de naamsbekendheid dient in de communicatie-uitingen veelvuldig de nieuwe merknaam gecommuniceerd te worden.
- Voor de aanbeveling met betrekking tot de associaties: zie de tweede aanbeveling bij het intrinsiek gepositioneerde merk.
- Consumenten zijn, bij extrinsiek gepositioneerde merken, van mening dat de kwaliteit van het product tevens zal veranderen na een verandering in de merknaam. Daarom dient de onderneming duidelijk te communiceren dat het merk alleen een verandering in de merknaam heeft doorgemaakt, maar dat het product verder hetzelfde is gebleven.

Onderzoeksaanbevelingen

Het hier gepresenteerde onderzoek kent naast de genoemde resultaten belangrijke beperkingen. Ten eerste betreft het een studie naar korte termijn effecten van merknaamveranderingen. Onderzocht zijn de 'primaire reacties' van respondenten op merknaamveranderingen. Het lijkt plausibel dat deze sterker zullen zijn dan reacties later. Immers, mensen wennen aan nieuwe namen. Ten tweede is het een experimentele studie waarbij respondenten werden 'gefocused' op de merknaamveranderingen en waarbij in beperkte mate ondersteunende communicatie werd toegevoegd. In werkelijkheid zullen merknaamwijzigingen soms gepaard gaan met veel meer communicatie en soms ook met helemaal geen communicatie. Dit hangt af van het doel dan een bedrijf heeft

met de naamswijziging. Het is interessant te onderzoeken wat het effect is van meer of minder communicatie op merkwaardeveranderingen na een naamswijziging. Op al deze punten lijkt vervolgonderzoek wenselijk, waarbij gezien de aard van de vraagstellingen experimenteel onderzoek het meest voor de hand ligt.

LITERATUUR

- Aaker, D.A. (1991), *Managing Brand Equity: capitalizing on the value of a brand name*, The Free Press, New York.
- Aaker, D.A. (1996), *Building Strong Brands*, The Free Press, New York.
- Aaker, D.A., K.L. Keller (1990), 'Consumer Evaluations of Brand Extensions', *Journal of Marketing*, 54, (January), 27-41.
- Ambler, T. (2000), *Marketing and the Bottom Line: the new metrics of corporate wealth*, Prentice Hall, Pearson Education.
- Baldinger, A.L. and Rubinson, J. (1996), 'Brand Loyalty; the link between attitude and behavior', *Journal of Advertising Research*, 36 (6), 22-34.
- Biswas and Sherrell (1993), 'The influence of Product Knowledge and Brand Name on internal price standards and confidence', *Psychology and Marketing*, 10 (1), 31-46.
- Bottomley, P.A., S.J. Holden (2001), 'Do We Really Know How Consumers Evaluate Brand Extensions? Empirical Generalisations Based on Secondary Analysis of Eight Studies', *Journal of Marketing Research*, 38, pp. 494-500.
- Feldwick, P. (1996), 'What is Brand Equity anyway, and how do you measure it?', *Journal of the Market Research Society*, 38 (2), 85-104.
- Floor, J.M.G. en Raaij, W.F. van (1998), *Marketingcommunicatiestrategie*, Educatieve Partners Nederland BV, Houten
- Franzen, G. en Goessens, C. (1998), *Merken & Reclame: hoe reclame-effectiviteit brand equity beïnvloedt*, Kluwer BedrijfsInformatie, Deventer.
- Franzen, G. and Bouwman, M. (2001), *The Mental World of Brands: mind, memory and brand success*, World Advertising Research Centre, UK.
- Keller, K.L. (2002), *Strategic Brand Management: Building, Measuring and Managing Brand Equity*, Upper Saddle River, Prentice Hall, New Jersey. Sec. ed.
- Kwast, P. van der (2001), 'Uitgebloeide merken', *Intermediair*, 17.
- Lane, V.R. ((2000), 'The Impact of Ad Repetition and Ad Content on Consumer Perceptions of Incongruent Extensions', *Journal of Marketing*, 64, pp. 80-91.
- McCracken, J.C. and Macklin, M.C. (1998), 'The role of brand names and visual cues in enhancing memory for consumer packaged goods', *Marketing Letters*, 9 (2), 209-226.
- Melick, W. van (1999), *Merktrouw: behavioristisch versus psychologisch definiëren, meten en stimuleren van merktrouw*, GVR-monografie, Genootschap voor Reclame, Amsterdam.
- Reddy, Holak, Bhat (1994), 'To Extend or not to Extend: Succes Determinants of Line Extensions', *Journal of Marketing Research*, 31, 243-262.
- Sen, S. and Morwitz, V.G. (1996), 'Is it better to have loved and lost than never to have loved at all? The Effect of Changes in Product Features over Time', *Marketing Letters*, 7 (3), 225-235.
- Szybillo, G.J. and Jacoby, J. (1974), 'Intrinsic versus extrinsic cues as determinants of perceived product quality', *Journal of Applied Psychology*, 59 (1).
- Teas, R.K. and Grapentine, T.H. (1996), 'Demystifying Brand Equity: framework guides research on the role brands play in purchase decisions', *Journal of Marketing Research*, 8 (2).
- Yoo, B., N. Donthu, S. Lee (2000), 'An Examination of Selected Marketing Mix Elements and Brand Equity', *Journal of the Academy of Marketing Science*, 28 (2), 195-211.

Bijlage

Tabel B1. In analyses gebruikte vragen (extrinsiek/kleine verandering) en codes

Betrokkenheid met categorie

BC1	Ik ben geïnteresseerd in bier
BC2	Over bier kan ik lang praten
BC4	Het kopen van het juiste bier is voor mij erg belangrijk

Mate van verschil

MG3	Ik vind Hertog Jan bier duidelijk anders dan andere merken bier
-----	---

Merkbekendheid voor verandering

MB01	Ik herken Hertog Jan bier tussen de andere merken bier
MB02	Ik kan enkele eigenschappen van Hertog Jan bier opnoemen
MB03	Ik heb het idee dat ik het merk Hertog Jan goed ken
MB04	Ik kan snel het logo of het symbool van Hertog Jan herkennen
MB05	Van Hertog Jan heb ik een helder beeld

Merkassociaties voor verandering

MA02	Betrouwbaar
MA03	Vernieuwend
MA04	Gewaagd
MA05	Genieten ('gezond' bij Zendium)
MA06	Modern ('jong' bij Zendium)

Gepercipieerde kwaliteit voor verandering

GK01	Ik waardeer Hertog Jan
GK02	Hertog Jan bier is van goede kwaliteit
GK03	Ik kan op Hertog Jan bier vertrouwen

Share of customer voor verandering

SC01	Ik beschouw mijzelf als een frequente gebruiker van Hertog Jan bier
SC02	Hertog Jan is het enige merk bier dat ik koop en drink (thuis)
SC03	De volgende keer dat ik bier koop, koop ik Hertog Jan bier

Attitude merk voor verandering

AT01	Over mijn laatste gebruikservaring met Hertog Jan bier ben ik tevreden
AT02	Ik heb een sterke voorkeur voor Hertog Jan bier
AT03	Ik vind het plezierig om Hertog Jan bier te kopen en te drinken

Betrokkenheid merk voor verandering

BM01	Als een ander merk in de aanbieding is, koop ik toch Hertog Jan bier
BM02	Ik zou Hertog Jan bier aanbevelen bij anderen
BM03	Hertog Jan past bij mij

Passendheid nieuwe merknaam

PV1	Ik vind de naam Prince Jan bij het merk passen
PV2	De naam Prince Jan sluit aan bij datgene waar het merk voor staat

Perceptie mbt mate van naamsverandering

PM1	Ik beschouw de naamsverandering van Hertog Jan in Prince Jan als een kleine naamsverandering
PM2	Ondanks de nieuwe naam is er weinig veranderd aan het merk
PM3	De nieuwe naam Prince Jan doet me sterk denken aan de oude naam

Merkbekendheid na verandering

MB11	Ik zal Prince Jan bier niet meteen herkennen in de supermarkt
MB12	Van Prince Jan bier heb ik een helder beeld

Merkassociaties na verandering

MA11	Behoudend
MA12	Betrouwbaar
MA13	Vernieuwend
MA14	Gewaagd
MA15	Genieten ('gezond' bij Zendium)
MA16	Ouderwets ('jong' bij Zendium)

Gepercipieerde kwaliteit na verandering

GK11	Ik waardeer Prince Jan bier evenveel als Hertog Jan bier
GK12	Prince Jan bier is van goede kwaliteit
GK13	Ik kan op Prince Jan bier vertrouwen

Share of customer na verandering

SC11	De volgende keer dat ik bier koop, koop ik Prince Jan bier
SC12	Prince Jan bier is het enige merk bier dat ik ga kopen en drinken (thuis)

Attitude merk na verandering

AT11	Ik heb een sterke voorkeur voor Prince Jan bier
AT12	Prince Jan roept bij mij een prettig gevoel op

Betrokkenheid merk na verandering

BM11	Als een ander merk in de aanbieding is, zal ik toch Prince Jan bier kopen
BM12	Ik zou Prince Jan bier aanbevelen bij anderen
BM13	Prince Jan past bij mij

Tabel B2. Betrouwbaarheid schalen (Cronbach alfa)

Vragen voor verandering	C. A. Zendium	C.A. Hertog Jan
BC1, BC2, BC4	0,6638	0,8319
MG3	0,6763	0,5808
MB01, MB02, MB03, MB04, MB05	0,8370	0,8162
MA02, MA03, MA04, MA05, MA06	0,5381	0,5179
GK01, GK02, GK03	0,7820	0,8577
SC01, SC02, SC03	0,9339	0,8000
AT01, AT02, AT03	0,8153	0,7864
BM01, BM02, BM03	0,7904	0,7374

Vragen na verandering	C. A. Zendium	C.A. Hertog Jan
PM1, PM2	0,7853	0,8356
PV1, PV2, PV3	0,6696	0,6272
MB11, MB12	0,5774	0,5774
MA11, MA12, MA13, MA14, MA15, MA16	0,6102	0,5991
GK11, GK12, GK13	0,6776	0,6269
SC11, SC12	0,7959	0,7966
AT11, AT12	0,6544	0,6135
BM11, BM12, BM13	0,7534	0,7982

12. *Hollandse tomaten in Duitse winkels*

Het afstemmen van imago en claimsterkte in advertenties¹

P.W.J. VERLEGH

SAMENVATTING

Labels als “made in Germany” of “made in Taiwan”, Italiaans klinkende merknamen als “Bertolli” en “Buitoni” en afbeeldingen van meisjes in klederdracht vertellen de consument iets over het land van herkomst van een product. Uitgebreid onderzoek heeft aangetoond dat land van herkomst een sterke invloed heeft op de evaluatie van producten door de consument. In dit artikel wordt het land-van-herkomst effect opgesplitst in een directe invloed (land van herkomst als extra product-attribuut), en een modererende invloed op de verwerking van productclaims. De resultaten laten zien dat de effectiviteit van advertenties wordt beïnvloed door de mate waarin de sterkte van productclaims is afgestemd op het imago van het land van herkomst van een product. Het onderzoek is van belang voor het ontwerp van effectieve marketingcommunicatie voor landenimago’s (bv. Nederlandse landbouwproducten in Duitsland), maar heeft ook implicaties voor communicatiestrategieën van merken en organisaties.

1. LAND-VAN-HERKOMSTEFFECTEN

Producten en merken kunnen op verscheidene manieren gerelateerd zijn aan een land van herkomst. Hierbij valt te denken aan de bekende “made in ...” labels, maar ook aan merknamen of (symbolische) verwijzingen op de verpakking of in advertenties. Voorbeelden van dergelijke verwijzingen zijn merknamen als “Café de Colombia” en “Clearly Canadian”, maar ook het kleurgebruik en de Scandinavisch aandoende productnamen van IKEA, of de slogan “Freude am Fahren” die door BMW gebruikt wordt in haar Amerikaanse en pan-Europese reclamecampagnes. Zulke verwijzingen zijn niet altijd gerelateerd aan de “werkelijke” herkomst. Zo is het elektronikamerk “Kenwood” bijvoorbeeld van Japanse origine, en wordt “Haägen-Dasz” ijs gemaakt in de Verenigde Staten. Leclerc, Schmitt en Dubé (1994) lieten zien dat dergelijke verwijzingen een sterke invloed hebben op de evaluatie van producten door de consument. Naast de verwijzingen die marketers maken naar het land van herkomst, associeert de consument zelf in veel gevallen producten en merken met een bepaalde nationaliteit: Nike is Amerikaans; Gucci, Fiat en Ferrari komen uit Italië, en Sony en Mitsubishi komen uit Japan.

¹ Dit artikel is gebaseerd op het proefschrift “Country-of-Origin Effects on Consumer Product Evaluations” van Peeter Verlegh, promotores M.T.G. Meulenberg en J.E.B.M. Steenkamp (Wageningen Universiteit). Het onderzoek is mogelijk gemaakt door medewerking van GfK Nederland en het Productschap Tuinbouw.

Nicholas Papadopoulos (uit Griekenland) en de Canadese Louise Heslop (1993, p xxii) stelden dat de associatie van een merk met een herkomstland dezelfde rol heeft als het kennen van het beroep van een gesprekspartner op een feestje: het geeft een indruk van wie je tegenover je hebt, en vormt de basis voor een eerste oordeel over zijn of haar karakter en kwaliteiten.

De relevantie van herkomstland voor de consument heeft geleid tot een groot aantal wetenschappelijke artikelen die onderzoek beschrijven naar de invloed van deze variabele op de beoordeling van producten door de consument. Op basis van deze studies kan geconcludeerd worden dat het land van herkomst een belangrijke rol speelt in de beoordelingen van producten door consumenten (Verlegh 2001). Maheswaran (1994) liet bijvoorbeeld zien dat consumenten land van herkomst kunnen gebruiken als een cognitieve heuristiek, dat wil zeggen dat consumenten op basis van het land van herkomst een (eerste) indruk van het product vormen, waardoor zij in staat zijn om snel en eenvoudig een oordeel te geven. In de meta-analyse van Verlegh en Steenkamp (1999) werd tevens aangetoond dat de grootte van het “land-van-herkomsteffect” afneemt naarmate de consument meer informatie heeft over de eigenschappen van het product. Hong en Wyer (1989, 1990) lieten zien dat wanneer consumenten op de hoogte zijn van het land van herkomst van een product, dit ertoe kan leiden dat informatie over de eigenschappen van het product op een andere manier door de consument wordt verwerkt en beoordeeld. Bijvoorbeeld: bij het beoordelen van de kwaliteit van een DVD-speler, kan een set van technische specificaties positief beoordeeld worden wanneer de consument weet dat het product van een Japanse of Amerikaanse makelij is, terwijl deze specificaties minder positief beoordeeld worden wanneer de DVD-speler geproduceerd is in Maleisië of Mexico. Deze indirecte (modererende) invloed van land van herkomst op productevaluaties komt overeen met de “boodschapper-effecten” die bekend zijn uit onderzoek naar de effectiviteit van communicatie. Binnen dit onderzoeksgebied is vastgesteld dat consumenten meer waarde hechten aan informatie die afkomstig is uit een meer betrouwbare bron (zie bijvoorbeeld Smidts, Rossiter en McOmish 1998). In het oorspronkelijke onderzoek naar deze bron-effecten werd met name gekeken naar de betrouwbaarheid van presentatoren, experts en media. Recent onderzoek in de marketing-literatuur liet echter zien dat ook bedrijven en merken als “bron” van productinformatie beschouwd kunnen worden, waarbij de overtuigendheid van productclaims afhankelijk is van de reputatie van het bedrijf dat het betreffende product op de markt brengt (Goldberg en Hartwick 1990, Brown en Dacin 1997). In de huidige studie wordt gekeken naar de rol van *land van herkomst* als “bron” van productinformatie, waarbij verondersteld wordt dat een land van herkomst met een positief imago geldt als een betrouwbare bron voor de consument. Daarnaast wordt uiteraard gekeken naar de directe invloed van land van herkomst in de context van adverteren. Er wordt bekeken of (1) consumenten het land van herkomst van een product beschouwen als relevante productinformatie, die zij als een extra attribuut meewegen in hun evaluatie van een product, en bovendien of (2) het land van herkomst van een product van invloed is op de evaluatie van productclaims in een advertentie (door consumenten).

2. THEORETISCH KADER

Bij het vormen van een algemeen oordeel over een product wordt de beschikbare informatie door consumenten gecombineerd tot een geheel. De informatie-integratie theorie

(Anderson 1971, 1981) beschrijft hoe dit proces in zijn werk gaat. De theorie stelt dat consumenten eerst elk stukje informatie afzonderlijk beoordelen, en vervolgens deze beoordelingen integreren tot een totaalindruk. Bij de evaluatie van een auto van Duitse makelij, wordt in de beoordelingsstap de Duitse afkomst positief beoordeeld, vanwege het kwaliteitsimago van de Duitse auto-industrie. Soortgelijke beoordelingen worden gemaakt voor het opgegeven verbruik en de motorinhoud, en deze oordelen worden vervolgens samengevoegd in de integratiestap, waarin elk van deze oordelen wordt gewogen naar het belang dat de consument hecht aan de betreffende eigenschap. Het is bijvoorbeeld mogelijk dat “land van herkomst” of “benzineverbruik” zwaarder weegt dan acceleratievermogen. Dit integratieproces kan het best beschreven worden als de berekening van het gemiddelde van de gewogen oordelen over de individuele product-eigenschappen (Anderson 1981, Johar et al. 1997). Dit komt tevens overeen met de bevinding dat de invloed van land van herkomst op productevaluaties afneemt naarmate de meer informatie heeft over de eigenschappen van het product (Verlegh en Steenkamp 1999).

De uitkomst van het hierboven beschreven integratie-proces is sterk afhankelijk van het gewicht dat wordt toegekend aan elk afzonderlijk stukje informatie: informatie waaraan een groter gewicht wordt toegekend zal immers meer invloed hebben op het totaaloordeel over het product. Het gewicht dat wordt toegekend aan bepaalde informatie wordt sterk beïnvloed door de geloofwaardigheid ervan (Birnbbaum en Stegner 1979). Goldberg en Hartwick (1990) toonden aan dat de geloofwaardigheid van informatie (bijvoorbeeld in een advertentieclaim) wordt bepaald door een interactie tussen de reputatie van de adverteerder (bron) en de extremiteit van de claim (inhoud). De claim dat een product als nummer 1 uit een smaaktest kwam, was weinig geloofwaardig als het product afkomstig was van een fabrikant met een matige reputatie. Dit effect was zo sterk, dat het eindoordeel over het product van de matige fabrikant zelfs gunstiger werd, wanneer hij claimde dat het product als nummer 3 uit de test was gekomen, en niet als nummer 1. Dit resultaat is in overeenstemming met het zogenaamde boemerangeffect: adverteren met te positieve claims kan een negatieve uitwerking hebben op de evaluatie van een product (Rossiter en Percy 1997, p. 252). De studie van Goldberg en Hartwick toonde aan dat dit verklaard kan worden doordat de geloofwaardigheid van een claim afneemt met de sterkte ervan, tenminste wanneer de adverteerder een negatieve reputatie heeft.

Het belang van dit resultaat ligt in de afweging tussen de geloofwaardigheid en de sterkte van een claim. Hoewel een sterkere claim (bijvoorbeeld “wij waren nummer 1 in een recente smaaktest”) meer positief beoordeeld zal worden dan een zwakkere claim (“wij waren nummer 5 in een recente smaaktest”), zal de consument de tweede claim meer geloofwaardig vinden, en er dus meer gewicht aan toekennen. Dit is met name van belang voor adverteerders met een minder gunstig imago. Bij het verbeteren van het imago van deze adverteerders en hun producten draait het dus om de balans tussen sterkte en geloofwaardigheid van claims in advertenties.

Een complicerende factor in dit geheel is de hoeveelheid aandacht die een consument besteedt aan een advertentie. In de literatuur over dit onderwerp wordt gesteld dat de verwerking van informatie beschreven kan worden aan de hand van zogenaamde duaalprocesmodellen (Petty en Cacioppo 1986, Chaiken, Liberman en Eagly 1989). Volgens deze modellen kan een consument advertenties verwerken op een meer systematische wijze – ook wel de “centrale route” genoemd, of op een meer heuristische wijze, die tevens bekend staat als de “perifere route”. Welke route gevolgd wordt is afhankelijk

van de motivatie, de gelegenheid en het vermogen van de consument om de beschikbare informatie te begrijpen en te beoordelen. Systematische verwerking vereist een hoge score op elk van deze drie factoren. Bij systematische verwerking van informatie wordt elk stukje informatie kritisch bekeken en beoordeeld, en vervolgens zorgvuldig gewogen om te komen tot een eindoordeel over het product. Heuristische verwerking vraagt minder van de consument. De verwerking van informatie is beperkt, en er wordt gebruik gemaakt van heuristieken (bijvoorbeeld “Franse wijn is lekker”) zodat men op eenvoudige wijze tot een oordeel kan komen. In het algemeen zal de verwerking van informatie plaatsvinden op een wijze die het midden houdt tussen deze twee uitersten, waarbij er meer of minder systematische verwerking kan plaatsvinden.

Een toename in betrokkenheid bij een advertentie (bijvoorbeeld omdat zij gaat over een product dat men interessant vindt), zal leiden tot een meer systematische verwerking ervan. De betekenis is dat de consument meer aandacht zal besteden aan de inhoudelijke claims die in de advertentie gemaakt worden. Dit zal er toe leiden dat de consument een groter gewicht toekent aan deze claims (zie Anderson 1971, 1981), en dat men minder geneigd zal zijn om het product te beoordelen op basis van simpele heuristieken die – bijvoorbeeld – gebaseerd zijn op het land van herkomst van het product (Maheswaran 1994). Dit hoeft echter niet altijd het geval te zijn: Op basis van de resultaten van Goldberg en Hartwick (1990) is het immers plausibel dat consumenten die meer aandacht besteden aan een advertentie, eerder geneigd zullen zijn tot het in twijfel trekken van sterke claims die afkomstig zijn van adverteerders met een matige reputatie. Hoewel zij de rol van betrokkenheid niet expliciet meenemen in hun onderzoek, schrijven Goldberg en Hartwick (1990, p. 178) dat het boemerangeffect waarschijnlijk alleen optreedt wanneer er sprake is van een redelijke mate (“reasonable level”) van betrokkenheid bij de advertentie. Dit suggereert dat de tweeweg interactie tussen claimsterkte en bron in werkelijkheid een drieweg interactie is tussen claimsterkte, bron en betrokkenheid.

Op basis van de hierboven beschreven theorie kan een aantal hypothesen opgesteld worden over de rol van het land van herkomst in een advertentie. De eerste hypothese heeft betrekking op het directe effect van land van herkomst op de waardering van producten.

H1: Producten uit een land van herkomst met een positief imago in de betreffende productcategorie worden positiever beoordeeld dan dezelfde producten uit een land van herkomst met een relatief negatief imago in de betreffende productcategorie.

Naast deze directe invloed van land van herkomst, bestaat er ook een indirect effect. In lijn met de hierboven besproken theorie valt te verwachten dat de invloed van land van herkomst en claimsterkte op productevaluatie wordt gemodereerd door de mate waarin de consument betrokken is bij de advertentie. Om de implicaties van deze drieweg-interactie tussen land van herkomst, claimsterkte en betrokkenheid beter te kunnen begrijpen, wordt een onderscheid gemaakt tussen de effecten van land van herkomst voor landen met een positief imago (hypothese 2), en landen met een minder positief imago (hypothese 3). Een land van herkomst met een positief imago geldt als een betrouwbare bron voor de consument. Bij toenemende aandacht voor de advertentie, heeft de consument geen reden om te twijfelen aan de geloofwaardigheid van de advertentie, en zal de consument de argumenten uit de advertentie zwaarder laten meewegen bij de evaluatie van het geadverteerde product.

H2: Wanneer een consument een product beoordeelt uit een land van herkomst dat een relatief positief imago heeft in de betreffende categorie, zal een grotere mate van betrokkenheid bij een advertentie voor dit product leiden tot een sterkere invloed van de (positieve) claims op de productevaluatie.

Een land van herkomst met een negatief imago geldt als een weinig betrouwbare bron voor de consument. Bij toenemende aandacht voor de advertentie, kan de consument gaan twijfelen aan de geloofwaardigheid van de argumenten uit de advertentie. De consument zal daarom sterk letten op de inhoud van de claims. Bij zeer sterke positieve claims zal hij of zij gaan twijfelen aan de geloofwaardigheid van de argumenten in de advertentie, waardoor de claims minder sterk (of zelfs helemaal niet) mee zullen tellen bij de beoordeling van het geadverteerde product, en de consument sterker vertrouwt op het (minder positieve) imago van het land van herkomst. Bij gematigd positieve claims consument komt de geloofwaardigheid minder sterk in het gevaar, waardoor deze claims wel worden meegenomen in het oordeel over het geadverteerde product. Deze combinatie van effecten leidt tot het boemerang-effect, waarbij gematigd positieve claims een positiever oordeel opleveren dan sterk positieve claims.

H3a: Wanneer een consument een product beoordeelt uit een land van herkomst dat een relatief negatief imago heeft in de betreffende categorie, en het product geadverteerd wordt met gematigd positieve claims, zal een grotere mate van betrokkenheid bij de advertentie voor dit product leiden tot meer positieve productevaluaties.

H3b: Wanneer een consument een product beoordeelt uit een land van herkomst dat een relatief negatief imago heeft in de betreffende categorie, en het product geadverteerd wordt met sterk positieve claims, zal een grotere mate van betrokkenheid bij de advertentie voor dit product niet leiden tot meer positieve productevaluaties.

3. HET “WASSERBOMBE PROBLEEM”

De hypothesen zijn getest in een onderzoek naar de evaluatie van Nederlandse en Spaanse tomaten door de Duitse consument. Wereldwijd is Nederland één van de grootste exporteurs van groenten en fruit. Veertig procent van de totale export in deze productcategorie gaat naar Duitsland, dat daarmee de belangrijkste bestemming van Nederlandse groenten en fruit is (Productschap Tuinbouw, 2001). In de afgelopen jaren heeft het imago van de Nederlandse land- en tuinbouw aanzienlijke kritiek gekregen vanuit Duitsland. Dieptepunt was de betiteling van Nederlandse tomaten als “Wasserbomben”. Dergelijke kritieken hebben ertoe geleid dat de Nederlandse land- en tuinbouw en haar producten (met name tomaten) een minder positieve reputatie hebben bij de Duitse consument, hoewel vermeld dient te worden dat de reputatie de laatste jaren langzaam maar zeker verbeterd is. Spaanse tomaten vormen de tegenpool van dit imago. Dit product is haast even bekend als de Nederlandse tomaat (88% versus 95% geholpen bekendheid – GfK 1996), en heeft een positief imago. Tabel 1 toont de gemiddelde evaluaties van Spaanse en Nederlandse tomaten, zoals gemeten in de Holland Imago Monitor, een door GfK uitgevoerd onderzoek onder 1500 Duitse consu-

menten. Het grote verschil in reputatie van Nederlandse en Spaanse tomaten, gekoppeld aan de grote bekendheid bij de Duitse consument, maakt het mogelijk om onze hypothesen over de rol van land van herkomst in de beoordeling van advertenties te onderzoeken in een praktische context met grote relevantie.

Tabel 1. Evaluatie van Spaanse en Nederlandse tomaten door Duitse consumenten (N=1500).*

	Gemiddelde evaluatie (st. dev.)	
	Nederlandse tomaten	Spaanse tomaten
smakelijk	56.3 (28.3)	81.5 (17.7)
natuurlijk	51.6 (25.3)	71.7 (19.6)
goede kwaliteit	63.3 (26.0)	79.2 (17.6)

* op basis van data verzameld door GfK (1998), eigenschappen beoordeeld op een schaal van 10 tot 100.

4. DATAVERZAMELING EN STEEKPROEF

4.1. Opzet van de studie

Stimulus materiaal en vragenlijst voor dit onderzoek werden, met medewerking van GfK en de afdeling marktonderzoek van het Productschap Tuinbouw, toegevoegd aan een groter onderzoek naar de mening van Duitse consumenten over land- en tuinbouwproducten uit Europese landen. Dit onderzoek werd uitgevoerd in een representatief panel van Duitse consumenten. Het stimulusmateriaal voor de hier beschreven studie bestond uit een advertentie, die in een aparte enveloppe werd meegestuurd met de vragenlijst. Deze advertentie was gemanipuleerd volgens een volledig factorieel 2 x 2 x 2 proefontwerp. De drie gemanipuleerde factoren in dit ontwerp waren *Land van Herkomst* (Nederland versus Spanje), *Claimsterkte* (Matig versus Sterk positief), en *Betrokkenheid bij de advertentie* (Laag versus Hoog). Het experiment werd uitgevoerd volgens een tussen proefpersonen ontwerp: iedere respondent kreeg slechts één van de acht verschillende stimuli te zien. De proefpersonen in de verschillende cellen in dit ontwerp waren vergelijkbaar op demografische factoren en geografische spreiding. De vragenlijst en het stimulusmateriaal werden gestuurd naar 1500 consumenten. Een totaal van 1009 vragenlijsten werd teruggestuurd. Vragenlijsten werden verwijderd uit de steekproef wanneer de vragenlijst niet volledig of onnauwkeurig ingevuld was (108 consumenten), of de consument had aangegeven dat hij of zij niet bekend was met Nederlandse of Spaanse tomaten (150 consumenten). Uiteindelijk bestond de steekproef hierdoor uit 751 consumenten, min of meer gelijk verdeeld over de acht cellen in het ontwerp (celgroottes varieerden tussen 88 en 101).

4.2. Vragenlijst

De vragenlijst voor dit onderzoek begon met een korte instructie, waarin de consument werd verzocht de enveloppe te openen, de advertentie te bekijken, en de antwoordschalen in te vullen. De instructie bevatte tevens de manipulatie van betrokkenheid, die in de volgende paragraaf besproken wordt. De instructie werd gevolgd door een aantal evaluatieve uitspraken (bijvoorbeeld “De geadverteerde tomaat smaakt lekker”), waar-

voor de respondent steeds diende aan te geven in hoeverre hij of zij het eens was met de uitspraak. Dit gebeurde op een tienpuntsschaal, lopende van 10 tot 100. Het linkeruiteinde van de schaal was gelabeld “helemaal niet mee eens”, en het rechteruiteinde was gelabeld “volledig mee eens”.

4.3. Stimulusmateriaal

Het stimulusmateriaal bestond uit fullcolour tijdschriftadvertenties op A4-formaat (zie bijlage I voor zwart-wit reproducties). Het bovenste deel (ongeveer 70 %) van de pagina bestond uit een kleurenfoto van een tuinder met een mand tomaten. Over het midden van de foto was een slogan gedrukt. Het onderste deel bestond uit een tekstblok, waarin minder of meer positieve claims over de geadverteerde tomaten gemaakt werden. Deze advertenties waren ontworpen door de onderzoeker. Om de advertenties zo realistisch mogelijk te maken werden zij getest in open interviews uitgevoerd in een Duits winkelcentrum. Het commentaar uit deze interviews werd gebruikt om de advertenties te verbeteren, en de uiteindelijke versie werd verbeterd en opgemaakt door professionals van een groot reclamebureau. Door het land van herkomst van de tomaten en de sterkte van de claims onafhankelijk van elkaar te variëren werden er vier versies van de advertentie gecreëerd.

Land van herkomst werd gemanipuleerd door in de advertentie op twee plaatsen de vlag van het land te tonen, en in de advertentie een fictieve merknaam te gebruiken die direct gerelateerd was aan het herkomstland van de tomaten (Hollandia voor Nederlandse tomaten en Spania voor Spaanse tomaten). In de tekst werd tevens tweemaal gemeld dat de tomaten afkomstig waren uit Nederland (Spanje), en verbouwd door Nederlandse (Spaanse) tuinders.

Claimsterkte werd gemanipuleerd door twee versies te schrijven van de copy van de advertentie, één gematigd positief, en één sterk positief. In de “gematigd positieve” versie werd gesteld dat door tuinders die Spania / Hollandia tomaten verbouwd aandacht werd besteed aan natuur en milieu, wat resulteerde in een kwaliteitsproduct met een goede smaak en aroma. De tekst sloot af met de mededeling dat uit een smaaktest gebleken was dat 6 van de 10 consumenten deze tomaten smakelijker vonden dan andere. In de “sterk positieve” versie werd gesteld dat door tuinders die Spania / Hollandia tomaten verbouwd *meer dan in enig ander land* aandacht werd besteed aan natuur en milieu, wat resulteerde in een *ongeëvenaard* kwaliteitsproduct met een *onovertroffen smaak en aroma*. De tekst sloot af met de mededeling dat uit een smaaktest gebleken was dat 9 van de 10 consumenten deze tomaten *veel* smakelijker vonden dan andere.

Betrokkenheid bij de advertentie werd gemanipuleerd in de instructietekst boven de vragen. De manipulatie was gebaseerd op de manipulaties die gebruikelijk is in onderzoek naar informatieverwerking (Petty en Cacioppo 1986). Er waren twee niveau's van betrokkenheid: hoog en laag. In de versie met lage betrokkenheid werd de consument verteld dat de geadverteerde tomaten niet verkrijgbaar waren, en dat de studie een pre-test was voor een groter onderzoek dat zou plaatsvinden in een ander land. Er werd gevraagd om de advertentie te bekijken, en vervolgens de vragen in te vullen. In de hoge betrokkenheid-versie werd de consument verteld dat de geadverteerde tomaten binnenkort geïntroduceerd zouden worden in de lokale supermarkten, en dat zij deel uit maakte van een geselecteerde groep van consumenten die de advertentie moest beoordelen. Er werd gevraagd om de advertentie zorgvuldig te bestuderen alvorens de vragen in te vullen.

4.4. Afhankelijke variabele

De afhankelijke variabele voor dit experiment was de aankoopintentie van de consument. Deze intentie werd gemeten met twee items, te weten “Als ik Hollandia / Spania tomaten in de winkel zou zien, zou ik ze beslist kopen” en “Ik zou liever Hollandia / Spania tomaten kopen dan tomaten van een ander merk”. De betrouwbaarheid van deze schaal was hoog (Cronbach $\alpha = 0,88$).

4.5. Covariaat: a-priori attitude

Omdat consumenten onderling kunnen verschillen in hun a-priori attitude ten opzichte van Spaanse / Nederlandse tomaten werd een covariaat gebruikt in de analyses van de data. Deze covariaat bestond uit een drietal items die aan de consumenten voorgelegd waren in het kader van de algemene imago-survey waar onze vragenlijst een onderdeel van was. De items waren “smakelijk”, goede kwaliteit”, en “natuurlijk”, en werden beantwoord voordat de consument begon aan de eigenlijke vragenlijst voor dit onderzoek. De betrouwbaarheid van de covariaat was 0.90 voor Nederlandse tomaten, en 0.87 voor Spaanse. De gemiddelde scores van de twee landen op deze schaal bevestigen het beeld dat geschetst werd in tabel 1, namelijk dat het imago van Nederlandse tomaten (gemiddelde = 58.4) significant slechter was dan dat van Spaanse tomaten (gemiddelde = 76.8; $F(1,656) = 125.82$, $p < 0.001$).

5. RESULTATEN

Tabel 2 toont de celgemiddelden uit het experiment. Er is duidelijk sprake van een land-van-herkomst effect, waarbij de aankoopintentie in alle condities voor Spania tomaten veel positiever is dan voor Hollandia tomaten. Tabel 2 laat verder zien dat er voor Spania tomaten die geadverteerd worden met gematigd positieve claims een negatief effect is van betrokkenheid op aankoopintenties (68.5 versus 61.6). Dit effect is significant bij $\alpha = .05$. Andere effecten voor Spania tomaten waren niet significant. Voor Hollandia tomaten toont tabel 2 juist een significant positief effect van betrokkenheid voor tomaten geadverteerd met gematigd positieve claims (46.0 versus 56.1). Voor advertenties met een sterk positieve claim vinden we een niet-significant negatief effect van betrokkenheid. Deze resultaten impliceren een omkering van het effect van claimsterkte onder condities van lage en hoge betrokkenheid: bij lage betrokkenheid leiden sterkere claims tot een meer positief oordeel over Hollandia tomaten, maar bij hogere betrokkenheid is er een negatief (maar niet significant) effect van claimsterkte (zie figuur 1).

De data is tevens geanalyseerd in een variantie-analyse (ANOVA) met aankoopintentie als afhankelijke variabele. De voorspellende variabelen waren claimsterkte, land van herkomst en betrokkenheid bij de advertentie. Er werd in deze analyse een significant hoofdeffect van land van herkomst gevonden ($F(1,743) = 54.30$, $p < .001$, $\eta^2 = .068$). Dit effect bevestigt hypothese 1. Wanneer het imago van het land van herkomst als covariaat werd meegenomen in de analyse, werd dit effect beduidend kleiner ($F(1,649) = 6.80$, $p = .009$, $\eta^2 = .010$), wat erop wijst dat het land-van-herkomsteffect grotendeels toe te schrijven is aan verschillen in het algemene imago van Nederland en Spanje als producent van tomaten. De overige analyses werden allen uitgevoerd met en zonder

Tabel 2. Aankoopintentie voor tomaten bij claims van verschillende sterkte, onder lage en hoge betrokkenheid.*

	SPANIA		HOLLANDIA	
	Lage betrokkenheid	Hoge betrokkenheid	Lage betrokkenheid	Hoge betrokkenheid
gematigd positief	68.5 ^a	61.6 ^a	46.0 ^{a,b}	56.1 ^a
sterk positief	64.3	65.6	54.4 ^b	52.1

^a Verschil tussen hoge en lage betrokkenheid is significant ($\alpha = .05$)

^b Verschil tussen gematigd positief en sterk positief is significant ($\alpha = .05$)

toevoeging van de a-priori attitude als covariaat, hetgeen geen noemenswaardig effect bleek te hebben op de resultaten. In het vervolg van dit artikel zullen daarom slechts de resultaten van analyses met covariaat worden besproken. Zoals eerder besproken in sectie 3 vertegenwoordigen hypothesen 2, 3a en 3b samen een drie-weginteractie tussen land van herkomst, betrokkenheid en claimsterkte. Deze interactie is inderdaad significant in de ANOVA ($F(1,649) = 11.28, p = .001, \eta^2 = .017$). Tabel 3 toont de resultaten van de ANOVA.

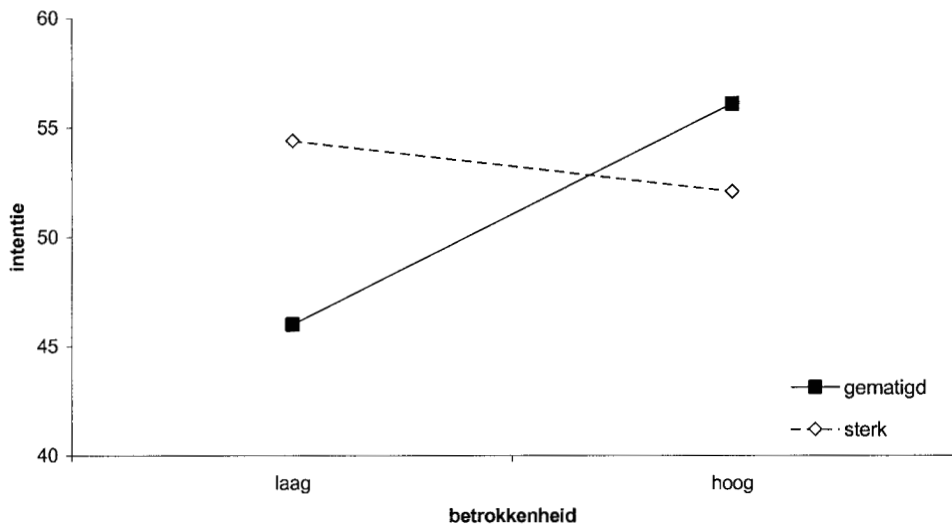
Tabel 3. Resultaten van de ANOVA.

	df	F	p	η^2
Imago (covariaat)	1	114.65	<.001	.150
Land van herkomst (L)	1	6.80	.009	.010
Claimsterkte (C)	1	.23	.635	.000
Betrokkenheid (B)	1	.15	.695	.000
L*C	1	.61	.435	.001
L*B	1	3.81	.051	.006
B*C	1	1.24	.265	.002
L*B*C	1	11.28	.001	.017
Error	649			
Totaal	658			

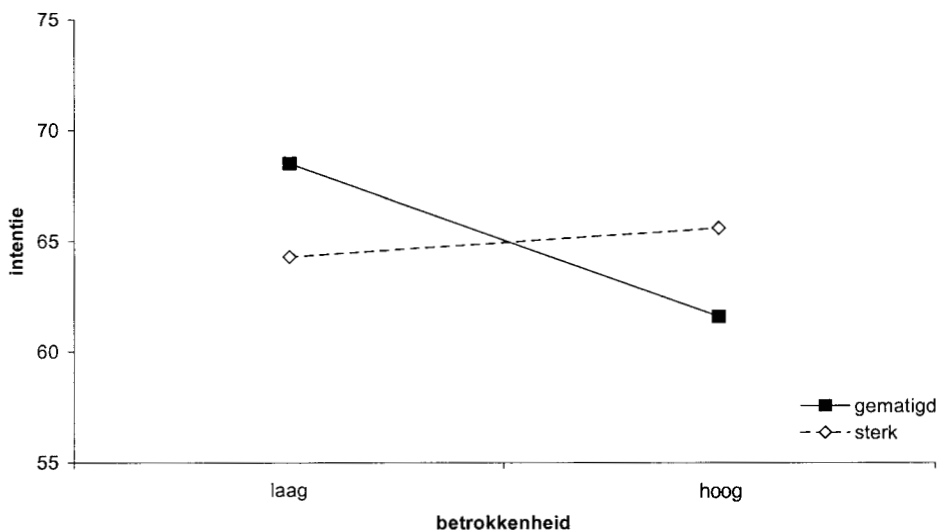
Hypothesen 2, 3a en 3b werden verder getoetst in aparte ANOVAs op de data voor respectievelijk Hollandia en Spania tomaten. Deze analyses waren gelijk aan de hierboven beschreven ANOVAs. Hypothese 2 werd bekeken in een ANOVA voor Spania tomaten, met aankoopintentie als afhankelijke variabele. Er werden geen hoofdeffecten gevonden van betrokkenheid en claimsterkte ($F(1,292) < 1$), en eveneens geen significante interactie tussen deze twee variabelen ($F(1,292) = 2.20, p = .14, \eta^2 = .007$). Kortom, Hypothese 2 werd niet bevestigd door de data, hoewel er voor de Spaanse tomaten geadverteerd met gematigde claims wel sprake is van een interessant (en significant) negatief verschil tussen de evaluatie onder lage en hoge betrokkenheid effect (zie tabel 2).

Voor Nederlandse tomaten voorspellen hypothesen 3a en 3b een negatieve interactie tussen betrokkenheid en claimsterkte: verwacht wordt dat gematigd positieve claims

HOLLANDIA TOMATEN



SPANIA TOMATEN



Figuur 1. Aankoopintenties voor Hollandia en Spania tomaten, geadverteerd met gematigd en sterk positieve claims, onder condities van lage en hoge betrokkenheid.

positiever beoordeeld worden naarmate de betrokkenheid toeneemt, en dat dit niet het geval is voor sterk positieve claims. Een ANOVA voor de data over Hollandia tomaten, met intentie als afhankelijke variabele, liet inderdaad zien dat er een significante interactie was tussen deze twee variabelen ($F(1,356) = 10.22, p = .002, \eta^2 = .028$). De richting van de interactie is af te leiden uit het bovenste deel van figuur 1. Betrokkenheid had een positief effect op de aankoopintentie voor Hollandia tomaten geadverteerd met gematigd positieve claims ($F(1,180) = 15.48, p < .001, \eta^2 = .079$), en geen significant effect op de aankoopintentie voor Hollandia tomaten geadverteerd met sterk positieve claims ($F(1,175) = .66, p = .42, \eta^2 = .004$), deze resultaten komen overeen met hypothesen 3a en 3b. Er was geen hoofdeffect van claimsterkte voor Hollandia tomaten ($F(1,356) < 1$), en een marginaal hoofdeffect van betrokkenheid ($F(1,356) = 3.16, p = .076, \eta^2 = .009$).

6. DISCUSSIE EN IMPLICATIES VOOR DE PRAKTIJK

Onderzoek naar de invloed van land van herkomst op productevaluaties wordt vaak gekenmerkt door een beperkte externe validiteit, en het toepassen van onderzoeksontwerpen die tot gevolg kunnen hebben dat het effect “kunstmatig” versterkt wordt (zie bijvoorbeeld Verlegh en Steenkamp 1999). In de huidige studie wordt (overigens niet voor het eerst) aan deze kritiek tegemoet gekomen: De invloed van land van herkomst wordt onderzocht in een praktisch relevante situatie, waarbij de consument een advertentie met een grote hoeveelheid aanvullende informatie (naast het herkomstland) getoond wordt, en een vergelijking gemaakt wordt tussen oordelen over twee concurrerende herkomstlanden met elk een grote naamsbekendheid. De resultaten van het onderzoek onder een groot aantal Duitse consumenten laten duidelijk zien dat land van herkomst een belangrijke invloed heeft op de evaluatie van producten door de consument.

Wellicht het belangrijkste resultaat in deze studie is echter dat land van herkomst naast een direct effect op productevaluaties, ook een indirect effect op evaluaties kan hebben. De variabele “land van herkomst” geeft hierbij een indicatie van de geloofwaardigheid van andere informatie in een advertentie. Op deze wijze heeft land van herkomst een modererende invloed op de uitwerking van claims die gemaakt worden in een advertentie. Met andere woorden, land van herkomst bepaalt op welke wijze een productclaim invloed heeft op de productevaluatie door de consument. Bij een land van herkomst met een positieve reputatie (in een bepaalde productcategorie) kunnen zowel gematigd positieve als sterk positieve claims een positief effect hebben op de productevaluatie. Wij veronderstelden dat deze impact zou toenemen met de hoeveelheid aandacht die de consument besteedt aan de advertentie, maar deze veronderstelling werd niet bevestigd door de data. Dit kan mogelijk verklaard worden doordat de gematigd positieve claim wellicht onvoldoende overtuigend was. Dat wil zeggen: deze claim is wellicht zo bescheiden dat zij niet goed past bij het ijzersterke imago van Spaanse tomaten in Duitsland. Rossiter en Percy (1997, p. 252) waarschuwen voor het gevaar van te zwakke claims, en stellen dat de “onderhandse boemerang” (onderclaimen) net zo desastreuus kan zijn als de “overhandse boemerang” (overclaimen). Deze studie levert een eerste empirisch bewijs voor deze stelling. Bij een land met een minder positieve reputatie is onderclaimen geen enkel probleem. Onze studie laat zien dat de invloed van gematigde claims voor producten uit een dergelijk land toeneemt naarmate de consument meer

aandacht aan de claim besteedt. Bij sterk positieve claims vinden we een negatief effect van betrokkenheid, dat echter niet significant is. Kortom: de resultaten suggereren dat overclaimen niet effectief is, maar ook niet schadelijk. Het is echter mogelijk dat bij een grotere betrokkenheid, of herhaaldelijke blootstelling aan een advertentie, de geloofwaardigheid van sterk positieve claims verder afneemt, waardoor de claims hun effect kunnen verliezen.

Voor de marketing van Nederlandse land- en tuinbouwproducten suggereert het resultaat op het eerste gezicht wellicht dat het effectief is om met behulp van sterk positieve claims te proberen het imago bij de Duitse consument te verbeteren. Er wordt in dit onderzoek immers nooit een significant negatief verschil gevonden tussen sterk positieve en matig positieve claims. Het is echter belangrijk om op te merken dat er voor dit onderzoek gebruik is gemaakt van een kunstmatige manipulatie van de mate van betrokkenheid van de consument. Dat wil zeggen dat het niet mogelijk is om de niveaus “laag” en “hoog” uit dit onderzoek te koppelen aan de hoeveelheid aandacht die de consument in werkelijkheid zal besteden aan advertenties voor Nederlandse producten. In het licht hiervan is het in de praktijk dan ook verstandiger om verschillende claims (variërend in sterkte) te toetsen in een opzet waarbij de consument op een natuurlijke wijze geconfronteerd wordt met de advertenties.

De resultaten van dit onderzoek laten zien dat het land van herkomst een complexe invloed kan hebben op de evaluatie van producten, en tevens op de verwerking van advertenties door de consument. Een soortgelijke invloed kan ook verwacht worden voor merken en organisaties, waar de reputatie van het bedrijf of merk van invloed kan zijn op de beoordeling van de claims die gemaakt worden in een advertentie.

Het onderzoek is bijzonder, omdat het een experimentele opzet combineert met dataverzameling onder een grote groep van consumenten. Een dergelijke aanpak is interessant vanuit het oogpunt van externe validiteit (zie ook Winer 1999), maar heeft als nadeel dat de resultaten maar moeilijk kunnen worden vergeleken met onderzoek in psychologische laboratoria, waar onder streng gecontroleerde omstandigheden een effect zeer precies bepaald kan worden, en het onderliggende proces nauwkeurig bekeken kan worden. De huidige onderzoeksopzet kan worden beschouwd als een interessante en waardevolle middenweg tussen de realiteit van de veldstudie en de controle van het laboratorium. Deze opzet leidt tot richtlijnen voor de marketingpraktijk die gebaseerd zijn op wetenschappelijke theorie, getoetst in een representatieve steekproef van een relevante populatie.

LITERATUUR

- Anderson, N.H., 1971. Integration theory and attitude change. *Psychological Review*, 78, 171-206.
- Anderson, N.H., 1981. *Foundations of Information Integration Theory*, New York: AP.
- Birnbaum, M.H., en S.E. Stegner, 1979. Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, 37, 48-74.
- Brown, T.J., en P.A. Dacin, 1997. The company and the product: Corporate associations and consumer product responses, *Journal of Marketing*, 61, 68-84.
- Chaiken, S., A. Liberman en A.H. Eagly, 1989. Heuristic and systematic information processing within and beyond the persuasion context. In: J.S. Uleman en J.A. Bargh (eds.) *Unintended Thought*. New York: Guilford, pp. 212-252.
- Fern, E.F. en K.B. Monroe, 1996. Effect-size estimates: issues and problems in interpretation.

Journal of Consumer Research, 23, 89-105.

GfK, 1996. *Image der Holländischen Agrarwirtschaft*. Onderzoeksrapport GfK Nederland, Dongen.

GfK, 1998. *Holland Imago Monitor 98*. Onderzoeksrapport GfK Nederland, Dongen.

Goldberg, M.E. en J. Hartwick, 1990. The effects of advertiser reputation and extremity of advertising claim on advertising effectiveness. *Journal of Consumer Research*, 17, 172-179.

Hong, S.-T. en R.S. Wyer jr., 1989. Effects of country of origin and product-attribute information on product evaluation: An information processing perspective. *Journal of Consumer Research*, 16, 175-187.

Hong, S.-T. en R.S. Wyer jr., 1990. Determinants of product evaluation: Effects of the time interval between knowledge of a product's country of origin and information about its specific attributes. *Journal of Consumer Research*, 17, 277-288.

Johar, G.V., K. Jedidi, en J. Jacoby, 1997. A varying-parameter averaging model of on-line brand evaluations. *Journal of Consumer Research*, 24, 232-247.

Leclerc, F., B.H. Schmitt en L. Dubé, 1994. Foreign branding and its effects on product perceptions and attitudes. *Journal of Marketing Research*, 31, 263-271.

Maheswaran, D., 1994. Country of origin as a stereotype: Effects of consumer expertise and attribute strength on product evaluations. *Journal of Consumer Research*, 21, 354-365.

Papadopoulos, N. en L.A. Heslop (eds.), 1993. *Product-Country Images: Impact and Role in International Marketing*, New York: Haworth Press.

Petty, R.E. en J.T. Cacioppo, 1986. *Communication and Persuasion*. New York: Springer

Rossiter, J.R. en L. Percy, 1997. *Advertising Communications and Promotion Management*. 2nd edition Oxford (UK): Blackwell

Smidts, A., J.R. Rossiter en M. McOmish, 1998. De presentator in reclame: Een test van het Vis-CAP-model. In: Bronner, A.E. et al. (Eds.), *Recente ontwikkelingen in het marktonderzoek, Jaarboek Nederlandse Vereniging voor Marktonderzoek en Informatiemangement 1998* (p. 37-54). De Vrieseborch, Haarlem.

Verlegh, P.W.J (2001), *Country-of-origin effects on consumer product evaluations*. Proefschrift, Wageningen Universiteit.

Verlegh, P.W.J., en J.B.E.M. Steenkamp (1999), A review and meta-analysis of country-of-origin research, *Journal of Economic Psychology*, 20, 521-546.

Winer, R.S. (1999), Experimentation in the 21st century: the importance of external validity, *Journal of the Academy of Marketing Science*, 27, 349-358.

Bijlage: Advertentie gebruikt in het experiment

Hollandia – sterk positief



Tomaten aus Holland heissen von nun ab **Hollandia Tomaten**. Beim Anbau von Hollandia Tomaten wird mehr als in jedem anderen Land auf Natur und Umwelt geachtet. Mit Hilfe von natürlichen Mitteln sorgt der holländische Bauer für ein Qualitätsprodukt ohnegleichen. Hollandia Tomaten haben einen unübertroffen vollen Geschmack und ein reiches Aroma. Forschung weist auf, dass 9 von 10 Konsumenten Hollandia Tomaten viel schmackhafter finden als andere Tomaten.

Hollandia Tomaten:
Schmecken so gut wie Tomaten aus eigenem Garten



Spania – sterk positief

Tomaten aus Spanien heissen von nun ab **Spania Tomaten**. Beim Anbau von Spania Tomaten wird mehr als in jedem anderen Land auf Natur und Umwelt geachtet. Mit Hilfe von natürlichen Mitteln sorgt der spanische Bauer für ein Qualitätsprodukt ohnegleichen. Spania Tomaten haben einen unübertroffen vollen Geschmack und ein reiches Aroma. Forschung weist auf, dass 9 von 10 Konsumenten Spania Tomaten viel schmackhafter finden als andere Tomaten.

Spania Tomaten:
Schmecken so gut wie Tomaten aus eigenem Garten



Hollandia - matig positief

Tomaten aus Holland heissen von nun ab **Hollandia Tomaten**. Beim Anbau von Hollandia Tomaten wird auf Natur und Umwelt geachtet. Mit Hilfe von natürlichen Mitteln sorgt der niederländische Bauer für ein Qualitätsprodukt. Hollandia Tomaten haben einen guten Geschmack und Aroma. Forschung weist auf, dass 6 von 10 Konsumenten Hollandia Tomaten schmackhafter finden als andere Tomaten.

Hollandia Tomaten:
Schmecken beinah so gut wie Tomaten aus eigenem Garten



Spania – matig positief

Tomaten aus Spanien heissen von nun ab **Spania Tomaten**. Beim Anbau von Spania Tomaten wird auf Natur und Umwelt geachtet. Mit Hilfe von natürlichen Mitteln sorgt der spanische Bauer für ein Qualitätsprodukt. Spabia Tomaten haben einen guten Geschmack und Aroma. Forschung weist auf, dass 6 von 10 Konsumenten Spania Tomaten schmackhafter finden als andere Tomaten.

Spania Tomaten:
Schmecken beinah so gut wie Tomaten aus eigenem Garten



13. *Verschillen stadswijken in onderdekking en nonrespons?*

Een verkennend onderzoek naar selectieve uitval in een telefonische enquête

F. JANSMA, H. VAN GOOR en R. VEENSTRA*

SAMENVATTING

Wij onderzochten twee verklaringen voor verschillen in onderdekking en nonrespons tussen woonwijken in een telefonische enquête onder de inwoners van de gemeente Groningen. 1. Zijn deze verschillen het gevolg van verschillen in bevolkingssamenstelling tussen woonwijken? Om deze vraag te beantwoorden maakten we gebruik van acht sociaal-demografische kenmerken van individuen, zoals leeftijd en geslacht. 2. Zijn deze verschillen het gevolg van verschillen in sociale desorganisatie tussen woonwijken? Sociale desorganisatie werd door ons gemeten aan de hand van een tiental wijkenmerken. Om deze verklaringen te toetsen pasten we een logistische multi-niveau-analysetechniek toe op een steekproef van 7000 personen en alle 14 woonwijken. Uit ons onderzoek blijkt dat de verschillen in onderdekking en nonrespons grotendeels verklaard kunnen worden door verschillen in bevolkingssamenstelling tussen woonwijken. Sociale desorganisatie had alleen een klein effect in geval van onderdekking: mensen in woonbuurten die fysiek en sociaal in verval zijn, hebben vaker geen telefoon of een geheim nummer dan mensen in andere woonbuurten.

1. INLEIDING

Deelname aan enquêtes varieert naar geografisch gebied. Zo bestaan er verschillen in nonrespons tussen landen, tussen stad en platteland en tussen stadswijken. In Nederland is de nonrespons vooral hoog in sterk verstedelijkte gebieden, met de drie grootste steden (Den Haag, Rotterdam en Amsterdam) als treurig hoogte- of zo men wil dieptepunt. Overigens doen in de gehele Randstad mensen minder vaak mee aan onderzoek dan in de rest van Nederland. Ook binnen steden treedt variatie op: in wijken met hoogbouw en met lage huren bijvoorbeeld is de deelname aan enquêtes geringer dan elders (Maas en De Heer, 1995; Burhenne en Van der Leest, 1997).

* De data zijn ons welwillend ter beschikking gesteld door het Bureau Onderzoek van de gemeente Groningen. In het bijzonder danken wij Arnold Veenkamp van dit bureau voor zijn hulp bij de samenstelling van het databestand. Voorts danken wij Annemiek van Goor voor enkele nuttige methodologische opmerkingen en Tom Snijders voor zijn adviezen over de gebruikte methode.

Wat de oorzaken zijn van de verschillen in nonrespons tussen geografische gebieden is nauwelijks onderzocht. Eén van de weinige studies op dit gebied is een onderzoek van Goyder et al. (1992) naar sociale desorganisatie als mogelijke verklaring voor verschillen in nonrespons (weigeringen) tussen en binnen drie steden in Canada. Verrassend genoeg bleken in hun onderzoek verschillen in respons tussen stadswijken niet samen te hangen met verschillen in sociale desorganisatie. Wel vonden zij verschillen tussen steden: bij toenemend inwonertal neemt de nonrespons toe.

Een belangrijke beperking van de studie van Goyder et al. (1992) is dat de auteurs slechts van één analyseniveau (het individuele niveau) gebruik maken en dat kenmerken van de context als individuele kenmerken worden opgevat. Gelet op het feit dat zij zowel effecten op individueel als op boven-individueel niveau willen aantonen, zou een multiniveau-analyse passender zijn geweest.

Een andere – meer algemene – beperking van onderzoek op dit terrein is dat men zich uitsluitend richt op nonrespons en een tweede belangrijke bron van selectieve uitval – onderdekking – negeert. Er zijn echter aanwijzingen dat ook onderdekking systematisch varieert naar geografisch gebied. Zo blijkt onderdekking, net als nonrespons, toe te nemen met urbanisatiegraad (Lavrakas, 1993: 34). Onderdekking en nonrespons vormen dus beide een bedreiging voor de representativiteit van de uiteindelijk gerealiseerde steekproef.

In aansluiting op het voorgaande rijst de vraag op welke wijze we voor deze geografische verschillen kunnen corrigeren. Kunnen onderzoekers afdoende voor deze verschillen corrigeren met behulp van individuele kenmerken van de onderzochten? Of dienen ze ook rekening te houden met kenmerken van de onderzochte geografische gebieden? Om deze vragen te kunnen beantwoorden is nader onderzoek nodig hoe verschillen tussen gebieden kunnen worden verklaard. Ons onderzoek is een eerste stap in deze richting. In dit artikel doen wij verslag van een verkennend onderzoek naar verschillen in onderdekking en nonrespons tussen wijken in een telefonische enquête onder de bewoners van de stad Groningen. Twee mogelijke verklaringen voor deze verschillen zijn door ons onderzocht.

1. Zijn verschillen in onderdekking en nonrespons tussen wijken het gevolg van verschillen in *bevolkingssamenstelling* tussen wijken? Anders gezegd: hangen de verschillen tussen wijken samen met kenmerken van individuen waarvan bekend is dat ze met onderdekking en nonrespons samenhangen? Is dat het geval, dan is er sprake van een compositie-effect: verschillen in bevolkingssamenstelling zijn dan verantwoordelijk voor verschillen in onderdekking en nonrespons tussen wijken. Om dit na te gaan maken we gebruik van een aantal sociaal-demografische kenmerken van individuen.

2. Zijn verschillen in onderdekking en nonrespons tussen wijken het gevolg van verschillen in *wijkenmerken*? In navolging van Goyder et al. (1992) gaan we na of verschillen in sociale desorganisatie tussen wijken samenhangen met verschillen in onderdekking en nonrespons.

2. LITERATUUR

2.1. Onderdekking en nonrespons op individueel niveau

Onderdekking bestaat uit de personen die buiten het steekproefkader vallen. In een telefonische enquête gebruiken onderzoekers gewoonlijk het telefoonboek als steekproef-

kader. De onderdekking bestaat dan uit individuen of huishoudens die wel tot de doelpopulatie behoren, maar geen telefoonaansluiting hebben, of een geheim nummer, of die KPN Telecom opdracht hebben gegeven hun nummer niet aan derden te verstrekken ten behoeve van telemarketing- en directmarketingdoeleinden.

Onderzoek naar onderdekking is betrekkelijk schaars. Uit het weinige onderzoek dat er is, komt sociaal-economische status als belangrijke factor naar voren. Personen met een lage sociaal-economische status hebben relatief vaak geen telefoonaansluiting of een geheim nummer. Voorts blijkt dat mensen die maatschappelijk minder sterk zijn geïntegreerd, eveneens vaker tot de onderdekkingscategorie behoren. Zo hebben zowel jongeren als ouderen minder vaak een bekend telefoonnummer, evenals alleenwonenden en gescheiden personen. Hetzelfde zien we bij mensen die werkloos zijn of een uitkering hebben en bij allochtonen (Cornelisse-Claassen 1987; Smith 1990; Engbersen 1990).

Nonrespons bestaat uit personen die niet bereikt worden of die medewerking aan onderzoek weigeren. Uit onderzoek naar de samenhang tussen individuele kenmerken en nonrespons blijkt dat sociaal-economische status en maatschappelijke integratie belangrijke verklarende factoren zijn. Mensen met een hogere sociaal-economische status of een hogere opleiding responderen vaker (Goyder, 1987; Hartmann en Schimpl-Neimanns, 1992). Dat maatschappelijke integratie van invloed is op nonrespons blijkt uit het onderzoek van Goyder (1987): gehuwden en werkenden doen vaker mee aan enquêtes, terwijl ouderen juist minder vaak meedoen.

Uit de hierboven beschreven onderzoeken blijkt dat zowel sociaal-economische status als maatschappelijke integratie van invloed zijn op onderdekking en nonrespons. In hoeverre geldt dit ook voor ons onderzoek? In een analyse van onderdekking en nonrespons op individueel niveau in het ook door ons gebruikte databestand vonden Rispen en Van Goor (2000) in lijn met bovenstaande uitkomsten statistisch significante effecten voor een aantal sociaal-demografische kenmerken, onder andere voor leeftijd, land van herkomst en type huishouden. Om vast te stellen of verschillen tussen wijken in onderdekking en nonrespons het gevolg zijn van compositie-effecten, zullen wij gebruik maken van dezelfde sociaal-demografische kenmerken.

2.2. Sociale desorganisatie als oorzaak van onderdekking en nonrespons op wijkniveau

De sociale-desorganisatietheorie is voor het eerst geformuleerd door Shaw en McKay (1942) in hun onderzoek naar jeugdcriminaliteit. Volgens deze theorie leiden een hoge residentiële mobiliteit, een lage sociaal-economische status en een grote etnische heterogeniteit in een stadswijk tot een afname in sociale cohesie. Deze afname in sociale cohesie leidt ertoe dat de wijkbewoners niet in staat zijn een voor de wijk algemeen geldend normen- en waardenstelsel te creëren en de benodigde sociale controle uit te oefenen ter handhaving van een dergelijk stelsel. Zo ontstaat er een voedingsbodem voor criminaliteit. Deze criminaliteit wordt voor een belangrijk deel veroorzaakt door groepen tieners die zonder toezicht van ouders op straat rondhangen.

In later onderzoek is herhaaldelijk een samenhang tussen sociaal-structurele buurtkenmerken en criminaliteit aangetoond. De veronderstelde samenhang met sociale desorganisatie wordt eveneens duidelijk bevestigd: lokale relatienetwerken, binding aan de buurt en informele sociale controle nemen in omvang of betekenis af onder invloed van factoren als residentiële mobiliteit en een lage sociaal-economische status (Kasarda en Janowitz, 1974; Sampson en Groves, 1988/89; Sampson en Raudenbush, 1999/2000).

Daarnaast blijkt dat sociaal-structurele buurtkenmerken en de daaruit voortvloeiende sociale desorganisatie en criminaliteit ook van invloed zijn op houding en gedrag van individuele buurtbewoners: ze leiden tot gevoelens van onveiligheid, gedragsbeperkingen (bijvoorbeeld 's avonds niet alleen over straat durven in de eigen buurt), gevoelens van machteloosheid en een gebrek aan zelfvertrouwen (Taylor en Covington, 1993; Taylor, 1996; Boardman en Robert, 2000).

Een interessante aanvulling op het onderzoek naar sociale desorganisatie verschaft ons het onderzoek naar de effecten van urbanisatiegraad op hulpvaardigheid. Een van de onderzochte vormen van hulpvaardigheid is deelname aan enquêtes. Nonrespons (met name wegeringen) vat men op als een indicator voor niet-hulpvaardig, wantrouwend gedrag tegenover onbekenden. Uit onderzoek komt een sterke samenhang naar voren tussen hulpvaardigheid (waaronder deelname aan enquêtes) en urbanisatiegraad: stedelingen zijn minder hulpvaardig dan plattelanders. Merkwaardig genoeg blijkt er niet een duidelijk lineair verband te bestaan tussen inwonertal en hulpvaardigheid (Stebly, 1987). Deze schijnbaar tegenstrijdige uitkomsten laten zich evenwel eenvoudig verklaren als rekening wordt gehouden met andere omgevingskenmerken. Zo tonen House en Wolf (1978) aan dat verschillen in nonrespons (wegeringen) niet samenhangen met inwonertal of urbanisatiegraad van iemands woonplaats, maar met verschillen in criminaliteit. Zij stellen dat deelname aan enquêtes samenhangt met de sociale organisatie van iemands woonplaats. Deze bevinding sluit goed aan bij de sociale-desorganisatietheorie.

Op basis van de besproken studies kan nu worden aangegeven waarom sociale desorganisatie van de wijk (door ons gemeten aan de hand van sociaal-structurele buurtkenmerken en criminaliteit) van invloed is op onderdekking en nonrespons. Het gaat dus om de vraag welke mechanismen het effect van sociale desorganisatie op onderdekking en nonrespons vermoedelijk bepalen. Wij kunnen deze mechanismen in ons onderzoek evenwel niet rechtstreeks toetsen, aangezien wij alleen weten of iemand tot de onderdekking, respectievelijk de nonrespons behoort of niet.

Om te beginnen is het mogelijk dat sociaal isolement als gevolg van sociale desorganisatie van de buurt er toe leidt dat mensen minder vaak een telefoonaansluiting hebben, maar wij verwachten dat sociale desorganisatie vooral leidt tot meer geheime telefoonnummers. Sociale desorganisatie en criminaliteit leiden immers tot gevoelens van onveiligheid, machteloosheid en wantrouwen. Geheime nummers nu bieden de mogelijkheid om zich van de buitenwereld af te schermen, zowel tegenover onverschillige, vijandige en gewantrouwde medebuurtbewoners, als tegenover buitenstaanders, waartoe ook officiële instanties gerekend kunnen worden (vergelijk Engbersen, 1990).

Vervolgens zouden in geval van nonrespons sociale desorganisatie en daarmee samenhangende gevoelens er toe kunnen leiden dat mensen minder vaak de telefoon opnemen. Zo is het denkbaar dat kwetsbare bevolkingscategorieën (alleenstaande vrouwen, ouderen) 's avonds als gevolg van onveiligheidsgevoelens de telefoon niet beantwoorden, of alleen opnemen als men weet, bijvoorbeeld aan de hand van een nummERMelder, dat een familielid of bekende belt. Het grootste effect verwachten wij evenwel op wegeringen: onveiligheidsgevoelens en wantrouwen, veroorzaakt door sociale desorganisatie en criminaliteit in de buurt, zullen er toe leiden dat mensen eerder weigeren een onbekende persoon die namens een officiële instantie belt, te woord te staan.

Wij verwachten dus dat naast eventuele compositie-effecten buurtkenmerken (sociaal-structurele antecedenten van sociale desorganisatie en criminaliteit) een zelfstandig effect zullen hebben op onderdekking en nonrespons.

3. DATA EN ANALYSEMETHODE

3.1. Data

De gegevens voor dit artikel zijn ontleend aan een telefonische enquête in 1996 onder inwoners van zestien jaar en ouder van de gemeente Groningen. De steekproef bestond uit 7.000 personen (namen en adressen) en is getrokken uit de Gemeentelijke Basisadministratie. Vervolgens zijn door KPN Telecom bij de adresgegevens de telefoonnummers gezocht. Als gevolg van onderdekking is 37 procent van de oorspronkelijke steekproef uitgevallen. Vervolgens kon van de personen met een bekend telefoonnummer 49 procent niet worden geïnterviewd. In totaal is door onderdekking en nonrespons samen 68 procent van de oorspronkelijke steekproef uitgevallen (zie verder Rispens en Van Goor, 2000).

De geanonimiseerde dataset bevat gegevens over de oorspronkelijke steekproef van 7.000 personen. Bekend is of iemand al dan niet behoort tot de onderdekking (geen telefoon, geheim nummer, nummer door KPN niet aan derden verstrekt), respectievelijk de nonrespons (telefoon niet opgenomen of weigering). Het gemeentelijke Bureau Onderzoek heeft sociaal-demografische gegevens uit gemeentelijke databestanden gekoppeld aan alle 7.000 personen. Deze gegevens zijn door ons gebruikt om de vertekeningen op individueel niveau als gevolg van onderdekking en nonrespons vast te stellen.

De indeling in 14 wijken die we hanteren in onze analyses, is gelijk aan de indeling zoals deze gehanteerd wordt door de gemeente Groningen (*Gronings geld*, 1997). De door ons gebruikte gegevens op wijkniveau, onder meer over inkomens, typen huishoudens en criminaliteit, zijn afkomstig uit gemeentelijke databestanden die ons door Bureau Onderzoek ter beschikking zijn gesteld.

3.2. Kenmerken

Op individueel niveau beschikken we over acht sociaal-demografische kenmerken. 1. *Geslacht*: 0 = vrouw, 1 = man. 2. *Burgerlijke staat* is geoperationaliseerd door middel van drie dummy-variabelen: verweduwd, gescheiden en ongehuwd. Gehuwd is als referentiecategorie gebruikt. 3. *Leeftijd* is een continue variabele en bestaat uit leeftijd in jaren. 4. *Land van herkomst* bestaat uit zes categorieën: Nederland; Westerse landen (West Europa exclusief Noord-Mediterrane landen, Noord-Amerika, Japan, Israël, Australië en Nieuw Zeeland); Indonesië (inclusief Molukken en Nieuw-Guinea); Suriname/Antillen (Suriname, Aruba en Nederlandse Antillen); Turkije/Noord-Afrika (Turkije, Marokko, Algerije, Tunesië, Libië en Egypte); Overige landen. De zes categorieën worden weergegeven door vijf dummy-variabelen. De categorie Nederland wordt als referentiecategorie gebruikt. 5. *Werkloosheid*: 0 = niet werkloos, 1 = werkloos. 6. *Uitkering van de Gemeentelijke Sociale Dienst*: 0 = geen uitkering, 1 = wel uitkering. 7. *Huishoudenomvang* is het aantal personen waaruit het huishouden bestaat. 8. *Huishoudentype* is geoperationaliseerd door middel van vier dummy-variabelen: éénoudergezin, standaardgezin, meer verzorgenden en alleenwonend. Samenwonend is als referentiecategorie gebruikt.

Voor de operationalisatie van sociale desorganisatie op wijkniveau gebruiken we elf indicatoren: de drie klassieke indicatoren van Shaw en McKay, zes wijkkenmerken en twee criminaliteitsvariabelen. 1. *Sociaal-economische status*. Voor de sociaal-economi-

sche status van een wijk beschikken we over tien inkomengerelateerde variabelen, zoals gemiddeld inkomen per huishouden en percentage huishoudens met huursubsidie. Factoranalyse, uitgevoerd op deze variabelen, levert twee factoren op (zie uitvoeriger Jansma, 2000). Aangezien de tweede factor alleen bestaat uit het percentage huishoudens met een minimuminkomen, laten we deze vervallen. De eerste factor verklaart 73,8 procent van de variantie. Voor de negen variabelen zijn Z-scores berekend die vervolgens zijn samengevoegd tot een maat voor de sociaal-economische status van de wijk. 2. *Residentiële stabiliteit*: het percentage van de wijkbevolking dat langer dan zes jaar woonachtig is op het huidige adres. 3 en 4. *Etnische heterogeniteit*: het percentage van de wijkbevolking dat niet-Nederlands is en het percentage van de wijkbevolking dat niet-Westers is. Niet-Nederlands = Westerse landen en Indonesië; niet-Westers = Suriname/Antillen, Turkije/Noord-Afrika, Overige landen. 5. *Het aantal oude woningen in de wijk*: het percentage van het totaal aantal woningen in de wijk dat gebouwd is voor 1945. 6. *Leegstand in de wijk*: het percentage van het totaal aantal woningen in de wijk dat langer dan één jaar onbewoond is. 7. *Bewoningsdichtheid*: de gemiddelde omvang van het huishouden, ofwel het gemiddelde aantal personen in een huishouden per wijk. 8. *Het voorkomen van jeugdgroepen*: het percentage van de wijkbevolking dat tussen de 10 en 20 jaar oud is. 9. *Gezinsontwrichting*: gezinsontwrichting is geoperationaliseerd aan de hand van drie variabelen: percentage éénoudergezinnen op alle huishoudens met kinderen in de wijk; percentage éénpersoonshuishoudens op het totaal aantal huishoudens in de wijk; het percentage gescheiden personen op het aantal gescheiden en gehuwde personen in de wijk. Deze drie variabelen komen bij factoranalyse in één factor terecht, die 84,6 procent van de variantie vastlegt. Ze zijn daarom samengevoegd tot één maat voor gezinsontwrichting. 10 en 11. *Criminaliteit*: criminaliteit is geoperationaliseerd aan de hand van twee variabelen: het percentage verdachten in de wijk in 1994-96 en het percentage delicten in de wijk in dezelfde periode. Meegeteld zijn verdachten, respectievelijk delicten met betrekking tot verstoring van de openbare orde, geweld, vermogensdelicten en vandalisme.

De wijkvariabelen zijn door ons op twee manieren gebruikt. Ten eerste hebben we zelf, op theoretische gronden, de variabelen op basis van hun Z-scores samengevoegd tot één variabele die als totaalmaat voor de sociale desorganisatie van de wijk fungeert. Dit samenvoegen is gedaan op basis van de verwachte samenhang van deze wijkenmerken met sociale desorganisatie.

Ten tweede hebben we een factoranalyse uitgevoerd op de wijkvariabelen om te kijken of onze totaalmaat in een aantal dimensies kan worden uiteengelegd. Op deze manier kunnen we nagaan of wellicht een specifiek aspect van sociale desorganisatie een centrale plaats inneemt bij het verklaren van onderdekking en nonrespons. De factoranalyse levert vier factoren op. De eerste factor verwijst naar de mate waarin een wijk gekenmerkt wordt door een *stabiele middenklasse-bevolking*. Na-oorlogse wijken met gezinnen met opgroeiende kinderen en met relatief weinig verloop in de wijkbevolking scoren hier hoog op. De tweede factor meet de mate van *sociale achterstand van een wijk*, zowel in sociaal-economisch opzicht als in termen van sociale uitsluiting. Hoog scoren wijken met een laag gemiddeld inkomen, met relatief veel alleenstaanden, eenoudergezinnen, allochtonen en personen verdacht van criminele activiteiten. De derde factor verwijst naar een algehele achteruitgang en *verval van de buurt in fysiek en sociaal opzicht*. In deze factor komen fysieke verloedering van de woonomgeving en delicten, waarvan vooral buurtbewoners het slachtoffer zullen zijn, samen. De vierde factor ten slotte geeft de mate van *mobilititeit van de wijkbevolking* weer.

3.3. Analysemethode

Voor de analyses maken we gebruik van multiniveau-analyse (Snijders en Bosker, 1999). Met deze methode kan men variabelen op zowel individueel niveau als op ‘hoger’ niveau onderzoeken: in ons geval individuen en stadswijken. Aangezien we in ons onderzoek te maken hebben met discrete (dichotome) afhankelijke variabelen – wel of niet een bekend telefoonnummer, respectievelijk wel of niet geïnterviewd –, maken we gebruik van de logistische variant van de multiniveau-analyse.

We toetsen aan de hand van deze logistische multiniveautechniek, voor zowel onderdekking als nonrespons, vier verschillende modellen. Model 0 bestaat uit het zogeheten ‘lege’ model. Aan de hand van dit model is te zien welk deel van de variantie zich bevindt op individueel niveau en welk deel op wijkniveau. In model 1 zijn de verklarende variabelen op individueel niveau opgenomen. Model 2 bestaat uit de individuele kenmerken en onze totaalmaat voor sociale desorganisatie in de wijk. In model 3 zijn de vier dimensies van sociale desorganisatie opgenomen die de factoranalyse hebben opgeleverd. Zo kunnen we nagaan welke van de dimensies het sterkste effect heeft op onderdekking, respectievelijk nonrespons. De dimensies worden afzonderlijk getoetst, zowel vanwege het kleine aantal wijken, als vanwege mogelijke multicollineariteitsproblemen.

4. RESULTATEN

4.1. De wijken in cijfers

In hoeverre verschillen de wijken onderling in onderdekking en nonrespons? Het percentage personen met een bekend telefoonnummer varieert aanzienlijk. In Nieuw-Oost heeft 78 procent van de wijkbevolking een bekend telefoonnummer tegen 50 procent in het Centrum. Het percentage geïnterviewde personen verschilt wat minder sterk. Het percentage geïnterviewden varieert tussen 45 procent (Korreweg/De Hoogte) en 61 procent (Nieuw-Oost). Onderdekking en nonrespons samen leiden tot een aanzienlijke uitval. Het uiteindelijke reponspercentage is het hoogst in Nieuw-Oost (47 procent), terwijl het minimum wordt behaald in het Centrum met slechts 22 procent.

4.2. Onderdekking

De resultaten van de multiniveau-analyse zijn per model weergegeven in Tabel 1. Het eerste model bestaat alleen uit de constante: het intercept. De variantie op individueel niveau is vastgezet op 3,29 (Snijders en Bosker, 1999: 224-225). De variantie op wijkniveau bedraagt 0,080. Om na te gaan hoe de onverklaarde variantie op individueel en op wijkniveau zich verhouden tot de verklaarde variantie gebruiken we de formule, voorgesteld door Snijders en Bosker (1999):

$$R_o^2 = \tau_o^2 / (\sigma_F^2 + \tau_o^2 + \sigma_R^2)$$

R_o^2 geeft de proportie onverklaarde variantie weer. τ_o^2 is de onverklaarde variantie op macro-niveau (de wijk). σ_R^2 is de onverklaarde variantie op micro-niveau (het individu). De verklaarde variantie σ_F^2 is de variantie van de afhankelijke variabele van het logistische multiniveaumodel, *logit* (P_{ij}).

Toepassing van deze formule voor model 0 toont aan dat in het lege model, zonder ver-

klarende variabelen, 2,4 procent van de totale variantie zich op het wijkniveau bevindt. 97,6 procent bevindt zich dus op het individuele niveau. We kunnen hieruit concluderen dat de te verklaren variantie op wijkniveau gering is.

In model 1 zijn de individuele kenmerken toegevoegd. We vinden statistisch significante negatieve effecten voor geslacht ($t = -2,81$), leeftijd ($t = -7,00$) en het behoren tot een standaardgezin ($t = -2,72$). Dit betekent dat mannen, ouderen en personen die deel uit maken van een standaardgezin, vaker een bekend telefoonnummer hebben. Statistisch significante positieve effecten worden gevonden voor de variabelen verweerd ($t = 1,73$), gescheiden ($t = 6,02$), Westers ($t = 1,77$), Indonesisch ($t = 5,96$), Surinaams/Antilliaans ($t = 6,73$), Turks/Noord-Afrikaans ($t = 4,90$), afkomstig uit overige landen ($t = 4,36$), werkloos ($t = 1,78$), het krijgen van een uitkering van de GSD ($t = 7,54$) en alleenwonend ($t = 3,69$). Personen uit deze categorieën hebben vaker geen telefoon of een geheim nummer. Met name komt het land van herkomst hier als belangrijke bron van vertekening naar voren.

Hoeveel van de variantie wordt nu verklaard door de individuele kenmerken? We zien in Tabel 1 voor model 1 dat de onverklaarde variantie op wijkniveau 0,036 bedraagt. De verklaarde variantie kan worden vastgesteld door allereerst *logit* (P_{ij}) te berekenen aan de hand van het gevonden regressiemodel. De waarde van *logit* (P_{ij}) bedraagt 0,387. Nu kan aan de hand hiervan de proportie verklaarde variantie (R_v^2) van de totale variantie berekend worden. Voor model 1 is dit 0,104: 10,4 procent van de totale variantie wordt verklaard door de individuele variabelen. De proportie onverklaarde variantie op individueel niveau bedraagt 0,886. De onverklaarde variantie op wijkniveau is 0,010. Dit betekent dat in vergelijking met model 0 de individuele kenmerken de onverklaarde variantie op het individuele niveau doen afnemen van 97,6 procent naar 88,6 procent en de onverklaarde variantie op wijkniveau van 2,4 procent naar 1,0 procent. De individuele variabelen leggen dus een groot deel van de proportie onverklaarde variantie vast, zowel op individueel niveau als op wijkniveau. Daarmee blijft er op wijkniveau weinig onverklaarde variantie voor de wijkvariabelen over.

Bij de toetsing van model 2, het model met de totaalmaat voor sociale desorganisatie, zien we dat onze totaalmaat een statistisch significant effect oplevert. De onverklaarde variantie op wijkniveau bedraagt nu 0,006. Wanneer we op dezelfde wijze te werk gaan als bij model 1 blijkt dat 11,4 procent van de totale variantie door model 2 wordt verklaard. 0,6 procent van de totale variantie is onverklaarde variantie op het wijkniveau en 87,9 procent is onverklaarde variantie op individueel niveau. Vergelijking met model 1 leert dat onze totaalmaat voor sociale desorganisatie de onverklaarde variantie op wijkniveau doet afnemen met 0,4 procent. Ook de onverklaarde variantie op individueel niveau daalt en wel met 0,7 procent onder invloed van de totaalmaat voor sociale desorganisatie.

Vervolgens zijn de effecten berekend voor de vier factoren. Zoals in paragraaf 3.4 reeds is opgemerkt, doen we dit voor elke factor afzonderlijk. De effecten voor factor 1: stabiele middenklasse-wijk ($t = -1,14$), factor 2: sociale achterstand ($t = 1,26$) en factor 4: residentiële mobiliteit ($t = 1,49$) zijn statistisch niet significant (resultaten niet weergegeven). De uitkomsten voor factor 3: verval van de wijk, zijn dat wel: $t = 2,38$ (zie Tabel 2, model 3). Mensen die in wijken wonen met veel langdurige leegstand en relatief veel criminaliteit behoren relatief vaker tot de onderdekkingscategorie. Wanneer we op dezelfde wijze als bij model 1 de proportie verklaarde en onverklaarde variantie op individueel en wijkniveau berekenen, vinden we dat model 3, met op wijkniveau alleen factor 3, 10,9 procent van de totale variantie verklaart. De proportie onverklaarde variantie

Tabel 1: Ongestandaardiseerde effectparameters van individuele en wijkenmerken op onderdekking (afhankelijke variabele: het hebben van een bekend telefoonnummer: 0 = ja, 1 = nee).

Variabelen	Model 0		Model 1		Model 2		Model 3	
	B	s.e.	B	s.e.	B	s.e.	B	s.e.
INDIVIDUELE KENMERKEN (N = 7.000)								
Geslacht (0 = vrouw; 1 = man)			-0,149**	0,053	-0,150**	0,053	-0,150**	0,053
Burgerlijke staat (gehuwd = ref.):								
verweduwd (0 = nee; 1 = ja)			0,234*	0,135	0,232*	0,129	0,235*	0,135
gescheiden (0 = nee; 1 = ja)			0,662**	0,110	0,655**	0,110	0,660**	0,110
ongehuwd (0 = nee; 1 = ja)			0,115	0,086	0,105	0,086	0,108	0,086
Leeftijd (in jaren)			-0,014**	0,002	-0,014**	0,002	-0,014**	0,002
Land van herkomst (Nederland = ref.):								
Westerse landen (0 = nee; 1 = ja)			0,227*	0,128	0,228*	0,129	0,227*	0,129
Indonesië (0 = nee; 1 = ja)			0,739**	0,124	0,741**	0,124	0,740**	0,124
Suriname/Antillen (0 = nee; 1 = ja)			1,117**	0,166	1,117**	0,167	1,126**	0,166
Turkije/Noord-Afrika (0 = nee; 1 = ja)			1,790**	0,365	1,785**	0,367	1,793**	0,364
Overige landen (0 = nee; 1 = ja)			0,610**	0,140	0,605**	0,141	0,612**	0,141
Werkloosheid (0 = nee; 1 = ja)			0,151*	0,085	0,148*	0,085	0,153*	0,085
Uitkering GSD (0 = nee; 1 = ja)			0,769**	0,102	0,770**	0,102	0,773**	0,102
Gezinsituatie (samenwonend = ref.):								
Eénoudergezin (0 = nee; 1 = ja)			-0,046	0,153	-0,045	0,153	-0,043	0,153
Standaardgezin (0 = nee; 1 = ja)			-0,400**	0,147	-0,397**	0,147	-0,394**	0,147
Meerdere verzorgenden (0 = nee; 1 = ja)			0,255	0,331	0,253	0,332	0,257	0,332
Alleenwonend (0 = nee; 1 = ja)			0,428**	0,116	0,425**	0,116	0,422**	0,116
Omvang huishouden (aantal)			0,077	0,085	0,080	0,085	0,078	0,085
WIJKKENMERKEN (N = 14)								
Totaalmaat sociale desorganisatie					0,329*	0,141		
Factor 3: verval van de wijk							0,124*	0,052
Constante	-0,568**	0,081	-1,300**	0,252	-1,307**	0,251	-1,305**	0,250
Variantie op individueel niveau	3,29	0,000	3,29	0,000	3,29	0,000	3,29	0,000
Variantie op wijkniveau	0,080*	0,035	0,036*	0,018	0,024	0,013	0,022	0,012

Toelichting: Eénzijdige t-toets; significantieniveau: * $\alpha = 0,05$; ** $\alpha = 0,01$.

op individueel niveau bedraagt 88,5 procent en op wijkniveau 0,6 procent. Vergeleken met model 1 bindt de wijkvariabele voor de mate van neergang en verval van de buurt dus een extra 0,4 procent van de onverklaarde variantie op wijkniveau.

4.3. Nonrespons

Uit de resultaten voor model 0 (zie Tabel 2) blijkt dat de schatting van de onverklaarde variantie op wijkniveau niet statistisch significant is. Dit betekent dat de wijken onder-

Tabel 2: Ongestandaardiseerde effectparameters van individuele en wijkenmerken op non-respons (afhankelijke variabele: wel of niet geïnterviewd: 0 = wel, 1 = niet).

Variabelen	Model 0		Model 1	
	B	s.e	B	s.e.
INDIVIDUELE KENMERKEN (N = 4.394)				
Geslacht (0 = vrouw; 1 = man)			0,157**	0,062
Burgerlijke staat (gehuwd = ref.):				
Verweduwd (0 = nee; 1 = ja)			0,391**	0,150
Gescheiden (0 = nee; 1 = ja)			0,393**	0,145
Ongehuwd (0 = nee; 1 = ja)			0,395**	0,099
Leeftijd (in jaren)			0,004*	0,002
Land van herkomst (Nederland=ref.):				
Westerse landen (0 = nee; 1 = ja)			0,066	0,156
Indonesië (0 = nee; 1 = ja)			-0,109	0,175
Suriname/Antillen (0 = nee; 1 = ja)			0,451*	0,269
Turkije/Noord-Afrika (0 = nee; 1 = ja)			1,709*	0,800
Overige landen (0 = nee; 1 = ja)			1,013**	0,222
Werkloosheid (0 = nee; 1 = ja)			0,071	0,117
Uitkering GSD (0 = nee; 1 = ja)			0,087	0,157
Gezinssituatie (samenwonend = ref.):				
Eénoudergezin (0 = nee; 1 = ja)			0,067	0,190
Standaardgezin (0 = nee; 1 = ja)			-0,070	0,166
Meerdere verzorgers (0 = nee; 1 = ja)			0,529	0,443
Alleenwonend (0 = nee; 1 = ja)			0,210	0,139
Omvang huishouden (aantal personen)			0,072	0,101
Constante	-0,061	0,045	-0,749**	0,307
Variantie op individueel niveau	3,29	0,000	3,29	0,000
Variantie op wijkniveau	0,014	0,010	0,003	0,006

Toelichting: Eénzijdige t-toets; significantieniveau: * $\alpha = 0,05$; ** $\alpha = 0,01$.

ling niet statistisch significant verschillen in responspercentage. Wanneer we de proportie onverklaarde variantie voor model 0 berekenen, dan blijkt dat 0,4 procent van de totale variantie onverklaarde variantie op wijkniveau is. 99,6 procent van de totale variantie is dus onverklaarde variantie op individueel niveau. Vanwege de verwaarloosbaar geringe variantie op wijkniveau zijn de twee vraagstellingen die betrekking hebben op het verklaren van verschillen tussen wijken, hier niet meer relevant. Toch kijken we in model 1 alsnog wat de effecten van de individuele kenmerken zijn. De variantie op wijkniveau daalt na de toevoeging van de individuele kenmerken van 0,004 naar 0,001. De berekening van de verklaarde variantie laat voorts zien dat slechts 2,7 procent van de totale variantie verklaard wordt door de individuele variabelen. 0,1 procent van de totale variantie is onverklaarde variantie op het wijkniveau. De individuele kenmerken doen dus de onverklaarde variantie op wijkniveau dalen van 0,4 procent naar 0,1 procent. 97,2 procent van de totale variantie is onverklaarde variantie op het individueel niveau!

5. CONCLUSIE

Centraal in dit artikel stond de vraag of verschillen tussen wijken in onderdekking en nonrespons vooral te verklaren zijn aan de hand van individuele kenmerken (compositie-effecten) of aan de hand van wijkenmerken, te weten de mate van sociale desorganisatie van de onderzochte wijken. In hoeverre is deze vraagstelling nu te beantwoorden?

Uit ons onderzoek blijkt dat bij zowel onderdekking als nonrespons de onverklaarde variantie op wijkniveau slechts gering is. Bij onderdekking bevindt 2,4 procent van de onverklaarde variantie zich op wijkniveau, bij nonrespons is dit maar 0,4 procent. Hierbij dienen we wel aan te tekenen dat logistische multiniveaumodellen in het algemeen slechts betrekkelijk geringe effecten te zien geven op collectief niveau, veel geringer althans dan lineaire multiniveaumodellen (Sniijders en Bosker, 1999).

Bij de onderdekking daalt na toevoeging van individuele kenmerken het percentage onverklaarde variantie op wijkniveau van 2,4 procent naar 1,0 procent. Een groot deel van de onverklaarde variantie op wijkniveau wordt dus verklaard door individuele kenmerken. In model 2 en model 3 worden nog wel significante effecten gevonden voor wijkenmerken, in model 2 voor onze totaalmaat voor sociale desorganisatie en in model 3 voor één component daarvan: de mate van verval van de wijk. De gevonden effecten zijn echter klein, in beide gevallen 0,4 procent. Uit deze resultaten komt naar voren dat met name bewoners van wijken die in verval verkeren, zoals blijkt uit langdurige leegstand en relatief veel criminaliteit, naar verhouding wat vaker niet over een telefoon beschikken of een geheim nummer hebben. Fysieke en sociale achteruitgang en verval van de wijk brengt de bewoners er kennelijk toe zich van de buitenwereld af te sluiten.

Bij de nonrespons is de onverklaarde variantie op wijkniveau met 0,4 procent nog veel kleiner dan bij de onderdekking. Na het toevoegen van individuele kenmerken daalt dit percentage tot 0,1 procent. Variabelen op wijkniveau blijken dus niet van belang te zijn voor het verklaren van verschillen in nonrespons tussen wijken. Overigens blijken de onderzochte sociaal-demografische kenmerken meer in het algemeen niet echt belangrijk te zijn voor het verklaren van nonrespons, aangezien zij slechts drie procent van de totale variantie verklaren. Vermoedelijk zijn andere factoren van meer belang, zoals belangstelling voor het onderwerp van studie, of de mate waarin het onderwerp als bedreigend wordt ervaren. Helaas beschikken wij niet over gegevens waarmee dit nader kan worden onderzocht.

Op basis van deze uitkomsten kunnen we concluderen dat de verschillen in onderdekking en nonrespons tussen de onderzochte stadswijken overwegend worden veroorzaakt door verschillen in bevolkingssamenstelling. Er is dus sprake van compositie-effecten; sociale desorganisatie van de wijk speelt slechts een zeer kleine rol en dan nog alleen in het geval van onderdekking.

Nu kunnen op basis van ons onderzoek geen verstrekkende conclusies worden getrokken. Ons onderzoek heeft een beperkt en verkennend karakter: we hebben slechts één, middelgrote provinciehoofdstad onderzocht. Wel blijkt dat onze uitkomsten overeenstemmen met die van Goyder et al. (1992). Er zijn dus meer aanwijzingen dat buurt-effecten betrekkelijk bescheiden zijn. Nader onderzoek zal moeten uitwijzen of deze conclusie een meer algemene geldigheid heeft. Wat ons onderzoek wel duidelijk maakt, is dat voor de correctie van eventuele vertekeningen tussen wijken in onderdekking en nonrespons individuele kenmerken van de onderzochte populatie verreweg het belangrijkste zijn. Overigens is het daarbij wel zaak over de 'goede' variabelen te beschikken,

want ons onderzoek maakt tevens duidelijk dat het niet mogelijk is om alleen met behulp van enkele sociaal-demografische basiskennmerken afdoende te corrigeren voor vertekeningen.

LITERATUUR

- Boardman, J.D. & S.S. Robert, 2000. Neighborhood socioeconomic status and perceptions of self-efficacy. *Sociological Perspectives* 43, 117-136.
- Burhenne, W. & J. van der Leest, 1997. *Respons en non-respons in het Aanvullend Voorzieningengebruik Onderzoek*. Dongen: GfK Nederland.
- Cornelisse-Claassen, T.M., 1987. "Het Bevolkingsregister als steekproefkader voor het telefonisch enquêteren". In: *Recente ontwikkelingen in het marktonderzoek: Jaarboek van de Nederlandse Vereniging van Marktonderzoekers 1987-'88*, pp. 35-49.
- Engbersen, G., 1990. *Publieke Bijstandsgeheimen: Het ontstaan van een onderklasse in Nederland*. Leiden/Antwerpen: Stenfert Kroese.
- Goyder, J., 1987. *The silent minority: Nonrespondents on sample surveys*. Cambridge: Polity Press.
- Goyder, J., J. Lock & T. McNair, 1992. Urbanization effects on survey nonresponse: A test within and across cities. *Quality and Quantity* 26, 39-48.
- Gronings Geld: Financiële positie van minima in de stad Groningen*, 1997. Groningen: Gemeente Groningen.
- Hartmann, P.H. & B. Schimpl-Neimanns, 1992. Sind Sozialstrukturanalysen mit Umfragedaten möglich?: Analysen zur Repräsentativität einer Sozialforschungsumfrage. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 44, 315-340.
- House, J.S. & S. Wolf, 1978. Effects of urban residence on interpersonal trust and helping behavior. *Journal of Personality and Social Psychology* 36, 1029-1043.
- Jansma, F.W.L., 2000. *Verschillen tussen wijken in onderdekking en nonrespons*. Doctoraalscriptie. Groningen: Vakgroep Sociologie, Rijksuniversiteit Groningen.
- Kasarda, J.D. & M. Janowitz, 1974. Community attachment in mass society. *American Sociological Review* 39, 328-339.
- Lavrakas, P.J., 1993. *Telephone survey methods: Sampling, selection, and supervision*. Second edition. Newbury Park, CA., etc.: Sage.
- Maas, C. & W. de Heer, 1995. Response developments and the field work strategy. *Bulletin de Méthodologie Sociologique* No. 47 (June), 36-51.
- Rispens, S. & H. van Goor, 2000. Een vertekend beeld van de maatschappij? Een onderzoek naar selectieve uitval in een telefonische enquête ten gevolge van onderdekking en nonrespons. *Sociologische Gids* 47, 448-474.
- Sampson, R.J. & W.B. Groves, 1988/89. Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology* 94, 774-802.
- Sampson, R.J. & S.W. Raudenbush, 1999/2000. Systemic social observation of public spaces: A new look at disorder in urban neighborhoods. *American Journal of Sociology* 105, 603-651.
- Shaw, C.R. & H.D. McKay, 1942. *Juvenile delinquency and urban areas: A study of rates of delinquency in relation to differential characteristics of local communities in American cities*. Chicago: The University of Chicago Press.
- Smith, T.W., 1990. Phone home? An analysis of household telephone ownership. *International Journal of Public Opinion Research* 2, 369-390.
- Snijders, T. & R. Bosker, 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Trowbridge/Wiltshire: The Cromwell Press.
- Stebly, N.M., 1987. Helping behavior in rural and urban environments: A meta-analysis. *Psychological Bulletin* 102, 346-356.
- Taylor, R.B., 1996. Neighborhood responses to disorder and local attachments: The systemic model of attachment, social disorganization, and neighborhood use value. *Sociological Forum* 11, 41-74.
- Taylor, R.B. & J. Covington, 1993. Community structural change and fear of crime. *Social Problems* 40, 374-395.

14. Steekproeftrekking bij onderzoek naar zeldzame uitkomsten ¹

B. DONKERS, PH. FRANSES en P.C. VERHOEF

SAMENVATTING

Door ontwikkelingen op het gebied van CRM worden bedrijven steeds vaker geconfronteerd met niet frequent geobserveerde uitkomsten, de zogenaamde zeldzame uitkomsten. Om zinvolle uitspraken te doen over deze zeldzame uitkomsten is het zinvol om gebruik te maken van de gestratificeerde steekproef, waarbij men de zeldzame uitkomst oververtegenwoordigt in de steekproef. Hiermee kan een onderzoeker zeer veel geld besparen. Deze steekproeftrekkingsmethode heeft echter wel gevolgen voor de statistische analyse. Dit geldt vooral als men met een logit model de zeldzame uitkomst wil verklaren. In dit artikel wordt uitgebreid stilgestaan bij deze gevolgen. Tevens wordt een toepassing gepresenteerd in de context van CRM.

1. INLEIDING

Steekproeftrekking lijkt in het marktonderzoek wel eens een ondergeschikte plaats in te nemen. In kwantitatief onderzoek is er standaard veel aandacht voor de vraagstelling, de dataverzamelmethode en de statistische analyses. Het trekken van een goede steekproef krijgt echter veel minder aandacht. Er lijken zich op dit gebied dan ook veel minder ontwikkelingen voor te doen. Dit blijkt wel uit het feit dat bij de meeste marktonderzoekboeken er alleen wordt gerefereerd aan studies van meer dan een decennium geleden (zie o.a. Aaker, Kumar and Day, 2001). Dat neemt niet weg dat er door nieuwe ontwikkelingen op het gebied van bijvoorbeeld Customer Relationship Management, er nieuwe belangrijke onderwerpen naar voren komen voor de theorie over steekproeftrekking in het marktonderzoek.

In dit artikel willen wij op één van deze onderwerpen ingaan. In het marktonderzoek kan men geconfronteerd worden met binaire uitkomsten die relatief weinig voorkomen. Een voorbeeld is de adoptie van een nieuw product. In het begin van de levensfase zal nog maar een beperkt aantal consumenten het product gekocht hebben. Om iets over deze groep consumenten te zeggen, zullen er een behoorlijk aantal mensen ondervraagd moeten worden wanneer een willekeurige steekproef genomen wordt. Stel bijvoorbeeld dat ongeveer 2% van de populatie het product geadopteerd heeft. In een willekeurige steekproef van 400 consumenten, zal men gemiddeld maar 8 kopers van het nieuwe

¹ Dit artikel is gebaseerd op Donkers, B., Franses Ph.H. en Verhoef P.C. (2001), "Using Selective Sampling for Binary Choice Models to Reduce Survey Costs", Working Paper, ERS-2001-67-MKT, Erasmus Reserach Institute in Management (www.erin.nl).

product aantreffen. Het zal duidelijk zijn dat het op basis hiervan lastig is om uitspraken te doen over de verklarende variabelen van product adoptie. De enige oplossing lijkt dus om de steekproef groter te maken, zodat er in totaal meer kopers in de steekproef komen. Een groot nadeel is echter, dat de steekproef dan al snel heel erg groot wordt. Als men bijvoorbeeld minimaal 50 kopers in de steekproef wil hebben, betekent dit bij een willekeurige steekproef een steekproefgrootte van 2500 respondenten. Een andere mogelijkheid is om de steekproeftrekking zo vorm te geven, dat het aantal kopers in de steekproef oververtegenwoordigd wordt. Deze manier van trekken blijft zonder gevolgen als men slechts interesse heeft in beschrijvende analyses over de verschillen tussen kopers en niet-kopers. Als echter bijvoorbeeld logit-modellen worden gebruikt om het groepslidmaatschap te verklaren kunnen er zich problemen gaan voordoen. Deze problemen worden verergerd als er ook nog geselecteerd wordt op verklarende variabelen.

In dit artikel willen wij ingaan op bovenstaande onderwerpen. Hiervoor zal eerst uitgebreid worden stilgestaan bij de theorie over steekproeftrekking. Vervolgens zal ingegaan worden op de mogelijke problemen bij analyses. Hierbij zal vooral aandacht worden besteed aan het logit-model. Tenslotte, zal een toepassing worden besproken¹. Wij eindigen het artikel met een conclusie.

2. STEEKPROEFTREKKING

Bij het trekken van een steekproef worden standaard een aantal stappen doorlopen (zie o.a. Aaker, Kumar en Day, 2001; Lehmann, Gupta en Steckel, 1998). De eerste stap betreft de populatiekeuze. Deze keuze is meestal gebaseerd op de gestelde onderzoeksvraag. Vervolgens wordt het steekproefkader gekozen. De keuze van dit kader is niet triviaal. Door een verkeerde keuze kan een bepaald gedeelte van de populatie niet of slechts beperkt in de uiteindelijke steekproef voorkomen. In dit artikel zal niet verder op deze belangrijke stappen van het steekproefproces worden ingegaan. In plaats daarvan beperken we ons tot de steekproeftrekkingsmethode en het bepalen van de steekproefgrootte. Daarnaast wordt er uitgebreid stilgestaan bij de manier waarop de steekproeftrekking de analysefase beïnvloedt. Het is opvallend dat in veel marktonderzoekboeken hier maar beperkt aandacht aan wordt besteed.

Traditioneel wordt er in het marktonderzoek onderscheid gemaakt tussen willekeurige en niet-willekeurige trekkingsmethoden. Hierbij onderscheidt de niet-willekeurige steekproef zich door het feit dat er weinig of geen invloed wordt uitgeoefend op de samenstelling van de steekproef, waarbij men kan denken aan de gemakssteekproef of de sneeuwbal methode. Van de willekeurige steekproeven is de volledig a-selecte of volledig willekeurige steekproef de meest bekende. Hierbij heeft ieder element uit de populatie dezelfde kans om getrokken te worden. Naast de volledig willekeurige steekproef kan men ook gestratificeerde willekeurige steekproeven trekken. Bij deze steekproefmethode vinden willekeurige trekkingen plaats, conditioneel op de strata. Het is natuurlijk ook mogelijk om in een gemakssteekproef met strata te werken. Aangezien de algemeen gangbare statistische technieken gebaseerd zijn op de aanname van willekeurige trekkingen (eventueel conditioneel op de strata), zullen wij de niet-willekeurige methoden in de rest van dit artikel buiten beschouwing laten. De trekkingsmethode die in dit artikel aanbevolen wordt, leidt tot een gestratificeerde willekeurige steekproef.

De volledig willekeurige trekkingsmethode is algemeen bekend, maar de ingewikkel-

der gestratificeerde steekproefmethode verdient nog enige verdere toelichting. Bij deze steekproefmethode worden vooraf strata gedefinieerd, waaruit willekeurig elementen worden getrokken. Zo zou een onderzoeker de steekproef kunnen verdelen in leeftijdsgroepen. Het trekken uit de gedefinieerde strata kan dan vervolgens proportioneel of disproportioneel gebeuren. Bij een proportionele steekproef zijn de verhoudingen tussen de aantallen observaties in elk stratum in de steekproef gelijk aan de werkelijke verhoudingen in de populatie. Bij een disproportioneel gestratificeerde steekproef is dat niet het geval. Voor een disproportionele steekproef wordt bijvoorbeeld gekozen wanneer men over een bepaalde deelgroep uit de populatie betrouwbaardere uitspraken wil doen. Daarnaast kan het gebruik hiervan leiden tot het gebruik van kleinere steekproeven zonder dat de betrouwbaarheid van de uiteindelijke conclusies in het gevaar komt. Een bekend voorbeeld van een disproportioneel gestratificeerde steekproef is het retailpanel van A.C. Nielsen. Hierin worden namelijk grote supermarkten relatief oververtegenwoordigd.

Hoewel de trekkingsmethode waarin de weinig geobserveerde uitkomst wordt gebruikt om strata te definiëren in principe te zien is als een gestratificeerde willekeurige steekproef, is een wat meer gedetailleerde onderverdeling gewenst. Om dit te verduidelijken is in figuur 1 een meer gedetailleerde classificatie van gestratificeerde steekproeven weergegeven. Hierin wordt een onderscheid gemaakt tussen stratificatie op x- en y-variabelen. x-variabelen zijn de variabelen die gebruikt worden om een andere geobserveerde grootte te verklaren. De variabele die verklaard gaat worden, wordt aangeduid als y-variabele. Traditioneel vindt stratificatie vooral plaats op x-variabelen. Zo stratificeren Verhoef, Franses en Hoekstra (2001) bijvoorbeeld bij een onderzoek onder klanten van een financiële dienstverlener hun steekproef op basis van relatieduur en het aantal afgenomen diensten. Deze variabelen worden vervolgens opgenomen als antecedenten (x-variabelen) in een model om de aankoop van nieuwe diensten te verklaren. Bij selectie op de y-variabelen wordt in de steekproef de verdeling van de te bestuderen uitkomst als criterium genomen bij de trekking van de steekproef. Dit wordt ook wel uitkomst-afhankelijke steekproeftrekking (“outcome-dependent sampling”) genoemd (Donkers, Franses en Verhoef, 2001). Deze selectie kan eveneens proportioneel en disproportioneel plaatsvinden. Disproportionele selectie zal vooral plaats vinden bij uitkomsten die weinig voorkomen, zoals bijvoorbeeld bij de al eerder gememoreerde productadoptie. Uit de classificatie in figuur 1 blijkt verder dat selectie op x- en y-variabelen ook gecombineerd kan worden. Hierbij is vooral de combinatie disproportionele stratificatie op zowel x- en y-variabelen interessant. We zullen het nut van deze combinatie toelichten met behulp van twee voorbeelden.

		Selectie op y-variabele (uitkomst)		
		Willekeurig	Stratificatie	
			Proportioneel	Disproportioneel
Selectie op x-variabelen (antecedent)	Willekeurig ^b			
	Stratificatie	Proportioneel		
		Disproportioneel		

Figuur 1. Classificatie van gestratificeerde steekproeven.

Eerder is al ingegaan op het nut van selectie op de uitkomsten bij de adoptie van nieuwe producten. Stel nu eens dat een mobiele telefonie aanbieder in het begin van de levenscyclus geïnteresseerd is in de factoren die adoptie van mobiele telefoons verklaren. Omdat er nog maar zo weinig mensen een mobiele telefoon hebben, is selectie op de adoptie uitkomst zinvol. Ervaring en de adoptie- en diffusieliteratuur (Rogers, 1983) lijken echter te suggereren dat juist jongeren heel snel geneigd zijn om mobiele telefoons te adopteren. Het totaal aantal bezitters is dus relatief klein, maar onder jongeren zal dit juist groot zijn. Om nu een inschatting van de effecten van leeftijd en andere variabelen, zoals inkomen, op de adoptie van de mobiele telefoon te kunnen bepalen, zou men een volledige willekeurige steekproef kunnen nemen, maar hierin zullen de aantallen “adopters” per leeftijdsgroep sterk variëren. Aangezien een steekproef over het algemeen informatiever is wanneer er veel variatie in de variabelen zit, is het in deze situatie interessant om een gestratificeerde willekeurige steekproef te gebruiken. Bij de stratificatie zou men bij het grootste deel van de populatie de “adopters” kunnen oververtegenwoordigen (hier zijn er relatief weinig van), terwijl op hetzelfde moment men de jongeren zonder mobiele telefoons ook wil oververtegenwoordigen (ook hier zijn er relatief weinig van). Er is dan dus sprake van stratificatie op zowel leeftijd als op het bezit van een mobiele telefoon, dus zowel de x- als de y-variabele speelt een rol bij de stratificatie.

Een tweede voorbeeld speelt in een customer relationship management context. In deze context zijn veel bedrijven geïnteresseerd in de vraag hoe het komt dat klanten de onderneming verlaten (Ganesh, Arnold en Reynolds, 2000). In bepaalde markten is het percentage klanten dat in een bepaalde periode (bijv. een jaar) een bedrijf verlaat zeer laag. In de financiële dienstverlening zijn bijvoorbeeld percentages tussen de 5 en 10% per jaar niet ongewoon. In dit soort markten is het weggaan van een klant dus een weinig geobserveerde uitkomst. Bij disproportionele stratificatie op de variabele “klantdefectie” kan dus een kleinere steekproef worden gebruikt. De onderzoeker kan immers het aantal klanten dat weggaat oververtegenwoordigen. Hierbij zou er bijvoorbeeld voor kunnen worden gekozen om evenveel klanten die blijven als weggaan te interviewen. Deze stratificatie kan er echter wel toe leiden, dat er geen goed antwoord wordt verkregen op de vraag waarom klanten de onderneming de rug toe keren. Uit ervaring weet de onderneming dat klanten met een korte relatieduur relatief veel vaker weggaan. Relatieduur lijkt dus een belangrijke verklarende variabele. Er zijn echter ook klanten met een korte relatieduur, die toch blijven. De vraag is dan waarom deze zelfde groep klanten ondanks hun korte relatieduur toch klant blijft. Hebben deze klanten bijvoorbeeld gemiddeld genomen een hogere tevredenheid? Klanten met een korte relatieduur komen echter relatief weinig voor in het klantenbestand. Er kunnen hierdoor onvoldoende klanten met een korte relatieduur in de steekproef voorkomen, om naast de invloed van relatieduur ook de invloed van andere variabelen vast te stellen. Om dit te voorkomen zullen klanten met een korte relatieduur oververtegenwoordigd moeten worden terwijl het tegelijkertijd ook interessant is om de klanten die weggaan te oververtegenwoordigen. Ook hier wordt dus een combinatie van een disproportionele stratificatie op x- en y-variabelen plaats.

Bovenstaande voorbeelden³ laten duidelijk zien wat de mogelijkheden zijn voor stratificatie op x- en/of y-variabelen. Disproportionele stratificatie op de y-variabele is zinvol bij uitkomsten die relatief weinig voorkomen. Deze stratificatie maakt het mogelijk om met veel kleinere steekproeven toch betrouwbare resultaten te krijgen. Disproportionele stratificatie op zowel de x- en y-variabele is zinvol wanneer zowel de y-variabele als de

x-variabele relatief weinig voorkomen en de x-variabele de uitkomst beïnvloedt. Het grote voordeel van dergelijke stratificatiemethoden is dus dat men aanzienlijk minder data hoeft te verzamelen, en dit leidt tot interessante kostenbesparingen. Een probleem is dat door deze stratificaties de data-analyse wordt beïnvloed. Wanneer hier geen rekening mee wordt gehouden kan dit leiden tot verkeerde conclusies. In de volgende sectie zal hier verder op worden ingegaan.

3. STRATIFICATIE EN ANALYSE

In de bespreking van de gevolgen van de stratificatie op de analyse maken we onderscheid tussen drie typen analyses:

- (1) beschrijvende analyses van één variabele (frequentie-tabellen, gemiddelden e.d.);
- (2) univariate toetsende analyses;
- (3) verklarende analyses met logit-modellen.

3.1. Beschrijvende analyses

Bij disproportionele stratificatie op zowel x- als y-variabelen zijn bepaalde groepen in de steekproef oververtegenwoordigd en andere groepen ondervertegenwoordigd. De beschrijvende analyses zijn hierdoor niet correct. Om dit probleem op te lossen, moet er een weging plaats vinden. In dit artikel behandelen we alleen de weging voor het berekenen van het gemiddelde en de variantie (gekwadrateerde standaarddeviatie). Stel er zijn een aantal strata (h) en de populatiegrootte is N. Op basis van de gemiddelden in de verschillende strata, kan het gemiddelde in de steekproef als volgt berekend worden (Churchill, 1999; p. 527):

$$\bar{x}_{st} = \sum_{h=1}^L \frac{N_h}{N} * \bar{x}_h \quad (1)$$

waarbij,

\bar{x}_{st} = gemiddelde voor de gestratificeerde steekproef,

\bar{x}_h = gemiddelde voor stratum h, en

N_h = populatiegrootte van stratum h.

De variantie (gekwadrateerde standaarddeviatie) van dit gemiddelde wordt met de volgende formule berekend (Churchill, 1999; p. 527):

$$s_{\bar{x}_{st}}^2 = \sum_{h=1}^L \left(\frac{N_h}{N} \right)^2 * s_{\bar{x}_h}^2 \quad (2)$$

waarbij,

$s_{\bar{x}_h}^2$ = variantie van x in stratum h.

Een restrictie voor het gebruik van disproportioneel gestratificeerde steekproeven is dat de onderzoeker de populatiegroottes van de verschillende strata dient te kennen.

3.2. Univariate toetsen

Dat stratificatie ook een invloed kan hebben op de uitkomst van univariate toetsen, zoals de t- en F-toets, wordt vaak vergeten. Deze invloed doet zich vooral voor, indien de afhankelijke variabele (y) samenhangt met de onafhankelijke variabele (x). Bij de toelichting van dit probleem bekijken we de t-toets, waarmee verschillen in gemiddelden van twee groepen getoetst kunnen worden. Hierbij zal in het bijzonder het geval besproken worden waarbij de twee groepen gerelateerd zijn aan een binaire uitkomstvariabele.

De formule voor de t-toets is als volgt:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_{\bar{x}_1 - \bar{x}_2}} \quad (3)$$

waarbij,

$s_{\bar{x}_1 - \bar{x}_2}$ = geschatte standaarddeviatie van $\bar{x}_1 - \bar{x}_2$,

\bar{x}_1 = gemiddelde waarde x in groep 1, dan wel 2,

t = t-waarde.

Aangezien de twee deelsteekproeven onafhankelijk van elkaar zijn, kan de standaarddeviatie als volgt worden berekend:

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}} \quad (4)$$

waarbij,

\hat{s}_1 = de standaarddeviatie van x in groep 1 (dan wel 2),

n_1 = de grootte van de groep 1 (dan wel 2) in de steekproef,

zijn. Het zal duidelijk zijn, dat de uitkomst van de t-toets wordt beïnvloed door de grootte van het verschil in gemiddelden, standaarddeviaties van de twee verschillende groepen en de grootte van de groepen. Voor de helderheid van de toelichting nemen we het eerder gegeven voorbeeld over retentie bij een financiële dienstverlener als uitgangspunt. Stel, de onderzoeker wil de verschillen weten tussen klanten die blijven en klanten die het bedrijf verlaten⁶. In dit geval is defectie dan wel retentie dus de afhankelijke variabele met twee groepen.

Stel nu eens dat er alleen gestratificeerd wordt op de x-variabele relatieduur. Hierbij worden relatief meer klanten met een korte relatieduur opgenomen in de steekproef,

omdat deze groep relatief weinig voorkomt. Stel nu dat we willen toetsen of klanten die weggaan een significant kortere relatieduur hebben dan klanten die blijven. Als er geen samenhang is zal de uitkomst van de t-toets niet beïnvloed worden door de stratificatie. Indien het veronderstelde verschil zich inderdaad voordoet in de populatie is er echter wel een probleem. In dat geval zal in de groep vertrekkende klanten relatief veel klanten met een korte relatieduur voorkomen. Hierdoor daalt de gemiddelde relatieduur in deze groep. Deze gemiddelde relatieduur daalt ook wel in de groep blijvers. Deze daling is echter minder groot. Hierdoor wordt het verschil in de gemiddelde relatieduur tussen de twee groepen groter dan in de populatie, waardoor de t-waarde onterecht hoger wordt. Voor de standaarddeviatie kan formule (2) gebruikt worden. Bovendien wordt door een betere spreiding van de observaties over de twee deelgroepen de standaarddeviatie kleiner. Hierdoor wordt de t-waarde ook nog eens vergroot. Kortom, stratificatie op de x-variabelen heeft geen invloed op de uitkomst van de t-toets als er geen samenhang is met de bestudeerde uitkomst. Wanneer dit echter wel het geval is leidt stratificatie op de x-variabele tot een snellere verwerping van de nulhypothese van de t-toets.

Een disproportionele stratificatie op alleen de y-variabele heeft ook een effect, maar dit effect heeft vooral positieve gevolgen. Stel dat vertrekkers (waarvan er relatief weinig zijn) worden oververtegenwoordigd in de steekproef. Er is wederom sprake van dezelfde onderzoeksvraag als bij de stratificatie op de x-variabele. Stratificatie op de y-variabele heeft echter geen effect op de gemiddelde relatieduur in de beide groepen. Aangezien er door de stratificatie een meer gelijke verdeling is van de observaties over de twee groepen, zullen n_1 en n_2 dichter bij elkaar liggen. Wanneer de varianties in de twee groepen niet al te verschillend zijn, zal dit leiden tot een kleinere standaarddeviatie van het verschil. Dit heeft tot gevolg dat eenzelfde verschil tussen de twee groeps-gemiddelden een grotere kans geeft om de nulhypothese te verwerpen. De toets krijgt dus meer onderscheidend vermogen.

Evenals bij de stratificatie op alleen de x-variabele zorgt een disproportionele stratificatie op zowel de x- als y-variabelen eveneens voor problemen, wanneer de x-variabele samenhangt met de y-variabele. Dit weet je in principe pas na de dataverzameling. Echter bij bepaalde verschijnselen heeft de onderzoeker vaak op basis van theorie of ervaring al een idee over mogelijke samenhang. In dit geval wordt evenals bij stratificatie op x-variabelen de gemiddelde waarde van de x-variabele in de twee groepen verkeerd ingeschat. Dit probleem kan opgelost worden door per groep het effect, dat de stratificatie op basis van de x-variabele heeft op het groepsgemiddelde, te corrigeren door de weging die hierboven is beschreven (zie formule (1)). De standaarddeviatie moet eveneens worden aangepast. Hiervoor kan formule (2) weer worden gebruikt. Het positieve effect van stratificatie op de y-variabele blijft dan behouden voor de toets.

3.3. Analyses met logit-modellen

Logit-modellen worden gebruikt om binaire uitkomsten te verklaren dan wel te voorspellen. De functionele vorm van een logit-model is als volgt (Franses en Paap, 2001, p. 54):

$$\Pr[Y_i = 1 | X_i] = \frac{\exp(X_i \beta)}{1 + \exp(X_i \beta)} \quad (5)$$

Bij stratificatie op één of meer van de x-variabelen in het logit-model doen zich in principe geen problemen voor bij de schattingsresultaten. Als gestratificeerd wordt op de y-variabele moet er wel een aanpassing plaatsvinden. Bij deze stratificatie is de geschatte constante in het logit-model niet juist (Cramer, Franses en Slagter, 1999). Tevens zijn alle standaardfouten niet correct. De schatters van de x-variabelen veranderen echter niet. Als men dus als onderzoeker de binaire uitkomst wil *verklaren* zijn niet veel aanpassingen in het model nodig. Alleen de constante in het model kan niet goed worden geïnterpreteerd. Wanneer men dit model echter wil gebruiken, om de binaire uitkomst te *voorspellen*, moet de constante wel worden aangepast. Donkers, Franses en Verhoef (2001) geven hiervoor de procedure.

Bij stratificatie op zowel de x- als y-variabelen, is er wel een probleem. Dit probleem doet zich voor als de x- en y-variabelen, waarop gestratificeerd wordt met elkaar samenhangen. Zowel de constante als de andere parameterschattingen zijn dan niet correct. Daarnaast zijn de standaardfouten niet goed. Hierdoor kunnen op basis van het model verkeerde conclusies worden getrokken. Ook hier zijn wederom aanpassingen nodig, die op zich redelijk eenvoudig zijn te implementeren in bijvoorbeeld een computerpakket als SAS (Donkers, Franses en Verhoef, 2001).

4. EMPIRISCHE TOEPASSING

4.1. Data

Om het nut van stratificatie op zowel de x- als y-variabelen aan te tonen, zullen we de volgende empirische toepassing laten zien. Het betreft een case bij een financiële dienstverlener. Deze dienstverlener is geïnteresseerd in de vraag waarom klanten de onderneming verlaten. Zoals al eerder aangegeven verlaat echter maar een zeer klein percentage van totale klantenbestand deze onderneming. In een panelstudie zijn in 1999 2300 klanten ondervraagd. In het jaar 2000 zijn deze 2300 klanten wederom ondervraagd. Hiervan wilden er 1374 meewerken. Van deze 1374 klanten verlieten 53 klanten de onderneming (3.9%). Om goede uitspraken te doen waarom klanten vertrekken zijn er nog 30 vertrekkende klanten extra ondervraagd.

Voor klanten kunnen er verschillende redenen zijn om te vertrekken. Bij dienstverleners is vooral de tevredenheid over de dienstverlening van belang (Reichheld, 1996). In de vragenlijst is dan ook naar de tevredenheid van klanten gevraagd. Daarnaast lijkt de relatieduur een belangrijke verklarende variabele. Klanten met een korte relatieduur ervaren vaak minder switchkosten en zijn vaak nog gevoeliger voor een slechte performance van de onderneming (Bolton, 1998; Verhoef, 2001). In het model om het weggaan van klanten te verklaren worden deze variabelen dan ook meegenomen⁴.

4.2. Toepassing

In de toepassing zullen we ons concentreren op het effect van stratificatie op de schatting van het logit-model. Hierbij hanteren we vier verschillende steekproeven. De eerste steekproef betreft de originele steekproef. In de tweede steekproef worden alle 30 extra klanten toegevoegd. Vervolgens zal in de derde steekproef maar een gedeelte van deze 30 extra klanten worden toegevoegd. Het betreft 6 klanten die vertrekken en een korte relatieduur hebben. Er wordt dan dus gestratificeerd op zowel de x- (relatieduur)

als y variabelen (weggaan). Tenslotte wordt in de laatste steekproef gewerkt met een veel kleinere steekproef. De grootte is nog maar 40% van de originele steekproefgrootte (549). Om dit te bereiken zijn uit de blijvers willekeurig een aantal respondenten uit de steekproef verwijderd. De steekproef bestaat aldus uit de extra 30 vertrekkende klanten, de reeds in de originele steekproef aanwezige vertrekkende klanten (53) en 466 blijvende klanten. Er is dus alleen gestratificeerd op de y-variabelen.

In het geschatte logit-model zijn zoals gezegd de tevredenheid over de dienstverlening meegenomen en de relatieduur. Hierbij is relatieduur gecodeerd als een dummy-variabele, waarbij 1 staat voor een korte relatieduur en 0 voor een lange relatieduur. De afhankelijke variabele "weggaan" is zo gecodeerd dat 1 voor vertrekken en 0 voor blijven staat. Een negatieve coëfficiënt voor tevredenheid betekent dus, dat klanten die tevreden zijn een kleinere kans hebben om weg te gaan.

Tabel 1. Onjuiste en juiste schattingsresultaten voor het logit-model bij verschillende steekproeven (standaardfouten tussen haakjes).

	Steekproef 1	Steekproef 2		Steekproef 3		Steekproef 4	
N	1374	1404		1380		549	
Schatting		Onjuist	Juist	Onjuist	Juist	Onjuist	Juist
<i>Constante</i>	-0.453 (0.812)	0.073 (0.685)	-0.648 (0.696)	-0.213 (0.795)	-0.470 (0.793)	0.238 (0.807)	-0.577 (0.807)
<i>Tevredenheid</i>	-0.903 (0.252)	-0.903 (0.212)	-0.850 (0.215)	-0.958 (0.247)	-0.898 (0.246)	-0.948 (0.246)	-0.894 (0.246)
<i>Dummy Relatieduur</i> (0 = lang, 1 = kort)	0.699 (0.335)	0.697 (0.282)	0.696 (0.278)	1.162 (0.282)	0.672 (0.294)	0.633 (0.314)	0.684 (0.311)

In tabel 1 staan de resultaten van de schattingen van het logit-model. Voor steekproef 2 tot en met 4 zijn er twee schattingsresultaten weergegeven. In de eerste kolom staan de schattingsresultaten van de ongecorrigeerde en dus onjuiste methode. Vervolgens worden in de tweede kolom de gecorrigeerde en dus juiste schattingsresultaten weergegeven. Uit de schattingsresultaten van de originele steekproef blijkt, dat klanten die tevreden zijn minder snel geneigd zijn om weg te gaan, terwijl klanten met een korte relatieduur sneller geneigd zijn om te vertrekken. In de tweede steekproef, waarbij dus het aantal vertrekkers is oververtegenwoordigd wordt de constante bij de ongecorrigeerde schattingen veel groter. De overige coëfficiënten blijven ongeveer hetzelfde. Dit bevestigt dus dat bij alleen stratificatie op de y-variabelen er geen significante veranderingen plaatsvinden in deze parameter schattingen. Door de correctie toe te passen wordt de juiste constante geschat. De andere coëfficiënten veranderen hierdoor bijna niet.

In de derde steekproef wordt zowel gestratificeerd op de x- als de y-variabelen. Als er geen correctie wordt toegepast veranderen de parameterschattingen voor tevredenheid en relatieduur. Dit geldt vooral voor die van relatieduur. Dit komt doordat deze groep oververtegenwoordigd is in de groep vertrekkers. Het effect van relatieduur wordt hierdoor overschat. Daarnaast is ook de geschatte constante groter geworden. Door de correctie toe te passen worden weer de juiste parameterschattingen verkregen. Deze schattingen komen ongeveer overeen met die van de schattingen in de originele steekproef. Alleen zijn nu de standaardfouten wat kleiner, omdat er meer vertrekkers in de steek-

proef aanwezig zijn. In de laatste steekproef is een groot aantal blijvers willekeurig verwijderd. Hierdoor is het aantal vertrekkers in de steekproef relatief hoog. Er heeft aldus een stratificatie op de y-variabele plaatsgevonden. Net zoals bij de tweede steekproef is de schatting van de constante in het ongecorrigeerde model weer te hoog. De andere coëfficiënten zijn ongeveer gelijk. Door de correctie toe te passen wordt wederom de juiste constante geschat. Een belangrijk gevolg van deze laatste steekproef is dat door selectie op de uitkomst de steekproef met ongeveer 60% wordt verkleind, terwijl de schattingsresultaten ongeveer hetzelfde blijven. Ofwel door gebruik te maken van een gestratificeerde steekproef en hiervoor te corrigeren bij de analyse kan een grote besparing in de dataverzamelingskosten worden verkregen.

5. SAMENVATTING

In dit artikel is uitgebreid stilgestaan bij het nut van stratificatie van de steekproef op basis van de te bestuderen binaire uitkomst. Deze stratificatie is vooral zinvol, wanneer deze uitkomst zeer weinig voorkomt. In onze empirische toepassing hebben we bijvoorbeeld laten zien, dat met een 60% kleinere steekproef schattingsresultaten met een hogere precisie verkregen kunnen worden. Bij telefonisch onderzoek zou bij deze applicatie bij gemiddelde gesprekskosten van ongeveer 10 euro per gesprek ruim 8000 Euro bespaard zijn. Er kan dus op deze manier behoorlijk bespaard worden op de onderzoekskosten. Zoals besproken levert deze stratificatie wel problemen op bij de analyse. In dit artikel hebben we stilgestaan bij deze problemen en zijn we specifiek ingegaan op het logit-model. In de toepassing hebben we laten zien dat met een aantal correcties de juiste schattingsresultaten worden verkregen.

De vraag is nu wanneer een marktonderzoeker kan stratificeren op binaire uitkomsten. De belangrijkste voorwaarde is dat de binaire uitkomst bekend is bij de onderzoeker en dat de totale aantallen in de populatie bekend zijn. Deze voorwaarden leiden soms tot problemen. Immers, vaak wordt in een vragenlijst gevraagd naar deze uitkomst en weet men deze niet op voorhand. Zo zal men bij een onderzoek naar de adoptie van duurzame goederen vaak vooraf niet weten of de respondent het duurzame goed gekocht heeft of niet. Een oplossing hiervoor is om de adoptie als toegangsvraag te stellen. Als er voldoende kopers in de steekproef zitten, wordt de rest van de enquête na de adoptievraag niet meer afgenomen bij kopers van het duurzame goed en wordt er alleen nog gezocht naar niet-kopers. Dit probleem doet zich niet voor in een CRM-context. Hier observeren bedrijven het gedrag van klanten en kan op basis van deze observatie gestratificeerd worden. Meer in het algemeen zal de methode beter toepasbaar zijn wanneer de doelgroep beschreven is in een database van een bedrijf. Wanneer men dan dit gedrag wil onderzoeken, kan men op basis van de database een stratificatie uitvoeren. Voor de praktijk is het verder relevant om te weten wat nu de optimale verdeling van de uitkomsten binnen een steekproef is. Onderzoek laat zien dat een gelijke verdeling van de twee uitkomsten gemiddeld genomen tot een optimaal resultaat leidt (Lancaster en Imbens, 1991). Daarnaast wordt vaak als vuistregel genomen, dat men minimaal tussen de 50 en 100 waarnemingen per uitkomst wil hebben. Dit aantal hangt echter ook af van het aantal verklarende variabelen in het logit-model. Bij meer verklarende variabelen zijn in het algemeen meer waarnemingen gewenst. Daarnaast hangt dit aantal ook af van de grootte van het effect en de gewenste precisie. Concluderend kan dus gesteld worden, dat stratificatie op basis van binaire uitkomsten vooral toepasbaar

is, wanneer één van deze uitkomsten weinig voorkomt en de onderneming of het marktonderzoeksbureau de binaire uitkomst observeert. In deze context kan deze stratificatie, mits de juiste analyses gebruikt worden, leiden tot de juiste analyse uitkomsten en significante besparingen op de onderzoekskosten.

NOTEN

1. Wij hebben er naar gestreefd het artikel niet te technisch te schrijven. De bedoeling is dat u als lezer inzicht verkrijgt in de mogelijke problemen bij het trekken van dit soort steekproeven. De benodigde technische aanpassing in het logit-model zijn makkelijk toe te passen in bijvoorbeeld SAS. De programma's hiervoor zijn gratis verkrijgbaar bij de auteurs
2. De cel met combinatie willekeurig – willekeurig is dus gewoon een a-selecte steekproef. Er vindt dan immers totaal geen stratificatie plaats.
3. Het hier besproken voorbeeld gaat er vanuit dat er een positief verband is tussen relatieduur en retentie. Dit hoeft niet altijd zo te zijn. Literatuur op het gebied van relatiemarketing ondersteunt dit verband echter wel (Verhoef, 2001).
4. In de literatuur worden er nog meer determinanten van klantretentie bediscussieerd (zie o.a. Verhoef, 2001). In dit artikel zullen we echter vanwege onze focus op de methode hiervan abstraheren.

LITERATUUR

- Aaker, D. V. Kumar en G. Day, 2001, *Marketing Research*, Chichester: John Wiley & Sons
- Bolton R.N., 1998. A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction, *Marketing Science*, 17 (1), 45-65
- Donkers, B., Franses Ph.H. en Verhoef P.C., 2001. Using Selective Sampling for Binary Choice Models to Reduce Survey Costs, Working Paper, ERS-2001-67-MKT, Erasmus Research Institute in Management, Rotterdam
- Churchill Jr., G.A., 1998. *Marketing Research Methodological Foundation*, The Dryden Press Harcourt Brace College Publishers, Fort Worth
- Cramer, M., P.H. Franses en E. Slagter, 1999. Censored Regression Analysis in Large Samples with Many Zero Observations, *Econometric Institute Report EI-9939/A*, Erasmus University, Rotterdam
- Franses, P.H. en R. Paap, 2001. *Quantitative Models in Marketing Research*, Cambridge: Cambridge University Press
- Ganesh, J, M.J. Arnold en K. Reynolds, 2000. Understanding the Customer Base of Service Providers: An Examination of the Differences between Switchers and Stayers, *Journal of Marketing*, 65 (Juli), 65-87
- Lancaster, T. en G.W. Imbens, 1991. *Choice Based Sampling: Inference and Optimality*, Department of Economics Working Paper, Brown University, Providence
- Lehmann, D.R., S. Gupta en J.H. Steckel, 1998. *Marketing Research*, Massachusetts Reading: Addison Wesley
- Reichheld, F.F., 1996. *The Loyalty Effect*, Boston: Harvard Business School Press
- Rogers, E.M., 1983. *The Diffusion of Innovations*, New York: The Free Press
- Verhoef P.C., 2001. *Analyzing Customer Relationships: Linking Relational Constructs and Marketing Instruments to Customer Behavior*, Dissertatie, Rotterdam: Tinbergen Instituut,
- Verhoef, P.C., P.H. Franses en J.C. Hoekstra, 2001. The Impact of Satisfaction and Payment Equity on Cross-buying: A Dynamic Model for a Multi-service Provider, *Journal of Retailing*, 77 (3), 359-379