

# **Straffen in context**

## **Het effect van referentiekaders op de publieke opinie over straf**

*Jan de Keijser en Henk Elffers*

Verdwijnt de in de literatuur vaak geconstateerde punitiviteitskloof, het verschijnsel dat leken veel zwaarder willen straffen dan rechters, als men leken over strafzaken laat oordelen in een context van andere, zwaardere en lichtere, strafzaken? Deze kwestie hebben wij onderzocht met behulp van een experimentele methode. Anders dan verwacht blijkt dat een referentiekader waarin ernstiger zaken zijn opgenomen bij leken juist een toename in de door hen voorgestane straffen teweegbrengt, vooral bij mensen die toch al een punitieve instelling hebben.

### **De punitiviteitskloof: een artefact?**

In de afgelopen jaren is veelvuldig onderzoek gedaan naar de verschillen in door strafrechters feitelijk opgelegde straffen en de door het publiek geprefereerde strafzwaarte, mede met het oog op de veronderstelde bedreiging die een dergelijke kloof met zich zou meebrengen voor de legitimiteit van de strafrechtpleging.<sup>1</sup> In de loop van de tijd werden daarbij steeds meer kritische vragen gesteld bij de manier waarop zo'n kloof is vastgesteld. Met name is beargumenteerd dat voor een faire vergelijking leken en rechters zich wel over identieke zaken zouden moeten uitspreken, wat niet in alle onderzoek het geval was. Wanneer men leden van het publiek veel en gedetailleerde informatie over concrete strafzaken verschaft, lijkt het oordeel over de passende straf een stuk milder en genuanceerder te worden. Naarmate de taak die men het publiek opdraagt meer lijkt op de taak van de rechter, blijken de oordelen meer op elkaar te lijken. De vaak geconstateerde kloof tussen rechters en publiek is dus, tot op zekere hoogte, een meetartefact. Vraag is natuurlijk: tot op welke hoogte?

In Nederland is door De Keijser e.a. (2006) onder gecontroleerde omstandigheden onderzocht of het publiek net zo oordeelt als de rechter, wanneer het wel over dezelfde informatie beschikt als de rechter. Concreet stond daarbij de volgende vraag centraal: is er sprake van een punitiviteitskloof tussen strafrechter en publiek wanneer beide groepen hun strafoordeel baseren op exact gelijke informatie zoals neergelegd in reële en gedetailleerde strafdossiers? Strafrechters en leken

1 We noemen hier voor referentiedoelinden: Bottoms, 1995; Sprott & Doob, 1997; Hough & Roberts, 1998; Beyens, 2000; Mattinson & Mirrlees-Black, 2000; Cullen e.a., 2002; Kuhn, 2002; Hough & Park, 2002; Mirrlees-Black, 2002; Sociaal en Cultureel Planbureau, 2002, 2005; Roberts e.a., 2003; Van Koppen, 2003; Ministerie van Justitie, 2003; Barber & Doob, 2004; Elffers & De Keijser, 2004; Hutton, 2005; Elffers e.a., 2007a; Hoge Raad voor de Justitie, 2007; Elffers & De Keijser, 2008; vergelijk ook Hessing e.a., 2003.

werd verzocht een oordeel te vormen omtrent passende straf op basis van een dossier van woninginbraak, van zware mishandeling, ofwel van eenvoudige mishandeling. Bij vergelijking van de oordelen van rechters en leken met betrekking tot dezelfde dossiers bleek dat de leden van het publiek (de leken) onder deze omstandigheden toch nog steeds veel strenger zeggen te willen straffen in vergelijking met de strafrechters. Geconcludeerd werd dan ook dat er (in ieder geval in Nederland) wel degelijk een substantiële kloof bestaat tussen strafrechters en publiek, die niet een artefact is van de methode, zoals men in het geval van onderzoek op basis van surveyvragen kan volhouden.

Is daarmee de existentie van de kloof eens en voor altijd aangetoond? Natuurlijk kende dat onderzoek ook beperkingen. De auteurs zelf noemden er verscheidene (vergelijk ook Elffers & De Keijser, 2007). Zo werd van de deelnemers, anders dan in een werkelijk strafproces, verwacht zich een oordeel te vormen enkel en alleen op basis van een schriftelijk dossier, zonder de verdachte te zien en te horen ter terechtzitting. Ook wisten de deelnemers dat hun beslissing, in tegenstelling tot levensechte zaken, geen consequenties zou hebben voor betrokkenen van vlees en bloed.

Op deze kanttekeningen gaan wij hier niet in, wij concentreren ons op de derde tegenwerping, die veel rechters naar voren brengen na lezing van het onderzoeksrapport (De Keijser e.a., 2006, 44-45). Volgens hen missen leken die deelnemen aan een dergelijk onderzoek een referentiekader, terwijl dat bij rechters uit hoofde van hun professie zonder meer aanwezig is. Veel rechters menen dat juist het ontbreken van zo'n referentiekader bij leken ten grondslag ligt aan de kloof. De argumentatie is dat rechters impliciet of expliciet een voorliggende zaak vergelijken met een scala aan andere strafzaken die zij, vanuit eigen of collegiale ervaring, paraat hebben. Bij gewone burgers is daar geen sprake van. In een onderzoek waarin gewone mensen een strafdossier te beoordelen krijgen, waarna dat oordeel vergeleken wordt met dat van echte rechters, kan dit betekenen dat de vergelijking eigenlijk niet fair is. Volgens die critici is te verwachten dat zo een referentiekader bij burgers een matigende werking heeft op de door hen geprefereerde strafmaat in een voorliggende zaak. We gaan hier dieper in op dit veronderstelde effect van een referentiekader en we presenteren twee gerelateerde experimentele studies, waarin onder gecontroleerde omstandigheden het effect van referentiekaders wordt onderzocht.

### **Het effect van een referentiekader**

Door opleiding en ervaring hebben rechters een referentiekader opgebouwd waarbinnen een voorliggende zaak wordt vergeleken met andere, meer of minder ernstige misdrijven, en waarbinnen men zich rekenschap geeft van uitspraken in soortgelijke zaken. Rechters zijn gebonden aan het gelijkheids- en het proportionaliteitsbeginsel (Von Hirsch, 1992; Kelk, 1999). Zij streven binnen hun reeds gevormde referentiekader proportionaliteit in de straftoemeting na (Corstens, 1998; De Keijser, 2000). Het proportionaliteitsbeginsel brengt matiging met zich mee voor veel gevallen. Men houdt ruimte voor nog ernstiger gevallen en legt zelden de wettelijke maximumstraf op: rechters voelen zich door het vóórkomen

van nog zwaardere zaken begrensd in het toemeten van straf. Hun referentiekader bestaat vooral uit hun zicht op wat er zo al aan zaken voorkomt, en wat voor straf daar door de bank genomen voor gegeven wordt. Niet verwonderlijk is dat initiatieven om tot consistente straftoemeting te komen zich vooral richten op het systematiseren van een vergelijkingsperspectief (Schoep & Schuyt, 2005; Van Tulder & Diephuis, 2007).

Leken missen een dergelijk referentiekader. In het hiervoor aangehaalde onderzoek beoordelen leken eens en slechts eens een strafdossier, en zij hebben daarom geen referentiepunt. Die beoordeling door leken van het voorgelegde strafdossier mist daardoor een reële verankering, terwijl bekend is hoe belangrijk ankerpunten voor de straftoemeting zijn (Fitzmaurice & Pease, 1986; English & Mussweiler, 2001). Voor zover de leken tevoren al enige belangstelling voor en kennis van strafrecht hebben, zal dat toch voornamelijk gebaseerd zijn op media-berichten, waarbij het veelal om uitzonderlijke zaken gaat, met vaak relatief hoge eisen en hoge straffen. Zonder reëel referentiekader is voor de oordelende leek in beginsel 'the sky the limit' bij het kiezen van een passende straf. Dat leken nogal eens gebruik zullen maken van die hen geboden ruimte kan als volgt worden beargumenteerd: een leek die, voor het eerst en voor de enige maal, geconfronteerd wordt met een in detail beschreven strafzaak, zal vooral getroffen zijn door het abjecte van het ten laste gelegde misdrijf. Daardoor zal deze leek geneigd zijn tot forse maatregelen, zonder zich daarbij te kunnen of hoeven storen aan overwegingen van proportionaliteit binnen een of ander referentiekader.

Deze redenering voedt de veronderstelling dat het confronteren van leken met een testzaak hen een rudimentair referentiekader aanbiedt te midden van diverse andere zaken die variëren in ernst. Dan zou men, net als rechters, proportionaliteit willen betrachten en de testcasus relatief lichter straffen om zwaardere straffen te kunnen reserveren voor de nog zwaardere gevallen, waarvan men dan inmiddels weet heeft. Er is geen empirisch onderzoek beschikbaar hoe leken omgaan met straftoemeting in een vergelijkende context, maar uit de algemene literatuur over besluitvorming onder onzekerheid (Kahneman e.a., 1982) is bekend dat bij het oordelen een vergelijking met andere (nabije) gevallen een veel gebruikte strategie is.

Hieraan gerelateerd is 'framing': de experimentele sociale psychologie heeft ampel evidentie geleverd van framingeffecten, dat wil zeggen dat het er veel toe doet hoe een casus wordt gepresenteerd. Onder andere het onderzoek van Kahneman en Tversky (1979) liet zien dat logisch equivalente gevallen, die zo werden gepresenteerd dat ze hetzij positief hetzij negatief overkomen, tot drastisch verschillende uitkomsten leiden indien men gevraagd wordt daarover te beslissen. Wat is hiervan de betekenis voor wat we kunnen verwachten als leken een strafzaak beoordelen? Duidelijk is dat een zorgvuldige formulering vereist is, en te verwachten is dat het nogal uitmaakt of de testcasus relatief ernstig of licht is te midden van de vergelijkingszaken. Ook is evident dat het er nogal toe zal doen of men de te beoordelen casus sequentieel laat afwerken, of dat men de zaken eerst laat bestuderen, om dan een oordeel over de gehele reeks tegelijk te vragen.

Het hierna besproken onderzoek toetst de hypothese dat het aanbieden van een referentiekader aan leken hun strafoordeel matigt ten opzichte van hun oordeel in een referentiekaderloze context. We presenteren twee experimentele studies waarin aan leken om een oordeel in een strafzaak wordt gevraagd, terwijl sommigen wel en anderen geen referentiekader – in verschillend geframede varianten – wordt aangereikt.<sup>2</sup> Tussen de twee studies is een belangrijke variatie aangebracht in de aard van het referentiekader.

### **Experiment 1: vijf mishandelingscasus met verschillende mate van letsel**

Onze onderzoeksvraag is of beschikbaarheid van een referentiekader en de daarvan uitgaande (impliciete en expliciete) wens proportioneel te beslissen, een matigend effect heeft op strafvoorkeuren van gewone burgers? De verwachting is dat leken die één enkele zaak beoordelen, *ceteris paribus*, gemiddeld strenger straffen dan leken die diezelfde zaak te midden van andere, ernstigere en minder ernstige zaken beoordelen. Om effecten van referentiekaders op de straftoemeting door leken met de nodige nuance in kaart te kunnen brengen, onderscheiden wij naast een volledig referentiekader ook een tweetal eenzijdige referentiekaders: één type met enkel lichtere zaken, en één type met enkele zwaardere zaken naast de testcasus.

#### *Onderzoeksmateriaal, opzet van het experiment en dataverzameling*

Het onderzoeksmateriaal is een verzameling casusbeschrijvingen in de vorm van ‘vignetten’. Ieder van de vijf fictieve, maar realistische vignetten beschrijft in tien tot vijftien regels een geval van mishandeling met verschillende niveaus van ernst. De vignetten beschrijven steeds een verschillend geval (verhaal) van mishandeling, dit omdat we aan deelnemers meerdere vignetten wilden kunnen voorleggen. Indien alleen de ernst van letsel zou variëren, was de bedoeling van het experiment te gemakkelijk te voorzien. De vignetten worden aangeduid met een subscript dat de ernst van het opgelopen letsel karakteriseert, ten opzichte van testcasus  $C_0$ : een negatieve index staat voor minder erg letsel, een positieve voor ernstiger gevolgen:

- $C_{-2}$ : Mishandeling met blauw oog en schaafwond ten gevolge.
- $C_{-1}$ : Mishandeling met gehechte lip en schouder uit de kom ten gevolge.
- $C_0$ : Mishandeling met drie gekneusde ribben, een gebroken kaak en een zware hersenschudding ten gevolge.
- $C_1$ : Mishandeling met blijvend verlamvend nekletsel ten gevolge.
- $C_2$ : Mishandeling met de dood ten gevolge.

Ter illustratie geven wij de integrale tekst van de testcasus  $C_0$ . De twee zwaardere ( $C_1$  en  $C_2$ ) en de twee minder zware casus ( $C_{-1}$  en  $C_{-2}$ ) beschrijven andere gevallen (verhalen), die in essentie slechts op de zojuist genoemde aspecten van ernst verschillen.

2 De eerste studie profiteerde van de medewerking van een aantal Leidse bachelorstudenten, die het veldwerk hebben uitgevoerd en bij opzet en analyse behulpzaam waren (Elffers e.a., 2007b).

*Maandag omstreeks 17.15 uur is Michel Stevens ter hoogte van de Voorweg in Zoetermeer in elkaar geslagen. Hij was onderweg van zijn werk naar huis en werd onverwachts aangevallen. De verdachte Thomas Tuinstra is een 23-jarige man uit dezelfde stad. De twee kenden elkaar van de plaatselijke voetbalclub 'FC Zoetermeer', waar zij in hetzelfde team spelen. Volgens hun trainer waren er al gedurende enige tijd spanningen tussen beide spelers. Deze spanningen namen toe toen Michel zondag werd opgesteld op de gebruikelijke positie van Thomas en Thomas op de reservebank kwam. Toen Thomas zijn teamgenoot een dag later tegen het lijf liep, escaleerde de situatie. Thomas stompte Michel in het gezicht, waardoor Michel zijn evenwicht verloor en op de grond viel. Toen Michel op de grond lag, heeft Thomas hem nog een paar trappen tegen de borst en het hoofd gegeven. Michel heeft aan de mishandeling drie gekneusde ribben en een gebroken kaak overgehouden. Daarnaast heeft hij een paar dagen in het ziekenhuis gelegen vanwege een zware hersenschudding. Thomas is even later aangehouden en overgebracht naar het politiebureau voor verhoor.*

Het experiment volgt een 'between subjects design' met vier condities, ieder daarvan een variant van het referentiekader (zie tabel 1). Deelnemers werden willekeurig toegewezen aan één van die condities. De eerste variant behelst het ontbreken van een referentiekader (aan te duiden als GR, i.e. Geen Referentiekader). Deelnemers in deze variant kregen slechts één casus te beoordelen: onze testcasus  $C_0$ . Deelnemers aan de tweede variant kregen alle vijf de casus te beoordelen: dat is dus de testcasus binnen een Volledig Referentiekader (VR). In de derde conditie bestond het referentiemateriaal naast casus  $C_0$  alleen uit zwaardere zaken, het Eenzijdig Zwaarder Referentiekader (EZR), terwijl in de laatste conditie  $C_0$  werd aangeboden naast uitsluitend lichtere zaken, het Eenzijdig Lichter Referentiekader (ELR).

De volgorde van de vignetten binnen de experimentele condities was willekeurig bepaald. De instructie aan deelnemers gebood eerst alle casus te lezen alvorens zich een oordeel over passende straffen te vormen. Ook vermeldde de instructie dat de straf enkel en alleen in de vorm van een gevangenisstraf uitgedrukt kon worden, door ons omgerekend naar dagen.

**Tabel 1: Onderzoeksontwerp, vier experimentele condities ( $C_0$  is de testcasus)**

	Experimentele condities: referentiekaders			
	Geen Referentiekader GR	Volledig Referentiekader VR	Eenzijdig Zwaarder Referentiekader EZR	Eenzijdig Lichter Referentiekader ELR
+	$C_0$	$C_2$	$C_2$	
		$C_1$	$C_1$	
		$C_0$	$C_0$	$C_0$
		$C_{-1}$		$C_{-1}$
		$C_{-2}$		$C_{-2}$
-				

In deze experimentele setting is het niet in de eerste plaats van belang of proefpersonen een doorsnee uit de bevolking vormen, maar vooral dat de samenstelling van de proefpersonengroepen tussen de verschillende experimentele condities geen systematische verschillen vertoont. Wel moet, om contrasteren met rechters mogelijk te maken, worden voorkomen dat de proefpersonen zelf rechters zijn of een nagenoeg rechterlijk referentiekader zouden kunnen hebben. Gekozen is voor studenten in de exacte wetenschappen aan de Universiteit Leiden. In mei 2006 werd aan het eind van hoorcolleges Farmacologie, Biomedische Wetenschappen, en Geneeskunde studenten door hun docent verzocht te blijven zitten om deel te nemen aan een onderzoek. Van de 428 aanwezige studenten hebben er 290 deelgenomen aan de studie.<sup>3</sup> De deelnemers werden door een toevalsmechanisme aan de verschillende condities toegekend.

*Bevindingen*

De specifieke verwachting was dat de testcasus  $C_0$  alléén zwaarder bestraft zou worden dan  $C_0$  naast zwaardere ( $C_1$  en  $C_2$ ) en lichtere ( $C_{-1}$  en  $C_{-2}$ ) zaken.<sup>4</sup> Tabel 2 geeft nu de gemiddelde straf voor onze testcasus ( $C_0$ ) binnen ieder van de vier experimentele condities weer.

**Tabel 2: Gemiddelde straf (dagen gevangenis) voor testcasus  $C_0$  vergeleken tussen de vier experimentele condities, Experiment I**

Versie van referentiekader		N	Gemiddelde straf voor $C_0$	Sd.	Post hoc paarsgewijze verschiltoets (LSD) $p < .05$
GR	Geen referentiekader alleen $C_0$	60	274	181	GR $\neq$ VR GR $\neq$ EZR
VR	Volledig referentiekader $C_{-2}$ , $C_{-1}$ , $C_0$ , $C_1$ , $C_2$	70	476	440	VR $\neq$ GR VR $\neq$ EZR VR $\neq$ ELR
EZR	Eenzijdig zwaarder referentiekader $C_0$ , $C_1$ , $C_2$	71	762	955	EZR $\neq$ GR EZR $\neq$ VR EZR $\neq$ ELR
ELR	Eenzijdig lichter referentiekader $C_{-2}$ , $C_{-1}$ , $C_0$	70	223	220	ELR $\neq$ VR ELR $\neq$ EZR

Tukey-uitbijters verwijderd, zie voetnoot 4

Overall  $F_{3,267} = 13.4, p < 0.001$

Tabel 2 laat grote verschillen zien in de bestraffing van  $C_0$  tussen de versies van het referentiekader. De straf voor precies dezelfde casus is sterk afhankelijk van aan- of afwezigheid en type referentiekader dat aan de respondent wordt aange-

3 Conditie GR: N=66. Conditie VR: N=74. Conditie EZR: N=74. Conditie ELR: N=76.

4 Alvorens aan te vangen met de analyses was het geboden om te corrigeren voor extreme waarnemingen ('uitbijters'), die een sterk vertekenend effect hebben op het rekenkundig gemiddelde, en daarom verwijderd worden. Uitbijters zijn bepaald volgens de definitie van Tukey (1977), hetgeen inhoudt dat ze groter zijn dan het derde kwartiel plus 1,5 maal de interkwartielafstand. Bij ieder vignet werden zo enkele waarnemingen verwijderd.

boden. In zoverre wordt de verwachting bevestigd dat referentiekaders voor het oordeel van leken over straf van belang kunnen zijn. Inspectie van de gemiddelden in tabel 2 leert ons dat we te maken hebben met een effect dat qua richting tegengesteld is aan wat wij verwachtten. Werd  $C_0$  zonder enig referentiekader met gemiddeld 274 dagen gevangenisstraf gewaardeerd, binnen het volledige referentiekader lag die straf ruim 200 dagen hoger: 476 dagen gemiddeld. Dat tegenovergestelde effect doet zich nog sterker gelden wanneer we alleen  $C_0$  vergelijken met  $C_0$  in een eenzijdig zwaardere kader. In dat laatste geval ligt de gemiddelde straf op niet minder dan 762 dagen. Opvallend is voorts dat vergelijking tussen alleen  $C_0$  met  $C_0$  binnen een eenzijdig lichter referentiekader geen significant verschil in de gemiddeld opgelegde straffen laat zien.

Een context waarin zwaardere zaken dan de testcasus voorkomen, trekt de gemiddelde straf dus sterk omhoog, terwijl van de lichtere aangeboden zaken geen duidelijk effect uitgaat. Vergelijken we het eenzijdig zwaardere kader met het volledige kader, dan valt op dat de straf voor  $C_0$  in het volledige kader lichter is (476 tegenover 762). Tezelfdertijd levert de toevoeging van alleen lichtere zaken op de versie 'Geen Referentiekader' geen verschil in de straftoemeting op. Een en ander leidt ons tot de voorlopige conclusie dat een referentiekader waarin ernstiger zaken zijn opgenomen vooral een *dramatiserend* effect op de straftoemeting heeft, dus ook bij de relatief lichtere casus  $C_0$ , en ondanks de eventuele aanwezigheid van nog lichtere zaken. Doordat de deelnemer aan het experiment, alvorens  $C_0$  te bestraffen, nog veel dramatischer zaken heeft gelezen, zet dat hem of haar wellicht in een negatieve mentale instelling, die de te nemen beslissingen kleurt en domineert. Het aanbod van lichtere, minder ernstige zaken kan dat niet compenseren: er lijkt geen sprake van een *matigend* effect.

### Intermezzo: zwaarste casus te zwaar?

Wij noemden de conclusie van Experiment 1 'voorlopig' omdat heroverweging van de manipulatie van ernst (via letsel) in het experiment laat zien dat de aangeboden referentiekaders wellicht erg uit balans waren: de zwaarste casus was wel erg veel zwaardere dan de andere. Een alles overstemmend dramatiserend effect van ernstiger casus is dan welhaast opgelegd pandoer. Dit wordt geïllustreerd door de linkerhelft van tabel 3. Die gegevens (m.b.t. Experiment 1) suggereren dat het referentiekader in ons eerste experiment toch erg uit balans was. De straf, en bij nader inzien ook het letsel, vormen bij de ernstiger casus veel grotere stappen ten opzichte van testcasus  $C_0$  dan de verschillen tussen  $C_0$  en de lichtere casus. Zo is het verschil in strafmaat tussen  $C_0$  en  $C_2$  351 dagen, maar tussen  $C_0$  en  $C_2$  bedraagt dat verschil niet minder dan 3.242 dagen. Achteraf menen wij dat de verschillen in letsel tussen die vignetten oog wel erg groot zijn tussen de zwaardere varianten, ten opzichte van de veel kleinere verschillen in letsel tussen de lichtere varianten. Het was zodoende wellicht onmogelijk voor de respondenten om aan het beschreven dramatiserend effect te ontsnappen.

Zou er nog steeds sprake zou zijn van een dominant dramatiserend effect, dat sterker is dan een 'matigingseffect', wanneer een referentiekader aangeboden wordt dat veel meer in balans is dan dat van Experiment 1? Om die reden is een

vervolgexperiment (Experiment 2) uitgevoerd waarin specifiek die vraag centraal stond.

**Experiment 2: vijf mishandelingscasus met kleinere verschillen in ernst van het letsel**

Onze tweede onderzoeksvraag was dan ook: wanneer een referentiekader aan deelnemers aan het experiment wordt aangeboden dat qua variatie in ernst veel meer in balans is dan het referentiekader uit Experiment 1, is er dan nog steeds sprake van een al-overstemmend dramatiserend effect van ernstiger zaken?

*Materiaal: herschalen van ernst*

Om tot een beter gebalanceerd referentiekader te komen, en tegelijkertijd vergelijkbaarheid met Experiment 1 te waarborgen, zijn de vignetten uit het eerste experiment als uitgangspunt genomen en zijn enkel de daarin beschreven letsel-niveaus aangepast. Tabel 3 toont de resulterende letselniveaus als opgenomen in de vignetten van Experiment 2, tezamen met de oorspronkelijke niveaus zoals die in Experiment 1 gebruikt zijn.

**Tabel 3: Aard van letsel en gemiddeld uitgedeelde straffen bij de vijf vignetten van het volledig referentiekader (VR), Exp. 1 en Exp. 2**

	Experiment 1		Experiment 2	
	Letsel	Gevangenisstraf in dagen (gem.)	Letsel	Gevangenisstraf in dagen (gem.)
C <sub>2</sub>	Dood	3.718	Verlamming	1.784
C <sub>1</sub>	Verlamd	1.604	Verlies van een oog	908
C <sub>0</sub>	Gekneusde ribben, gebroken kaak, zware hersenschudding	476	Gebroken ribben, gebroken kaak	428
C <sub>-1</sub>	Gescheurde lip, schouder uit kom	145	Gescheurde lip, schouder uit kom, beschadigde tanden	206
C <sub>-2</sub>	Blauw oog, schaafwond	125	Blauw oog	114

Data zijn getrimd volgens de Tukey-methode (zie voetnoot 4), resulterende in N=70 in experiment 1 en N=32 in experiment 2.

*Procedure en design, dataverzameling, manipulatie-check*

Procedure en design waren identiek aan die zoals beschreven bij Experiment 1, zie tabel 1 voor het onderzoeksdesign. Gegevens zijn verzameld in mei en juni 2007 bij 143 eerstejaars studenten criminologie aan de Universiteit Leiden (N=76) en de Vrije Universiteit Amsterdam (N=67), die aan het eind van een hoorcollege hun medewerking aan het experiment verleenden. Zij werden door een toevalsmechanisme aan de experimentele condities toegewezen. Weliswaar hadden zij reeds een cursus inleiding strafrecht/strafprocesrecht gevolgd, maar wij menen dat zij



daarmee toch geenszins kunnen worden gezien als personen die een rechterlijk referentiekader benaderen.

Om te zien of de herschaling het beoogde doel bereikt, bezien we in tabel 3 de gemiddelde straffen uitgedeeld voor de vijf vignetten binnen volledig referentiekader.<sup>5</sup> Terwijl, in absolute zin, de toename in dagen gevangenisstraf van ongelijke omvang is tussen vignetten, laat tabel 3 ook een duidelijke evenredigheid zien: ieder vignet van grotere ernst levert een straf op die telkens ruwweg een factor twee zwaarder is dan de voorgaande. Die evenredigheid was ver te zoeken in Experiment 1.

### Bevindingen

Wanneer niet langer sprake is van een dominant dramatiserend effect door verwijdering van de extreem ongebalanceerde ernstiger zaken binnen een referentiekader, dan zou *in ieder geval* gevonden moeten worden dat de gemiddelde straf voor alleen  $C_0$  van gelijke omvang is aan de gemiddelde straf voor  $C_0$  binnen een volledig referentiekader.

**Tabel 4: Gemiddelde straf (dagen gevangenis) voor testcasus  $C_0$  vergeleken tussen de vier experimentele condities in Experiment 2**

Versie van referentiekader		N	Gemiddelde straf $C_0$	Sd.	Post hoc paarsgewijze verschiltoets (LSD) $p < .01$
GR	Geen referentiekader alleen $C_0$	32	182	162	GR $\neq$ VR GR $\neq$ EZR
VR	Volledig referentiekader $C_{-2}, C_{-1}, C_0, C_1, C_2$	30	428	352	VR $\neq$ GR VR $\neq$ ELR
EZR	Eenzijdig zwaarder referentiekader $C_0, C_1, C_2$	30	371	343	EZR $\neq$ GR EZR $\neq$ ELR
ELR	Eenzijdig lichter referentiekader $C_{-2}, C_{-1}, C_0$	32	137	98	ELR $\neq$ VR ELR $\neq$ EZR

Tukey-uitbijters verwijderd, zie voetnoot 5

Overall  $F_{3,120} = 9.2, p < 0.001$

In tabel 4 staat de gemiddelde gevangenisstraf voor de testcasus  $C_0$  in ieder van de vier experimentele condities in Experiment 2.<sup>6</sup> De tabel laat zien dat er ook in Experiment 2 sprake moet zijn geweest van een dramatiserend effect van ernstiger zaken binnen een referentiekader. Zelfs nu de vignetten herschaald zijn, en het referentiekader veel meer in balans te noemen is dan het referentiekader in Experiment 1, zijn de patronen in tabel 4 nagenoeg hetzelfde als die in tabel 2. De gemiddelde straf voor de testcasus binnen volledig referentiekader is met 428

5 Ook nu werden enkele uitbijters per conditie verwijderd, volgens de methode van Tukey, cf. voetnoot 4.

6 De gemiddelden zijn wederom gecorrigeerd voor uitbijters volgens dezelfde procedure als in Experiment 1.

dagen ook hier weer veel zwaarder dan de straf voor precies dezelfde casus zonder referentiekader, namelijk 182 dagen. Wederom verschilt de straf voor  $C_0$  niet tussen het eenzijdig licht referentiekader en geen referentiekader: we observeren geen ‘matigingseffect’ van lichtere zaken in een referentiekader. Tussen de straf voor  $C_0$  in het volledige referentiekader en de straf voor dezelfde casus binnen het eenzijdig zwaardere referentiekader bestaat geen significant verschil.

Het overstemmende dramatiserende effect kan, ten slotte, nog verder geïllustreerd worden aan de hand van vergelijking tussen de referentiekaders van gemiddelde straffen voor andere vignetten dan testcasus  $C_0$ . In tabel 5 staan de straffen voor de casus  $C_{-2}$  en  $C_{-1}$  vergeleken tussen het volledige referentiekader en het eenzijdig lichtere referentiekader, in Experiment 2. In dezelfde tabel staan de straffen voor  $C_1$  en  $C_2$  vergeleken tussen het volledige kader en het eenzijdig zwaardere kader. Bij de twee lichte casus is een substantieel en significant effect te zien van de aanwezigheid van zwaardere casus in een referentiekader. Zo bedraagt de gemiddelde straf voor de lichtste casus  $C_{-2}$  bij afwezigheid van de twee zwaarste casus ( $C_1$  en  $C_2$ ) 55 dagen, terwijl bij aanwezigheid van de zware casus in het volle referentiekader de straf voor datzelfde vignet is toegenomen tot 114 dagen gevangenisstraf. Wij spreken ook hier van een dramatiserend effect. Maar wat gebeurt er met de straffen voor de zwaardere casus  $C_1$  en  $C_2$  wanneer die met en zonder de lichtste vignetten worden aangeboden? Tabel 5 laat zien dat tussen de betreffende varianten van het referentiekader geen significante verschillen waar te nemen zijn tussen de gemiddelde straffen voor die zware casus. Wij kunnen daarvoor dus niet spreken van een matigingseffect.

**Tabel 5: Gemiddelde straf (dagen gevangenis) voor casus  $C_{-2}$  en  $C_{-1}$  vergeleken tussen condities VR en ELR; voor  $C_1$  en  $C_2$  vergeleken tussen condities VR en EZR in experiment 2**

Versie van referentiekader		Gemiddelde straf in dagen (tussen haakjes de standaarddeviatie)			
		$C_{-2}$	$C_{-1}$	$C_1$	$C_2$
VR	Volledig referentiekader $C_{-2}, C_{-1}, C_0, C_1, C_2$	114 (98)	206 (211)	908 (825)	1.784 (1.244)
EZR	Eenzijdig zwaarder referentiekader $C_0, C_1, C_2$			923 (946)	1.543 (1.349)
ELR	Eenzijdig lichter referentiekader $C_{-2}, C_{-1}, C_0$	55 (53)	51 (42)		
Verschiltoets tussen referentiekaders binnen een kolom (casus)		$t=3.0$ $p<0.01$	$t=4.1$ $p<0.01$	$t=0.0$ NS	$t=0.8$ NS

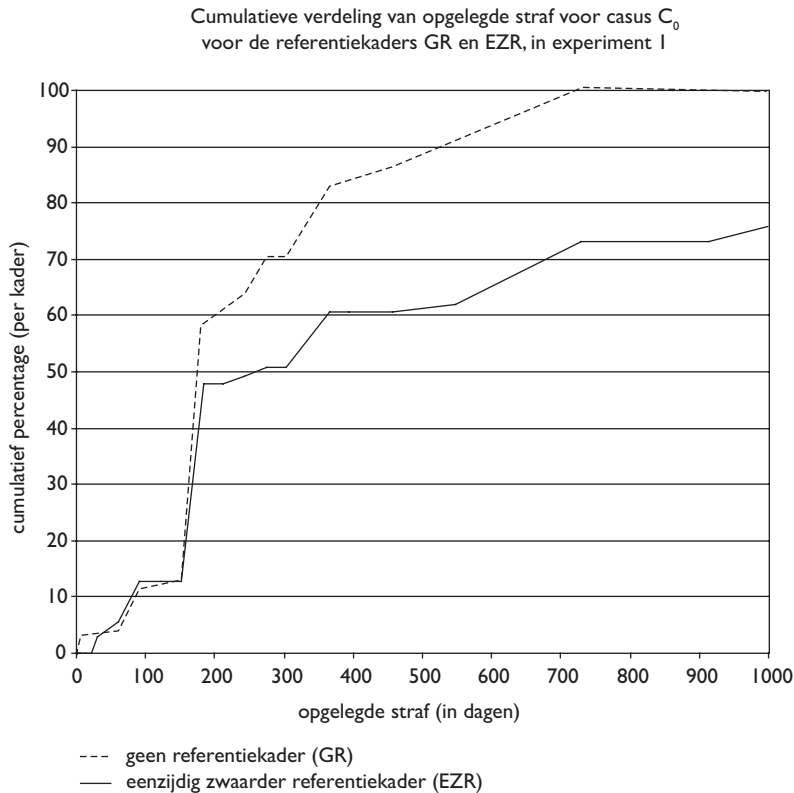
Tukey-uitbijters verwijderd, zie voetnoot 5

Noot: alle t-toetsen gecorrigeerd voor ongelijke varianties

## Dramatiseringseffect vooral bij punitief ingestelden

We zagen dat in de referentiekaders met (ook) zwaardere zaken de aan  $C_0$  uitgedeelde straffen gemiddeld hoger komen te liggen dan in de beide andere referentiekaders. Terugkerend naar Experiment 1 merken we op dat tabel 2 nog meer toont: bij de referentiekaders met zwaardere zaken is niet alleen het gemiddelde hoger, maar ook de spreiding. Er is daarom aanleiding iets nauwkeuriger te onderzoeken wat er aan de hand is. Dat doen we aan de hand van de volgende figuur, waarin we kijken naar de cumulatieve verdeling van uitgedeelde straffen in de beide referentiekaders GR en EZR, de kaders die qua spreiding het meest van elkaar verschillen.

### Cumulatieve verdeling van strafzwaarte per kader



De figuur laat zien dat de beide kaders nauwelijks verschillen voor de lage strafzwaarten: mensen die minder dan een half jaar gevangenisstraf willen geven komen in even groten getale voor in beide kaders. Pas vanaf de mediaan gaan de groepen uiteenlopen: de mensen binnen de EZR-groep die relatief (in vergelijking met anderen in hun groep) zwaar willen straffen, straffen verhoudingsgewijs veel zwaarder dan hun tegenpolen binnen de GR-groep: zo geldt bijvoorbeeld dat

in die laatste groep de zwaarste uitgedeelde straf op twee jaar ligt, terwijl nog bijna 30 procent van de EZR-groep daar vaak fors bovenuit wil gaan. Onze interpretatie van dit fenomeen is dat er een interactie-effect bestaat van algemene punitiviteit en gevoeligheid voor het dramatiseringseffect van een zwaarder referentiekader. Anders gezegd: wie niet zo erg punitief is ingesteld, blijft niet zo gevoelig voor dramatisering door een zwaarder kader, wie sterk punitief is juist wel.<sup>7</sup> Deze interpretatie, dat slechts een bepaald deel van het publiek gevoelig is voor dramatisering, sluit goed aan bij de bevindingen van De Keijser e.a. (2006; vgl. ook Kuhn, 2002), die lieten zien dat ook de kloof tussen rechter en publiek het sterkste is bij hen die zich grote zorgen maken omtrent criminaliteit en de veronderstelde toename daarvan.

Deze interpretatie van het gevondene nuanceert dus de eerdere conclusies: een zwaarder referentiekader trekt inderdaad de strafzwaarte gemiddeld omhoog, maar we moeten dit dramatiseringseffect vooral lokaliseren onder de toch al punitief ingestelde mensen. Een en ander laat uiteraard onverlet dat de initiële hypothese, dat een referentiekader tot matiging van de uitgedeelde straffen zou leiden, moet worden verworpen.

## Conclusies en discussie

In dit artikel beschouwden wij of onderzoek naar straftoemeting door leken in concrete casus een eerlijker vergelijking met rechters oplevert, wanneer die leken de te beoordelen strafzaken in de context van andere zaken beoordelen. De verwachting was dat een context in de vorm van een referentiekader met zaken van variërende ernst tot een milder oordeel voor een gegeven zaak leidt dan wanneer diezelfde zaak in isolatie wordt beschouwd. Het vergelijkend perspectief van rechters, door hun ervaring en de hen ter beschikking staande instrumenten, ontbeert de gemiddelde leek immers wanneer deze een enkele keer een straf toemeet. Wij beargumenteerden dat voor rechters het proportionaliteitsbeginsel matiging met zich meebrengt voor veel gevallen en dat leken een reële context missen, waarbinnen zij proportioneel kunnen straffen.

De experimenten die wij uitvoerden, lieten zien dat er inderdaad een effect van referentiekaders uitgaat op het oordeel van leken, althans voor de in ons onderzoek bestudeerde studenten; dat effect was echter tegenovergesteld aan de verwachting. Er bleek een *dramatiserend* effect te zijn van het gelijktijdig ook aanbieden van zwaardere casus, terwijl van enig *matigend* effect van het gelijktijdig ook aanbieden van lichtere zaken in een referentiekader geen sprake bleek. Trachtten wij dat effect nog te wijten aan het wel erg ongebalanceerde referentiekader van Experiment 1, ook in Experiment 2 werden dezelfde sterke aanwijzingen voor dramatisering ten gevolge van het opnemen van zwaardere zaken gevonden. Op de keper beschouwd konden we laten zien dat het dramatiseringseffect vooral diegenen tot nog zwaarder straffen verleidt die van aanvang af punitief zijn ingesteld, terwijl minder punitief ingestelden daarvoor relatief ongevoelig zijn.

7 Vergelijkbare grafieken kunnen worden gepresenteerd voor een vergelijking van volledig kader (VR) met geen kader (GR), alsook voor Experiment 2. Korthedshalve laten we die resultaten onbesproken.

Kunnen we achteraf begrijpen waarom de hypothese wordt verworpen, dat een referentiekader tot matiging leidt? In onze interpretatie reikt een referentiekader aan leken niet alleen het verhoopte mechanisme van proportioneel straffen aan, met de noodzaak ruimte te houden voor nog ernstiger gevallen. Daarnaast, of zelfs in de eerste plaats, activeert zo'n referentiekader juist een ander proces, het zich realiseren wat een abjecte misdrijven zoal worden gepleegd. Deze verklaring past heel goed in het framingsparadigma. Echter, zowel de rechters die het argument van het referentiekader naar voren brachten, als wij als onderzoekers, die dat vervolgens hebben overgenomen, hadden een te restrictieve interpretatie van wat het frame aanbrengt. Een meervoudig kader richt niet alleen de aandacht op de onderlinge *verhouding* van de zaken, maar verandert ook de mentale instelling van de beoordelaars in *absolute zin*, doordat het hen doordringt van de ernst van de te berechten misdrijven. Het referentiekader zorgt niet alleen voor het benadrukken van proportionaliteit, maar ook voor dramatisering. Rechters zijn, door gewenning, vermoedelijk niet meer gevoelig voor dit laatste effect, maar leken, die voor het eerst met zulke zaken worden geconfronteerd, zijn dat wel.

Al met al lijkt een diverse malen gesuggereerde bedenking aangaande de validiteit van het eerder besproken experimenteel kloofonderzoek hiermee ontkracht te zijn, aan de hand van concreet en gedetailleerd casusmateriaal. Naar aanleiding van onze experimenten zou de stelling zelfs te verdedigen zijn, dat de kloof tussen rechters en publiek nog veel groter wordt wanneer we de mening van het publiek binnen een referentiekader meten, in plaats van gegeven één enkele geïsoleerde casuspositie. De theoretische verwachtingen die ten grondslag lagen aan het referentiekadereffect, dienen op basis van de empirische bevindingen zonder meer verworpen te worden. Het mechanisme van matiging ten gevolge van proportionaliteit binnen een aangeboden referentiekader snijdt *voor leken* geen hout.

Natuurlijk kan men tegenwerpen dat de referentiekaders uit onze experimenten vrij bescheiden van aard zijn, en nog in de verste verte niet lijken op de veel uitgebreidere en meer verankerde referentiekaders die echte rechters impliciet of expliciet hanteren. Men kan dan de stelling betrekken dat de hypothese 'het referentiekader verkleint de kloof' pas ontkracht is als vergelijkbare resultaten worden gevonden in een studie waarin leken eerst een 'echt' referentiekader hebben gevormd of 'aangeleerd', bijvoorbeeld door de leken bij herhaling uitgebreide en zeer variërende zaken en kaders aan te bieden. Pas door zo'n opzet zou een echt eerlijke vergelijking met rechters tot stand kunnen komen.

Het is natuurlijk waar dat onze experimentele referentiekaders maar een flauwe afspiegeling zijn van de ervaringskaders die echte rechters al doende opbouwen. Toch wijzen wij de suggestie van de hand dat fair onderzoek alleen mogelijk is door leken net zolang te trainen tot ze ook zo'n echt referentiekader hebben. In die opzet gooien we namelijk het kind met het badwater weg: al wat we dan zouden aantonen is dat leken die getraind worden tot ze nagenoeg rechters zijn, ook qua straftoemeting niet meer van rechters te onderscheiden zijn. Uiteraard zijn ze dan echter ook geen leken meer te noemen.

## Literatuur

- Barber, J. & Doob, A.N. (2004). An analysis of public support for severity and proportionality in the sentencing of youthful offenders. *Canadian Journal of Criminology and Criminal Justice*, 46(3), 327-342.
- Beyens, K. (2000). *Straffen als Sociale Praktijk: Een Penologisch Onderzoek naar Straftoemeting*. Brussel: VUBpress.
- Bottoms, A.E. (1995). The philosophy and politics of punishment and sentencing. In: C.M.V. Clarkson & R. Morgan (eds.). *The politics of sentencing reform*. Oxford: Clarendon Press, 17-50.
- Corstens, G.J.M. (1998). Eenheid van rechtspraak, in het bijzonder in de straftoemeting *Nederlands Juristenblad*, 73, 297-302.
- Cullen, F.T., Fisher, B.S. & Applegate, B.K. (2000). Public opinion about punishment and corrections. In: M. Tonry (ed.). *Crime and Justice: A review of research* (Vol. 27). Chicago: University of Chicago Press, 1-79.
- Elffers, H. & Keijser, J.W. de (2004). Het geloof in de kloof: Wederzijdse beelden van rechters en publiek. In: J. W. de Keijser & H. Elffers (red.). *Het maatschappelijk oordeel van de strafrechter: De wisselwerking tussen rechter en samenleving*. Den Haag: Boom Juridische uitgevers, 53-84.
- Elffers, H. & Keijser, J.W. de (2007). De strafrechter en de burger. Zij konden bijeen niet komen ... *Rechtstreeks*, 2, 7-30.
- Elffers, H. & Keijser, J.W. de (2008). Different perspectives, different gaps. Does the general public demand a more responsive judge? In: H. Kury (ed.). *Fear of crime – Punitivity. New Developments in Theory and Research (Crime and Crime Policy Vol. 3)*. Bochum: Universitätsverlag Dr. N. Brockmeyer, 447-470.
- Elffers, H., Keijser, J.W. de, Koppen, P.J. van & Haeringen, L. van (2007a). Newspaper juries: A field experiment concerning the effect of information on attitudes towards the criminal justice system. *Journal of Experimental Criminology*, 3(2), 163-182.
- Elffers, H., Keijser, J.W. de, Haeringen, L. van, Kanning, M., Koesse, M., Kruijff, M. de, Vreeling, J., Wijngaarden, S. van (2007b). *De punitiviteitskloof tussen rechters en publiek. Effect van referentiekader en van media-aandacht*. NSCR rapport 2007/4. Leiden: NSCR.
- Englich, B. & Mussweiler, T. (2001). Sentencing under uncertainty: anchoring effects in the courtroom. *Journal of applied social psychology* (31)7, 1535-1551.
- Fitzmaurice, C. & Pease, K. (1986) *The psychology of judicial sentencing*. Manchester: Manchester University Group.
- Hessing, D.J., Keijser, J.W. de & Elffers, H. (2003). Explaining capital punishment support in an abolitionist country: The case of the Netherlands. *Law and Human Behavior*, 27(6), 605-622.
- Hoge Raad voor de Justitie. (2007). *De Belgen en Justitie in 2007*. Brussel: Bruylant.
- Hough, M. & Park, A. (2002). How malleable are attitudes to crime and punishment? Findings from a British deliberative poll. In: J.V. Roberts & M. Hough (eds.). *Changing attitudes to punishment: Public opinion, crime and justice*. Cullompton, Devon: Willan, 163-183.
- Hough, M. & Roberts, J.V. (1998). *Attitudes to punishment: Findings from the British Crime Survey*. Home Office Research Study, vol. 179. London: Home Office.
- Hough, M. & Roberts, J.V. (1999). Sentencing trends in Britain. *Punishment and Society*, 1(1), 11-26.
- Hutton, N. (2005). Beyond populist punitiveness? *Punishment & Society*, 7(3), 243-258.
- Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.

- Kahneman, D., Slovic, P. & Tversky, A. (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.
- Keijser, J.W. de (2000). *Punishment and purpose, from moral theory to punishment in action*. Amsterdam: Thela Thesis.
- Keijser, J.W. de & Elffers, H. (2007) Onbegrip voor het strafoordeel bedreigt het draagvlak voor het rechtssysteem niet. In: M. Malsch & N. van Manen (red.). *De begrijpelijkheid van de rechtspraak*. Den Haag: Boom Juridische uitgevers, 115-128.
- Keijser, J.W. de, Koppen, P.J. van & Elffers, H. (2006). *Op de stoel van de rechter: Oordeelt het publiek net zo als de strafrechter?* (Research Memoranda; vol. 2, afl. 2). Den Haag: Raad voor de rechtspraak.
- Keijser, J.W. de, Koppen, P.J. van & Elffers, H. (2007). Bridging the gap between judges and the public? A multi-method study. *Journal of Experimental Criminology*, 3(2), 131-161.
- Kelk, C. (1999). Proportionaliteit in het strafrecht. In: J. Blad e.a. (red.). *Crimineel jaarboek 1999*. Nijmegen: Ars Aequi Libri, 15-25.
- Koppen, P.J. van (2003). *Verankering van rechtspraak: Over de wisselwerking tussen burger, politie, justitie en rechter*. Deventer: Kluwer (oratie Vrije Universiteit).
- Kuhn, A. (2002). Public and judicial attitudes to punishment in Switzerland. In: J.V. Roberts & M. Hough (eds.). *Changing attitudes to punishment: Public opinion, crime and justice*. Cullompton, Devon: Willan, 115-127.
- Mattinson, J. & Mirrlees-Black, C. (2000). *Attitudes to crime and criminal justice: Findings from the 1998 British Crime Survey*. London: Home Office.
- Ministerie van Justitie (2003). *Justitie Issue Monitor* (ongepubliceerd rapport).
- Mirrlees-Black, C. (2002). Improving public knowledge about crime and punishment. In: J.V. Roberts & M. Hough (eds.), *Changing attitudes to punishment: Public opinion, crime and justice*. Cullompton, Devon: Willan, 184-197.
- Roberts, J.V. & Hough, M. (2002). Public attitudes to punishment: The context. In: J.V. Roberts & M. Hough (eds.). *Changing attitudes to punishment: Public opinion, crime and justice*. Cullompton, Devon: Willan, 1-14.
- Roberts, J.V. & Hough, M. (2005). *Attitudes to sentencing and the courts. Understanding public attitudes to criminal justice*. Berkshire: Open University Press, 68-87.
- Roberts, J.V., Stalans, L.J., Indermaur, D. & Hough, M. (2003). *Penal populism and public opinion: Lessons from five countries*. New York: Oxford University Press.
- Schoep, G.K. & Schuyt, P.M. (2005). *Instrumenten ter ondersteuning van de rechter bij de straftoemeting*. Leiden: Meijers Instituut.
- Sociaal en Cultureel Planbureau (2002). *Sociaal en cultureel rapport 2002: De kwaliteit van de quatrijke sector*. Den Haag: Sociaal en Cultureel Planbureau.
- Sociaal en Cultureel Planbureau (2005). *De sociale staat van Nederland 2005*. Den Haag: Sociaal en Cultureel Planbureau.
- Sprott, J.B. & Doob, A.N. (1997). Fear, victimization, and attitudes to sentencing, the court and the police. *Canadian Journal of Criminology*, 39(3), 275-291.
- Tukey, J.W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison Wesley.
- Tulder, F. van & Diephuis, B. (2007) *Afgewogen straffen: Analyse en verbetering van de Databank Consistente Straftoemeting* (Vol. 3). Den Haag: Raad voor de rechtspraak.
- Von Hirsch, A. (1992). Proportionality in the philosophy of punishment, *Crime and justice: A review of research* (Vol. 16). Chicago: Chicago University Press, 55-98.