

Criteria-instabiliteit bij MADM-problemen

Jelte Orij (s0174947)
26-04-2012

Universiteit Twente

Bacheloropdracht Technische Bedrijfskunde

Begeleider:

Dr. J.M.G. Heerkens

Meelezer:

Dr. Ir. L.L.M. Van der Wegen

Inhoudsopgave

1. Multi attribute decision-making	1
2. Het principe “criteria-instabiliteit”	1
3. Doel van het onderzoek	2
4. Gewichten	2
4.1 De veranderlijkheid van gewichten	2
4.2 De meetmethode van gewichten	3
5. Onderzoek inzake de veranderlijkheid van gewichten	4
5.1 Dimensies van verandering in gewichten	4
5.1.1 Het subject van vergelijking: wat gaan we vergelijken?	4
5.1.2 De vergelijkingsstandaard: waarmee gaan we vergelijken?	5
5.2 Van “perfect rationeel” naar “keuze-specifieke variantie”	7
5.2.1 De onhoudbaarheid van “perfecte rationaliteit”	7
5.2.2 Procedurele variantie	8
5.2.3 Descriptieve variantie	10
5.3 Een nieuw fenomeen van verandering in gewichten	12
6. Onderzoeken met nader te verklaren veranderingen in gewichten	12
6.1 Verandering van gewichten bij beredeneren ofwel het bedenken van criteria	13
6.1.1 Belangrijkheid van attributen: lange termijn waarde versus keuze-specifiek gewicht	15
6.1.2 Variabiliteit van criteria en gewichten	16
6.1.3 Kan het selectie/weeg-proces van criteria verklaard worden?	18
6.1.4 Het selectie/weeg-proces van criteria als hersen-activiteit	19
6.2 Verandering van gewichten bij vooraf aangereikte criteria	24
6.3 Een model voor de variatie in gewichten	26
7. De definitie van criteria-instabiliteit	28
7.1 Overzicht van verandering in gewichten	29
7.2 Systematische en willekeurige invloeden op keuzegedrag	30
7.2.1 Systematische invloeden op keuzegedrag	30
7.2.2 Onsystematische invloeden op keuzegedrag: criteria-instabiliteit	31
7.2.3 Een overzicht van locale effecten op keuze-gedrag en criteria- instabiliteit	32
7.3 Discussiepunten	33
7.3.1 Systematisch/onsystematisch en verklaarbaar/onverklaarbaar	33
7.3.2 “Anchoring” met betrekking tot de score	34
8. Evaluatie	35
8.1 Wat hebben we gedaan?	35
8.2 Aannames en validiteit	35
8.3 Aanbevelingen voor vervolgonderzoek	37
Literatuurlijst	38
Definitielijst	41
Bijlage A	43
Bijlage B	44

1. Multi attribute decision-making

In dit eerste hoofdstuk wordt de term MADM uit de titel van dit verslag nader toegelicht. Dit verslag gaat namelijk over beslissingen waarbij door een persoon (ook wel beslisser) uit meerdere alternatieven het meest aantrekkelijke alternatief moet worden gekozen. Bij het maken van de beslissing maakt de beslisser gebruik van criteria. Deze vorm van beslissen wordt MADM (multi-attribute decision-making) genoemd en wordt in dit verslag als uitgangspunt genomen. Het feit dat MADM-beslissingsproblemen veelvoorkomend zijn in bijvoorbeeld economische vraagstukken, alsmede het dagelijkse leven, maakt onderzoek op dit gebied een belangrijk punt van beslissingsanalyse (Weber, Borchering, 1992).

Omdat bij de meeste beslissingen alternatieven betrokken zijn die zowel voor- als nadelen hebben, richten veel onderwerpen binnen het onderzoek naar MADM zich op *trade-offs* tussen de attributen, of criteria, van de alternatieven (Goldstein, 1990). Zo kan het bijvoorbeeld zijn dat men een afweging maakt tussen de criteria “kosten” en “kwaliteit”. Een bepaald alternatief zal hoger scoren op het ene criterium, terwijl het lager scoort op het andere criterium.

Er bestaan verschillende wiskundige modellen waarbij deze *trade-off* in kaart wordt gebracht. Deze kunnen als hulpmiddel dienen om, al dan niet met subjectieve invoer van de beslisser, tot een voor de beslisser optimale keuze van het MADM-probleem te komen. Een generieke eigenschap van deze methoden is dat de eindkeuze van preferente alternatieven altijd afhankelijk is van zowel de scores op als de gewichten van de criteria (Yeh, 2002). Het onderzoek in dit verslag richt zich specifiek op de gewichten, of relatieve belangrijkheid, van de criteria. Naast de scores spelen deze gewichten in de meeste MADM-problemen een belangrijke rol (Weber, Borchering, 1992). Zo ook in dit onderzoek.

2. Het principe “criteria-instabiliteit”

Dit verslag richt zich specifiek op criteria-instabiliteit binnen de in het vorige hoofdstuk beschreven MADM-problemen. In dit hoofdstuk wordt deze term nader verklaard. We beperken hier ons echter tot een onvolledige omschrijving, want later in dit verslag zal deze verder uitgewerkt worden.

De gewichten die de een persoon aan bepaalde criteria toekent, zijn nogal eens aan verandering onderhevig. Hoewel criteria-instabiliteit zou staan voor een bepaalde, schijnbaar onverklaarbare verandering in gewichten, is er nog geen algemeen geaccepteerde definitie van. Ten einde een voorlopige werkdefinitie van criteria-instabiliteit op te stellen, maak ik hier nu onderscheid tussen verklaarbare en onverklaarbare verandering van gewichten. Ter toelichting volgt een voorbeeld.

Stel een beslisser krijgt een reeks achtereenvolgende keuze-situaties waarbij steeds de voor de beslisser preferente auto gekozen moet worden uit een lijst van vijf auto's. De auto's worden onder andere gespecificeerd door “kleur” en “prijs”. Los van de overige criteria, gaan we ervan uit dat de beslisser het relatief belangrijk vindt dat de auto een donkere kleur heeft. Het criterium “kleur” heeft dus een relatief hoog gewicht. Stel nu dat na twee gemaakte keuzes extra informatie aan de beslisser wordt getoond. Deze informatie maakt duidelijk dat

een actie van het garagebedrijf het mogelijk maakt om voor honderd euro de auto over te laten spuiten naar een gewenste kleur. Vanaf nu zal de beslisser een stuk minder waarde hechten aan de gespecificeerde kleur van een auto. Het gewicht voor “kleur” is verklaarbaar veranderd door de extra informatie.

Wat echter wanneer tijdens deze identieke reeks keuzes de extra informatie niet verschaft zou zijn, maar er wel een duidelijke verandering in de belangrijkheid van het door de beslisser gebruikte criterium “kleur” geconstateerd zou worden? De verandering wordt nu erkend als onverklaarbaar.

Dit is het gebied waarin criteria-instabiliteit zich afspeelt. Criteria-instabiliteit is dus, grofweg, de verandering in gewichten die geheel onverklaarbaar optreedt¹. Met name in situaties waar er een gering verschil is tussen de preferentie van de alternatieven, zoals in veel complexe beslissingsproblemen het geval is, kan een geringe variatie in gewichten al snel een verandering in de keuze teweegbrengen (Palmeira, Krishnan, 2008). Dit gegeven maakt criteria-instabiliteit een belangrijk onderzoeksgebied.

3. Doel van het onderzoek

Het onderzoek naar criteria-instabiliteit is relatief nieuw en zoals eerder gezegd bestaat er nog geen algemeen geaccepteerde definitie van het fenomeen. Ook de positie van criteria-instabiliteit ten opzichte van de vele onderzoeken over verandering in gewichten in de literatuur van de besliskunde is tot op heden nog onduidelijk. Zo is er geen duidelijke grens tussen verklaarbare en onverklaarbare veranderingen in gewichten en daarom is het niet goed te zeggen wanneer iets valt onder criteria-instabiliteit. Het doel van het door mij gedane onderzoek is dan ook om de zojuist genoemde onduidelijkheden te verduidelijken ter ondersteuning van verder onderzoek naar criteria-instabiliteit. De twee doelstellingen van het door mij gedane onderzoek die hieruit resulteren zijn als volgt:

- Het principe van “criteria-instabiliteit” te definiëren.
- Een gestructureerd overzicht te maken van de onderzoeken in de literatuur omtrent verandering in gewichten waarbij onderzoeken met betrekking tot criteria-instabiliteit worden afgebakend.

4. Gewichten

Met de term “gewicht” wordt in dit verslag de belangrijkheid van een criterium bedoeld. Wanneer men spreekt over het gewicht van een criterium kan dit op verschillende manieren geïnterpreteerd worden. Het belangrijkste onderscheid in deze opvattingen komt tot uiting in de veranderlijkheid en de meetmethode van de gewichten.

4.1 De veranderlijkheid van gewichten

Ten aanzien van verandering in gewichten zijn er volgens Goldstein (1990) twee benaderingen te onderscheiden, namelijk “globale” en “locale” interpretaties van gewichten. Bij globale interpretaties van gewichten worden gewichten gezien als een representatie van de waarden van een persoon. Hoewel deze op lange termijn aan verandering onderhevig zijn,

¹ Verderop in dit verslag zal toegewerkt worden naar een gedetailleerdere omschrijving van criteria-instabiliteit.

zijn ze op korte termijn tussen verschillende keuze-situaties in vaststaand. Verder zijn ze niet afhankelijk van keuze-specifieke stimuli². Een voorbeeld van een globale verandering in gewichten is wanneer een persoon na het winnen van een grote geldprijs niet meer op de prijs van kleine producten let bij het winkelen, terwijl hij dat voor het winnen van de geldprijs wel deed. Hier is een globale verandering in het gewicht van het criterium “productprijs” opgetreden, welke nu minder in acht wordt genomen.

Locale interpretaties van gewichten gaan ervan uit dat de gewichten tussen verschillende keuzes juist wel verschil kunnen vertonen door persoonlijke karakteristieken en de stimuli. Goldstein refereert naar enkele mogelijke contextuele oorzaken die de locale interpretatie van gewichten onderbouwen. Deze en andere invloedsfactoren van verandering in gewichten worden later in dit verslag besproken. Voor nu echter is het van belang het onderscheid te zien tussen gewichten die op korte termijn tijdens een reeks achtereenvolgende keuze-situaties kunnen veranderen (de locale interpretatie) en de verandering die zich slechts zou beperken tot de lange termijn en onafhankelijk is van de keuze-situatie (globale interpretatie). Later in dit verslag zullen we zien dat de globale interpretatie geen stand houdt en dat gewichten wel degelijk afhankelijk zijn van keuze-specifieke stimuli.

4.2 De meetmethode van gewichten

Wanneer een persoon een keuze maakt hanteert deze dus bepaalde gewichten bij criteria. Om erachter te komen welke gewichten precies gebruikt zijn, zijn er volgens Goldstein twee meetmethoden. Enerzijds is er het “subjectieve gewicht”, dat staat voor de beslisser zijn perceptie van de gewichten en kan achterhaald worden door de beslisser te vragen naar de waardering van de relatieve belangrijkheid van een criterium. Anderzijds is er het zogenaamde “paramorfische gewicht”, waarbij uit de scores die de beslisser aan de alternatieven op de criteria heeft gegeven met behulp van wiskundige modellen de gewichten worden berekend. Zo kan er bijvoorbeeld gebruik gemaakt worden van meervoudige regressie-analyse, wat gebruikt wordt in de onderzoeken van Levine welke ik in dit verslag zal behandelen. Meervoudige regressie werkt kortweg als volgt.

$$Y_i = w_0 + w_1 x_1 + \dots + w_m x_m + s_i$$

Van een beslisser worden verschillende scores op verschillende alternatieven verzameld. Y_i staat voor de totaalscore van alternatief i . x_1 tot en met x_m staan voor de scores op de afzonderlijke m criteria, welke bekend zijn. S_i staat voor de storingsterm behorende bij elk alternatief met een verwachtingswaarde van 0. Met behulp van de bekende data uit de steekproef en wiskundige benaderingsmethoden zoals de kleinste-kwadratenmethode worden de parameters w voor elk criterium geschat. In feite probeert het paramorfische model dus het daadwerkelijke beslissingsgedrag van de beslisser in kaart te brengen, terwijl het subjectieve gewicht staat voor de mening van de beslisser over de door hem gebruikte gewichten. Het dan ook van belang om in te zien dat de paramorfische gewichten niet staan voor een “objectieve” waarheid. Een objectief, oftewel feitelijk juist gewicht zou een gewicht zijn dat in een bepaalde keuze-situatie precies voldoet aan de regels van een normatieve theorie binnen de besliskunde. Het paramorfische gewicht is echter een benadering van het daadwerkelijk gebruikte subjectieve gewicht. Eigenlijk zou het hier gemaakte onderscheid dus beter gedefinieerd kunnen worden als een verschil tussen paramorfisch-subjectief en

² De “stimuli” staan onder andere voor de manier van vraagstelling, of de manier hoe de alternatieven worden beschreven. Zie het deel “Onderzoek inzake de veranderlijkheid van gewichten” voor meer hierover.

subjectief-subjectieve meetmethoden. Voor het gemak zijn de termen in dit verslag afgekort en het voldoet als eenmaal in acht genomen wordt dat deze termen gaan over de expressie van de besliser en niet een objectieve waarheid. In dit verslag worden zowel subjectieve als paramorfische gewichten onderzocht.

5. Onderzoek inzake de veranderlijkheid van gewichten

Binnen de paramorfische en subjectieve gewichten komt verandering op een verschillende manier tot uiting. Later in dit verslag zal echter blijken dat de grondslag hiervan eenzelfde principe omvat. Zo wordt er bij beide meetresultaten onderscheid gemaakt tussen zowel verklaarbare als onverklaarbare verandering. Paramorfische gewichten zijn bedoeld om beslis-gedrag te modelleren. Met name dit onverklaarbare deel maakt het lastig om tot een descriptief model te komen van beslis-gedrag. Subjectieve gewichten behoren daarentegen niet tot een wiskundig model. Het is een mening die over verschillende keuze-situaties heen vergeleken kan worden. We zullen in dit verslag zien dat er net zoals bij paramorfische gewichten zowel verklaarbare redenen zijn die van invloed zijn op de subjectieve gewichten (Barlas, 2003), als onverklaarbare redenen (Palmeira en Krishnan, 2008).

5.1 Dimensies van verandering in gewichten

Zoals we in het vorige hoofdstuk gezien hebben is er een onderscheid tussen interpretaties van de veranderlijkheid van gewichten en een onderscheid tussen meetmethoden. Dit gezamenlijk zorgt ervoor dat de term “gewicht” verschillend kan worden opgevat. Het fenomeen “verandering in gewichten” kan ook verschillend worden opgevat. Voordat we verder in gaan op de mogelijke veranderingen in gewichten en het onderscheid tussen verklaarbare en onverklaarbare verandering, is het dus van belang dat we eerst specificeren wat we in dit verslag precies bedoelen met verandering in gewichten. Ook de relatieve vergelijkingsstandaard waaraan we verandering meten wordt hier behandeld.

5.1.1 Het subject van vergelijking: wat gaan we vergelijken?

In het vorige hoofdstuk kwam naar voren dat de subjectieve gewichten door de besliser zelf expliciet vermeld worden na een bepaalde keuze-situatie. Ze geven dus een perceptie van de locale gewichten waarmee de besliser na redenering een keuze-situatie mee aangaat. Paramorfische gewichten daarentegen worden afhankelijk van de gekozen alternatieven en de bijbehorende score van die alternatieven op de criteria door middel van wiskundige methoden, zoals regressie-analyse, naderhand bepaald. Deze gewichten zouden dus gezien kunnen worden als een benadering van de werkelijke gewichten die tijdens de keuze worden gehanteerd. Een grotere steekproef zal zorgen voor een nauwkeurigere benadering. Volgens Barlas (2003) is er nogal eens een verschil tussen de gewichten die daadwerkelijk tijdens een keuze gebruikt worden en de perceptie van de besliser over de gewichten die hij gebruikt. De daadwerkelijk gebruikte gewichten staan hier voor een nauwkeurige benadering van het paramorfische gewicht, terwijl de perceptie van de besliser over de gebruikte gewichten staat voor het subjectieve gewicht.

De achterliggende gedachte hierachter is dat een besliser bij een beslissing te maken krijgt met zowel een “accuracy goal” als een “justification goal”. De besliser wil zowel een accurate keuze maken als de keuze kunnen verantwoorden aan anderen. Het maken van een accurate

keuze wordt door de onderzoekers beschreven als “het maken van de keuze die je het liefst wil”. Gezien de sterke effecten van verleidelijke attributen tijdens de keuze, denkt de beslisser dat deze een accurate keuze maakt wanneer op deze criteria gefocust wordt, terwijl de beslisser in feite misleid wordt. Verleidelijke attributen krijgen dus een hoger gewicht in keuzes dan in de subjectieve evaluaties. De subjectieve evaluatie is waar men naderhand het gewicht wat hij of zij in de keuze gebruikt denkt te hebben specificceert.

Het doel om een beslissing te verantwoorden zorgt er ook nog eens voor dat rationele attributen in subjectieve evaluaties een hoog gewicht krijgen en een lager gewicht tijdens de keuze. Dit effect wordt nog eens versterkt wanneer beslissers de ervaringen die zij in hun leven hebben opgedaan in acht nemen. Tijdens het nemen van de beslissing versterken de ervaringen over het algemeen namelijk de perceptie van een positief gevoel bij verleidelijke criteria en een negatief gevoel bij rationele criteria. Ten slotte is de beschikbare informatie over de criteria nog van invloed. Het blijkt dat beslissers, wanneer ze zelf kunnen beslissen over welk criterium ze informatie willen hebben, over het algemeen eerst informatie over rationele criteria willen zien. Deze informatie kan gezien worden als de score van de alternatieven op de criteria. Het gevolg hiervan is dat na het zien van enkele rationele criteria de keuze hier al op gebaseerd wordt, zodat de overige verleidelijke criteria waar geen informatie over gevraagd is weerstaan worden. Hierdoor zullen de paramorfische gewichten van de rationele criteria hoger uitvallen dan de subjectieve gewichten.

Al met al impliceert de veranderlijkheid in deze dimensie van gewichtanalyse dat het van belang is om hier ook een afbakening te maken. Gebleken is dat de beslisser geen goed inzicht heeft in de door hem gebruikte gewichten van criteria tijdens de keuze-situatie.

Criteria-instabiliteit is echter niet iets wat zich voordoet tussen paramorfische en subjectieve gewichten, maar het gaat om paramorfische ofwel subjectieve gewichten zelf. Het is voor criteria-instabiliteit niet van belang in hoeverre de benadering van de daadwerkelijke gewichten bij een keuze overeenkomen met de perceptie van de beslisser van gebruikte gewichten, maar juist met de verschillen van enerzijds de daadwerkelijk gebruikte gewichten en anderzijds de perceptieve evaluaties.

De vraag rijst nu waarmee deze afzonderlijke “soorten” gewichten dan mee vergeleken moeten worden om verandering te constateren.

5.1.2 De vergelijkingsstandaard: waarmee gaan we vergelijken?

Er zijn verschillende manieren om verandering van gewichten te definiëren. In de literatuur wordt over het algemeen gesproken over de consistentie en inconsistentie van het keuzegedrag van de beslisser. Consistentie is hier een relatief begrip: keuzes zijn al dan niet consistent vergeleken met een bepaalde standaard. Dit houdt in dat het beslis-gedrag al dan niet voldoet aan bepaalde regels of een bepaalde regelmaat. Kleinmuntz (1990) onderscheidt twee soorten van consistentie die verschillen in hun relatieve standaard waar het beslis-gedrag mee vergeleken wordt. Zo is er “logische consistentie”. Hierbij wordt de consistentie van de beslisser vergeleken met de regels van een bepaalde normatieve theorie binnen de besliskunde. Als voorbeeld nemen we de utiliteitstheorie, die een norm stelt voor het gedrag van een beslisser in situaties van onzekerheid. Deze theorie heeft meerdere axioma's, waarvan een het axioma van “transitiviteit” is. Deze stelt dat wanneer bij een keuze-situatie tussen alternatieven L, M en N, M wordt verkozen boven L en N wordt verkozen boven M, N uiteindelijk verkozen moet worden boven L. De beslisser neemt op een bepaald tijdstip een

beslissing die wel of niet consistent is met deze normatieve standaard. Het is dus een alles of niets maatstaf.

Een andere vorm van consistentie kan gemeten worden wanneer men het gedrag van de beslisser over een bepaalde tijdsperiode analyseert. De beslisser kan tijdens verschillende keuze-situaties in een bepaalde mate gebruik maken van een bepaald mentaal proces. Bij “proces-consistentie” is dus wel sprake van een gradatie van consistentie, in dit geval de consistentie tussen de door de beslisser gebruikte denkschema's op verschillende tijdstippen. Een voorbeeld waar de verandering van het mentale proces tot uiting komt zijn framing-effecten, welke later in dit verslag nader worden beschreven. In het kort komen deze framing-effecten erop neer dat men een verschillend denkkader toepast en andere gewichten aan attributen toekent al naar gelang het beslissingsprobleem op een andere manier beschreven is³. Framing zorgt dus voor proces-inconsistentie, oftewel een beslisser die niet steeds hetzelfde mentale proces volgt bij het maken van beslissingen.

Onder proces-consistentie valt ook de consistentie van het keuze-gedrag tussen situaties waar de beslisser precies hetzelfde keuzeprobleem voor zich krijgt. Volgens Kleinmuntz wordt inconsistentie hier mogelijk veroorzaakt door “onverklaarbare veranderingen in oplettendheid of in (mentale) verwerkingsstrategie”.

Het onderzoek in dit verslag houdt zich bezig met proces-consistentie. We richten ons specifiek op de verandering in gewichten die tijdens een of meerdere achtereenvolgende keuze-situaties optreedt, ervan uitgaande dat externe factoren die niet behoren tot de context van de vraagstelling geen rol spelen. Tot de externe factoren behoort bijvoorbeeld het tussen de keuzes in overleggen met andere personen. Om deze externe factoren uit te sluiten, dient de beslisser zich uitsluitend met de keuze-situatie bezig te houden. Bij de in dit verslag behandelde onderzoeken wordt binnen een bepaalde reeks keuzes daarom ook geen tussentijdse pauze afgenomen. Bij meerdere achtereenvolgende keuze-reeksen wordt gebruik gemaakt van een korte pauze van bijvoorbeeld een kwartier bestaande uit ongerelateerde activiteiten, zoals bij de onderzoeken van Palmeira en Krishnan (2008). Het tijdsinterval waarop in dit verslag gewichten geanalyseerd worden is maximaal ongeveer een uur.

5.1.3 De onderzoeksdimensie

Aangezien we in de vorige paragraaf hebben gespecificeerd dat we paramorfische gewichten ofwel subjectieve gewichten vergelijken om verandering waar te nemen, volgt uit het gegeven dat we proces-consistentie onderzoeken, dat paramorfische ofwel subjectieve gewichten onderling vergeleken worden. Zo kan bijvoorbeeld verandering worden waargenomen tussen paramorfische gewichten van meerdere achtereenvolgende keuze-situaties.

Nu we vastgelegd hebben wat in dit verslag precies bedoeld wordt met verandering in gewichten, gaan we nu de ontwikkeling van onderzoeken inzake verandering in gewichten binnen deze dimensie behandelen. Deze onderzoeken hoeven echter nog niet per se onder onze voorlopige definitie van criteria-instabiliteit te vallen. Ze kunnen ons echter wel op weg helpen naar de einddefinitie.

³ In dit geval is er ook overlap met de logische consistentie, namelijk de overtreding van de normatieve regel van “descriptieve invariantie” (Tversky, Kahneman; 1986), welke stelt dat beslissingen tussen twee gelijke, doch verschillend omschreven keuzes, gelijke uitkomsten zouden moeten hebben.

5.2 Van “perfect rationeel” naar “keuze-specifieke variantie”

Hoewel er vanaf het laatste deel van de 20e eeuw tot op heden veel onderzoeken in de gedrags en neuro-wetenschappen zijn die het thema gewichtsverandering bij keuzes behandelen, is criteria-instabiliteit is nog niet zo lang erkend als een apart onderzoeksgebied. Een lange tijd ging men er zelfs van uit dat gewichten niet per keuze-situatie veranderden en dat als deze veranderingen optraden, dat ze dan als “ongelukjes” werden bestempeld waar verder geen aandacht aan zou moeten worden besteed omdat ze geen afbreuk zouden doen op het algemeen geaccepteerde principe. Dit veranderde toen er verschillende methoden werden ontwikkeld om verandering van gewichten te verklaren. Hierna werden er echter nog wel onverklaarbare veranderingen geconstateerd, welke apart worden behandeld als criteria-instabiliteit. De ontwikkeling van onderzoek naar veranderlijkheid in gewichten is dus als volgt samen te vatten:

1. Perfect rationele beslisser (paragraaf 5.2.1)
2. Onderkende oorzaken van verandering in gewichten (paragraaf 5.2.2 en 5.2.3)
3. Criteria-instabiliteit (paragraaf 5.3 en verder)

Deze ontwikkeling wordt nu nader toegelicht. De belangrijkste methoden van punt 2 worden gegeven. Toegewerkt wordt naar punt 3.

5.2.1 De onhoudbaarheid van “perfecte rationaliteit”

Tot en met de helft van de twintigste eeuw ging men er over het algemeen van uit dat er slechts sprake was van globale gewichten en dat een beslisser deze persoonlijke waarden in keuze situaties ten alle tijde perfect rationeel wist te vertalen in belangrijkheid van de gewichten. Een voorbeeld hiervan zijn de klassieke theorieën over beslissingen en economische vraagstukken van Von Neumann et al. (1947) en Savage, L.J. (1954), waar uitgegaan wordt van strikte normatieve methoden. Pas vanaf 1960 wordt toegewerkt naar een onderkenning van de locale interpretatie van gewichten. Zo is een belangrijke ontdekking binnen de besliskunde, die mede verwezenlijkt is door empirische bevindingen tussen omstreeks 1960 en 1990, het onderkennen van “procedurele variantie” en “descriptieve variantie” (Payne, Bettman, 1992; Shafer, 1986). Procedurele variantie houdt in dat bij verschillende, doch normatief gezien gelijke methoden van vraagstelling om de preferentie van alternatieven te achterhalen, voorkeuren voor alternatieven kunnen verschillen. Zo is het bijvoorbeeld mogelijk om de toegekende gewichten aan criteria die te maken hebben met een opbrengst in geld te beïnvloeden. Dit kan bewerkstelligd worden door bij een keuze tussen twee loterijen met een gelijke verwachte opbrengst de vraagstelling te veranderen zodat de beslisser een monetaire prijs moet verbinden aan aan de loterijen (zie “Compabiliteit van de schaal” later in dit verslag). Verder valt het gebruik van verschillende methoden uit de besliskunde om gewichten te bepalen onder een verschil in vraagstelling. Descriptieve variantie gaat niet over de manier van vraagstelling, maar over de representatie van het beslissingsprobleem. Het principe houdt in dat bij verschillende, normatief gelijke, methoden van omschrijving van het beslissingsprobleem, keuze-preferentie kan variëren. Dit verschilt in zoverre met procedurele variantie dat er bij descriptieve variantie wel telkens een manier van vraagstelling gebruikt wordt als “welk alternatief kies je?”. Slechts de omschrijving van de alternatieven verschilt, en niet de vraag zelf. Een voorbeeld hiervan zijn “Framing

effecten”, welke tot uiting komen in situaties waar alternatieven enerzijds als winst en anderzijds als verlies omschreven kunnen worden. Deze verschillende manieren van omschrijven zorgen ervoor dat men een ander soort gedrag aanneemt om dit beslissingsprobleem aan te pakken, wat zorgt voor een verandering in toegekende gewichten.

De bevindingen die aan deze ontdekking ten grondslag liggen zijn de eerste die deze verandering in gewichten onderkennen, in tegenstelling tot de klassieke methoden die met hun normatieve methoden uitgingen van een perfect rationele beslisser, wiens voorkeur niet zou afhangen van “procedure” en “descriptie”. Nu volgen de belangrijkste bevindingen op dit gebied.

5.2.2 Procedurele variantie

In deze paragraaf zullen de belangrijkste bevindingen worden beschreven die zorgen voor procedurele variantie.

Inter- en intra-methodische inconsistenties

Zoals in de vorige paragraaf beschreven kunnen toegekende gewichten verschillen wanneer gebruik gemaakt wordt van verschillende methoden voor bepaling van de gewichten (onderzoek is gedaan tussen ratio, swing, trade-off en pricing-out methoden, zie Weber, Borchering, 1992).

Onderzoek naar consistenties binnen methoden van gewichtbepaling zelf toont aan dat gewichten bij de ratio-, swing- en trade-off-methoden per methode kunnen verschillen. Borchering, Eppel en von Winterfeldt (1991) hebben over de genoemde methoden een studie gedaan waarbij respectievelijk 30%, 50% en 67% van de gewichtbepalingen als inconsistent werden gezien. Inconsistentie werd gemeten met de mate waarin ordinale en cardinale meningen binnen een bepaalde methode verschilden. Bij een ordinale mening wordt een ranking-lijst van criteria opgesteld, terwijl bij de cardinale mening de getallen die staan voor de belangrijkheid van de gewichten daadwerkelijk worden bepaald. Mijns inziens zou dit echter ook tot inter-methodische inconsistenties kunnen worden gerekend, aangezien het gebruik van een ordinale- dan wel cardinale schaal gezien kan worden als het gebruik van verschillende methoden. In ieder geval blijkt hieruit dat procedurele invariantie niet geldt, aangezien een perfect rationele beslisser altijd gelijke gewichten zou toekennen, ongeacht de methode waarop de gewichten worden samengesteld.

Choice vs Matching

Om het verschil tussen “choice” en “matching” te laten zien volgt een voorbeeld.

Er is sprake van een keuze-situatie tussen beleidsvormen over verkeersongevallen.

Beleidsvorm A leidt tot 570 doden en kost 12 miljoen euro. Beleidsvorm B leidt tot 500 doden en kost X miljoen euro. Bij “matching” zou worden gevraagd om tot een bedrag te komen voor X zodat je tot een gelijke preferentie komt voor beleidsvorm A en B, terwijl bij “choice” het bedrag voor X gegeven zou zijn en gevraagd zou worden de preferente beleidsvorm te kiezen. Het blijkt dat ten einde de beslisser voor B te laten kiezen, bij “choice” een substantieel hoger bedrag voor X kan worden gevraagd dan dat de beslisser in het geval van “matching” zelf zou aandragen. Het blijkt hier dus dat het voor de beslisser belangrijkste

attribuut, in dit geval het aantal doden, meer gewicht toegewezen krijgt wanneer gebruikt gemaakt wordt van een “choice” ten opzichte van een “matching” manier van vraagstelling. Deze verandering in gewicht wordt ook wel het “prominence effect” genoemd (Tversky et al., 1988), omdat het verschil in gewicht zich voordoet bij het voor de beslisser meest belangrijke (prominente) attribuut. Dit effect wordt onder andere toegewezen aan het gebruik van verschillende berekenschema's in het brein. Er zijn vele onderzoeken die deze berekenschema's trachten te beschrijven. Ik behandel hier het overzicht van Fischer en Hawkins (1993) die het meest globale beeld lijkt te scheppen.

Volgens Fischer en Hawkins kijkt de beslisser bij “choice” eerst of alternatief A minstens zo goed scoort op alle criteria en ten minste beter scoort op een criterium. In dat geval wordt A gekozen. Als er nog geen dominante optie is, wordt gekeken of een van de alternatieven een veel hogere score heeft op een bepaald attribuut, zodanig dat dit voordeel leidt tot de preferentie van dit alternatief. Als er geen alternatief is met een cruciaal voordeel op een criterium, wordt ten slotte gebruik gemaakt van een lexicografische strategie. Dit houdt in dat de beslisser altijd het alternatief kiest op basis van een specifiek attribuut, ofwel een set subattributen hiervan en de trade-offs tussen de overige attributen buiten beschouwing laat (Campbell et al., 2006). Fischer en Hawkins stellen dat hier alleen de belangrijkste attributen in acht worden genomen, terwijl de overige criteria buiten beschouwing worden gelaten.

De theorie van Fischer en Hawkins wordt dus gebruikt om het eerder beschreven prominence effect te verklaren. In de onderzoeken over het prominence effect bij “choice” en “matching” is telkens sprake van een uitgangspositie waarbij slechts twee criteria aan de orde zijn. Wanneer er 3 of meer criteria in acht worden genomen, worden volgens Tversky et al. (1988) de gewichten van de criteria minder bij de keuze betrokken. Er wordt steeds meer gebruik gemaakt van de zogenaamde “meerderheidsregel”, die door Zhang (2005) gedefinieerd wordt als de keuze van het alternatief dat het hoogste scoort op het grootste aantal criteria. Al met al blijkt het dat de cognitieve analyse bij “choice” beperkt is, omdat slechts een deel van de criteria in acht genomen wordt.

In tegenstelling tot de voorgaande voorbeelden van “choice”, zou bij “matching” de cognitieve analyse rekening houden met intervallen van attribuutwaarden, welke door middel van *trade-offs* onderling worden afgewogen. Zo kan het bijvoorbeeld zijn dat de beslisser afweegt in hoeverre hij bereid is een bepaalde hoeveelheid extra geld uit te geven per aantal verkeersdoden dat hierdoor daalt. Mogelijk zijn er naast het prominente attribuut nog andere criteria van toepassing ten opzichte waarvan ook wordt afgewogen. Omdat nu ook deze minder belangrijke criteria in acht genomen worden, zorgt dit volgens Tversky et al. uiteindelijk voor een uniformere verdeling van gewichten onder de attributen, in tegenstelling tot de situatie bij “choice”, waar juist een bepaald prominent criterium een extra hoog gewicht krijgt. Er ontstaat dus een verschil in gewichten wanneer de ene keer volgens een “choice”-methode en de andere keer volgens een “matching”-methode de preferentie van alternatieven wordt gevraagd.

“Choice” taken zouden dus beredeneerd worden volgens een ordinale manier, terwijl “matching” meer kwantitatieve analyse zou vergen, wat lijkt op de cardinale methode (Payne et al., 1992). In die zin lijkt het dus overlap te hebben met de hiervoor beschreven inter-methodische inconsistenties van gewichtsbepaling.

Compatibiliteit van de schaal

Volgens Tversky en Thaler (1990) is het mogelijk de gewichten van de beslisser te beïnvloeden wanneer verschillende eenheden, of schalen, gebruikt worden in de vraagstelling. Het blijkt dat wanneer een persoon tussen twee loterijen met een gelijke verwachte opbrengst moet kiezen, waar loterij G een grote kans heeft op een kleine opbrengst en loterij K een kleine kans op een grote opbrengst, de persoon over het algemeen risicomijdend gedrag vertoont en loterij G boven K verkiest. Echter wanneer gevraagd wordt een zekerheidsequivalent aan beide loterijen te verbinden, valt deze bij K over het algemeen hoger uit, wat tegenstrijdig is met het risicomijdende gedrag van de eerste keuze⁴. Het zekerheidsequivalent van een loterij is een monetaire waarde van een zodanige hoeveelheid dat het met zekerheid verkrijgen van deze waarde een gelijke preferentie heeft ten opzichte van de loterij. Verwacht zou worden dat wanneer men in eerste instantie voor G kiest, deze ook een hogere zekerheidsequivalent zou ontvangen. Dat dit blijkbaar niet het geval is, ligt volgens de onderzoekers aan het feit dat wanneer een monetaire zekerheidsequivalent gevraagd wordt, het criterium dat ook monetair is (in dit geval de opbrengst) een hoger gewicht krijgt. Het relatief hogere gewicht van dit criterium volgt uit het feit dat het gedrag van de beslisser van risicomijdend meer naar risicozoekend verschuift, wat impliceert dat het criterium “kans” een lager gewicht krijgt.

Het hiervoor beschreven effect zou volgens Tversky en Thaler niet alleen voor een monetaire schaal gelden, maar voor alle schalen die in de vraagstelling gebruikt worden en al dan niet overeenkomstig is met de schaal van het criterium. Een voorbeeld van een andere schaal komt uit een van de studies van de onderzoekers. Hierin lieten zij beslissers een voorspelling doen van het cijfer van een student in een bepaald vak, gegeven de huidige resultaten van die student van twee andere gerelateerde vakken. De resultaten werden zowel uitgedrukt in “beoordeling” (A tot D+) als in een “rangorde ten opzicht van de klas” (1 t/m 100). Het bleek dat de groep die de resultaten in “beoordeling” moest voorspellen, meer waarde hechtte aan de vooraf gegeven data over de “beoordeling” en dit compatibiliteitseffect was ook aanwezig bij de groep die een voorspelling moest maken in termen van de rangorde. Hier werd een hoger gewicht gegeven aan het criterium van de rangorde.

Al met al is het nu duidelijk dat de manier van vraagstelling een effect kan hebben op de gewichten die een persoon aan attributen verbindt. Dit gegeven maakt dat het principe van “procedurele invariantie” niet houdbaar is.

5.2.3 Descriptieve variantie

In deze paragraaf zullen de belangrijkste bevindingen worden beschreven die zorgen voor descriptieve variantie.

Framing-effecten

De manier van “framing” gaat over de presentatie van het beslissingsprobleem. De effecten hiervan vinden vooral hun oorsprong in keuze-situaties waar risico en kansen aan de orde zijn, ook wel loterijen genoemd. Deze loterijen zijn al eerder aan bod geweest bij “compatibiliteit van de schaal”. Hier veranderde het beslisgedrag van risicomijdend naar

⁴ Zie Bijlage 1 voor een nadere omschrijving van de termen “risicomijdend” en “risicozoekend”.

risicozoekend wanneer de ene keer gevraagd werd naar een keuze tussen twee loterijen en de andere keer naar het zelf formuleren van het tweede alternatief met een gelijke preferentie, namelijk de zekerheidsequivalent. Bij “framing” zijn bij beide achtereenvolgende keuze-situaties de alternatieven reeds verstrekt en hoeft alleen een keuze te worden gemaakt. Hier kan echter ook een verschuiving van risicomijdend naar risicozoekend (of andersom) worden veroorzaakt, namelijk door middel van het gebruik van verschillende “frames”.

Over het algemeen worden twee soorten frames onderscheiden. Enerzijds frames die ervoor zorgen dat alternatieven als *winsten* worden gezien en anderzijds frames die als gevolg hebben dat alternatieven als *verliezen* worden gekenmerkt (Payne et al., 1992). Volgens Tversky en Kahneman (1986) is het zo dat bij representaties van het beslissingsprobleem waar alternatieven als winsten worden gekenmerkt, de beslisser over het algemeen risicomijdend gedrag vertoont, terwijl bij alternatieven die gezien worden als verliezen men juist risicozoekend te werk gaat. Er volgt een voorbeeld uit de onderzoeken van Tversky en Kahneman. De fractie proefpersonen die voor een bepaalde optie heeft gekozen is tussen vierkante haken achter het alternatief aangegeven.

Keuze-situatie 1 (126 proefpersonen):

Ga ervan uit dat je in totaal 300 euro bezit. Je moet nu kiezen tussen:

- een zekere *winst* van 100 euro [72%]
- 50% kans om 200 euro te *winnen* en 50% kans om niets te *winnen* [28%]

Keuze-situatie 2 (128 proefpersonen):

Ga ervan uit dat je in totaal 500 euro bezit. Je moet nu kiezen tussen:

- een zeker *verlies* van 100 euro [36%]
- 50% kans om niets te *verliezen* en 50% kans om 200 euro te *verliezen* [64%]

Hoewel in beide instanties de uitkomst van de keuze bestaat uit een zekere 400 euro enerzijds en een gelijke kans tussen 300 en 500 euro anderzijds, is er in keuze-situatie 1 duidelijk sprake van risicomijdend gedrag, terwijl in keuze-situatie 2 meer risicozoekend gedrag vertoond wordt. In keuze-situatie 2, waar het referentiepunt van de zekere optie als verlies wordt beschreven, gaat men een stuk minder waarde hechten aan zekerheid. Deze verandering in risico-benadering representeert dus in zoverre een verandering in gewichten dat het attribuut “kans” een verschillende mate van belangrijkheid krijgt toegewezen.

Informatie-presentatie effecten

Het blijkt uit te maken of attribuutwaarden in woorden of in getallen worden uitgedrukt. Ook de manier waarop bijvoorbeeld grafieken worden weergegeven (attributen of alternatieven op de verschillende assen) en de schijn van compleetheid van de informatie heeft effect op het beslis-gedrag en dus onder andere de relatieve belangrijkheid van de criteria (Payne et al., 1992).

Ten slotte hebben Fischer et al. (1987) aangetoond dat wanneer een attribuut weergegeven wordt als een proxy-attribuut, dit proxy-attribuut meer gewicht krijgt dan zou moeten volgens de normatieve en rationele utiliteits-theorie. Een proxy-attribuut is een indirecte graadmeter voor de evaluatie van een alternatief en het zou in feite vervangen kunnen worden door een

meer fundamenteel attribuut. Zo is de “respons-tijd” bij de brandweer bijvoorbeeld een proxy-attribuut van het meer fundamentele attribuut “voorkomen brandschade”. Het blijkt dat het cognitief meer moeite vergt om proxy-attributen te evalueren door de verbanden die moeten worden gelegd met het fundamentele attribuut dat het proxy-attribuut representeert. Als de relatie met het fundamentele lastig te herleiden is, wordt nogal eens gebruik gemaakt van een heuristiek die het proces versimpelt. De gedachte achter deze heuristiek, aangeduid als het *relative importance model*, is dat de beslisser vaak geen perfect inzicht heeft in de onderliggende kansen van de relaties tussen het proxy- en het fundamentele attribuut. Zo kan de beslisser het proxy-attribuut zien als een directe vertaling van het fundamentele attribuut, terwijl het proxy-attribuut in werkelijkheid maar 50% van de gevallen van het fundamentele attribuut representeert. Zo werd in een onderzoek van Fischer et al. het criterium “ziekte door milieuvervuiling” als fundamenteel attribuut gebruikt en “mate van milieuvervuiling” als proxy-attribuut. Dit proxy-attribuut kreeg een relatief hoger gewicht toegewezen door de proefpersonen.

Het blijkt dat de onderlinge relatie van het proxy-attribuut ten opzichte van het fundamentele attribuut vaak te hoog wordt ingeschat; het proxy-attribuut wordt als meer representatief gezien voor het fundamentele attribuut dan het daadwerkelijk is. Hierdoor krijgt het proxy-attribuut een hoger gewicht toegewezen.

5.3 Een nieuw fenomeen van verandering in gewichten

De hiervoor besproken voorbeelden van procedurele en descriptieve variantie worden in de literatuur veelal bestempeld als “response mode biases”, oftewel denkfouten die zich voordoen door verschillende methoden van vraagstelling en/of beschrijving van het beslissingsprobleem. Deze fouten ontstaan onder andere doordat gebruik gemaakt wordt van een cognitief proces waar meer emotionele overwegingen in acht genomen worden. Dit gaat ten koste van de rationaliteit, zoals bij framing-effecten het geval is (De Martino et al., 2006). Hieruit blijkt dus dat er niet zoiets bestaat als de perfect rationele mens. We zijn al makkelijk om de tuim te leiden door simpele methodieken.

Wat echter wanneer een reeks keuzes twee identieke keuzes bevat? Hier zou de mens toch rationeel genoeg moeten zijn om deze gelijk te beoordelen? We zullen nu enkele onderzoeken behandelen waaruit het tegendeel blijkt. Blijkbaar zijn er nog andere effecten die ervoor zorgen dat de beslisser tijdens een reeks keuzes zijn of haar gewichten verandert. We zijn een stapje dichter bij criteria-instabiliteit.

6. Onderzoeken met nader te verklaren veranderingen in gewichten

Nu zullen we de resultaten van enkele onderzoeken behandelen die te maken lijken te hebben met onze voorlopige werkdefinitie van criteria-instabiliteit en ons wellicht verder kunnen helpen naar de uiteindelijke definitie. De onderzoeken hebben dus veelal betrekking op de eerder omschreven willekeurige variatie in gewichten. Ze zijn echter niet altijd specifiek gewijd aan dat onderwerp. De titel “Onderzoek over criteria-instabiliteit” kan dus eigenlijk gezien worden als “Onderzoek waar onder andere het principe van onze voorlopige werkdefinitie van criteria-instabiliteit behandeld wordt”.

Het eerste onderzoek van Levine et al. (1996) bouwt voort op onderzoeken naar de invloed

van het overdenken, of beredeneren, van beslissingen. Dit onderzoek wordt als eerste behandeld, omdat het naast een mogelijke indicatie van criteria-instabiliteit ook het eerder genoemde onderscheid tussen globale- en locale gewichten nader verklaart, wat ons zal brengen op de eerste subdefinitie van criteria-instabiliteit.

6.1 Verandering van gewichten bij beredeneren ofwel het bedenken van criteria

Hoewel men ertoe geneigd is aan te nemen dat het overdenken zorgt voor een betere afweging van criteria, wat voor kwalitatief betere beslissingen zou zorgen, weerleggen een aantal genoemde studies in Levine et al. deze aanname door te stellen dat het beredeneren juist een verstoring zou hebben op het beslissingsproces. Dit zou onder andere tot uiting komen in een verandering van gewichten die de beslisser hanteert. De begrippen “kwaliteit” en de “verstoring” hiervan worden als volgt gedefinieerd:

Kwaliteit werd onder andere gemeten op basis van vergelijking van de genomen beslissingen met de beslissingen van experts op dat gebied en de evaluatie van de mate van tevredenheid met de genomen beslissingen na een bepaalde tijdsperiode. De verstoring zorgt voor een verminderde kwaliteit van de beslissing. Specifiek wordt er op gewezen dat de beredenerende beslisser, wanneer deze na het maken van een keuze na bijvoorbeeld 2 of 3 weken gevraagd wordt over de tevredenheid van de gemaakte keuze, vaak meer ontevreden blijkt over de gemaakte keuze dan de niet beredenerende beslisser. De spijt die de beredenerende beslisser heeft van de gemaakte keuze, staat hier dus voor de genoemde verstoring op de kwaliteit.

De studies van Levine et al. bouwden hierop voort. In hun onderzoeken behandelden ze de analyse van gezichten van mensen. Door de beslisser moesten keuzes worden gemaakt die gingen over in hoeverre foto's van menselijke gezichten gelijkenis vertoonden en in hoeverre de gezichten “leuk” gevonden werden. Verder moest door de beredenerende beslisser aangegeven worden wat de onderbouwing was van de keuze, of anders gezegd, de gebruikte criteria moesten worden aangegeven. We zullen verder in dit verslag wanneer we spreken over “beredeneerde keuzes” keuzes bedoelen waar criteria door de beslisser zelf moeten worden onderscheiden.

Levine et al. omschrijven het door hen geobserveerde effect door middel van het “variable weight model” voor de variabiliteit van gewichten. Dit model neemt als uitgangspunt dat mensen bij het nemen van beslissingen over het algemeen relatief variabel en inconsistent zijn met het selecteren en wegen van criteria. Zo zouden mensen die hun beslissingen beredeneren volgens Levine en Wilson (1991) slechts een subset van alle mogelijke criteria in acht nemen. Deze subset zou een relatief hoog gewicht krijgen. Aan het gewicht van de overige criteria wordt in het verslag echter geen aandacht besteed. Uit de onderzoeken van Levine is echter te herleiden dat ze er vanuit gaan dat de overige criteria niet in acht worden genomen, wat lijkt op een toekenning van een gewicht van 0. Aangezien dit punt onderbelicht is, maar wel van belang is voor ons onderzoek, gaan we hier later nog op in.

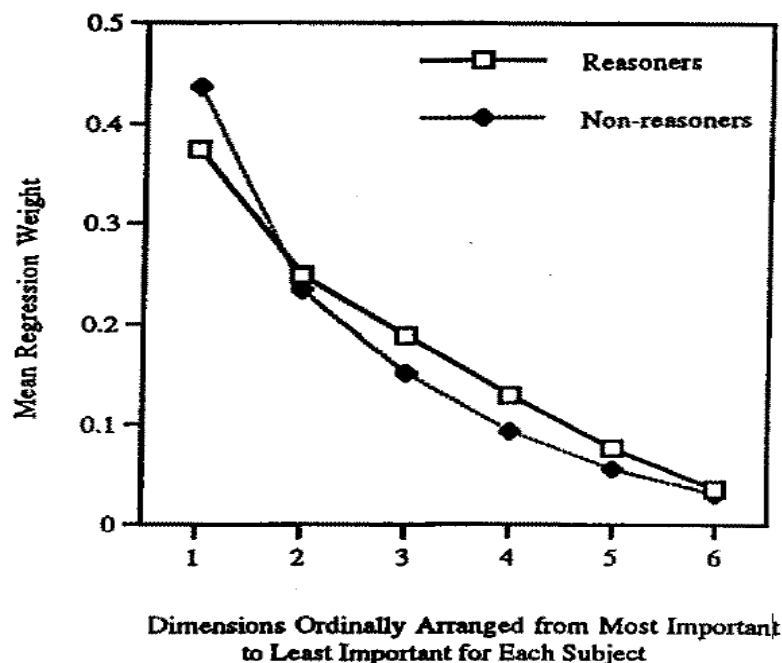
Een perfect rationele beslisser zou het gewicht van een criterium wat hij of zij over het algemeen belangrijk vindt niet laten variëren bij verschillende alternatieven. Het blijkt echter dat de beredenerende beslisser juist wel varieert in de toewijzing van belangrijkheid. Door

deze variatie in gewicht die per keuze per criterium optreedt, zouden over meerdere keuze-situaties gezien de verschillen tussen de gemiddelde gewichten tussen de criteria juist minder moeten zijn. Een voorbeeld kan dit nader toelichten:

Een bepaalde persoon kijkt bij de beoordeling van een gezicht op “leukheid” over het algemeen naar de mooiheid van de ogen en de lach. Na tien gezichten te hebben beoordeeld, blijkt dat de niet-beredenerende beslisser bij nagenoeg elke beslissing de ogen en de lach als zeer belangrijk beschouwt, terwijl bijvoorbeeld de mond en neus minder belangrijk worden geacht. Over het algemeen gezien zijn de ogen en de lach hier dus een stuk belangrijker vergeleken met de neus en de mond. Daarentegen is de beredenerende beslisser variabel met het toewijzen van gewichten per keuze-situatie. De ene keer wordt beredeneerd dat de ogen en lach wel van belang zijn, maar de andere keer blijken toch de oren een grote rol te spelen. Omdat de criteria de ene keer een hoog en de andere keer een laag gewicht toegewezen krijgen, is het algemene resultaat na de tien keuzes dat de criteria onderling minder verschil in belangrijkheid vertonen: er is een kleinere kans dat de waarde die de persoon globaal gezien aan criteria hecht ook daadwerkelijk in zijn keuze naar voren komt.

In een overzicht van de resultaten van het onderzoek (Figuur 1) is inderdaad te zien dat de gewichten van “Reasoners” na meerdere keuze situaties gemiddeld een kleiner verschil vertonen, wat te zien is aan het vlakkere patroon in de grafiek.

Figuur 1



Hoewel dit vlakkere patroon van gewichten wel wordt verklaard door het “variable weight model”, is het meer een onderbouwing van de verminderde kwaliteit van de keuze (zie hiervoor ook het deel over “lange termijn waarde versus keuze-specifiek gewicht” verderop in dit verslag). Een meer treffende conclusie die gegeven wordt ter ondersteuning van het

“variable weight model”, wordt gedaan wanneer de onderzoekers de resultaten van gelijke (als door de onderzoekers onderkend “gelijke”) alternatieven vergelijken. Het blijkt namelijk dat de beredenerende beslisser significant meer variabel is in de waardering van nagenoeg gelijke alternatieven.

Wanneer we het “variable weight model” en de gerelateerde resultaten uit het onderzoek van Levine et al. zien als vorm van criteria-instabiliteit, is het belangrijk in te zien hoe in dit onderzoek de term, al dan niet indirect, beschreven wordt. Zoals zojuist gesteld wordt de variabiliteit van waardering van gelijke alternatieven gezien als onderbouwend voor het model. Dit lijkt te gaan om een combinatie van gewichten en scores. Onderzoek werd uitgevoerd met foto's van gezichten van mensen als alternatieven. Wanneer bij hetzelfde gezicht de ene keer “neus” en “ogen” als belangrijk worden beschouwd voor de eindbeslissing en de andere keer “ogen” en “mond”, kan het namelijk best zijn dat naast het specifieke gewicht van de “mond”, ook de score is verhoogd. Van de eerder gegeven voorbeelden van procedurele- of descriptieve variantie lijkt hier echter geen sprake, aangezien de identieke alternatieven twee keer hetzelfde zijn gepresenteerd.

Al met al wordt de variabiliteit van gewichten in dit onderzoek verklaard door onder andere de subset in criteria die de beredenerende beslisser selecteert. We zullen hier in hoofdstuk 6.1.2 verder op in gaan. Eerst zullen we naar aanleiding van de voorgaande onderzoeken toewerken naar de eerste sub-definities van criteria-instabiliteit.

6.1.1 Belangrijkheid van attributen: lange termijn waarde versus keuze-specifiek gewicht

Het hiervoor besproken onderzoek van Levine et al. sprak over de kwaliteit van keuzes en de mogelijke verstoring hiervan. De kwaliteit van keuzes lijkt in het onderzoek gerelateerd te zijn aan de “waarden” van personen. Zo kan een persoon over het algemeen veel waarde hechten aan het milieu. Vervolgens mag aangenomen worden dat deze persoon het van belang acht dat het attribuut “milieuvriendelijkheid”, of afgeleiden daarvan, bij beslissingen een hoog gewicht krijgen. Dit zou immers zorgen voor een “kwalitatief goede keuze”, als door Levine et al. omschreven. Hun onderzoek wees echter uit, dat deze keuze-specifieke gewichten van elkaar, en dus ook van de “waarde”, kon verschillen, wat ervoor zorgde dat een persoon later spijt had van de genomen beslissing.

Onderscheid in belangrijkheid van attributen kan dus gemaakt worden tussen een vorm van lange-termijn waarde en een keuze-specifiek gewicht. De lange-termijn waarde kunnen tussen personen verschillen omdat zijn verschillende achtergronden hebben die deze waarden hebben gevormd. Tijdens een leven kunnen deze waarden per persoon veranderen (Milton Rokeach, 1979). De gewichten op het niveau van keuzes kunnen afwijken van deze waarden.

Het gaat echter niet alleen om waarden waar deze keuzes van af kunnen wijken. Stel namelijk dat iemand de loterij wint, dan gaat zijn lange termijn gewicht van het criterium “kosten” omlaag. Dit zijn waarden op zeer lange termijn. Er zijn echter ook waarden op kortere termijn. Zo kan iemand een keer haast hebben en snel nog even wat te eten willen kopen omdat hij of zij veel trek heeft. Gemak en tijdsbesparing gaan hier voor, waardoor het

gewicht voor “kosten” omlaag gaat. In dit verslag richten we ons echter op keuze-situaties waar deze achtergrond-context niet meer verandert. De stimuli waar de beslisser aan is blootgesteld beperken zich louter tot de tijdspanne waarin de beslissingen worden genomen. Dit is in overeenstemming met de conclusies van Goldstein eerder in dit verslag ten aanzien van globale en lokale gewichten, en het brengt ons op de eerste graadmeter van criteria-instabiliteit:

Criteria-instabiliteit speelt zich af op het niveau van locale keuze-specifieke gewichten en laat verandering in globale lange-termijn waarden buiten beschouwing.

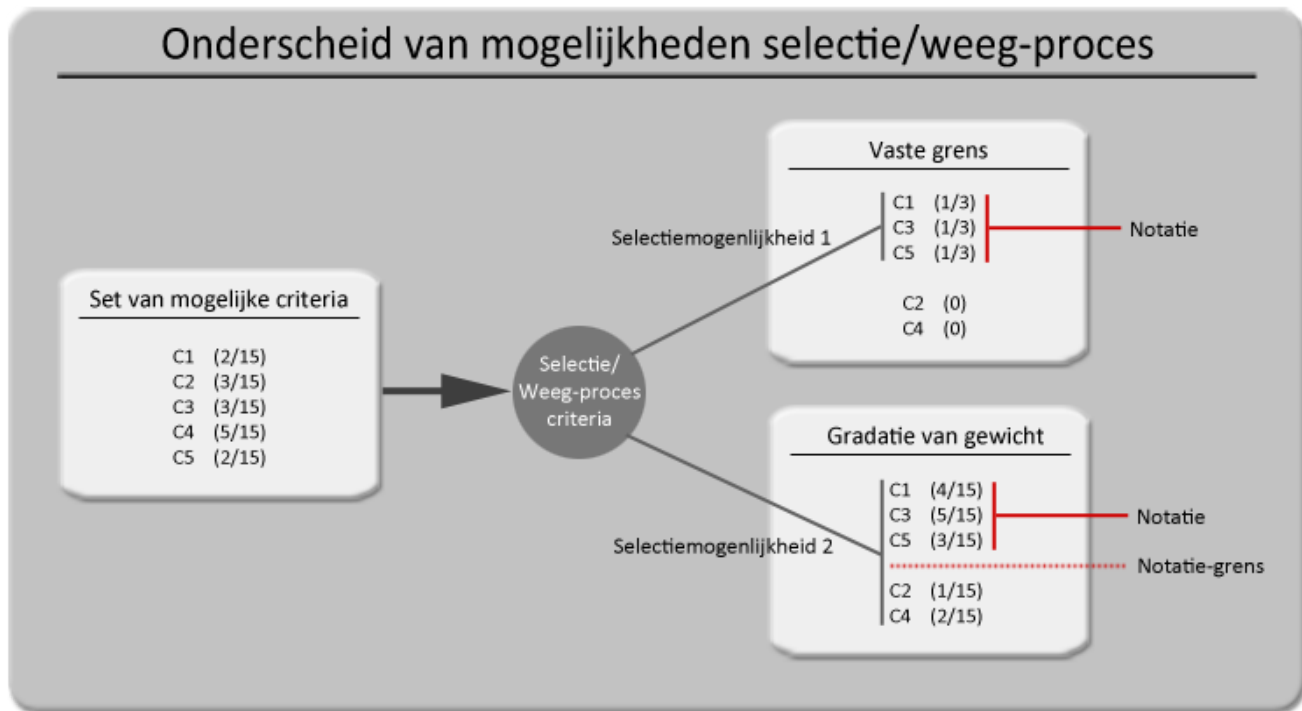
6.1.2 Variabiliteit van gebruikte criteria en hun gewichten

In deze paragraaf zullen we de selectie van criteria vertalen in een vorm van verandering in gewichten, waardoor deze selectie in feite een kandidaat wordt voor criteria-instabiliteit. Al eerder is vermeld dat Levine et al. stellen dat het “variable weight model” een verklaring is voor het irrationele keuze-gedrag van de beredenerende beslisser. De beredenerende beslisser focust zich op een subset van criteria en geeft deze op lokaal niveau een hoger gewicht dan deze op globaal niveau zou bezitten. Het is van belang om in te zien wat hier bedoeld wordt met de term “focussen op”. In de experimenten die Levine et al. uitvoerden, vroegen zij aan een deel van de participanten om tenminste drie criteria te noteren waarop hun keuze gebaseerd was. Per keuze-situatie werd echter niet gevraagd naar een gradatie van belangrijkheid.

De term “focussen” kan hier dus betekenen dat alleen de door de beslisser genoteerde subset van criteria tijdens de beslissing in acht werden genomen werd, terwijl de rest van de set van mogelijke criteria compleet buiten beschouwing blijft, en dus een gewicht van 0 toegewezen krijgt. Aangezien er niet gevraagd werd om een gradatie van gewicht van de genoteerde n criteria, wordt er dan vanuit gegaan dat elk gebruikt criterium een gewicht van $1/n$ ontvangt.

Wanneer we echter de onderzoeken bestuderen waar Levine et al. op voortbouwen, namelijk die van Wilson (1991), wordt het volgende verteld over de gekozen subset van criteria: “Er wordt bij een beslissing die wordt beredeneerd dikwijls gefocust op criteria die bij vorige beslissingen minder gewicht toegewezen kregen, terwijl criteria die bij de vorige beslissingen veel gewicht ontvingen bij deze beslissing relatief worden ondergewaardeerd.”. Dit lijkt al veel meer op een situatie waarbij er tussen verschillende keuze-situaties gezien sprake is van een gradatie van gewichten van waarden tussen de 0 en 1, in plaats van een 0 voor “niet in de gekozen subset” en een $1/n$ van “wel in de gekozen subset”. Dit onderscheid is visueel gemaakt in de afbeelding hieronder.

Figuur 2



Toelichting bij Figuur 2:

Aan de linker kant staat de lijst van criteria die de beslisser bij deze beslissing mogelijk in acht kan nemen. De gewichten van deze criteria gaan uit van een bepaalde situationele context, bijvoorbeeld de situatie dat je op dat moment weinig geld beschikbaar hebt. Ook zijn de descriptieve- en procedurele-variantie hier niet van belang, aangezien de achtereenvolgende keuze-situaties identiek zijn. Het enige effect waar we hier vanuit gaan is dus het “focussen” op criteria, waar de onderzoeken van Levine over spreken.

Gegeven deze set van criteria, stelt Levine dat door invloed van beredeneren een beperkt aantal criteria geselecteerd en genoteerd wordt (“Selectiemogelijkheid 1”). De mogelijkheid die ik hier naar voren wil brengen is die van een gradatie in toegewezen gewicht (“Selectiemogelijkheid 2”). Hier worden bepaalde criteria bewust onderscheiden, omdat de belangrijkheid hiervan boven een bepaalde grens uitkomt (“Notatie-grens”) en vervolgens genoteerd. De andere criteria worden mogelijk nog wel in acht genomen, maar met een zeer laag gewicht.

Volgens Fasolo et al. (2006) en Heerkens (2003, p172), die de relatieve belangrijkheid van in acht genomen criteria bestudeerden, kennen beslissers veelal gewichten toe die waarden aannemen in de vorm van de “stap-functie”. Deze functie heeft enkele hoge en meerdere kleine waarden, met weinig tot geen waarden daartussen. Dit is een mogelijke verklaring voor de notatie-grens, waarboven de criteria dermate belangrijk worden gevonden dat ze genoteerd worden.

Hoewel de “selectie” van criteria, zoals er in de studies van Levine over wordt gesproken, in eerste instantie dus geen overlap lijkt te hebben met de verandering van gewichten zoals bij onze werkdefinitie van criteria-instabiliteit, brengt het voorgaande onderscheid hier verandering in. Het blijkt dat wanneer men vanuit een andere invalshoek kijkt op de door

Levine gegeven definitie van het “focussen” op criteria, er wel degelijk sprake lijkt te zijn van het fenomeen van criteria-instabiliteit zoals wij tot nu toe hebben besproken, namelijk criteria-instabiliteit als verandering in gewichten van criteria.

De vraag rijst nu wat de oorzaak is voor het verschil in criteria waar op gefocust wordt en het bijkomende verschil in toegewezen gewicht.

6.1.3 Kan het selectie/weeg-proces van criteria verklaard worden?

Aansluitend op Levine wijden de onderzoeken van Palmeira en Krishnan (2008) zich onder andere ook aan de variabiliteit van criteria bij beredeneerde keuzes, oftewel keuzes waar criteria door de beslisser zelf onderscheiden moeten worden. Naast dat zij evenals Levine het bestaan van instabiliteit van de geselecteerde set van criteria erkennen, geven zij een globale omschrijving van het selectie/weeg-proces zelf. Voordat we hierop ingaan, behandelen we eerst de strekking van hun onderzoek.

Palmeira en Krishnan wijden een deel van hun onderzoek aan wat zij beschouwen als criteria-instabiliteit. De onderzoekers stellen dat zowel bepaalde variabiliteit in de gebruikte criteria (onderzoek 4a) alsmede in de toegewezen gewichten (onderzoek 4b) vallen onder dit principe. In dit opzicht lijkt het op het eerder behandelde onderzoek van Levine, waar de willekeurige selectie van een subset van criteria per keuze-situatie wordt vertaald in de variabiliteit van gewichten. In de onderzoeken van Palmeira en Krishnan beperkt criteria-instabiliteit met betrekking tot de gewichten zich specifiek tot keuze-situaties waar de criteria zowel qua omschrijving als de vooraf gegeven score gelijk zijn. Met de nieuwe inzichten uit de vorige paragraaf weten we nu ook dat bij onderzoek 4a sprake geweest kan zijn van verandering in gewichten en niet alleen van gebruikte criteria.

Ten aanzien van criteria-instabiliteit op het niveau van geselecteerde criteria, wordt aangetoond dat een beslisser op twee verschillende tijdstippen (interval van 5 minuten) wel degelijk verschillende criteria selecteert ter evaluatie van eenzelfde alternatief.

In het onderzoek wordt verder de notie gemaakt dat de keuzes van een beslisser zowel een vaststaande- als een random component bevatten. De vaste waarde zou kunnen worden verkregen door een gemiddelde te nemen van een serie van toegekende gewichten, terwijl de random waarde staat voor de criteria-instabiliteit ten aanzien van gewichten die zelfs over zeer korte tijdsintervallen kan optreden. Later in dit verslag zullen we een wiskundige benadering van criteria-instabiliteit behandelen die gebaseerd is op dit principe.

Een punt van discussie is in hoeverre de onderzoeken van Levine enerzijds en Palmeira en Krishnan anderzijds daadwerkelijk verschillen in hun omschrijving van criteria-instabiliteit. Hoewel Palmeira en Krishnan de variabiliteit in de selectie van criteria als een apart staand fenomeen behandelen, zou je ook kunnen zeggen dat “selectie” staat voor “toewijzing van belangrijkheid” en dus samenhang vertoont met situaties waar de criteria reeds vooraf gegeven zijn en niet hoeven te worden gespecificeerd door de beslisser. Bij het onderzoek van Palmeira en Krishnan werd gevraagd de drie belangrijkste attributen te geven die van invloed waren op het nemen van de beslissing. Dit houdt echter in dat er nog andere attributen in het spel geweest konden zijn. Deze worden in dit onderzoek buiten beschouwing gelaten, terwijl ze best bij de beslissing van belang geweest konden zijn, echter met een lager gewicht. In die zin kan “het verschil in gebruikte criteria” dus vervangen worden door “het

verschil in belangrijkheid van criteria”, wat staat voor verandering in gewichten. Dit is in feite ook de manier waarop in het onderzoek van Levine de geselecteerde subset van criteria wordt vertaald met de meest belangrijke criteria op dat moment, in acht nemende andere criteria die een lager gewicht krijgen.

Nu we vastgesteld hebben dat in onderzoek 4a van Palmeria en Krishnan criteria-instabiliteit eenzelfde principe omvat als bij Levine, kunnen we stellen dat de oorzaken die Palmeira en Krishnan geven ook van toepassing zijn op Levine.

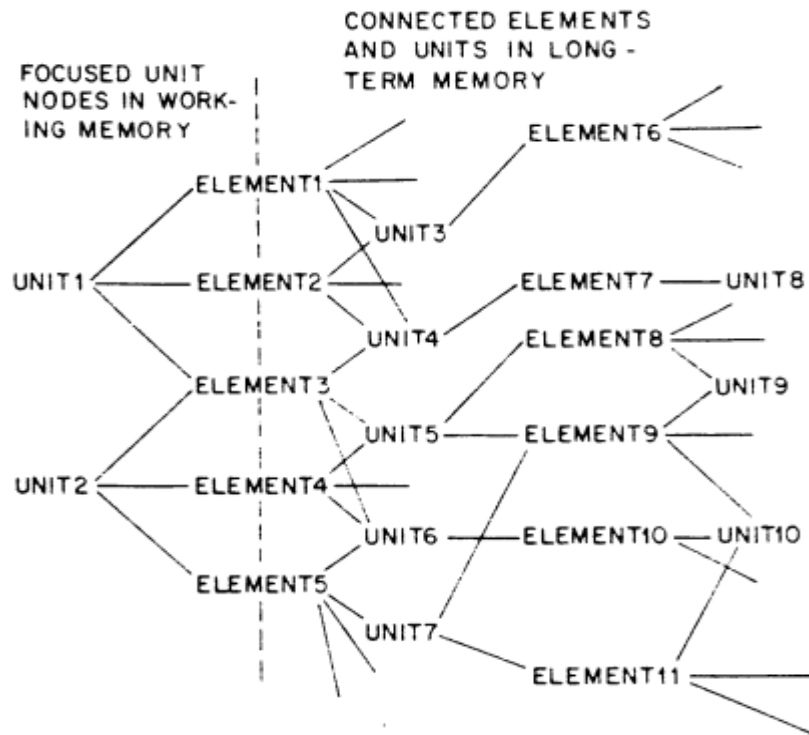
Palmeira en Krishnan stellen dat het selectie/weeg-proces afhangt van de “toegankelijkheid” van de criteria. Een toegankelijker criterium wordt tijdens een keuze-situatie eerder geselecteerd en ontvangt dus een hoger gewicht. Deze toegankelijkheid zou afhangen van de onderscheidendheid van de alternatieven. Zo zou een beslisser wanneer deze de twee vakantiesoorten “zonvakantie” en “wintersport” vergelijkt, vrijwel altijd de criteria “weertype” en “sportiviteit” in acht nemen. Palmeira en Krishnan onderscheiden naast deze toegankelijke criteria echter ook criteria die willekeurig geselecteerd worden. Waarom treedt de willekeurige verhoogde toegankelijkheid van deze overige criteria hier op?

6.1.4 Het selectie/weeg-proces van criteria als hersen-activiteit

In deze paragraaf zullen we beschrijven welke hersen-activiteiten een rol spelen bij het maken van een beslissing. Verschillende invloeden op deze activiteiten kunnen een verklaring zijn voor het eerder behandelde selectie/weeg-proces.

Bij de term “toegankelijkheid” in het onderzoek van Palmeria en Krishnan verwijzen zij naar de onderzoeken van Nedungadi (1990). Hierin wordt de “toegankelijkheid” omschreven als de mate waarin de hersenen bepaalde knooppunten aanspreken bij het maken van een beslissing. Deze knooppunten representeren in de onderzoeken van Nedungadi de mogelijke alternatieven die in de “consideration set” terecht kunnen komen tijdens het beslissingsproces. Deze set is volgens Palmeira en Krishnan te vergelijken met de geselecteerde criteria waar wij het over hebben. Anderson (1983) geeft een overzicht van het betreffende cognitieve proces.

Figuur 3



Een onderscheid tussen het werkgeheugen en het lange-termijngeheugen, beide bestaande uit knooppunten (units) welke onderling door elementen verbonden zijn.

Aan de linkerkant is het werkgeheugen te zien. Dit deel van het geheugen houdt zich bezig met de situationele gebeurtenis, in dit geval de beslissing. Het verwerkt alle informatie omtrent de beslissing die op dat moment beschikbaar is, zoals contextuele en externe omgevingsfactoren, informatie over het te bereiken doel en ook sporen van informatie uit het lange-termijngeheugen. Deze overlap met het lange-termijngeheugen komt in Figuur 2 tot uiting in de elementen op de stippellijn. De "toegankelijkheid" van de elementen zorgt voor een grotere kans dat relevante data uit het lange-termijngeheugen wordt verkregen.

Naast de onderscheidendheid van de alternatieven, die Palmeria en Krishnan opmerkten, stelt Nedungadi dus dat de kans dat de knooppunten in de "consideration set" belanden, afhangt van de sterkte van de elementen, oftewel verbindingen. Een van de belangrijkste factoren die Nedungadi en Anderson geven voor het sterker maken van de verbindingen, is het zogendaamde "priming". Dit houdt in dat de beslisser voor zijn beslissing uit wordt blootgesteld aan een stimulus die een sterke associatie heeft met, of een representatie is van het knooppunt. Dit komt in het dagelijks leven bijvoorbeeld voor in de vorm van reclames. Men versterkt bij het zien van een merk, al dan niet onbewust, de verbindingen naar het knooppunt dat in je hersens dat merk representeert. Bij een volgende keuze zal je sneller aan dat merk denken.

Hoewel dit principe louter over alternatieven lijkt te gaan, is eerder al verteld dat Palmeira en Krishnan stellen dat dit ook opgaat voor het selectie- en weegproces van criteria bij een keuze-situatie. Volgens hen wordt de verandering van het selectie- en weegproces tussen verschillende achtereenvolgende keuze-situaties veroorzaakt door een verandering in de

“toegankelijkheid” van de criteria. Dit houdt dus in dat de elementen die verbonden zijn met de knooppunten die deze criteria in de hersens representeren, tussen keuze-situaties in van sterkte kunnen veranderen. Verandering in gewichten zou in een meer elementaire vorm dus gezien kunnen worden als een verandering in sterktes van de elementen.

Nedungadi en Anderson stellen dus dat dit voornamelijk door “priming” bewerkstelligd kan worden. Het effect van “priming” is sterker wanneer het meer recentelijk is opgetreden.

Wanneer we dit op de onderzoeken van Palmeira en Krishnan betrekken, zou het dus zo kunnen zijn dat zelfs op het korte tijdsinterval van zo'n 15 minuten tussen de achtereenvolgende beslissingen in, “priming” zou kunnen zorgen voor een verandering in gewichten. Helaas geraken we hier echter op een dwaalspoor. Hoewel het niet te weerleggen valt, wijst er ook niets op dat de knooppunten tussentijds geactiveerd worden, aangezien de beslissers bij Palmeira en Krishnan tussen de beslissingen in bezig waren met “ongerelateerde activiteiten”.

Wellicht dat Levine meer helderheid kan scheppen. Bij Levine wordt een verklaring voor het verschil in selectie en weging van criteria gezocht in de mogelijkheid dat het beredeneren zorgt voor een verstoring van een automatisch proces dat zich zou voordoen in de hersens. Dit is te vergelijken met een bepaalde handeling die iemand zo vaak gedaan heeft, dat het na een tijdje nagenoeg automatisch gaat. Wanneer men dan aan deze persoon vraagt hoe de handeling precies in zijn werk gaat, kan het voor deze persoon lastig zijn het stap voor stap uit te leggen. De persoon voert deze handeling nu telkens uit zonder te beredeneren. Wanneer hij of zij tijdens de handeling gaat beredeneren, kan dit het automatische proces verstoren.

Strack en Deutsch (2004) geven een gedetailleerder beeld van dit verschijnsel. Zij maken onderscheid tussen een reflectief en een impulsief systeem binnen de hersenen. Kennis binnen het impulsieve systeem zou zijn ontstaan over een lange tijdsperiode, naar aanleiding van opgedane ervaringen. Het bestaat voornamelijk uit associatieve verbindingen, waarbij bijvoorbeeld de entiteit van een “oud persoon” verbonden is met de karakteristiek “traagheid”. Het reflectieve systeem daarentegen is van een meer beredenerende aard. Gebaseerd op betekenissen, begrippen, feiten en data uit het impulsieve systeem wordt hier deductief naar een beslissing toegewerkt.

Strack en Deutsch wijzen op het belang van de interactie tussen het reflectieve en het impulsieve systeem. Hoewel beslissingen wel alleen door het impulsieve systeem genomen kunnen worden, is voor een reflectieve beslissing altijd data uit het impulsieve systeem nodig. Er zijn meerdere punten in het reflectieve beredeneerproces waar interactie is met het impulsieve systeem (zie Figuur 3).

mate waarmee bij reflectieve beslissingen informatie uit het impulsieve systeem wordt gehaald, zelfs op korte tijdsintervallen nogal variabel kan zijn. Dit is bij een impulsieve beslissing veel minder het geval. Dit komt dus overeen met de visie van Levine, wanneer we een reflectieve beslissing als een beredeneerde beslissing zien en een louter impulsieve als een niet beredeneerde.

Het blijkt dat de interactie tussen het reflectieve en het impulsieve systeem afhangt van de eerder genoemde “toegankelijkheid” van deze verbindingen. Dit is nauw verwant aan de eerder behandelde onderzoeken van Anderson (1983) en Nedungadi (1990). Strack en Deutsch komen echter nog met een oorzaak van het verschil in “toegankelijkheid” tussen verschillende keuze-situaties in, namelijk het zogenaamde “anchoring”. Dit fenomeen houdt in dat een bepaalde beslissing beïnvloed kan worden door zowel externe informatie als ook genomen beslissingen voorafgaand aan deze beslissing (Tversky, Kahneman, 1974). Voortbouwend op dit onderwerp en van belang voor ons onderzoek is het “Selective Accessibility Model” van Mussweiler en Strack (1999), welke dieper ingaat op de “toegankelijkheid” van data in de hersenen. Het model verklaart onder andere dat beslissingen die, ook in een korte tijdspanne, na mekaar genomen worden, invloed kunnen hebben op de “toegankelijkheid” van data die bij deze beslissingen in acht genomen wordt. Een onderzoek van Mussweiler en Strack geeft een goed voorbeeld van “anchoring” waar dit principe tot uiting komt.

Eerst wordt aan de beslisser gevraagd of de gemiddelde temperatuur in Duitsland hoger of lager is dan 20 graden Celcius. Gezien het feit dat 20 graden Celcius voor Duitsland aan de hoge kant lijkt, gaat de beslisser volgens de “anchoring”-theorie over het algemeen op zoek naar bewijs dat de gemiddelde temperatuur hoog is. Volgens het “Selective Accessibility Model” beïnvloedt dit de “toegankelijkheid” van de data die met dit bewijs geassocieerd wordt. Wanneer de beslisser vervolgens als taak kreeg in een veelheid aan lettercombinaties bepaalde woorden te onderscheiden, werden de woorden “zon”, “strand” en “heet”, welke geassocieerd worden met de ondersteuning van het bewijs van een hoge temperatuur, sneller herkend dan woorden zoals “winter”, “kou” en “sneeuw”.

Tenslotte wordt een verklaring gegeven waarom we vast liepen bij het eerder besproken “priming”. Volgens Mussweiler en Strack gaat het eerder besproken “priming” namelijk om externe effecten, zoals (sluik)reclame en kunstmatig verkregen stimulerende data. Bij “anchoring” echter wordt het “priming” door de beslisser zelf tijdens het beslisproces bewerkstelligd en kan vervolgens invloed hebben op achtereenvolgende keuzes. Met name omdat het “priming”, oftewel het vergroten van de “toegankelijkheid”, bij zelf-generatie veel intensiever en dieper in de hersens verwerkt wordt, blijkt dit veel meer effect te hebben op achtereenvolgende beslissingen dan dat externe factoren zouden hebben.

Samenvattend blijkt dus dat een bepaalde beslissing ook invloed kan hebben op volgende beslissingen. Het “anchoring” principe zorgt voor een verschil in “toegankelijkheid” van de data in het impulsieve systeem. Bij beredeneren kunnen dus steeds verschillende data uit het impulsieve systeem gehaald worden. Dit komt overeen met onze eerdere bevindingen omtrent het criteria-selectie en weeg-proces, waarbij de subset van criteria waar op gefocust werd verschild. Waar priming over externe effecten gaat, biedt “anchoring” een verklaring voor de interne effecten bij de beslisser, oftewel het beredeneren. Bij het niet beredeneren wordt langs de impulsieve weg gegaan. De geselecteerde data, ofwel criteria, zijn hier niet

afhankelijk van de “toegankelijkheid” van relaties welke in korte tijdspanne veranderen, maar zijn juist zeer stabiel.

Wanneer we dit betrekking laten hebben op het onderzoek van Levine, zou het “anchoring” principe als volgt zijn werking kunnen hebben. Stel iemand beoordeelt achtereenvolgens een aantal gezichten. De eerste 4 gezichten hebben duidelijk een blijde gelaatsuitdrukking, waardoor de beslisser deze bij de “leuke” gezichten indeelt. Blijde gezichten worden over het algemeen herkend door de stand van de mond en de wenkbrauwen. De “anchor” is nu gericht op deze criteria, waardoor deze een hoger gewicht hebben gekregen⁵. De “toegankelijkheid” van deze kenmerken is toegenomen, wat tot uiting komt in de verandering van gewichten. Stel nu dat na deze reeks blijde gezichten een neutraal gezicht getoond wordt. Hoewel de stand van de mond en de wenkbrauwen niet veel verschilt van de vorige blijde gezichten, kan het geringe verschil in score al zorgen voor een verschil in preferentie, aangezien de gewichten verhoogd waren. Het nieuwe gezicht wordt nu al snel als “redelijk boos kijkend” gezien en dus door de beslisser bij de “niet leuke” gezichten ingedeeld. Dit zelfde geldt wanneer dit neutrale gezicht na 4 achtereenvolgende boze gezichten getoond worden. Het gezicht wordt nu sneller bij de “leuke” ingedeeld.

Gezien het feit dat het onderzoek van Levine niet op de manier is opgebouwd dat er per se een reeks op elkaar lijkende gezichten na elkaar weergegeven wordt, is de “anchoring” zeer variabel en wellicht onvoorspelbaar. We zullen hier later verder op in gaan. We kunnen er echter wel zeker van zijn dat “anchoring” überhaupt optreedt. Er zijn namelijk vele voorbeelden van “anchoring” waarbij de kenmerken van een enkel alternatief binnen een keuze de volgende keuze beïnvloeden. Een hiervan is het eerder gegeven voorbeeld over de gemiddelde temperatuur van Mussweiler en Strack.

Nu we beredeneerde beslissingen, oftewel beslissingen waarbij gebruikte criteria zelf moesten worden onderscheiden, behandeld hebben en de mogelijke oorzaak gevonden hebben in “anchoring”, gaan we nu kijken naar mogelijke criteria-instabiliteit bij identieke keuze-situaties waar de criteria wel degelijk vooraf zijn aangereikt.

6.2 Verandering van gewichten bij vooraf aangereikte criteria

Al eerder in dit verslag werden de onderzoeken van Palmeira en Krishnan (2008) genoemd. Zij verdelen criteria-instabiliteit in twee mogelijkheden. Enerzijds het gebruik van verschillende criteria bij situaties waar deze door de beslisser zelf onderscheiden moeten worden (studie 4a), en anderszijds verandering in gewicht van reeds gegeven criteria (studie 4b). Zie hiervoor ook bijlage A. Wij hebben bediscussieerd dat de eerste optie van studie 4a, namelijk de variabiliteit van gebruikte criteria, ook kan vallen onder criteria-instabiliteit met betrekking tot gebruikte gewichten wanneer dit vanuit een ander oogpunt bekeken wordt. Wat nu van belang is, is dat Palmeira en Krishnan ook situaties hebben onderzocht waar criteria en scores reeds vooraf aangereikt zijn.

Variatie op het niveau van gewichten werd onderzocht in studie 4b, waar gebruik gemaakt

⁵ Al eerder in dit verslag, aan het eind van paragraaf 6.1, is de interpretatie van “belangrijkheid van een criterium” als combinatie van gewicht en score naar voren gekomen. “Anchoring” zou in weze dus ook effect kunnen hebben op de perceptie van de score. Deze overweging wordt aan het eind van dit verslag, in paragraaf 7.2.4, nader behandeld. Voor nu gaan we uit van het effect dat zich louter in de gewichten manifesteert.

werd van dezelfde alternatieven, echter met een verschillende schaal die voor de score gebruikt werd. De vooraf gegeven score van het alternatief op elk attribuut werd de eerste keer namelijk gegeven op een schaal van 1 t/m 100, terwijl bij de tweede keuze-situatie een schaal van 1 t/m 5 gebruikt werd. Zo werd een score van 80 in de eerste vragenlijst vertaald naar een score van 4 in de tweede, om zo de indruk te wekken dat de beslissing nieuw was en niet gelijk aan een van de vorige. Het blijkt dus dat de gelijkenis in score betwist kan worden, aangezien gebruik gemaakt werd van verschillende schalen.

Als we er echter net als Palmeira en Krishnan van uit gaan dat dit verschil in schaal niet zorgt voor afwijkingen die hun onderzoeksresultaten zouden kunnen beïnvloeden, is ook hier sprake van criteria-instabiliteit aangezien er werd aangetoond dat de toegewezen gewichten verschil vertoonden. De vraag rijst of “anchoring” ook hier een mogelijke verklaring voor criteria-instabiliteit is.

Bij het beredeneren van criteria hebben we gesteld dat “anchoring” kan optreden bij het achtereenvolgens beoordelen van gezichten. We hebben tot nu toe nergens voorwaarden gesteld voor “anchoring” en ook in de literatuur is dit niet terug te vinden. We kunnen echter met de informatie tot nu toe het volgende stellen. “Anchoring” treedt tussen achtereenvolgende keuze-situaties op aan de hand van criteria van de alternatieven bij die keuze-situaties. Het is van belang dat de alternatieven enigszins vergelijkbaar zijn. Zo zijn twee gezichten goed met elkaar te vergelijken, maar wanneer achtereenvolgens een gezicht en een auto op leukheid moeten worden ingedeeld, slaat de set van mogelijke criteria op hele andere kenmerken. De vraag waar de grens precies ligt vanaf wanneer “anchoring” optreedt zodanig dat gebruikte gewichten significant veranderen voert te ver voor dit verslag. Ik ga er echter van uit dat achtereenvolgende alternatieven redelijk vergelijkbaar moeten zijn in hun karakteristieken. Zo vallen een houten stoel en een houten tafel van dezelfde meubulair-set net wel binnen deze grens.

We hebben bij beredeneren en Livine gesteld dat het zelf overdenken van criteria zorgt voor een versterkt effect van “anchoring”. Echter kan volgens onze beschrijving “anchoring” ook optreden bij vooraf aangereikte criteria. Er is namelijk nog steeds sprake van dezelfde focus op bepaalde criteria die meespelen bij de genomen beslissing. Vanwege het feit dat hier geen sprake is van zelf-generatie van criteria zou het effect hier wel minder en mogelijk niet significant zijn.

Om te bepalen of bij Palmeira en Krishnan ook sprake kan zijn van “anchoring”, dienen we nog een voorwaarde van “anchoring” op te stellen. Deze is meer voor de hand liggend dan de vorige en vereist geen bepaalde grenswaarde. De voorwaarde houdt in dat er een alternatief in de reeks achtereenvolgende keuzes moet zijn dat verschil vertoont op een of meerdere criteria. Een verklaring van verandering in gewichten door “anchoring” is namelijk niet mogelijk wanneer achtereenvolgens alleen gelijke gezichten worden getoond. Er moet na een reeks blanke gezichten een neutraal gezicht worden getoond om de focus op de neus en de wenkbrauwen tot uiting te laten komen.

Nu naar het onderzoek 4b van Palmeira en Krishnan. Hier werd op twee momenten zes keuzes afgenomen, waarbij tussen de momenten een pauze van vijf minuten gehanteerd werd. De zes keuzes bestonden uit keuzes tussen twee producten van een bepaalde productgroep. Deze productgroep varieerde per keuze. Zie hiervoor de tabel in bijlage A. Ten

eerste kan er geen “anchoring” optreden tussen de achtereenvolgende keuzes binnen de keuze-reeks. De productgroepen zijn namelijk niet vergelijkbaar. Bij het tweede keuzemoment werd precies dezelfde lijst afgenomen, maar zoals eerder gezegd met een verschil in schaal. We kunnen stellen dat de alternatieven identiek zijn. Dit voldoet niet aan de opgestelde tweede voorwaarde, welke stelde dat “anchoring” alleen kan optreden wanneer er een alternatief is dat verschilt van het andere alternatief.

De oorzaak bij Palmeria en Krishnan is dus niet direct te herleiden uit het “anchoring” principe, zoals het geval was bij de eerder besproken situaties waar criteria door de beslisser zelf onderscheiden moesten worden. Ook als we de oorzaak zoeken in de “toegankelijkheid” in bredere zin, zijn de invloedsfactoren hierop in dit geval onduidelijk. Ze worden ook niet door Palmeira en Krishnan vermeld, en staan dus nog open voor verder onderzoek. De onderzoekers stellen echter wel dat de hier genoemde verandering in gewichten willekeurig optreedt en slechts te modelleren is met een kansenfunctie, zoals in de volgende paragraaf zal worden behandeld.

Nu we hebben aangetoond dat criteria-instabiliteit zowel kan optreden bij situaties waar criteria door de beslisser zelf onderscheiden moeten worden als situaties waar de criteria reeds vooraf gegeven zijn, rest ons nog onze werkdefinitie van het fenomeen in meer detail uit te werken. Hiertoe kijken we om te beginnen naar de manier waarop huidige modellen die het beslis-gedrag van een beslisser modelleren rekening houden met criteria-instabiliteit. Een treffend voorbeeld hiervan zijn de recente vindingen van Fischer, Luce en Jia (2010), waarbij criteria-instabiliteit wordt gerepresenteert door een parameter die staat voor de “onsystematische” kansenvariabele. We zullen hier nu nader op in gaan.

6.3 Een model voor de variatie in gewichten

Een in de besliskunde vaak gebruikt model om het afwegingsproces van de beslisser in kaart te brengen is het “weighted additive value model” (Payne, Bettman, 1992).

$$U_i = \sum_j w_j v_{i,j}$$

Met $0 \leq w_j \leq 1$ (1) en $\sum_j w_j = 1$ (2)

Voor elk alternatief i , kan zo de samengestelde waarde U worden berekend door voor elk criterium j het gewicht van dat criterium w_j te vermenigvuldigen met de score v_{ij} van het alternatief i op criterium j . Een variant hierop is het “additive error model”, waar voor elk alternatief bij de uitkomst van de som een random fout-waarde wordt opgeteld. Deze fout-waarde is echter niet afhankelijk van de keuze-specifieke stimuli. Fischer, Luce en Jia (2010) komen vervolgens met een variatie op deze formule waar de fout-waarde wel degelijk gerelateerd is aan scores en gewichten. In feite lijkt deze ontwikkeling dus inherent aan de eerder besproken ontwikkeling in de onderkenning van procedurele variantie. Eerst werden afwijkingen van de perfecte rationaliteit als onverklaarbaar en onafhankelijk van keuze-specifieke stimuli gezien (random fout-waarde), terwijl later werd ingezien dat veranderingen in toegewezen gewicht en score wel degelijk kunnen afhangen van keuze-specifieke stimuli. Ze zijn dan ook tot een bepaald punt nauwkeuriger te modelleren in vergelijking met de

eerdere onafhankelijke random foutwaarde.

De formule die de onderzoekers hebben ontwikkeld is als volgt:

$$U_i = \sum_j W_j v_{i,j}^a$$

In dit RandMAU – Random weights Multi Attribute Utility – model dient a als een parameter die rekening houdt met de onzekerheid met betrekking tot de toewijzing van scores. Zo stellen de onderzoekers dat het cognitief makkelijker is om een score toe te wijzen aan een criterium van een alternatief dat een extreme waarde heeft dan dat deze een waarde in het midden heeft. Het kan bijvoorbeeld zijn dat iemand een medische behandeling moet kiezen, waarbij een criterium staat voor de mate van pijn waar je de rest van je leven onder moet lijden. De extreme waarden zijn “heel veel pijn” en “geen pijn” en in het midden is er “gemiddelde pijn”. De extreme waarden zijn makkelijk te beoordelen op score, maar de score van “gemiddelde pijn” is onduidelijk. Zo kan het zijn dat “gemiddelde pijn” meer richting “heel veel pijn” ofwel meer richting “geen pijn” wordt beoordeeld. Deze structurele afwijking in beoordeling van score van een criterium wordt met de parameter a gemodelleerd.

Wat voor ons echter van belang is is de nieuwe kijk die hier op parameter W van de gewichten wordt gegeven. Deze is in de laatste formule met een hoofdletter genoteerd om het verschil met het standaard model aan te geven. In het model dat de onderzoekers geven wordt namelijk uitgegaan van een random variatie in het gewicht per criterium.

In het onderzoek wordt het fenomeen onderzocht dat door de onderzoekers “preferentie-onzekerheid” wordt genoemd. Het wordt omschreven als de onzekerheid over welke “waarde” zou moeten worden toegewezen aan een alternatief. Deze waarde is een combinatie van scores en gewichten zoals zojuist behandeld in het door de onderzoekers opgestelde model. Aangetoond wordt verder dat preferentie-onzekerheid zorgt voor inconsistente beslissingen. Dit houdt in dat dezelfde beslisser op twee verschillende tijdstippen met dezelfde criteria een andere waarde verbindt aan eenzelfde alternatief. Bij een gering verschil tussen waarden van alternatieven kan een gering verschil in de waarde van een bepaald alternatief al zorgen voor een andere keuze. Ook stellen zij dat preferentie-onzekerheid wordt weergegeven door onder andere de random variatie in gewichten in het RandMAU model.

De rest van het onderzoek van Fischer, Luce en Jia gaat in op versterkende factoren van preferentie-onzekerheid. Er wordt echter niet verder in gegaan op de herkomst van de genoemde random variatie in gewichten, hoewel dit voor mijn onderzoek naar criteria-instabiliteit wel degelijk van belang is. Pas als we de literatuur achter het RandMAU-model bestuderen, komen we uiteindelijk uit op een onderzoek Jia et al. (1998). Hierin stellen de onderzoekers dat de onzekerheid met betrekking tot de gewichten die een beslisser hanteert het beste weer te geven is als een kansverdeling van mogelijke gewichten. De verdeling van het gewicht W , als in de RandMAU-formule, kan gezien worden als het resultaat van een random fout-waarde opgeteld bij het “echte” gewicht. Dit “echte” gewicht kan beschouwd worden als het globale gewicht dat iemand hanteert, in combinatie met de effecten van procedurele en descriptieve variantie. Deze effecten zijn context-gerelateerd, ook wel “systematisch” en dus te modelleren. De random fout-waarde staat voor de rest aan effecten dat “onsystematisch”, minder goed te modelleren is en dit lijkt dus meer richting criteria-instabiliteit te gaan.

Dit onsystematische effect zou volgens de onderzoekers het beste tot uiting komen in de zogenaamde Dirichlet-model. Hierin hebben de afzonderlijke gewichten een Beta-kansverdeling, welke verschillende kenmerken heeft van de random variatie in gewichten. De functie van de Beta-verdeling heeft vorm-parameters α en β . Het gebruik van een Dirichlet-model in combinatie met een specifieke Beta-kansverdeling die de juiste waarden voor α en β heeft brengt belangrijke voordelen met zich mee waardoor deze combinatie zich leent voor het modelleren van gewichten. De belangrijkste voordelen die ter onderbouwing gegeven worden zijn:

- De Beta-verdeling is gedefinieerd over het interval $(0,1)$. Dit houdt in dat $0 \leq w_j \leq 1$. Verder moet de som van de gewichten in de Dirichlet-verdeling uitkomen op 1. Dit betekent dat de Dirichlet-gewichten voldoen aan de voorwaarden (1) en (2) van het additive value model.
- De specifieke Beta-verdeling heeft een enkele top in de kansdichtheidsfunctie, welke staat voor het gemiddelde, of de verwachtingswaarde. Deze waarde komt overeen met het “echte” gewicht van de beslisser.
- Bij een specifieke Beta-verdeling kan bijvoorbeeld bereikt worden dat wanneer het “echte” gewicht klein is, bijvoorbeeld 0.1, de verdeling een scheefheid naar rechts vertoont. Dit houdt in dat er meer ruimte is voor random fouten die zorgen voor een keuze voor hogere gewichten dan lagere dan 0.1. Dit geldt ook wanneer de juiste parameters gebruikt worden bij een gewicht dat de 1 nadert, zodat er een scheefheid naar links is. In het midden ($w = 0.5$) is de kansdichtheid symmetrisch met parameters $\alpha = \beta$.

De ontwikkeling van de foutwaarde bij modellen die beslis-gedrag beschrijven is dus als volgt: zonder fout, onafhankelijke random fout, RandMAU met criterium-afhankelijke fout. Het belangrijkste punt waarop de RandMAU methode van preferentie modelleren vernieuwend is is dus dat de fout-waarde van een gewicht per criterium wordt gemodelleerd en niet onafhankelijk bij de totale waarde van score en gewicht wordt opgeteld. Hoewel de foutwaarde per criterium in het nieuwe model nog steeds onderhevig is aan willekeurigheid, kunnen de parameters zo worden ingesteld dat de random variatie zich begeeft rond een bepaald “echt” gewicht, dat van tevoren bijvoorbeeld kan worden geschat na een aantal proeven. Ook de scheefheid in de Beta-verdeling lijkt meer overeen te komen met het daadwerkelijke gedrag van een beslisser.

Volgens Jia et al. (1998) is het door hen opgestelde model en de achterliggende gedachte van een random-foutwaarde die ontstaat uit preferentie-onzekerheid een aanvulling op het reeds gedane onderzoek naar de “systematische effecten” van de response-mode in MADM. Deze worden echter niet helder gedefinieerd. Wat zijn precies systematische effecten? En is de random fout-waarde in het model dan niet systematisch? We zijn bijna bij de definitie van criteria-instabiliteit.

7 De definitie van criteria-instabiliteit

Om tot de definitie van criteria-instabiliteit te komen maken we eerst een overzicht van waar binnen de in dit verslag behandelde onderzoeken criteria-instabiliteit mogelijk voorkomt (paragraaf 7.1) en vervolgens gaan we verder in op het onderscheid tussen systematisch en

niet-systematische verandering (vanaf paragraaf 7.2).

7.1 Overzicht van verandering in gewichten

We hebben in dit verslag een aantal fenomenen behandeld waarbij verandering in gewichten optreedt. Om duidelijk te maken waar precies criteria-instabiliteit op kan treden, is een visueel overzicht gemaakt.

De afbakening van criteria-instabiliteit wordt in de onderstaande figuur weergegeven. Hier zijn ter verduidelijking ook de behandelde onderzoeken ten aanzien van criteria-instabiliteit ingedeeld. De dimensie van beschikbaarheid van criteria, welk onderscheid is gemaakt in paragraaf 6.1 en 6.2, is ook in acht genomen (het onderscheid tussen de rijen in de tabel). De kolommen zijn onderverdeeld in dimensies van verschil in gewichten ten aanzien van de methoden van het achterhalen van gewichten, in lijn met de onderzoeksdimensie die we in paragraaf 5.1.3 hebben vastgesteld. Zo kunnen paramorfische gewichten onderling worden vergeleken (linker kolom), paramorfische gewichten met subjectieve (middelste kolom) en subjectieve onderling (rechter kolom).

Figuur 5

	Paramorfisch / paramorfisch	Paramorfisch / subjectief	Subjectief / subjectief
Aangereikte criteria	Procedurele en descriptieve variantie	Barlas (2003)	Palmeira en Krishnan (2008, 4b)
Criteria zelf onderscheiden	Levine (1996) beredenerende beslisser Palmeira en Krishnan (2008, 4a)	?	?
Niet aangereikt en niet onderscheiden	Levine (1996) niet-beredenerende beslisser		



Mogelijkheid criteria-instabiliteit

Zoals te zien is in Figuur 5 zijn de veranderingen tussen paramorfische en subjectieve gewichten uitgesloten als mogelijke gebieden van criteria-instabiliteit. Rechtonder in de tabel zijn twee velden niet van toepassing, omdat er geen sprake kan zijn van subjectieve analyse in situaties waar geen sprake is van bewustwording van criteria. De velden met een vraagteken slaan op combinaties die niet worden behandeld in de onderzoeken in dit verslag. Ze staan mogelijk nog open voor verder onderzoek. De afbakening van dimensies zoals gegeven in Figuur 5 brengt ons op de volgende deel-definitie van criteria-instabiliteit:

Criteria-instabiliteit staat voor een locale verandering in gewichten tussen meerdere berekende paramorfische gewichten, ofwel meerdere gegeven subjectieve gewichten van een bepaald criterium behorende bij een reeks identieke keuzes.

In de onderzoeken over verandering in gewichten wordt telkens gesproken over de globale gewichten, ofwel de lange termijn waarden van personen, waar tijdens de beslissing van afgeweken wordt. Als deze afwijking een bewuste handeling zou zijn, zou men over voldoende rationaliteit moeten beschikken om tijdens deze handeling te herleiden dat de gemaakte beslissing ten aanzien van gewichten afwijkt van zijn of haar persoonlijke waarden. Het eerder aangetoonde verschil tussen paramorfische en subjectieve gewichten bewijst echter dat dit vaak niet het geval is. We kunnen dus stellen dat deze verandering in gewichten optreedt zonder dat de beslisser zich hiervan bewust is.

Criteria-instabiliteit is een niet-bewuste handeling.

Nu we de mogelijkheden van criteria-instabiliteit in de groene gebieden in Figuur 5 hebben afgebakend, rest ons alleen nog een selectie te maken tussen de onderzoeken binnen deze gebieden. Welke onderzoeken raken daadwerkelijk het principe van criteria instabiliteit? Hiervoor maken we een laatste onderscheid binnen het keuze-gedrag, namelijk tussen systematisch en onsystematisch gedrag.

7.2 Systematische en willekeurige invloeden op keuzegedrag

Ten einde proces-consistentie, oftevel de veranderlijkheid van gewichten tussen achtereenvolgende keuze-situaties, te analyseren maakt men dikwijls onderscheid tussen een systematische en een random component van het keuzegedrag van een beslisser. Al eerder in dit verslag kwam dit tot uiting in de formule die trachtte het gedrag van een beslisser te modelleren (Fischer et al., 2010), maar ook in veel andere onderzoeken komt dit onderscheid ter sprake (Kleinmuntz, 1990). Hier is de interpretatie van het systematische deel de “ware” mening van de beslisser, terwijl de random component een verstoring is hierop die optreedt wanneer de beslisser zijn ware mening probeert uit te drukken.

Deze ware mening wordt in feite benaderd door het paramorfische gewicht. We hebben gezien dat deze ware mening slecht te modelleren is, vanwege de vele oorzaken van verandering in gewichten bij beslissingen. Ten einde een gestructureerder overzicht te scheppen van deze veranderingen in gewichten, zullen we in dit verslag twee categorieën definiëren, namelijk systematisch en willekeurig keuzegedrag. Hoewel deze categorieën nauw verwant zijn aan het onderscheid binnen de gewicht-parameters van het eerder behandelde RandMAU model, geeft het model slechts een beschrijving terwijl wij ons nu focussen op de oorzaken van het systematische en random component van keuzegedrag.

7.2.1 Systematische invloeden op keuzegedrag

Systematisch wordt door de van Dale beschreven als “orderlijk en samenhangend volgens een systeem”. Wanneer we dit principe toepassen op verandering in gewichten bij beslissingen, zijn er twee mogelijkheden waar het aan de orde kan zijn. Ten eerste kunnen gewichten in een bepaalde richting veranderen. Zo kan een persoon van de een op de andere keuze een criterium ineens meer of minder belangrijk gaan vinden. Systematische invloeden op veranderingen in gewichten zijn dus invloeden waarvan bekend is in welke richting zij de gewichten van criteria zullen veranderen. Ten tweede is van belang in welke

mate het gewicht verandert. Een gewicht kan veel of weinig veranderen. Systematische invloeden zullen dus ook met een vooraf bekende hoeveelheid de gewichten veranderen.

Binnen de in paragraaf 7.1 gespecificeerde dimensies blijken procedurele en descriptieve variantie uit paragraaf 5.1.2 en 5.1.3 te voldoen aan dit principe. Bij het overzicht van de belangrijkste factoren in dit genre van variantie, is telkens duidelijk van welke criteria een verandering in gewichten optreedt, dan wel omhoog dan wel omlaag⁶. Hoewel de mate van verandering vaak niet expliciet in acht is genomen bij de behandelde onderzoeken, zou dit alsnog verkregen kunnen worden door bijvoorbeeld meerdere achtereenvolgende tests uit te voeren. Wanneer over een groot aantal tests heen de gemiddelde verandering van gewicht van een bepaald criterium wordt berekend, kan hieruit de mate van verandering afgeleid worden.

7.2.2 Onsystematische invloeden op keuzegedrag: criteria-instabiliteit

Onsystematische effecten treden in tegenstelling tot systematische effecten juist willekeurig op. Het is onvoorspelbaar in welke richting de verandering in gewichten zich manifesteert en ook de mate waarin verandering optreedt is vooraf niet bekend. Hieronder zouden volgens het overzicht in Figuur 5 de bevindingen van Levine en Palmeira en Krishnan vallen, aangezien we procedurele en descriptieve variantie hebben ingedeeld als zijnde systematische effecten. In hoofdstuk 6 hebben we geconcludeerd dat de oorzaak van de verandering in gewichten bij zowel de onderzoeken van Levine, alsmede een deel van de onderzoeken van Palmeira en Krishnan⁷ te maken heeft met de toegankelijkheid van elementen in de hersenen die bepaalde criteria representeren ofwel elementen die hiermee verbonden zijn. In paragraaf 6.1.4 hebben we gesteld dat “anchoring” de hoofdoorzaak is voor de verschillen in toegankelijkheid die zelfs op zeer korte tijdsintervallen op kan treden. Aan het einde van de paragraaf concludeerden we dat de hierdoor veroorzaakte verandering in gewichten over het algemeen zeer onvoorspelbaar is, omdat de wijze van “anchoring” hier onbekend is door de willekeurige reeks van alternatieven. Dit geldt zowel voor de mate als de richting van verandering. Samenvattend is er dus sprake van onsystematische invloeden bij zowel de onderzoeken van Levine als onderzoek 4a van Palmeira en Krishnan. In beide gevallen is er namelijk sprake van beredeneren en zijn de alternatieven op een willekeurige wijze geordend, zodanig dat geen systematiek in de “anchoring” gevonden kan worden.

In paragraaf 6.2 hebben we gesteld dat oorzaken van verandering in gewichten bij vooraf aangereikte criteria nog niet aan het licht zijn gebracht. Wel hebben Palmeira en Krishnan de resultaten van onderzoek 4b gekarakteriseerd met een principe van willekeurige variatie. Ook hier kan de richting en de mate van gewichtsverandering nog niet aan een bepaald systeem verbonden worden. Dit komt tot uiting in een bijna willekeurige kansverdeling, die zoals in paragraaf 6.3 slechts een grove benadering geeft van het keuzegedrag, maar wel onderhevig is aan willekeurigheid.

6 Hoewel niet expliciet vermeld binnen paragraaf 5.1.2, is er bij inter- en intra-methodische inconsistenties ook sprake van een structurele richting en mate van verandering in gewichten. Voor meer hierover refereer ik naar de onderzoeken van Weber en Borcharding (1992).

7 Palmeira en Krishnan wijden twee onderzoeken aan criteria-instabiliteit. Onderzoek 4a betreft de situatie waarbij criteria door de beslisser moeten worden onderscheiden, terwijl bij onderzoek 4b de criteria reeds gegeven zijn.

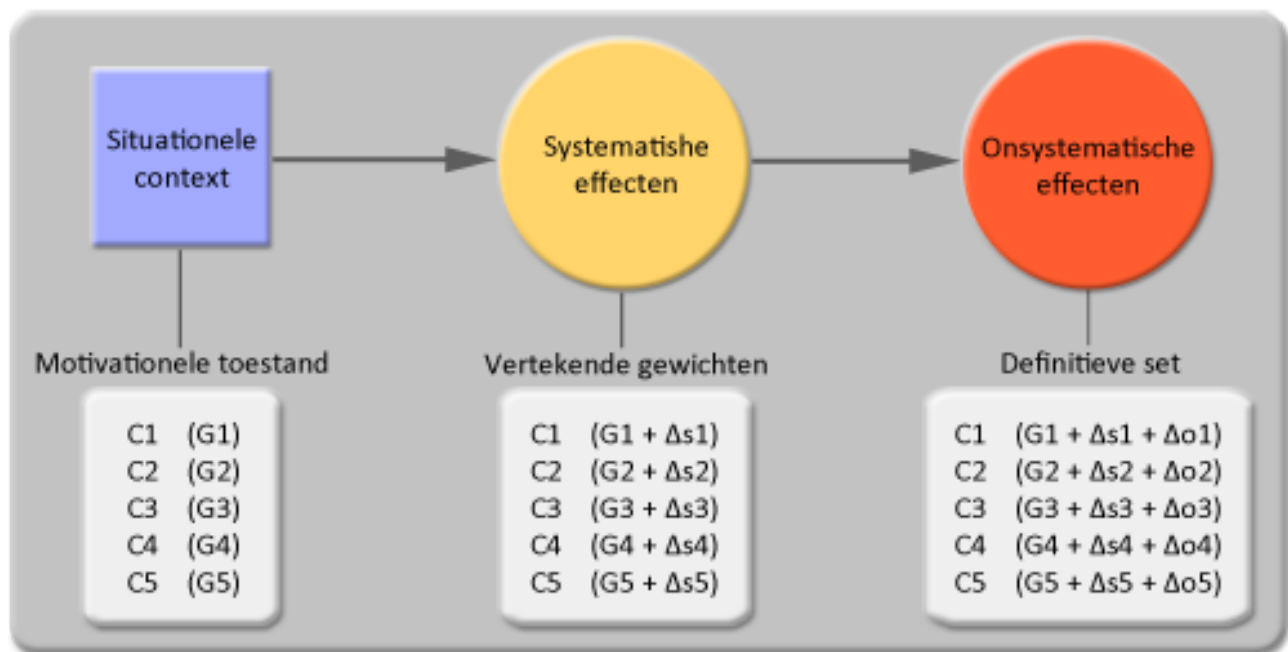
Criteria-instabiliteit delen wij naar aanleiding van ons onderzoek in bij de categorie van onsystematische invloedsfactoren op keuzegedrag. In combinatie met de eerder in dit verslag geformuleerde sub-definities, brengt dit ons op de volledige definitie van criteria-instabiliteit.

Criteria-instabiliteit staat voor een locale onsystematische niet-bewuste verandering in ofwel het paramorfische gewicht, ofwel het subjectieve gewicht van een bepaald criterium behorende bij een reeks identieke keuzes.

7.2.3 Een overzicht van locale effecten op keuze-gedrag en criteria-instabiliteit

In deze paragraaf wordt een samenvattend overzicht gegeven van de locale interpretatie van gewichten en de bijbehorende effecten zoals ze in dit verslag zijn ingedeeld. In het overzicht wordt gebruik gemaakt van een driedeling, namelijk een begin-set van gewichten van criteria, gewichten die veranderd zijn door systematische effecten en gewichten die (verder) veranderd zijn door onsystematische effecten. Dit wordt na de volgende figuur nader toegelicht.

Figuur 6



In de bovenstaande figuur zijn de locale effecten weergegeven die van invloed zijn op de gewichten binnen keuzes naar aanleiding van de conclusies van dit verslag. Als initiële toestand heb ik die van Ditrich en List (2010) overgenomen. Zij stellen dat een beslisser met een bepaalde set van mogelijke criteria aan een beslissing begint. De samenstelling van deze set is afhankelijk van de motivationale toestand van de beslisser. Zo kan het bijvoorbeeld zijn dat de beslisser onlangs te horen heeft gekregen dat hij of zij lijdt aan overgewicht. De beslisser zal zich bij een keuze tussen etenswaren nu mogelijk eerder focussen op de mate van calorie-gehalte dan op de aangenaamheid van de smaak. Dit komt overeen met het

lange termijn gewicht, ofwel globale gewicht, dat we eerder in dit verslag gedefinieerd hebben. Vervolgens zorgen de locale systematische effecten voor vertekende gewichten. Deze gewichten zijn niet de intentie van de beslisser: ze worden geleid door invloeden van procedurele en/of descriptieve aard. Het verschil in gewicht per criteria door systematische invloeden, Δ_s , is voorspelbaar in richting en hoeveelheid. Ten slotte zorgt criteria-instabiliteit voor een onsystematische verandering, Δ_o , waardoor de uiteindelijk set van criteria die daadwerkelijk in de keuze gebruikt wordt gevormd is. Δ_o is niet voorspelbaar in richting noch in mate, en kan het beste gemodelleerd worden met de in paragraaf 6.3 besproken Beta-verdeling uit het RandMAU-model.

7.3 Discussiepunten

Nu volgen enkele punten ter discussie omtrent een aantal hoofdzaken die binnen dit verslag behandeld zijn. Deze punten zijn niet van dien aard dat ze de hoofddefinitie van criteria-instabiliteit, zoals in dit verslag beschreven, verwerpen. Ze kunnen echter wel leiden tot nieuwe inzichten.

7.3.1 Systematisch/onsystematisch en verklaarbaar/onverklaarbaar

We hebben in dit verslag telkens wanneer we spraken over criteria-instabiliteit, dit ingedeeld bij onsystematische effecten. We hebben gesteld dat systematische effecten optreden met een vooraf bekende hoeveelheid en mate. Eigenlijk betekent dit dus dat systematische effecten niet per definitie verklaarbaar hoeven te zijn. Wanneer eenmaal een besliskundig model opgesteld kan worden van een beslisser naar aanleiding van een reeks keuzes in een bepaalde context, zou dit model over het algemeen geaccepteerd worden wanneer het telkens in staat is een volgende beslissing te voorspellen. Het beschrijvende karakter van deze modellen verplicht ons dus niet om een verklaring te zoeken voor het systematische deel van het model.

In tegenstelling tot de besliskunde is het belang van voorspelbaarheid bij menig andere situaties veel groter. Wat bijvoorbeeld wanneer een straaljager een constante lichte afwijking naar rechts vertoont. Je kunt hier net als bij het besliskundig descriptieve model stellen dat wanneer de afwijking constant naar rechts is, hier op ingespeeld kan worden en dit dus niet tot problemen hoeft te leiden. Echter kan de oorzaak van het probleem in de straaljager wel een scheur in de vleugel zijn, waardoor deze een grote kans heeft om ineens door te scheuren. Het is duidelijk dat het achterhalen van de oorzaak hier veel meer van belang is en dat de afwijking naar rechts hier eerder zou worden ingedeeld bij criteria-instabiliteit.

In dit verslag heb ik mij gehouden aan een descriptieve interpretatie van criteria-instabiliteit, waarbij criteria-instabiliteit staat voor de afwijking van een bepaalde opgestelde standaard, in dit geval bijvoorbeeld een model dat na vele achtereenvolgende keuzes is opgesteld. Dat er binnen deze standaard ook sprake is van een systematisch deel, wat bijvoorbeeld niet in te delen valt binnen procedurele of descriptieve variantie en dus onverklaarbaar is, leidt er niet toe dat dit gezien wordt als criteria-instabiliteit. Dit omdat het met een vrijwel 100% zekerheid de volgende keuze kan voorspellen. Al met al komt het erop neer dat binnen de systematische effecten ook onverklaarbare effecten kunnen optreden, maar dat criteria-instabiliteit zoals in dit verslag beschreven zich beperkt tot onsystematische verandering, die

per definitie tot een bepaald niveau onverklaarbaar is (zo kwamen we niet verder dan “willekeurige anchoring” als een mogelijke verklaring).

7.3.2 “Anchoring” met betrekking tot de score

Tot slot volgt in deze paragraaf nog een nadere uitwerking van de opmerkingen die eerder in dit verslag zijn gemaakt, namelijk aan het eind van paragraaf 6.1 en in voetnoot 5. Deze opmerkingen betroffen het feit dat “belangrijkheid” in het onderzoek van Levine, maar ook bij Palmeira en Krishnan en vele andere onderzoeken in de besliskunde, al snel wordt gezien als het “gewicht” van het criterium in de vorm van de gewicht-parameter in het “weighted additive value model” (ofwel varianten daarop, zoals het RandMAU model in paragraaf 6.3). Bij deze stellingname wordt echter de rol van de score vaak niet of te weinig in acht genomen. Vandaar dat bij “anchoring”, wat wij hebben neergezet als een mogelijke oorzaak van criteria-instabiliteit bij beslissingen waar criteria zelf moeten worden onderscheiden, een kanttekening moet worden gemaakt. Dit heeft geen complicaties voor de einddefinitie van criteria-instabiliteit, welke los staat van of “anchoring” al dan niet de oorzaak is van de onsystematische variatie.

“Anchoring” zoals wij het behandeld hebben bij beredeneerde beslissingen, kan dus ook een verschil in de perceptie van scores betekenen. Het is namelijk zo dat de “belangrijkheid van het criterium bij de gemaakte keuze” gevraagd wordt. Dit gaat om de samengestelde eindwaarde van een alternatief op een criterium; een combinatie van het gewicht van en de score op dat criterium. De vraag bij Levine en Palmeira en Krishnan (4a) om de belangrijkste criteria te noteren die mee speelden in de keuze, is dus feitelijk een combinatie van gewicht en score. Wat dat betreft zou je dus kunnen zeggen dat onsystematische variatie in enkel de gewichten zich alleen heeft afgespeeld bij studie 4b van Palmeira en Krishnan, waar criteria en scores reeds vooraf gegeven zijn en specifiek over de belangrijkheid, oftewel parameter w moest worden beoordeeld. En zelfs daar valt te betwisten of de perceptie van de score niet verschilt van de beschreven score.

Dit is afhankelijk van wat je precies verstaat onder belangrijkheid van het criterium bij de gemaakte keuze. Bij Levine en Palmeira en Krishnan (4a) was dit een combinatie van gewicht en score zoals in het RandMAU-model, en bij studie 4b was het specifiek op het gewicht gericht. Dit onderscheid is van belang wanneer de precieze mentale processen van het toewijzen van een score en het toewijzen van een gewicht moeten worden verklaard. Het is echter in de besliskunde gewoon om te komen tot een descriptief model. Hier is het niet van belang of “anchoring” zich bij dan wel scores, dan wel gewichten, dan wel allebei afspeelt. Het gaat om de uitkomst van het model op keuze-niveau. Zie bijlage B voor een treffende anekdote hierover.

Al met al kunnen we dus stellen dat “anchoring” niet precies te meten is op het niveau van enerzijds gewichten en anderzijds scores zoals gedefinieerd in het weighted additive value model. Het gaat om een combinatie van de descriptieve variabelen. Verschillende interpretaties van de effecten van anchoring op parameter w en/of s kunnen er al dan niet toe leiden dat “anchoring” als een mogelijke oorzaak wordt onderkend van criteria-instabiliteit wanneer criteria-instabiliteit staat voor een variatie in parameter w . Wanneer we criteria-instabiliteit zien als variatie in de belangrijkheid van een criterium, zoals persoonlijk gespecificeerd door de beslisser (bijvoorbeeld dat hij of zij de belangrijkheid van een criterium

na een beslissing noteert) is “anchoring” hier wel degelijk bij betrokken.

8 Evaluatie

In dit laatste onderdeel van het verslag wordt een samenvatting gegeven van de belangrijkste onderzoeksresultaten, alsmede een overzicht van de aannames die bij de totstandkoming van deze resultaten zijn gedaan. Ook worden aanbevelingen gedaan voor verder vervolgonderzoek met betrekking tot criteria-instabiliteit.

8.1 Wat hebben we gedaan?

Na de inleiding in het onderzoek begonnen we vanaf hoofdstuk 4 met zowel een beschrijving van de verschillende opvattingen met betrekking tot de term gewichten, alsmede een overzicht van de ontwikkeling in denkwijze ten aanzien van de veranderlijkheid van gewichten. Deze ontwikkeling kwam tot uiting in de erkenning van de veranderlijkheid van locale gewichten, waarvan we de in de literatuur meest bekende voorbeelden hebben behandeld onder de noemer “procedurele en descriptieve variantie”. Daarna gingen we in op onderzoeken waar men zelfs bij schijnbaar identieke keuze-situaties variatie in toegekende gewichten vertoonde. We hebben het beredeneren, oftewel het door de beslisser zelf onderscheiden van criteria, verder uitgezocht en kwamen via het selectie-en weegproces van criteria op “priming” en “anchoring”. Het “anchoring” leek te zorgen voor onvoorspelbare effecten ten aanzien van verandering in gewichten. Deze onvoorspelbaarheid kwam verder tot uiting bij de omschrijving van het RandMAU model. Ook werd hier een voorspelbaar deel onderscheiden. Nadat we in paragraaf 7.1 hadden vastgesteld binnen welke in dit verslag behandelde onderzoeken sprake kan zijn van criteria instabiliteit, namelijk Levine en Palmeira en Krishnan, gingen we verder in op het onderscheid tussen voorspelbaarheid en onvoorspelbaarheid. We definieerden het voorspelbare deel als “systematisch”, oftewel verandering van gewichten met een bekende hoeveelheid en mate. De overige onvoorspelbare factoren zijn “onsystematisch” en vallen onder de noemer van criteria-instabiliteit. Ten slotte gaven we een samenvattend overzicht van de locale keuze-factoren die in dit verslag onderzocht zijn, waarin duidelijk gemaakt werd dat een beslisser de keuze in gaat met een bepaalde motivationele toestand en dat systematische effecten en onsystematische effecten de daadwerkelijk gebruikte gewichten vormen. Ook hebben we de interpretatie van “anchoring” in de score als bijkomend effect nader beschreven en hebben gesteld dat dit geen consequenties heeft voor de definitie van criteria-instabiliteit.

8.2 Aannames en validiteit

In dit verslag zijn een aantal aannames gemaakt die een nadere verklaring vereisen ten behoeve de validiteit van deze aannames te waarborgen.

Een van de belangrijkste aannames die we hebben gemaakt is dat we er in paragraaf 7.2.1 vanuit zijn gegaan dat bij procedurele en descriptieve variantie verandering in gewichten optreedt met een bekende hoeveelheid. De aanname dat de vormen van variantie aan deze voorwaarde zouden voldoen, heeft namelijk tot de conclusie geleid dat we procedurele en descriptieve variantie bij de systematische effecten konden indelen. Dit heeft bijgedragen aan het gestructureerde overzicht binnen de literatuur, een van de doelstellingen van dit verslag. Ten eerste dient een kanttekening te worden gemaakt bij de voorwaarden van systematische

effecten van verandering in gewichten. De voorwaarden van “mate” en “richting” zijn descriptief. Zij hoeven geen indicatie te geven van wat een willekeurige beslisser voor gedrag zal vertonen als zijn gedrag niet van tevoren geanalyseerd is. Wanneer een enkele of een groep van meerdere beslissers over meerdere beslissingen heen geanalyseerd wordt, kan een systematiek in de richting en mate herleid worden. Bij een volgende keuze-situatie met identieke proefpersonen en omgevingsfactoren kan het effect van het al dan niet gebruik maken van procedurele of descriptieve variantie-methoden voorspeld worden. Dit is voldoende om aan deze voorwaarden te voldoen.

Het feit of deze mate al dan niet beïnvloedbaar is, zoals bijvoorbeeld dat het verschil in gewichten bij choice en matching vergroot wordt wanneer de relatieve belangrijkheid van het prominente attribuut ten opzichte van het andere attribuut vergroot wordt (Carmon, Simonson, 1998), is hier dus niet van belang. Er is namelijk sprake van een inductieve redenering om op de factor “mate” te komen. Effecten die van invloed zijn op de mate kunnen bijvoorbeeld dienen om een voorspelling te maken in situaties waar het keuze-gedrag van beslissers nog niet is geanalyseerd. Een voorbeeld bij het beschreven effect is de voorspelling dat mensen over het algemeen een steeds hoger gewicht toewijzen aan het belangrijkste attribuut bij “choice” ten opzichte van “matching” wanneer de relatieve belangrijkheid van het prominente attribuut groter is. De voorwaarden voor systematisch gedrag richting zich echter specifiek op het paramorfische descriptieve model en de voorspellende effecten zijn dus ook niet van belang wanneer we systematisch gedrag onderscheiden. Vandaar dat ze in dit verslag buiten beschouwing zijn gelaten.

De tweede belangrijke aanname gaat over het paramorfische gewicht. In dit verslag wordt de notie gemaakt dat dit het “werkelijke” gewicht van de beslisser zou zijn. De vraag is echter in hoeverre het paramorfische gewicht het “werkelijke” gewicht benadert. Aangezien de proeven kunstmatig zijn, meestal in grote collegezalen worden gehouden met vragen die de beslisser vaak niet persoonlijk aan het hart gaan, rijst de vraag of er wel aandachtig genoeg aan de keuze wordt deelgenomen om de werkelijke waarden of gewichten van de persoon te ontlokken. Wanneer de beslisser bijvoorbeeld in een andere situatie een beslissing moet nemen tussen zeer persoonlijke kwesties, zal de beslisser waarschijnlijk veel aandachtiger te werk gaan. Aandacht gaat mogelijk ten koste van consistentie. Hoewel dit lijkt op het eerder gemaakte onderscheid tussen impulsieve en beredeneerde beslissingen, richt dit principe zich specifiek op in hoeverre de beslisser de keuze op zich laat inwerken. Dit speelt zowel bij impulsieve als beredeneerde beslissingen een rol.

Ten slotte nog enkele opmerkingen bij aannames die minder fundamenteel zijn voor het onderzoek.

- In de onderzoeken van Palmeira en Krishnan heb ik de aanname van de onderzoekers gevolgd dat het verschil in schaal in proef 4b geen invloed heeft op hun onderzoeksresultaten. Deze aanname is echter niet blindelings gevolgd. Ik heb namelijk ook geen indicatie kunnen vinden in de literatuur, zelfs die omtrent procedurele en descriptieve variantie, dat het verschil in numerieke schaal van invloed zou zijn op het keuze-gedrag. Dit verschil in numerieke schaal, bijvoorbeeld een tien-puntsschaal ten opzichte van een honderd-puntsschaal, dient niet verward te worden met de procedurele variantie “compatibiliteit van de schaal”, waar het gaat om een eenheidsschaal van bijvoorbeeld procenten of geld.

- Voor degenen die de onderzoeken van Palmeira en Krishnan hebben doorgenomen vermeld ik ter verduidelijking nog de relevantie van de kern van hun onderzoek, wat eigenlijk ging over het effect van een zogenaamde “isolated option”. Dit is een alternatief dat wordt toegevoegd nadat een keuze gemaakt is. Dit alternatief zou nu eerder gekozen worden dan wanneer deze in de groep van alternatieven van de eerste keuze zat. Dit heeft echter niet met criteria-instabiliteit van doen en is dus buiten beschouwing gelaten.

8.3 Aanbevelingen voor vervolgonderzoek

Al in de vorige paragraaf is de aandacht van de beslisser met betrekking tot de keuze en de visie op het “werkelijke gewicht” aan de orde geweest. Verder onderzoek zou kunnen uitwijzen in hoeverre aandacht ten koste gaat van de consistentie van het keuze-gedrag. Er zou kunnen worden getoetst of de setting van een massa-vraagprocedure in een college-zaal, waar veel onderzoeken in de gedragswetenschappen van gebruikmaken, wel de juiste manier van onderzoeken is wanneer onderzoek wordt gedaan naar consistentie van keuze-gedrag. Wellicht is een deel van de inconsistentie te wijten aan een verschil in aandacht dat aan de taak besteed wordt. Het verschil in aandacht kan tot uiting komen in bijvoorbeeld een weloverwogen of een vlugge beslissing. Al eerder bij de onderzoeken van Levine hebben we gezien dat gewichten kunnen veranderen naarmate een meer beredeneerde keuze wordt gemaakt. Het verschil in aandacht bij de taak kan hier dan gezien worden als een verschil in “weloverwogenheid” of “beredenerendheid” en dit kan dus verschillende effecten met zich meebrengen die de onderzoeksresultaten kunnen beïnvloeden.

Ten aanzien van de kern van dit onderzoek, criteria-instabiliteit, zijn we beland op een niveau van “anchoring”. Verdere onderzoeken die de werking van de hersenen met betrekking tot de toegankelijkheid van elementen en “anchoring” op zeer korte termijn verder verklaren kunnen helpen het model van willekeurig keuze-gedrag te verfijnen. Dit zal ertoe leiden dat steeds meer effecten systematisch worden. Een volledig kloppend model zou echter rekening moeten houden met een bijna oneindig aantal omgevingsfactoren, en zal wellicht nooit haalbaar zijn. Oplossingen om “anchoring” op korte termijn te kunnen beheersen kunnen echter zeer van dienst zijn in situaties waar criteria-instabiliteit dient te worden geminimaliseerd.

Literatuurlijst

- Anderson, John R. (1983): "A Spreading Activation Theory of Memory"
Journal of Verbal Learning and Verbal Behavior, 22 (June), pp. 261-295
- Barlas (2003): "When choices give in to temptations: Explaining the disagreement among importance measures"
Organizational Behavior and Human Decision Processes 91 (2003) pp. 310–321
- Borcherding, Katrin; Eppel, Thomas; von Winterfeldt, Detlof (1991): "Comparison of Weighting Judgements in Multiattribute Utility Measurement"
Management Science, Vol. 37, No. 12 (Dec., 1991), pp. 1603-1619
- Campbell, Hutchinson, Scarpa (2006): "Lexicographic decision-making rules in discrete choice experiments: implications on individual-specific willingness to pay estimates"
Working Papers 2006.128, Fondazione Eni Enrico Mattei.
- Carmon, Ziv; Simonson, Itamar (1998): "Price-quality trade-offs in Choice versus Matching: new insights into the Prominence Effect"
Journal of Consumer Psychology (1998) Volume: 7, Issue: 4, pp. 323-343
- Ditrich, List (2010): "Where do preferences come from?"
International Journal of Game Theory, forthcoming
- Fasolo, Barbara; McClelland, Gary H.; Todd, Peter M. (2006): "Escaping the Tyranny of Choice: When fewer attributes make choice easier"
Marketing Theory March 2007 vol. 7 no. 1 pp. 13-26
- Fischer, Gregory W.; Hawkins, Scott A. (1993): "Strategy compatibility, scale compatibility, and the prominence effect."
Journal of Experimental Psychology: Human Perception and Performance, Vol 19(3), Jun, 1993. pp. 580-597
- Fischer, G. W., Damodaran W., Laskey K.B., Lincol D. (1987): "Preferences for Proxy Attributes"
Management Science Vol. 33, No. 2, Feb., 1987, pp. 198-214
- Fischer, J. Jia, M. F. Luce (2000): "Attribute conflict and preference uncertainty: The RandMAU model."
Management Science, May 2000, Volume: 46 Issue: 5 pp.669-684
- Fischer, Luce, Jia (2000): "Attribute Conflict and Preference Uncertainty: Effects on Judgment Time and Error"
Management Science 46, 1, 88-103, Jan. 2000
- Goldstein (1990): "Judgments of relative importance in decision making: Global vs local interpretations of subjective weight"
Organizational Behavior and Human Decision Processes, 47, pp. 313-336.
- Heerkens, J.M.G. (2003): "Modeling importance assessment processes in non-routine decision problems."
(dissertation)
- Hoffman (1960): "The paramorphic representation of clinical judgement"
Psychological Bulletin, Vol. 57, No. 2, 1960
- Jia et al. (1998): "Attribute Weighting Methods and Decision Quality in the Presence of Response Error: A Simulation Study"
Journal of Behavioral Decision Making 11:pp. 85–105
- Kleinmuntz (1990): "Decomposition and the control of errors in decision analytic models"
Working paper (Sloan School of Management) -- 2023-88

- Levine, Gary M.; Halberstadt, Jamin B.; Goldstone, Robert L. (1996): "Reasoning and the Weighting of Attributes in Attitude Judgments"
Journal of Personality and Social Psychology, Vol 70(2), Feb 1996, pp. 230-240
- Liu, Chun Chu (2003): "Simulating Weights Restrictions in Data Envelopment Analysis by the Subjective and Objective Integrated Approach"
Web Journal of Chinese Management, Vol 6, No 1
- de Martino, Benedetto; Kumaran, Dharshan; Seymour, Ben; Dolan, Raymond J. (2006): "Frames, Biases, and Rational Decision-Making in the Human Brain"
Science. 2006 August 4; 313(5787): pp. 684–687.
- Monat, Jamie P. (2009): "The benefits of global scaling in multi-criteria decision analysis"
Judgment and Decision Making, vol. 4, no. 6, October 2009, pp. 492-508
- Nedungadi, P. (1990): "Recall and Consumer Consideration Sets: Influencing Choice without Altering Brand Evaluations"
Journal of Consumer Research, 1990, vol. 17, issue 3, pp. 263-76
- Palmeira, Krishnan (2008): "Criteria instability and the isolated option effect"
Organizational Behavior and Human Decision Processes, Vol. 106 pp. 153-167
- Payne, Bettman (1992): "Behavioral decision research"
Annual Review of Psychology, 43, pp. 87-131
- Milton, Rokeach (1979): "Understanding Human Values: Individual and Societal"
Partly retrieved from <http://books.google.com/>
- Mussweiler, Strack (1999): "Comparing Is Believing: A Selective Accessibility Model of Judgmental Anchoring"
European Review of Social Psychology, 10:1, pp. 135-167
- von Neumann, John; Morgenstern, Oskar; Rubinstein, Ariel; Kuhn, Harold William (1947): "Theory of games and economic behavior"
Partly retrieved from <http://books.google.com/>
- Savage, L.J. (1954): "The foundations of statistics"
Partly retrieved from <http://books.google.com/>
- Shafer, Glenn (1986): "Savage revisited"
Statistical Science. 1986, Vol. 1, No. 4, pp. 463-501
- Stillwell, von Winterfeldt, John (1987): "Comparing Hierarchical and Nonhierarchical Weighting Methods for Eliciting Multiattribute Value Models"
Management Science, Volume 33 Issue 4, April 1987
- Strack, Deutsch (2004): "Reflective and Impulsive Determinants of Social Behavior"
Pers Soc Psychol Rev August 2004 vol. 8 no. 3 pp. 220-247
- Tversky et al (1988): "Contingent Weighting in Judgment and Choice"
Psychological Review, Vol 95(3), Jul 1988, pp. 371-384.
- Tversky, Kahneman (1986): "Rational Choice and the Framing of Decisions"
Journal of Business, 1986, vol.59, no.4, pt.2.
- Tversky, Kahneman (1974): "Judgement under uncertainty: Heuristics and Biases".

Science, New Series, Vol. 185, No. 4157 (Sep. 27, 1974), pp. 1124-1131

Tversky A.; Thaler R.H. (1990): "Preference Reversals"
Journal of Economic Perspectives, Vol. 4, No. 1, pp. 193-205.

Weber, Borcherdig (1992): "Behavioral influences on weight judgments in multiattribute decision making"
European Journal of Operational Research, Volume 67, Issue 1, 28 May 1993, pp. 1–12

Wilson, Timothy D. (1991): "Thinking too much: Introspection can reduce the quality of preferences and decisions."
Journal of Personality and Social Psychology, Vol 60(2), Feb, 1991.pp.181-192

Yeh, Chung-Hsing (2002): "A problem-based selection of multi-attribute decision-making methods"
International Transactions in Operational Research, Volume 9, Issue 2, March 2002, pp. 169–181

Zhang, Jiao; Hsee Christopher K.; Xiao, Zhixing (2005): "The majority rule in individual decision making"
Organizational Behavior and Human Decision Processes, 2006, 99(1), pp. 102-11

Definitielijst

Descriptieve variantie

Een verschil in keuze-gedrag wanneer gebruik gemaakt wordt van een andere beschrijving van dezelfde vraag. Dit verschil in keuze-gedrag komt in elk geval tot uiting in de gebruikte gewichten. Het kan bijvoorbeeld zijn dat een bepaald attribuut is weergegeven door middel van een proxy-attribuut, oftewel een attribuut dat het hoofdattribuut al dan niet deels representeert.

Logische consistentie

In hoeverre keuze-gedrag overeenkomt met bepaalde axioma van normatieve theorieën binnen de besliskunde.

Onsystematische verandering in gewichten

Zie systematische verandering. Het tegenovergestelde hiervan, oftewel het deel van verandering dat nagenoeg random optreedt.

Paramorfisch model

Een paramorfisch model is bedoeld om het beslis-gedrag van een of meerdere beslisers binnen een bepaalde context (omgevingsfactoren) te beschrijven. Nadat de groep beslisers steeds meer geanalyseerd wordt en er steeds meer data beschikbaar is over de keuzes, wordt het paramorfisch model steeds nauwkeuriger. Dat wil zeggen dat het beter kan voorspellen en een beter beeld schept van de daadwerkelijk gebruikte gewichten. De term “paramorf” staat ook wel voor “een bepaalde omschrijving van een fenomeen” (zie Bijlage B).

Paramorfisch gewicht

Het paramorfisch model wat door dit verslag heen gebruikt wordt, namelijk het RandMAU-model, heeft zowel parameters voor de score als voor het gewicht van een criterium. Het paramorfisch gewicht staat voor de parameter van het gewicht in deze.

Procedurele variantie

Een verschil in keuze-gedrag wanneer gebruik gemaakt wordt van een andere vraagmethode. Dit verschil in keuze-gedrag komt in elk geval tot uiting in de gebruikte gewichten. Een verschil in vraagmethode is bijvoorbeeld wanneer de ene keer gevraagd wordt een keuze te maken tussen twee alternatieven, terwijl de andere keer gevraagd wordt een prijs aan de alternatieven te verbinden.

Procesconsistentie

In hoeverre keuze-gedrag tussen verschillende keuze-situaties met elkaar overeenkomt.

Subjectief gewicht

Het gewicht van een criterium dat de beslisser na het maken van een beslissing kenbaar maakt (meestal door middel van het noteren van het gewicht). Het blijkt dat er nogal eens verschil zit tussen het subjectieve en het paramorfische gewicht. Wanneer een groot aantal keuzes van een beslisser in een bepaalde context is geanalyseerd kan het paramorfische gewicht een zeer nauwkeurige benadering geven van het daadwerkelijk gebruikte gewicht van een criterium. Daarom wordt gesteld dat beslisers vaak geen inzicht hebben in de door

hen gebruikte gewichten, zodat het subjectieve gewicht in principe niet het daadwerkelijk gebruikte gewicht representeert.

Systematische verandering in gewichten

Descriptieve benadering van verklaarbare mate en richting van verandering in gewichten. Dit houdt in dat een effect dat zorgt voor verandering in gewichten niet per se voor alle personen hoeft te gelden. In dit geval zou voor een bepaalde keuze-situatie de uitkomst voorspeld kunnen worden voor elke persoon. Het is slechts van belang dat wanneer het beslis-gedrag van een bepaalde persoon tijdens achtereenvolgende keuze-situaties wordt geanalyseerd, een model kan worden opgesteld waarin de systematiek tot uiting komt. Dat wil zeggen dat er na veel verwerkte data een model ontstaat dat voorspellingen kan maken van een nieuwe beslissing zonder dat er sprake is van random variatie.

Bijlage A

Palmeira en Krishnan (2008) hebben onderzoeken gedaan naar onder andere criteria-instabiliteit. Ze stellen dat criteria-instabiliteit zowel tot uiting kan komen in het gebruik van verschillende criteria bij identieke keuzes (studie 4a), als het gebruik van verschillende criteria bij identieke keuzes met vooraf aangereikte gewichten (studie 4b). Nu volgt een korte samenvatting van beide onderzoeken.

Studie 4a

De vragenlijst bestaat hier uit drie keuzes tussen twee opties. De drie keuzes bestonden uit twee alternatieven waar uit gekozen moest worden. Bij elke keuze was sprake van een andere categorie van alternatieven. Zo moest er gekozen worden uit universiteiten, steden om te werken en auto's. Vijfenveertig studenten namen deel aan de proef en aan hen werd gevraagd om telkens na een keuze de drie attributen te noteren die het belangrijkste werden gevonden in de beslissing. Na de drie keuzes was er een pauze van vijftien minuten met ongerelateerde activiteiten. Daarna werd dezelfde vragenlijst nogmaals afgenomen. Er werd statistisch aangetoond met $p < 0.01$ dat men over het algemeen niet gebruik maakte van precies dezelfde criteria.

Studie 4b

Negenenvijftig studenten maakten zes keuzes tussen producten uit een productgroep die per keuze verschilde (zie de onderstaande tabel).

Table 8
Study 4b: Stimuli

Product	Attributes	A	B
Mp3	Features	85	60
	Ease of use	60	80
Sofa	Spring durability	60	85
	Cushion softness	80	60
Cell phone	Durability	56	70
	Battery life	70	50
Light bulb	Durability	80	90
	Energy saving	94	80
Shirt	Wrinkle-resistant	56	74
	Stain-resistant	80	50
Vitamin water	Endurance	66	84
	Energy	86	70

Zoals te zien is zijn de criteria en scores van de alternatieven A en B vooraf gegeven. Na elke keuze in een productgroep moesten de deelnemers noteren welk gewicht zij gebruik dachten te hebben voor elk van de twee criteria, door honderd punten te verdelen tussen deze criteria. Na de zes gemaakte keuzes werd een pauze van vijf minuten afgenomen bestaande uit ongerelateerde activiteiten. Vervolgens werd dezelfde vraagprocedure herhaald, echter met een vooraf gegeven score van elk criterium op een schaal van 1 tot en met 5 in plaats van 1 tot en met 100. Dit om de indruk te wekken dat er sprake was van een andere keuze-situatie zodat men niet simpelweg dezelfde gewichten in zou vullen. Een score van 60 in de productgroep Mp3 van alternatief B in de bovenstaande tabel werd bij de tweede keuze-set dus weergegeven als een score van 3. Met $p < 0.01$ werd de nulhypothese dat gewichten van de zes productgroepen niet zouden veranderen verworpen.

Bijlage B

De volgende tekst is verkregen uit Hoffman (1960), p124-125.

Mathematical models are designed to provide a scheme whereby one set of events may be satisfactorily predicted from another, and whereby testable derivations may lead to more complete theoretical understanding of the phenomena. Such models therefore constitute a level of description and explanation which suffices for scientific purposes. It is not required of models that they bear any semblance of some "actual" state of affairs, either within the organism or elsewhere, nor would this necessarily lead to a better understanding of nature. This may be more clearly seen in relation to an example from the physical sciences.

A chemist has the task of describing a substance. He performs a number of operations (or tests) on the substance and determines its chemical composition. It turns out to be a relatively simple task. The substance is described, chemically, as CaCO_3 . But the work is not complete, for a different set of operations (or tests) might have produced a different level of description; one which may not be necessary for the chemist, but which is more suitable for other contexts. The mineralogist, for example, would perform a

series of tests to measure the optical properties of the crystal, or he might examine its hardness, solubility, and other characteristics. Two crystals identified by the chemist as CaCO_3 might in fact be somewhat different from one another to the mineralogist, one being Aragonite, and the other Calcite. These two crystals do in fact have the same chemical structure, but they differ in molecular structure; this being revealed by optical and other tests. It is apparent that different levels of description are possible with regard to substances, and that each has its peculiar advantages and shortcomings. In mineralogy, Calcite is commonly described as a paramorph of aragonite. This word is used generally to describe a substance having crystalline structural properties which differ from those of another substance with the identical chemical composition.

We have borrowed the term paramorphic from mineralogy and employ it in relation to representations of human judgment. The analogy may not be complete, but its limitations are not serious. The mathematical representation of the judgment process is a level of description that approaches the chemical description of minerals. The formula helps to account for or "explain" what is observed concerning certain properties or characteristics of the judge, just as the chemical formula "explains" many, though not all, properties or characteristics of the substance. In addition, the formulae are useful in making predictions concerning the outcomes of certain other tests which may later be employed. But as with chemical analysis, the mathematical description of judgment is inevitably incomplete, for there are other properties of judgment still undescribed, and it is not known how completely or how accurately the underlying process has been represented. The term "paramorphic representation," used in relation to judgment, would seem adequately to indicate this state of affairs.