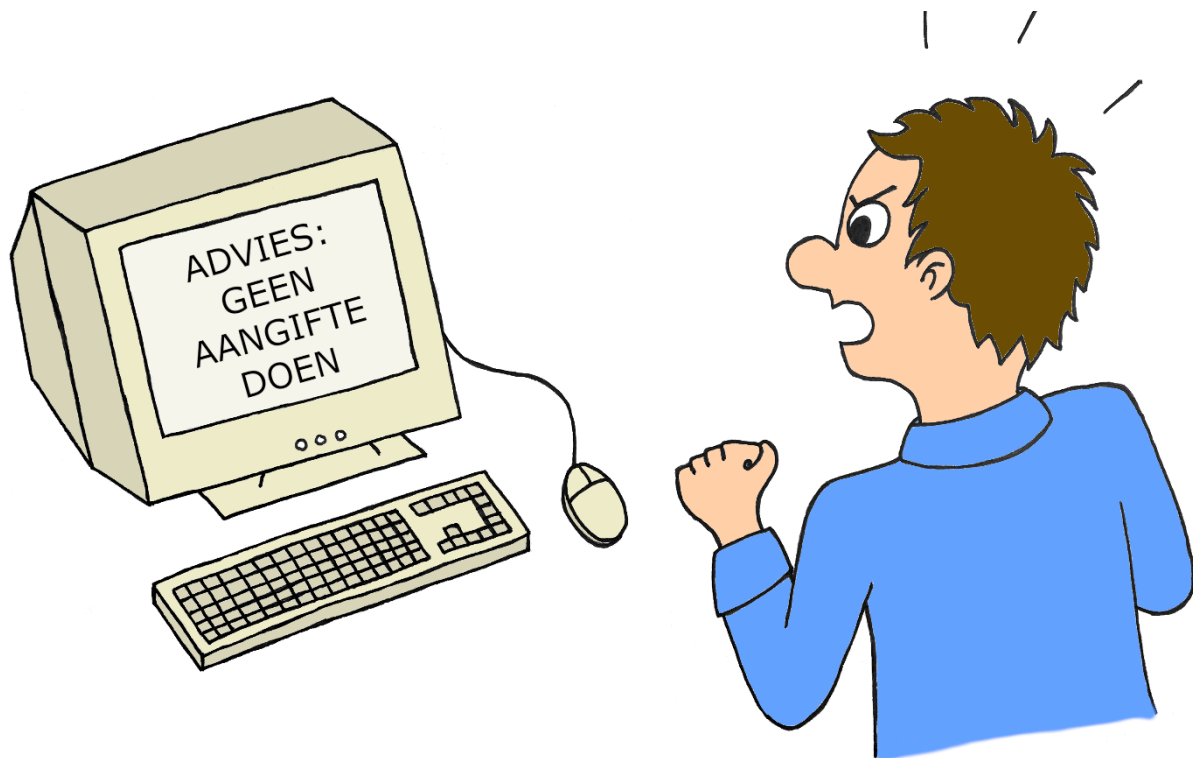


Artificiële intelligentie, is dat wel te vertrouwen?

Een experimentele studie naar het effect van uitleg over beslissingen van het intelligente aangiftesysteem van de politie op het vertrouwen van burgers in deze beslissingen.



Esther Nieuwenhuizen



Universiteit Utrecht



Artificiële intelligentie, is dat wel te vertrouwen?

Een experimentele studie naar het effect van uitleg over beslissingen van het intelligente aangiftesysteem van de politie op het vertrouwen van burgers in deze beslissingen.

Masteropleiding Publiek Management

Esther Nieuwenhuizen

5714796

Universiteit Utrecht

Departement Bestuurs- en Organisationswetenschap

2 juli 2020

Begeleiding

Eerste lezer: dr. Stephan Grimmelikhuijsen

Tweede lezer: Machiel van der Heijden MSc

Politie: Carlos Soares

Voorwoord

Utrecht, juli 2020

Voor u ligt mijn masterscriptie, waarmee er een eind komt aan mijn studententijd. Ik herinner me nog goed dat ik in november 2019 een keuze moest maken voor een scriptieonderwerp. Toen alle begeleiders hun onderzoeksthema's presenteerden, trok het verhaal van Stephan Grimmelikhuijsen meteen mijn aandacht. Zijn enthousiasme voor zijn onderzoeksthema (*AI, big data en algoritmen in het openbaar bestuur*) en voor het doen van experimenteel onderzoek werkte aanstekelijk. Het leek me een mooie uitdaging om een nieuwe manier van onderzoek doen te leren. En zo geschiedde.

Ruim een half jaar later ben ik helemaal ondergedompeld in de wereld van uitlegbare artificiële intelligentie (AI) en heb ik veel kennis opgedaan over experimentele methoden. Het lezen van talloze artikelen over (de werking van) algoritmen en urenlang filmpjes bekijken op Youtube om mijn statistiekkennis te verdiepen hebben hun vruchten afgeworpen. Ik ben trots op het eindresultaat! Al had ik dat natuurlijk niet kunnen bereiken zonder de hulp van velen.

Allereerst wil ik mijn scriptiebegeleider Stephan bedanken. Bedankt voor alle concrete en adequate feedback. Ik heb veel van en aan je gevraagd, omdat het doen van experimenteel onderzoek helemaal nieuw voor mij was. Zowel op inhoudelijk als methodologisch vlak heb ik ontzettend veel van je geleerd. Ook wil ik Erwin, mijn groepsgenoot, bedanken. Het was fijn om iemand te hebben waarmee ik inhoudelijk kon sparren en die altijd meedacht. Verder gaat mijn dank uit naar Machiel, mijn tweede lezer, die me gestimuleerd heeft om kritischer naar de theorie te kijken en daarmee mijn stuk naar een hoger niveau heeft gebracht. Ook bedank ik Carlos, Daphne en Floris voor hun hulp en begeleiding vanuit de Politie en het Nationaal Politielab AI. Tenslotte wil ik mijn familie en vrienden bedanken, die me elk op hun eigen manier geholpen hebben. Zonder jullie was het nooit gelukt om meer dan 500 respondenten te krijgen voor mijn onderzoek. Bedankt daarvoor!

Het waren bijzondere maanden om aan zo'n groot afstudeeronderzoek te werken. Zo zorgde Covid-19 ervoor dat al het onderwijs op afstand moest worden gegeven. Dit betekende dat al het contact met mijn begeleider en andere studenten digitaal plaatsvond. In het begin was dit even wennen, maar al snel hadden we onze draai hierin gevonden. Daarnaast heeft de *Black Lives Matter*-beweging veel stof doen opwaaien de afgelopen maanden. Dit onderstreept het belang van verantwoord gebruik van algoritmen bij de politie en motiveert mij om me hier de komende jaren voor in te zetten. Ik kijk er daarom erg naar uit om in september te beginnen als promovenda aan de USBO, waarbij ik me verder zal specialiseren in transparante algoritmisering bij de politie.

Esther Nieuwenhuizen

Samenvatting

In de publieke sector vindt een verandering plaats waarbij steeds meer beslissingen die impact hebben op burgers met behulp van algoritmen tot stand komen. Hoewel algoritmische modellen geprezen worden om hun objectiviteit en nauwkeurigheid, kleeft er ook een gevaar aan het gebruik van zulke modellen. Door het gebruik van algoritmen bij besluitvormingsprocessen in de publieke sector bestaat het risico op gebrek aan transparantie, waardoor burgers hun vertrouwen in overheidsbesluiten kunnen verliezen.

Om deze reden wordt zowel in de wetenschap als in de maatschappij gepleit voor uitlegbare AI, zodat burgers begrijpen waarom zij een bepaalde beslissing krijgen. In de wetenschap zijn al enkele onderzoeken gedaan over de relatie tussen uitlegbaarheid en het vertrouwen dat men heeft in een algoritmische beslissing. Deze onderzoeken richten zich echter niet op algoritmische besluiten in de publieke sector, waardoor nog onduidelijk is wat voor effect verschillende soorten uitleg kunnen hebben op het vertrouwen van burgers in zulke besluiten. Dit onderzoek probeert daar inzicht in te verkrijgen door de volgende vraag centraal te stellen: *“Wat is het effect van uitlegbaarheid van algoritmische besluiten op het vertrouwen van burgers in deze besluiten?”*.

Uit bestaand onderzoek komt naar voren dat twee soorten uitleg bij kunnen dragen aan het vertrouwen in algoritmische besluiten, namelijk:

- Procedurele uitleg: informatie over de gehele beslisprocedure.
- Inhoudelijke uitleg: motivering van het specifieke besluit.

Aan de hand van een surveyexperiment zijn de effecten van deze soorten uitleg onderzocht. Hiervoor is gebruik gemaakt van een casus: de slimme keuzehulp. Dit is een beslissysteem van de politie dat gebruik maakt van algoritmen om meldingen van internetfraude te beoordelen. 536 respondenten hebben in het surveyexperiment de slimme keuzehulp doorlopen. Hierbij zijn vier groepen onderscheiden, die ieder een andere uitleg kregen aangeboden. Vervolgens hebben de respondenten vragen ingevuld waaruit kon worden opgemaakt of zij vertrouwen hebben in het advies van de slimme keuzehulp. Hierdoor kon de causale relatie tussen uitlegbaarheid en vertrouwen in algoritmische besluiten worden onderzocht.

Uit de resultaten komen drie belangrijke bevindingen naar voren:

1. Een procedurele uitleg heeft een klein, positief effect op vertrouwen.
2. Een inhoudelijke uitleg heeft een groot, positief effect op vertrouwen.
3. Een inhoudelijke uitleg maakt een procedurele uitleg overbodig.

Mensen willen weten waarom zij in hun specifieke geval een afwijzing ('geen aangifte doen') krijgen. Een inhoudelijke motivering van een beslissing levert het meeste vertrouwen op in een algoritmische beslissing. Met de originele uitleg van de slimme keuzehulp kiest 72% van de mensen dat een afwijzend advies krijgt ('geen aangifte doen') ervoor om alsnog aangifte te doen. Dit onderzoek laat zien dat het geven van een inhoudelijke uitleg dit percentage kan terugbrengen naar 9,4%.

Om verantwoord door te kunnen gaan met het gebruik van algoritmen bij publieke besluitvormingsprocessen is het dus van belang dat er -waar mogelijk- altijd een inhoudelijke uitleg wordt gegeven. Indien een inhoudelijke uitleg niet mogelijk is, kan een procedurele uitleg ook voor een toename in vertrouwen zorgen, al is dit effect veel kleiner. Het is daarom raadzaam om na te denken of het wel wenselijk is om algoritmische beslissystemen te gebruiken waarbij het niet mogelijk is om het besluit inhoudelijk toe te lichten.

Inhoud

1	Inleiding.....	9
1.1	Toenemende impact van AI	9
1.2	Uitlegbare AI en de slimme keuzehulp	10
1.3	Probleem-, vraag- en doelstelling	12
1.4	Deelvragen	12
1.5	Maatschappelijke relevantie	13
1.6	Wetenschappelijke relevantie	14
1.7	Leeswijzer	14
2	De slimme keuzehulp	16
2.1	Tekstanalyse	16
2.2	Tekstclassificatie	17
2.3	Vervolg vragen	17
2.4	Advies	18
3	Theoretisch kader	20
3.1	Algoritmische besluiten	20
3.2	Uitlegbaarheid	23
3.3	Vertrouwen	31
3.4	Relatie uitlegbaarheid en vertrouwen	34
4	Methoden	39
4.1	Wetenschapsfilosofische benadering	39
4.2	Onderzoeksmethoden	39
4.3	Operationalisatie	44
4.4	Selectie respondenten	53
4.5	Data-analyse	55
4.6	Betrouwbaarheid en validiteit	56
4.7	Ethische verantwoording	58
5	Resultaten.....	60
5.1	Vertrouwende houding	60
5.2	Op vertrouwen gebaseerd gedrag	64
5.3	Slotsom	70
6	Conclusie en discussie	72
6.1	AI, is dat wel te vertrouwen?	72
6.2	Academische implicaties	74
6.3	Maatschappelijke implicaties	80

7	Literatuur.....	83
8	Bijlagen.....	93
8.1	Beschrijvende statistiek respondenten	93
8.2	Factoranalyse	94
8.3	Contrasten	94
8.4	Experiment	95
8.5	Aanvullende analyses	105

1 Inleiding

1.1 TOENEMENDE IMPACT VAN AI

Sinds september 2019 maakt de politie in Nederland gebruik van een 'slimme keuzehulp', die werkt op basis van artificiële intelligentie (AI) (Politie, 2019a). Bij het doen van aangifte van internetoplichting moeten burgers eerst de slimme keuzehulp doorlopen. Door het beantwoorden van een aantal vragen met betrekking tot de mogelijke fraudezaak kan de slimme keuzehulp snel vaststellen of er sprake is van een strafbaar feit, en direct de aangever adviseren om al dan niet aangifte te doen. Hierdoor kunnen onterechte aangiftes voorkomen worden (Politie, 2019a). De slimme keuzehulp is ontwikkeld in samenwerking met het Nationaal Politielab AI (ICAI, z.d.). In dit lab wordt onderzocht hoe AI het politiewerk kan ondersteunen. Met behulp van AI kan de politie haar eigen operationele slagkracht en efficiëntie vergroten, maar ook de dienstverlening verbeteren en servicegericht werken (Politie, 2019b).

Dit sluit aan bij de globale trend waarbij AI-ontwikkelingen een grote rol gaan spelen in ons dagelijks leven (Fry, 2018). Steeds meer beslissingen die een impact hebben op de maatschappij worden niet meer gemaakt door menselijke experts, maar worden nu door algoritmen ondersteund of zelfs gemaakt (Lepri, Oliver, Letouzé, Pentland & Vinck, 2018, pp. 612). Dit kan zorgen voor een verbeterde dienstverlening door de overheid (Lepri et al., 2018, p. 612). Ook binnen de Nederlandse overheid wordt op allerlei terreinen gebruik gemaakt van AI-ontwikkelingen om de dienstverlening te vergroten, maar ook om effectiever en efficiënter te werken (Rijksoverheid, 2019). Mona Keizer, staatssecretaris van Economische Zaken en Klimaat, presenteerde op 8 oktober 2019 de speerpunten (en mogelijke valkuilen) van de koers die Nederland inzet omtrent AI, in het *Strategisch Actieplan Artificiële Intelligentie (SAPAI)*. Eén van de speerpunten van het SAPAI is het waarborgen van een inclusieve aanpak die de mens centraal stelt, waarbij de overheid streeft naar betrouwbare AI. Belangrijk hierbij is dat AI zo wordt ingezet, dat iedereen het kan vertrouwen (Rijksoverheid, 2019).

Bij het gebruik van algoritmen binnen de overheid is dit zeer relevant. Aan de hand van algoritmen worden voorspellende modellen ontworpen die ambtenaren gebruiken bij het maken van beslissingen die voor burgers impact hebben (Guidotti et al., 2018). Hoewel deze modellen geprezen worden om hun objectiviteit en nauwkeurigheid, kleven er ook gevaren aan het gebruik van algoritmen. Zo bestaat het risico op een gebrek aan transparantie en verantwoordelijkheid, waardoor, zoals eerder benoemd, burgers en bedrijven het vertrouwen in de overheid kunnen verliezen (Lepri et al., 2018, pp. 612-614).

Om het vertrouwen in AI te behouden, is het van belang dat het transparant is hoe partijen AI toepassen. Met name in het geval van beslissingen die rechtsgevolgen hebben, een grote impact hebben op mens of maatschappij, of wanneer beperkt sprake is van mogelijkheden voor menselijke interventie en controle (Rijksoverheid, 2019, p. 44). Bij beslissingen die een individu raken, is het voor het vertrouwen in het beslissysteem volgens De Laat (2018, p. 526) extra relevant dat besluiten uitlegbaar zijn. Daarbij wordt in Europese wetgeving, *General Data Protection Regulation* (GDPR), gesteld dat burgers het recht hebben om een uitleg te krijgen over beslissingen waarbij gebruik wordt gemaakt van geautomatiseerde processen (De Laat, 2018, p. 539). Naast een wettelijke verantwoordelijkheid om inzicht te geven in het gebruik van AI, is het dus ook van belang voor het vertrouwen van burgers en bedrijven dat zo'n AI-toepassing transparant en uitlegbaar is (Diakopoulos, 2016; Kizilcec, 2016; Kroll et al., 2017).

1.2 UITLEGBARE AI EN DE SLIMME KEUZEHULP

Zoals eerder genoemd, is een gebrek aan transparantie een risico bij het gebruik van algoritmen binnen de overheid. De Laat (2018, p. 526) stelt dat het eerder uitzondering dan regel is dat algoritmen transparant en uitlegbaar gemaakt worden. Hij zegt dan ook dat het niet verantwoord is om verder te gaan met het invoeren van algoritmische besluitmachines zonder na te denken over de transparantie en uitlegbaarheid van deze systemen. Zowel in de wetenschap als in de maatschappij wordt een debat gevoerd over 'uitlegbare AI' (Mittelstadt, Russell & Wachter, 2019). Dit verwijst naar een uitleg die een rechtvaardiging geeft voor een beslissing van de ene entiteit (bijvoorbeeld het algoritme) naar de andere entiteit (bijvoorbeeld de burger) en vervolgens gebruikt kan worden om die beslissing te vertrouwen (Hind, 2019, p. 17). De uitleg moet dus inzicht geven in de motivering op grond waarvan AI een beslissing neemt of een conclusie trekt (Doran, Schulz & Besold, 2017).

Een uitleg is echter niet altijd noodzakelijk. Zo onderscheidt Hind (2019, p. 19) een spectrum aan algoritmische beslissingen, waarin de impact die ze hebben op een individu uiteenloopt. Hoe sterker de impact, des te groter de behoefte aan uitleg. Bij de beslissing welk medicijn iemand moet nemen, of iemand een promotie krijgt of bij het bepalen van de hoogte van een straf, is een uitleg gewenst. Hierbij hebben de beslissingen namelijk grote gevolgen voor de individuen die hier direct of indirect bij betrokken zijn. Aan de andere kant zijn er ook algoritmische beslissingen op basis waarvan je een bepaalde advertentie te zien krijgt of op basis waarvan je wordt aangeraden naar een specifieke film te kijken. Dit zijn geen beslissingen die het leven van iemand drastisch veranderen. Dit maakt de noodzaak voor een uitleg veel kleiner (Hind, 2019, p. 19).

Beslissingen van de slimme keuzehulp hebben impact op het leven van de individu die de beslissing krijgt. Ze zijn daarom te plaatsen op het spectrum van Hind (2019) aan de

kant waar besluiten een uitleg behoeven. De slimme keuzehulp is een beslissysteem van de politie dat gebruik maakt van algoritmen om aangiftes van internetfraude te beoordelen. Het geeft een positief of negatief advies met betrekking tot het al dan niet doen van aangifte bij internetfraude. Op dit moment kunnen slachtoffers ondanks een negatief advies toch nog aangifte doen (Politie, 2019a). Deze mogelijkheid zal op termijn verdwijnen, zodra de slimme keuzehulp geoptimaliseerd is. Bij zaken van internetfraude moet de burger dan dus de beslissing van de slimme keuzehulp accepteren. Uit onderzoek van Kunst & Van Dijk (2009) blijkt dat fraudezaken veel impact hebben op slachtoffers. Zo hebben slachtoffers vaak (veel) geld verloren en een negatief zelfbeeld ontwikkeld (zichzelf "dom" vinden) als gevolg van de fraude (Kunst & Van Dijk, 2009, p. 51). Daarnaast beoordeelt 90% van de mensen die via de slimme keuzehulp aangifte willen doen hun zaak van internetoplichting als 'ernstig' of 'zeer ernstig' (Jorritsma, Cremers & De Jong, 2020). Dit onderstreept de impact die het besluit van de slimme keuzehulp kan hebben op het leven van het slachtoffer.

Op dit moment kiest 72% van de mensen dat een afwijzend advies krijgt ('geen aangifte doen') er toch voor om aangifte te doen (Soares, 2020). De huidige uitleg wordt dus nog onvoldoende vertrouwd door gebruikers van de slimme keuzehulp. Daarbij komt dat sinds de coronacrisis er een enorme groei heeft plaatsgevonden in onlinefraude, omdat mensen meer online hun aankopen zijn gaan doen en criminelen hier handig op inspelen (Voskuil, 2020). Dit vergroot de noodzaak van uitlegbare beslissingen van de slimme keuzehulp.

Tot op heden is echter nog geen onderzoek gedaan naar de uitlegbaarheid van beslissingen van de slimme keuzehulp (Bex, 2019), terwijl dit cruciaal is voor het vertrouwen van burgers in overheidsbesluiten (Lepri et al., 2018). Ook vanuit de algemene beginselen van behoorlijk bestuur is de overheid verplicht om haar besluiten te motiveren (neergelegd in artikel 3:46 en 3:47 van de Algemene wet bestuursrecht). Bij algoritmische besluiten, zoals die van de slimme keuzehulp, moet de politie dus in staat zijn het besluit uit te leggen (Poorter & Goossens, 2019).

Hoe zo'n uitleg over een algoritmische beslissing het beste vormgegeven kan worden, is echter nog onduidelijk. Al sinds 1970 blijkt het uitlegbaar maken van algoritmische beslissingen problematisch (Wachter, Mittelstadt & Floridi, 2017a). Zoals Lipton (2018, p. 1) namelijk stelt: *"The academic literature has provided diverse and sometimes non-overlapping motivations for interpretability and has offered myriad techniques for rendering interpretable models"*. Er bestaat geen academische eenduidigheid over hoe zo'n uitleg het beste gegeven kan worden (Kim & Routledge, 2018; Kizilcec, 2016; Lepri et al., 2018; Lipton, 2018; Rader, Cotter & Cho, 2018). Toch is het, zoals hiervoor geconcludeerd, van belang dat algoritmische besluiten die een impact hebben op het leven van iemand, uitlegbaar gemaakt worden. Met name omdat die uitleg kan zorgen

voor vertrouwen in de beslissing (Hind, 2019, p. 17). De academische ambiguïteit met betrekking tot de uitlegbaarheid van algoritmische besluiten vormt hiermee de aanleiding van dit onderzoek.

1.3 PROBLEEM-, VRAAG- EN DOELSTELLING

Op dit moment is in de publieke sector weinig bekend over het effect van een uitleg over een algoritmische beslissing op het vertrouwen van burgers in die beslissing. Hoe publieke algoritmische besluiten het beste uitgelegd kunnen worden aan burgers is nog onduidelijk. Dit is problematisch, omdat 1) de overheid verplicht is om haar besluiten te motiveren en 2) besluiten die een impact hebben op iemand een grote behoefte om uitleg teweegbrengen. Dit leidt tot de volgende onderzoeksvraag:

“Wat is het effect van uitlegbaarheid van algoritmische besluiten op het vertrouwen van burgers in deze besluiten?”

Het doel van dit onderzoek is dus om inzicht te verschaffen in wat voor soort uitleg over algoritmische besluiten bijdraagt aan het vergroten van het vertrouwen van burgers in deze besluiten. In dit onderzoek wordt dit effect aan de hand van de casus van de slimme keuzehulp concreet gemaakt en onderzocht.

1.4 DEELVRAGEN

Om de hoofdvraag te kunnen beantwoorden, zijn zeven deelvragen opgesteld. De eerste vier vragen zijn theoretisch en worden in het theoretisch kader behandeld. De laatste drie zijn empirisch en worden beantwoord in het resultatenhoofdstuk op basis van de verzamelde data.

Theoretische deelvragen

1. Wat zijn algoritmische besluiten in de publieke sector?
2. Hoe kunnen algoritmische besluiten uitgelegd worden?
3. Hoe kan ‘vertrouwen’ gedefinieerd worden in de context van algoritmische besluiten?
4. Wat zijn mogelijke relaties tussen uitlegbaarheid en vertrouwen van burgers in algoritmische besluiten?

Bij de beantwoording van de theoretische deelvragen wordt een onderscheid gemaakt in vertrouwen als *houding* en op vertrouwen gebaseerd *gedrag*. Vanuit deze theoretische bevindingen zijn drie empirische deelvragen opgesteld.

Empirische deelvragen

1. Wat is de invloed van uitlegbaarheid op het vertrouwen van gebruikers in het advies van de slimme keuzehulp?
2. Wat is de invloed van uitlegbaarheid op de bereidheid van gebruikers om naar het advies van de slimme keuzehulp te handelen?
3. Wat is de invloed van uitlegbaarheid op het daadwerkelijke gedrag van gebruikers van de slimme keuzehulp?

1.5 MAATSCHAPPELIJKE RELEVANTIE

In het huidige tijdperk, waarin de overheid inzet op 'verantwoorde mensgerichte AI' (Rijksoverheid, 2019, p. 43), is het van belang dat gekeken wordt naar de uitlegbaarheid van AI-toepassingen die een grote impact hebben op mens of maatschappij. AI-toepassingen worden nu in de publieke sector ingezet als instrument om effectiever en efficiënter te kunnen werken, terwijl het van belang is om eerst na te denken over en te zorgen voor transparant gebruik van algoritmen (De Fine Licht & De Fine Licht, 2020; Zarsky, 2016). Transparantie kan namelijk bijdragen aan verantwoord gebruik van AI-toepassingen (Arrieta et al., 2020). Zonder verantwoord gebruik van algoritmen, kunnen de eventuele voordelen van het gebruik van algoritmen (efficiënter en effectiever werken) niet benut worden.

Op dit moment loopt beleid echter achter de op technologie (Lepri et al., 2018, p. 613). Dit houdt in dat er nog weinig concreet wordt nagedacht over verantwoord gebruik van algoritmen binnen de overheid. Hoffman (2019) benadrukt dat er bij overheden een urgente behoefte is aan handvatten over hoe (onbedoelde) negatieve effecten van (het gebruik van) algoritmen voorkomen kunnen worden, om het vertrouwen van burgers te waarborgen. Een onderzoek naar het effect van uitlegbaarheid van algoritmische besluiten op het vertrouwen van burgers in het aangiftesysteem kan een goede stap zijn in het concreet nadenken over verantwoord gebruik van algoritmen.

Daarnaast is nog weinig onderzoek gedaan naar de huidige uitlegbaarheid van beslissingen van de slimme keuzehulp. De focus lag voornamelijk op de accuratie, aanstuurbaarheid en efficiëntie van het kunstmatige aangiftesysteem (Bex, 2019, p. 32; Politie, 2019a). Op dit moment treedt de slimme keuzehulp alleen in werking bij het doen van aangifte naar internetfraude. Indien de werking van dit systeem als succesvol kan worden beschouwd, wordt het concept achter de slimme keuzehulp mogelijk ook voor andere aangifteterreinen ingezet (Bex, 2019). Om het systeem te kunnen kopiëren naar andere terreinen is het van belang dat duidelijk wordt hoe en in welke mate de beslissingen van de slimme keuzehulp uitgelegd moeten worden aan de burger, teneinde het vertrouwen van burgers in deze beslissingen te vergroten. Dit onderzoek zal daar handvatten voor bieden.

1.6 WETENSCHAPPELIJKE RELEVANTIE

Dit onderzoek is niet alleen voor de maatschappij, maar ook voor de wetenschap relevant. Lepri et al. (2018) benadrukken in hun werk de noodzaak van onderzoek naar transparantie over AI-toepassingen, zoals algoritmische besluiten in de publieke sector, zodat vertrouwen in publieke besluiten gewaarborgd kan worden. De Laat (2018, p. 527) stelt zelfs dat “*a call for transparency*” essentieel is om verantwoord door te kunnen gaan met het gebruik van algoritmen. Zoals eerder benoemd, is in de literatuur geen eenduidigheid over hoe uitleg over algoritmische besluiten het beste gegeven kan worden (Kim & Routledge, 2018; Kizilcec, 2016; Lepri et al., 2018; Lipton, 2018; Rader et al., 2018). De onuitlegbaarheid van algoritmen blijft tot op heden een groot probleem, dat zowel theoretische als politieke inspanningen vereist voordat we verder kunnen gaan met het gebruik van algoritmen bij publieke besluitvorming (De Laat, 2018, p. 540). Wetenschappelijk gezien is een onderzoek naar de uitlegbaarheid van algoritmische besluiten door de overheid dus noodzakelijk en relevant tegelijkertijd.

De link die in dit onderzoek gemaakt wordt tussen uitlegbaarheid en vertrouwen in algoritmische beslissingen is nog weinig onderzocht in de publieke sector. Eerder empirisch onderzoek in de computerwetenschappen naar deze relatie spitste zich toe op algoritmen gebruikt door Facebook (Rader et al., 2018) en algoritmen gebruikt bij het becijferen van studenten (Kizilcec, 2016). In beide onderzoeken wordt een link gelegd met *procedurele rechtvaardigheid*, een concept ontwikkeld door Tyler (1990). Volgens Tyler (1990) accepteren mensen sneller een uitkomst als het proces rechtvaardig is geweest. Vertrouwen in een algoritmische beslissing is per definitie hoger als de uitkomst overeenkomt met de verwachting van een individu (Kizilcec, 2016). In de bestaande onderzoeken wordt daarom geconcludeerd dat transparantie over de procedure van een beslissing (*procedurele uitleg*) alleen effectief is bij individuen waarbij hun verwachting niet bevestigd werd. Extra informatie leidt de aandacht af van de procedurele rechtvaardigheid van het besluitproces, waardoor een individu gevoed wordt in zijn gevoel van een onterechte uitkomst (Kizilcec, 2016; Rader et al., 2018). Het is wetenschappelijk gezien relevant om te kijken of de uitkomsten van bovengenoemde onderzoeken uit de computerwetenschappen over de relatie tussen uitlegbaarheid en vertrouwen in algoritmische beslissingen stand houden in de context van de publieke sector. Juist omdat in deze sector besluiten uitgelegd móéten worden vanwege het motiveringsbeginsel en vanwege het feit dat algoritmische overheidsbesluiten een grote impact kunnen hebben op burgers.

1.7 LEESWIJZER

Allereerst volgt een contexthoofdstuk, waarin de werking van de slimme keuzehulp toegelicht wordt. Vervolgens komt het theoretisch kader aan bod. Hierin worden de belangrijkste inzichten uit de literatuur met betrekking tot het onderwerp weergegeven

en worden tevens de eerste vier deelvragen van het onderzoek beantwoord. Daarna wordt in het methodenhoofdstuk besproken welke methoden en technieken in dit onderzoek zijn gehanteerd. Het resultatenhoofdstuk behandelt vervolgens de laatste drie empirische deelvragen. In de conclusie wordt de hoofdvraag beantwoord en in de discussie, tot slot, is aandacht voor academische en maatschappelijke implicaties van dit onderzoek.

2 De slimme keuzehulp

De slimme keuzehulp is een beslissysteem van de politie dat algoritmen gebruikt bij de beoordeling van aangiftes van internetfraude. In dit hoofdstuk worden de verschillende stappen toegelicht waaruit de slimme keuzehulp is opgebouwd. Dit geeft inzicht in de casus die in dit onderzoek gebruikt zal worden.

2.1 TEKSTANALYSE

De eerste stap van de slimme keuzehulp bestaat uit een open tekstveld waarin de burger kan beschrijven wat er gebeurd is. Vervolgens analyseert de slimme keuzehulp de door de burger ingevoerde tekst. Hiervoor wordt gebruik gemaakt van reguliere expressies. Een reguliere expressie is een manier om patronen in de tekst te beschrijven. De slimme keuzehulp zoekt naar (varianten op) veelvoorkomende formuleringen in de tekst, die relevant kunnen zijn voor internetoplichting, zoals *“ik heb niet ontvangen”* of *“ik heb het geld overgemaakt”* (Odekerken, 2020, p. 4).

Figuur 2.1: Tekstanalyse (Politie, 2020)

The screenshot shows a web form with a dark blue header containing the text 'Stap 1: Wat is er gebeurd?' and a small upward-pointing arrow icon. Below the header, the main content area has a light blue background. It starts with the bold heading 'Hier kunt u vertellen wat er is gebeurd.' followed by the instruction 'Het is belangrijk dat u zo volledig mogelijk bent. Denk bij het beschrijven aan:'. A bulleted list follows, containing eight items: 'of u betaald heeft voor wat u wilde kopen', 'of er door de verkoper iets geleverd is', 'of u heeft ontvangen wat u had verwacht', 'hoe lang u gewacht heeft voor u aangifte ging doen', 'wat u zelf al heeft gedaan om de zaak op te lossen', 'hoe de verkoper hierop heeft gereageerd', 'wat u verder nog opviel', and 'Als u bij een webwinkel heeft besteld, is het belangrijk dat u de url opgeeft zoals u deze gezien hebt en niet zoals de webwinkel algemeen bekend is (bijvoorbeeld: als u bij Coolblue iets besteld heeft en de url was coolblue.ru, voer dan coolblue.ru in en niet coolblue.nl)'. Below the list is a large, empty text input field with a thin blue border. In the bottom right corner of this field, the text '0 / 4000' is visible. At the bottom left of the form, there is a dark blue button with the white text 'Volgende'.

2.2 TEKSTCLASSIFICATIE

De tweede stap betreft een tekstclassificatie. Aan de hand van de tekstanalyse stelt de slimme keuzehulp één of meerdere observaties vast. Deze worden aan de burger voorgelegd. Indien de observaties onjuist zijn, dan worden deze observaties niet meegenomen in de verdere analyse.

Figuur 2.2: Tekstclassificatie (Politie, 2020)



Stap 2: Bevindingen van de Keuzehulp

Dit is er gebeurd:
Verwijder de vink bij de onjuiste bevinding(en).

- U heeft een product proberen te kopen. ?
- U heeft een redelijke tijd gewacht voor u aangifte wilde doen. ?
- U verwachtte een pakket. ?

Volgende

2.3 VERVOLGVRAGEN

Met een argumentatiemodule bepaalt de slimme keuzehulp of de gevonden observaties voldoende zijn om te concluderen of er sprake is van fraude. Dat gebeurt op basis van een aantal regels, die gebaseerd zijn op de wet (Artikel 326 van het Wetboek van Strafrecht), en in samenwerking met domeinexperts zo concreet mogelijk zijn gemaakt zodat de argumentatiemodule ermee kan redeneren (Odekerken, 2020).

Indien de slimme keuzehulp nog onvoldoende informatie uit het verhaal van de burger heeft kunnen halen om tot een conclusie te komen, stelt hij vervolgvragen. De slimme keuzehulp gebruikt dan dezelfde, hierboven genoemde regels om te bepalen welke vragen nog relevant zijn om te stellen aan de burger. Deze veelal gesloten vragen worden vervolgens één voor één gesteld. Na de beantwoording van elke vraag controleert slimme keuzehulp of er inmiddels genoeg informatie is om een conclusie te kunnen trekken. Dit voorkomt het stellen van meer vragen dan noodzakelijk (Odekerken, 2020).

Figuur 2.3: Vervolg vragen (Politie, 2020)

Stap 3: Vragen over wat er is gebeurd ^

Naar aanleiding van uw verhaal heeft de Keuzehulp nog enkele vragen.

Heeft u besteld bij een webwinkel? ? Ja Nee

Zo ja, vul dan hier de url in: 20 / 200

Zou u een product per post toegestuurd krijgen? ? Ja Nee

Eventuele extra toelichting: 0 / 200

Heeft u een Track en Trace code van de verkoper ontvangen voordat u betaald heeft? ? Ja Nee

Eventuele extra toelichting: 0 / 200

Gebruikte de verkoper een adres (straatnaam, huisnummer, postcode of woonplaats)? ? Ja Nee

Eventuele extra toelichting: 0 / 200

Volgende

2.4 ADVIES

Zodra de slimme keuzehulp voldoende informatie opgehaald heeft, wordt er een advies getoond aan de burger. Dit advies geeft aan of, en waarom, het zinvol is om al dan niet aangifte te doen. Daarnaast worden ook andere mogelijke vervolgstappen getoond. Om daadwerkelijk aangifte te doen, moet de burger vervolgens inloggen met zijn of haar DigiD. De informatie die de burger heeft opgegeven tijdens het doorlopen van de slimme keuzehulp, wordt automatisch op het aangifteformulier ingevuld (Odekerken, 2020).

Kortom, de slimme keuzehulp combineert verschillende technieken uit de kunstmatige intelligentie. Op basis van reguliere expressies wordt de tekst van de burger geanalyseerd. De argumentatiemodule die vervolgens die observaties classificeert, is een op regels-gebaseerd systeem, waardoor het mogelijk is om uit te leggen waarom de slimme keuzehulp een bepaald advies geeft, dat in overeenstemming is met de wet (Odekerken, 2020).

Figuur 2.4: Advies (Politie, 2020)

Stap 4: Ons advies ^

Ons advies: Aangifte doen ?

Op basis van uw verhaal lijkt de kans dat uw aangifte tot een strafzaak leidt klein. Toch raden wij u aan om door te gaan met uw aangifte.

Klik hieronder op 'Doorgaan met aangifte'. U krijgt dan een formulier te zien waarin u verdere details, zoals adressen en betaalgegevens, kunt invullen.

Let op: uw aangifte is pas afgerond nadat u het aangifteformulier heeft ingevuld en verzonden.


Wat u verder nog kunt doen ?

Naast het doen van aangifte (als dat van toepassing is) kunt u zelf proberen een oplossing te vinden met de verkoper.

- Mogelijk kan het [Juridisch loket](#) of de [Rechtstelefoon](#) u hierbij helpen.
- Op [consuwijzer.nl](#) vindt u praktische informatie over uw rechten.
- Op de [pagina "internetoplichting" van de politie](#) vindt u meer informatie over internetoplichting.
- Op de [pagina "internetoplichting" van de politie](#) vindt u informatie over uw mogelijkheden als u een rechtsbijstandverzekering heeft.

Ons advies volgt uit wat er is gebeurd:

- U heeft een redelijke tijd gewacht voor u aangifte wilde doen. ?
- Er is geen product geleverd. ?
- U heeft een product proberen te kopen. ?
- U heeft betaald voor het product. ?
- U heeft gehandeld via een webwinkel. ?
- U verwachtte een pakket. ?
- De andere partij gebruikte geen adres. ?

 **Doorgaan met aangifte** Stoppen

3 Theoretisch kader

In dit hoofdstuk wordt antwoord gegeven op de vier theoretische deelvragen, te weten:

1. Wat zijn algoritmische besluiten in de publieke sector?
2. Hoe kunnen algoritmische besluiten uitgelegd worden?
3. Hoe kan 'vertrouwen' gedefinieerd worden in de context van algoritmische besluiten?
4. Wat zijn mogelijke relaties tussen uitlegbaarheid en vertrouwen van burgers in algoritmische besluiten?

Allereerst wordt gekeken naar hoe algoritmische besluiten in de publieke sector worden toegepast. Hierna volgt een bespreking van mogelijke manieren om algoritmische besluiten uit te leggen. Dan wordt aandacht besteed aan hoe vertrouwen gekoppeld is aan algoritmische besluiten. Dit wordt gevolgd door een overzicht van mogelijke relaties tussen uitlegbaarheid van algoritmische besluiten en het vertrouwen van burgers in deze besluiten. Hieruit vloeien vier hypothesen voort, die de basis vormen voor het empirische onderzoek.

3.1 ALGORITMISCHE BESLUITEN

Het is van belang om aandacht te besteden aan algoritmische besluiten, om zo een volledig beeld te krijgen van waar dit onderzoek zich op richt. De eerste theoretische deelvraag luidt dan ook: *"Wat zijn algoritmische besluiten in de publieke sector?"*. Deze vraag zal beantwoord worden door allereerst kort uiteen te zetten wat 'algoritmen' zijn en hoe dit begrip past binnen de kunstmatige intelligentie. Vervolgens wordt gekeken naar algoritmische besluiten in de publieke sector. Dit vormt de fundering voor het vervolg van dit theoretisch kader, waarin algoritmische besluiten gelinkt worden aan uitlegbaarheid en vertrouwen.

3.1.1 Algoritmen

Simpel gezegd is een algoritme een mechanisme om een bepaalde vorm van gegevens op te nemen (*input*) en vervolgens een andere vorm van gegevens uit te voeren (*output*). Een algoritme ontvangt een input: een reeks instructies, waarmee een output geproduceerd kan worden (Hogan, 2015, p. 104). Het is dus een op zichzelf staande stapsgewijze reeks die computers (en andere 'slimme' apparaten) uitvoeren om berekeningen, gegevensverwerking en geautomatiseerde redeneertaken uit te werken. Ook worden algoritmen in toenemende mate gebruikt bij besluitvormingsprocessen, waarbij ze betekenisvolle patronen in gegevens ontdekken en analyseren (US Public Policy Council, 2017, p. 1).

In onze huidige economie en maatschappij zijn algoritmen alomtegenwoordig (Rader et al., 2018, p. 1). Ze worden gebruikt bij allerlei beslissingen waar we (dagelijks) mee te maken krijgen, zoals bij het bepalen van de toegang tot krediet (Kool, Timmer & Van Elst, 2015), bij advertenties die men op Facebook te zien krijgt (Rader et al., 2018), bij toepassingen in het onderwijs (Prins, 2016) of bij de besturing van een zelfrijdende auto's (Nagtegaal, 2019). Algoritmen zijn niet meer weg te denken uit de huidige samenleving.

Algoritmen in de context van AI

Algoritmen vallen onder het begrip van kunstmatige intelligentie, omdat algoritmen op basis van gegevens of signalen uit de omgeving in staat zijn om zelfstandig beslissingen te nemen en hiervan te leren. Hiermee wordt dus het denkvermogen van de mens nagebootst. Algoritmen kunnen zelfstandig leren en beslissingen nemen (Dignum & Van den Hoven, 2016; Harbers, 2018, p. 25).

Hierbij kan een onderscheid gemaakt worden tussen sterke en zwakke varianten van kunstmatige intelligentie (Brynjolfsson & McAfee, 2014). *Zwakke kunstmatige intelligentie* gaat over intelligente machines die één specifiek intellectueel probleem goed kunnen uitvoeren. Het klassieke voorbeeld hierbij is de schaakcomputer 'Deep Blue' die won van de beste menselijke schaker Kasparov (Campbell, Hoane & Hsu, 2002). Een meer algemeen voorbeeld is het analyseren van zoekgegevens door middel van zoekalgoritmen (Nagtegaal, 2019, p. 38).

Sterke kunstmatige intelligentie wordt toegepast bij intelligente machines die ons kunnen evenaren of zelfs voorbijstreven op bepaalde taken waarbij intelligentie vereist is. Hierbij kan de machine dus werkelijk problemen oplossen in complexe en onzekere omgevingen (Nagtegaal, 2019, p. 38). Een voorbeeld van sterke kunstmatige intelligentie is de 'supercomputer' Watson, die een in spreektaal gestelde vraag kan interpreteren. Vervolgens geeft de machine binnen enkele seconden het goede antwoord na zelf gezocht te hebben in verschillende boeken, tijdschriften, encyclopedieën en wetenschappelijke artikelen. (Ferruci, 2012; Rachlin, 2012).

Algoritmen in de context van de slimme keuzehulp

Zoals in het vorige hoofdstuk uitgelegd is, maakt de slimme keuzehulp eerst gebruik van reguliere expressies om de door de burger ingevoerde tekst te analyseren en vervolgens van een op regels-gebaseerd systeem om die analyses te classificeren. Hierbij is geen sprake van het evenaren of voorbijstreven van de mens wat betreft intelligentie bij deze taken. De algoritmen van de slimme keuzehulp vallen dus binnen de categorie van zwakke kunstmatige intelligentie. Dit maakt duidelijk wat voor algoritmen gebruikt worden binnen de context van de slimme keuzehulp. In de volgende paragraaf wordt specifiek ingegaan op het gebruik van algoritmische besluiten in de publieke sector.

3.1.2 Algoritmische besluiten in de publieke sector

In dit onderzoek wordt specifiek gekeken naar algoritmische besluitvorming. Dit is volgens Yeung (2018, p. 507): *“the use of algorithmically generated knowledge systems to execute or inform decisions, which can vary widely in simplicity and sophistication”*. Algoritmen worden dus gebruikt als ondersteuning bij het maken van beslissingen of zijn in staat zelf beslissingen te nemen. In de publieke sector zijn algoritmen bij besluitvormingsprocessen niet meer weg te denken. Op allerlei terreinen, zoals zorg, de rechtspraak en het onderwijs, worden algoritmen gebruikt bij de uitvoering van taken of voor (de ondersteuning van) besluitvormingsprocessen (Hutink & Wouda, 2019; Van Schouwenburg & Smits, 2019; Van Wetering, 2020).

Van street-level naar system-level bureaucratie

Binnen de overheid zelf groeit ook het gebruik van algoritmen bij besluitvormingsprocessen. Dit was vroeger niet vanzelfsprekend. Bovens en Zouridis (2002) spreken van een overgang van de *street-level* bureaucratie via een *screen-level* bureaucratie naar een *system-level* bureaucratie. In de *street-level* bureaucratie werden beslissingen, zoals het toekennen van een bouwvergunning of een uitkering, genomen door ambtenaren die in direct contact staan met burgers en discretionaire ruimte hebben in hun werk.

Vervolgens werd in toenemende mate ICT ingezet om processen te automatiseren, waardoor een overgang plaatsvond naar de *screen-level* bureaucratie. Besluitvormingsprocessen werden aan de hand van ICT steeds meer gestandaardiseerd, terug te zien in standaard tekstblokken met motivering. De *street-level* bureaucraat wordt een *screen-level* bureaucraat, omdat het contact met burgers afneemt en hij voornamelijk bezig is met het invoeren van formulieren in een geautomatiseerd systeem (Bovens, Zouridis & Van Eck, 2018, p. 3009). De discretionaire ruimte van de *screen-level* bureaucraat is daarmee aanzienlijk kleiner dan die van de *street-level* bureaucraat.

In het begin van de 21^{ste} eeuw vond een ontwikkeling plaats waarbij ICT-systeemontwikkelaars de plaats innamen van de oorspronkelijke *street-level* bureaucraten. Door ICT is de discretionaire ruimte van professionals die zich met individuele gevallen bezighouden, geheel verdwenen. In de *system-level* bureaucratie ligt de discretionaire ruimte bij de programmeurs die de beslisregels programmeren en de ICT ontwerpen (Bovens, Zouridis & Van Eck, 2018, p. 3010).

Conclusie

Kortom, steeds meer overheidsbeslissingen worden tegenwoordig niet meer gemaakt door menselijke beslissers, maar door geautomatiseerde systemen. Op basis van data en vooraf gedefinieerde algoritmen worden besluiten genomen over individuele klanten en casussen (Busch & Henriksen, 2018). Dit vormt het antwoord op de eerste theoretische deelvraag: *“Wat zijn algoritmische besluiten in de publieke sector?”*. In de volgende paragraaf zal ingegaan worden op hoe zulke besluiten uitgelegd kunnen worden.

3.2 UITLEGBAARHEID

In dit onderdeel zal de tweede deelvraag “*Hoe kunnen algoritmische besluiten uitgelegd worden?*” stapsgewijs beantwoord worden. Allereerst wordt het vraagstuk van uitlegbare algoritmische besluiten wettelijk en maatschappelijk ingekaderd. Vervolgens zal het concept transparantie uitgediept worden, van waaruit het concept uitlegbaarheid voortvloeit. Hierbij worden verschillende soorten uitleg over algoritmische besluiten academisch geduid.

3.2.1 Wettelijke verplichting

Sinds 25 mei 2018 is de GPDR in de Europese Unie van kracht. Dit betreft een uitwerking van verordening 2016/976 van het Europees Parlement en de Raad van Europa betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrije verkeer van die gegevens. De GDPR harmoniseerde de bestaande nationale wetgevingen inzake de bescherming van privacy en kende ook nieuwe rechten toe aan de burgers, zoals bijvoorbeeld het recht op informatie, correctie of verwijdering van foute of strikt private informatie (Berbers et al., 2018, p. 19). Een belangrijk element in de verordening is *uitlegbaarheid*: een individu heeft recht op logische uitleg van iedere algoritmische besluitvorming (Berbers et al., 2018, p. 19). De overheid is dus verplicht om burgers uitleg te verschaffen daar waar een besluit wordt gemaakt met behulp van of door algoritmen. Dit sluit aan bij het eerdergenoemde motiveringsbeginsel, waarbij de overheid verplicht is haar besluiten te motiveren (neergelegd in artikel 3:46 en 3:47 van de Algemene wet bestuursrecht). Ook bij algoritmische besluiten moet de overheid in staat zijn de motivering van het besluit uit te leggen (Poorter & Goossens, 2019).

3.2.2 Principes van goed bestuur

Niet alleen heeft de overheid een wettelijke verplichting, ook maatschappelijk gezien moet zij haar besluiten uitleggen. In de Nederlandse Code Goed Bestuur staat omschreven wat goed openbaar bestuur betekent voor besturen van individuele organisaties in het openbaar bestuur in Nederland, zowel op centraal als decentraal niveau (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2009, p. 7). Meijer, Schäfer & Branderhorst (2019, p. 13) geven in hun onderzoek een nieuwe invulling aan die principes door deze tegen het licht te houden. In de huidige debatten over waardenconflicten met betrekking tot de toepassing van nieuwe technologieën in de publieke sector is namelijk een nieuwe duiding van die principes gewenst (Meijer, Schäfer & Branderhorst, 2019).

Eén van de principes is *procedurele rechtvaardigheid*. Dit gaat over het correct – legaal en legitiem – functioneren van het openbaar bestuur (Meijer et al., 2019, p. 13). Onderdeel hiervan is de begrijpelijkheid van (algoritmische) beslissingen. Wanneer de

overheid een algoritme gebruikt voor een beslissing, dient het algoritme te motiveren wat de grondslag is voor deze beslissing (Meijer et al., 2019, p. 18). Dit betreft dus niet het afleggen van verantwoording over algoritmen die gebruikt worden, maar de beslissing van het algoritme begrijpelijk maken, zodat het principe van procedurele rechtvaardigheid gewaarborgd wordt. Algoritmische besluiten moeten dus uitgelegd kunnen worden. Het uitlegbaar maken van algoritmische besluiten houdt verband met *transparant* zijn over wat je doet (Van Eck, 2018; Meijer et al., 2019). Indien de overheid transparant is over haar besluiten, kan die procedurele rechtvaardigheid dus verzekerd worden. Wat *transparantie* precies is, en hoe dit zich verhoudt tot het concept *uitlegbaarheid* (van overheidsbesluiten) komt in de volgende paragraaf aan bod.

3.2.3 Transparantie

Algemeen bekende metaforen van transparantie zijn van Davis (1998, p. 121) "*lifting the veil of secrecy*" en van Den Boer (1998, p. 105) "*the ability to look clearly through the windows of an institution*". Transparantie gaat dus over het idee dat iets dat 'achter het gordijn gebeurt' openbaar wordt gemaakt en onder de loep genomen kan worden zodra 'de gordijnen naar beneden gehaald worden' (Meijer, 2009, p. 258). Bij overheidstransparantie wordt in de literatuur vaak verwezen naar de mate waarin een overheidsorganisatie relevante informatie onthult over haar functioneren, eigen besluitvormingsprocessen, procedures en prestaties (Grimmelikhuijsen, Porumbescu, Hong & Im, 2013, p. 576). Dit maakt het voor individuen en belangengroepen buiten de organisatie mogelijk om activiteiten en besluiten te volgen die binnen de organisatie genomen worden (Grimmelikhuijsen et al., 2013, p. 576). Om overheidstransparantie te realiseren, moet bepaald worden waarover men transparant is: het *subject* van transparantie. Vervolgens is het van belang om een keuze te maken hoe dat gedaan wordt: het *middel* om transparantie te bewerkstelligen. Hieronder wordt toegelicht welke keuzes voor dit onderzoek gemaakt zijn wat betreft het subject en het middel van transparantie.

Subject

Met betrekking tot het *subject* van transparantie, onderscheiden Grimmelikhuijsen et al. (2013, p. 576) drie onderwerpen van overheidsactiviteiten of -processen waar informatie over gegeven kan worden. Men kan transparant zijn over de 1) de besluitvormingsprocessen, 2) de beleidsinhoud en 3) de beleidsuitkomsten of -effecten. Bij het eerste onderwerp gaat transparantie over de mate waarin openheid is over de stappen die gezet worden bij het maken van een beslissing en de motivering hierachter. Het tweede onderwerp slaat terug op de door de overheid verstrekte informatie over het beleid zelf: wat de genomen maatregelen zijn, hoe deze geïmplementeerd zullen worden en wat implicaties daarvan kunnen zijn voor burgers en andere belanghebbenden. Het derde onderwerp van transparantie verwijst tot slot naar het geven van informatie over beleidsuitkomsten en in welke tijdsframe deze verwacht kunnen worden

(Grimmelikhuijsen et al., 2013, pp. 576-577). In het geval van het algoritme dat de politie gebruikt bij de beoordeling van aangiften van internetoplichting is sprake van een besluitvormingsproces. Het algoritme besluit tot een negatief of positief advies met betrekking tot het doen van aangifte. Als *subject* van transparantie is het in het kader van dit onderzoek dus relevant om focussen op besluitvormingsprocessen.

Middel

Naast de keuze waarover men transparant is, kan ook een onderscheid gemaakt worden op welke manier dat bereikt wordt: het *middel* waarmee transparantie bewerkstelligd wordt. Voortbouwend op de keuze hierboven om te focussen op transparantie over besluitvormingsprocessen, kunnen twee manieren onderscheiden worden om transparant te zijn. Informatie geven over algoritmische besluitvormingsprocessen kan door toegankelijkheid en door uitlegbaarheid (Lepri et al., 2018, p. 621; Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016, p. 6). Het eerste middel om transparantie over een besluitvormingsproces te bewerkstelligen is door het zorgen dat burgers, onafhankelijke experts en belanghebbenden toegang hebben tot het algoritme. Dit houdt in dat zij toegang hebben tot de codes die gebruikt worden en dat het mogelijk is om deze te onderzoeken en te beoordelen (Burrell, 2016; Kroll et al., 2017; Rader et al., 2018). Het tweede middel, uitlegbaarheid, verwijst naar het realiseren van transparantie door het verschaffen van juiste en begrijpelijke informatie over het besluitvormingsproces (Lepri et al., 2018; Mittelstadt et al., 2016; Rader et al., 2018).

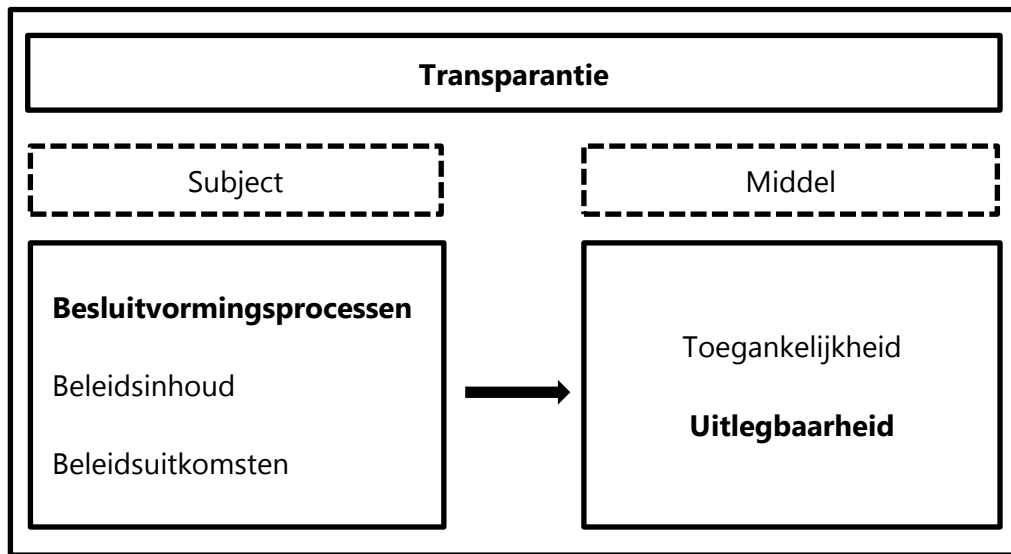
In dit onderzoek wordt gefocust op de tweede vorm van transparantie. De slimme keuzehulp moet volgens Bex (2019, p. 29) namelijk aan vier voorwaarden voldoen:

- *Accuraat*: zo min mogelijk fouten;
- *Transparant*: uitlegbaarheid van belangrijke besluiten;
- *Aanstuurbaar*: kunnen ontdekken waar fouten zitten en blijven verbeteren;
- *Efficiënt*: zo min mogelijk overbodige handelingen.

De tweede voorwaarde van Bex (2019) is dat de slimme keuzehulp transparant is. De manier waarop transparantie gewaarborgd wordt over besluiten van de slimme keuzehulp, is door uitlegbaarheid van belangrijke besluiten en niet door externe actoren toegang te verschaffen tot de codes die gebruikt worden.

In dit onderzoek gaat algoritmische transparantie dus over de beschikbaarheid van juiste en begrijpelijke informatie over beslissingen door geautomatiseerde systemen waardoor externe actoren de stappen kunnen volgen die gezet worden bij het maken van een beslissing en de motivering hierachter begrijpen. Dit legt een koppeling tussen transparantie en de uitlegbaarheid van besluiten door geautomatiseerde systemen, zoals de slimme keuzehulp. Allereerst is in het kader van dit onderzoek voor besluitvormingsprocessen gekozen als *subject* van transparantie. Het *middel* om dit te bereiken is via uitlegbaarheid. In figuur 3.1 is deze relatie schematisch weergegeven.

Figuur 3.1: Dimensies van overheidstransparantie



Nu uitgewerkt is dat uitlegbaarheid als middel genomen wordt om transparantie over besluitvormingsprocessen te bewerkstelligen, zal in het kader van dit onderzoek dieper ingegaan worden op specifiek de uitlegbaarheid van algoritmische besluiten.

3.2.4 Uitlegbare algoritmische besluiten

Definitie

Zoals hierboven beschreven moeten algoritmische besluiten uitlegbaar zijn. Mittelstadt et al. (2016, p. 17) omschrijven dit als volgt: *“the explanation of a decision must be comprehensible not only to data scientists or controllers, but to the lay data subjects (or some proxy) affected by the decision”*. Burgers die aangifte doen, moeten dus in staat zijn om de beslissing van de slimme keuzehulp te begrijpen. Dit sluit aan bij wat Rader, Cotter en Cho (2018, p. 3) zeggen. Zij stellen namelijk dat de uitlegbaarheid van een algoritmisch systeem en haar beslissingen maakt dat gebruikers meer vertrouwen hebben in de uitkomst van een beslissysteem. Wat voor uitleg moet je geven over algoritmische besluiten zodat de gebruiker de uitkomst begrijpt?

Wetenschappelijk debat

In de wetenschap bestaat nog geen eenduidigheid over hoe algoritmische beslissingen het beste uitlegbaar gemaakt kunnen worden (Kim & Routledge, 2018; Kizilcec, 2016; Rader et al., 2018; Wachter, Mittelstadt & Floridi, 2017a). In academische literatuur worden namelijk verschillende soorten uitleg beschreven, die niet altijd overeenkomen (Lipton, 2008). Sommige onderzoekers leggen de nadruk op het belang van de *timing* van de beslissing: het moment waarop de uitleg gegeven wordt (voorafgaand aan of na afloop van het besluit) (Wachter et al, 2017b). Andere auteurs besteden hier minder aandacht aan, maar focussen zich juist op wat er in die uitleg moet komen (Kim &

Routledge, 2018; Kizilcec, 2016; Rader et al., 2018). Ook daarbinnen bestaat geen overeenstemming over wat voor uitleg het beste gegeven kan worden. De soorten uitleg lopen uiteen van het simpel kenbaar maken dat er een algoritme gebruikt wordt bij het beslisproces of het geven van informatie over de redenering achter een specifieke algoritmische beslissing tot uitleg over de achterliggende ideeën en doelen van het beslissingsysteem. Kortom, er zijn verschillende manieren om uitleg te geven over algoritmische besluiten. Deze zullen hieronder worden toegelicht.

Soorten uitleg over algoritmische beslissingen

Bij een algoritmisch besluit moet nagedacht worden over de *timing* waarop de uitleg gegeven wordt en de inhoud van de uitleg (Kim & Routledge, 2018; Kizilcec, 2016; Rader et al., 2018; Wachter, Mittelstadt & Floridi, 2017b).

Timing van de uitleg

Wat betreft de *timing* van een algoritmische beslissing zijn er twee mogelijkheden (Kim & Routledge, 2018; Wachter et al., 2017b). Ten eerste kan uitleg *ex ante* gegeven worden, wat betekent dat de uitleg voorafgaand aan de beslissing gegeven wordt. Deze vorm van uitleg kan alleen gaan over de beslisprocedure. Het is namelijk niet mogelijk om uitleg te geven over de redenering achter een beslissing voordat deze genomen is. Ten tweede kan uitleg *ex post* worden gegeven, wat inhoudt dat de uitleg na afloop van de beslissing wordt gegeven. Deze vorm van uitleg kan zowel gaan over de inhoud van een beslissing als over de beslisprocedure (Kim & Routledge, 2018; Wachter et al., 2017b).

Een *ex ante* uitleg is op zichzelf nooit genoeg, omdat dan geen informatie gegeven kan worden over de specifieke beslissing. Daarnaast heeft een *ex ante* uitleg over de functionaliteiten van een algoritmisch beslissingsysteem als mogelijk nadeel dat "*gaming the system*" optreedt (Kroll et al., 2017, p. 658). Als burgers weten hoe het systeem zijn beslissingen maakt, kan strategisch gedrag optreden bij gebruik van het systeem, zodat burgers de voor hen gewenste uitkomst krijgen. Hieruit kan afgeleid worden dat het in de context van dit onderzoek voor de hand ligt om *ex post* uitleg te verschaffen, zodat burgers bij het invullen van de vragen van de slimme keuzehulp niet strategisch gaan antwoorden.

Inhoud van de uitleg

Een uitleg over een algoritmische beslissing kan zich op verschillende aspecten focussen. Hieronder worden drie onderzoeken besproken waarin gekeken wordt naar verschillende visies op wat er in zo'n uitleg kan staan. Het eerste onderzoek dat bespreekt waar zo'n uitleg zich op kan richten, is dat van Kim en Routledge (2018). Zij onderscheiden twee soorten uitleg. Allereerst een uitleg over de *systeemfunctionaliteit*. Dit betreft uitleg over de logica, betekenis, gevolgen en functionaliteit van een geautomatiseerd besluitvormingssysteem. Dit kan bijvoorbeeld uitleg zijn over de beslisbomen die het systeem gebruikt, of de classificatiestructuren. Daarnaast

onderscheiden ze uitleg over de *specifieke beslissing*, waarin informatie wordt gegeven over de logica, redenering en individuele omstandigheden van een beslissing door een geautomatiseerd besluitvormingssysteem. Dit kan bijvoorbeeld uitleg zijn over de weging van bepaalde eigenschappen of kenmerken. In feite maken de auteurs hier dus een onderscheid tussen een procedurele uitleg (systeemfunctionaliteit) en een inhoudelijke uitleg (specifieke beslissing). De auteurs geven enkel de mogelijkheden weer en maken nog geen normatief onderscheid met betrekking tot welke uitleg (in welke situatie) te prefereren is.

Een tweede onderzoek waarin verschillende soorten uitleg over algoritmische besluiten onderzocht worden is dat van Kizilcec (2016). Ook hier wordt het onderscheid tussen een inhoudelijke en een procedurele uitleg ondersteund. In een onderzoek naar het becijferingsproces van studenten (door een combinatie van medestudenten en een algoritme) maakt de auteur een onderscheid tussen een *procedurele uitleg* (de procedure waarop het cijfer tot stand is gekomen) en een *inhoudelijke uitleg* (de achterliggende argumenten voor het desbetreffende cijfer). Ook koppelt hij het soort uitleg aan de verwachting die studenten hadden over hun cijfer. Enkel wanneer het cijfer niet voldoet aan hun verwachting, lijkt het soort uitleg verschil te maken in het vertrouwen dat studenten in besluitproces hebben (Kizilcec, 2016, p. 2392). De auteur concludeert dat wanneer iemand een uitkomst krijgt die hij niet had verwacht, een procedurele uitleg zorgt voor het meeste vertrouwen in die beslissing (Kizilcec, 2016, p. 2393). Een inhoudelijke uitleg leidt dan af van het gevoel dat studenten hebben over de rechtvaardigheid van het beoordelingsproces (Kizilcec, 2016; Lind & Tyler, 1988).

De bovengenoemde bevindingen raken aan het concept van procedurele rechtvaardigheid. Grondleggers van dit concept, Thibaut, Walker, LaTour en Houlden, vonden al in 1974 dat als mensen het gevoel hebben dat een proces eerlijk is, zij de uitkomst van dat proces makkelijker kunnen accepteren. Ook uit het onderzoek van Kizilcec (2016) blijkt een voorkeur voor een procedurele uitleg bij een beslissing waarbij de verwachting van iemand niet overeenkomt met het daadwerkelijke besluit. In paragraaf 3.4 zal nog dieper ingegaan worden op de relatie tussen procedurele rechtvaardigheid en het vertrouwen dat burgers in een algoritmisch beslissing hebben.

Tot slot staat ook in het derde onderzoek, van Rader et al. (2018), de uitlegbaarheid van algoritmische besluiten centraal. De auteurs kijken naar de effecten van verschillende soorten uitleg over een algoritmisch beslissingsstelsel dat Facebook gebruikt. Ze ontwaren vier soorten uitleg (Rader et al., 2018 p. 2). Ten eerste een *wat-uitleg*, waarin kenbaar wordt gemaakt dat bij de besluitvorming gebruik wordt gemaakt van een algoritme. Ten tweede een *objectieve uitleg*, die gaat over de totstandkoming van het beslissingsalgoritme. Ten derde een *hoe-uitleg*, waarin de stappen worden toegelicht die gezet zijn om tot een bepaalde uitkomst te komen en dus hoe het systeem functioneert.

Ten vierde beschrijven de auteurs een *waarom-uitleg*, waarin de achterliggende ideeën en doelen van het beslissysteem worden uitgelegd. Dit maakt het voor gebruikers mogelijk om na te gaan in hoeverre hun eigen ideeën en doelen matchen met die van het systeem. Deze vorm van uitleg gaat niet in op hoe het systeem werkt.

De auteurs concluderen dat enkel de *hoe-uitleg* invloed heeft op begrip van het algoritme en ervoor zorgt dat gebruikers vertrouwen hebben in de uitkomsten (Rader et al., 2018, pp. 9-10). Dit betreft een procedurele uitleg omdat de procedure waarin de beslissing tot stand komt aan de gebruiker uitgelegd wordt. Van een inhoudelijke uitleg (motivering van de beslissing van het algoritme), zoals in de eerste twee onderzoeken duidelijk naar voren kwam, is hier geen sprake. Dit kan verklaard worden door de aard van het onderzochte beslisalgoritme. In het onderzoek zijn er beslissingen onderzocht die geen specifieke inhoudelijke uitleg vereisen (welke advertenties krijgt iemand te zien op Facebook). De drie andere soorten uitleg die in dit onderzoek genoemd worden (*wat, objectief en waarom*) kunnen wel bijdragen aan bewustwording van het gebruik van een algoritme (Rader et al., 2018, p. 6) of aan het gevoel dat het systeem correct en nauwkeurig is (Rader et al., 2018, pp. 7-8), maar vergroten het begrip van de algoritmische beslissing bij de gebruiker niet. Vanuit dit onderzoek kan dus aangenomen worden dat een procedurele uitleg bijdraagt aan de acceptatie van een besluit.

Overheidsbesluiten

De drie hierboven besproken onderzoeken richten zich op algoritmische besluiten in het algemeen en niet zo zeer op besluiten in de publieke sector. Er zijn nog weinig onderzoeken waarin concrete handvatten geboden worden voor hoe algoritmische besluiten in de publieke sector uitgelegd kunnen worden. Bij 'normale' overheidsbesluiten wordt in verschillende onderzoeken een tweedeling aangehaald met betrekking tot wat er in de uitleg over deze besluiten moet staan (De Fine Licht, Naurin, Esaiasson & Gilljam, 2014; Mansbridge, 2009). In deze onderzoeken komt naar voren dat overheden gebruik kunnen maken van transparantie over het besluitproces (*transparency in process*) en transparantie over de motivering van een besluit (*transparency in rationale*) (De Fine Licht et al., 2014, p. 113; Mansbridge, 2009, p. 386). Dit komt overeen met een procedurele en een inhoudelijke uitleg, die ook uit de eerdere besproken onderzoeken naar voren kwamen. Het literatuuronderzoek van De Fine Licht en De Fine Licht (2020) bouwt voort op deze bevinding, door te beargumenteren dat de tweedeling ook relevant is bij de uitleg van algoritmische overheidsbesluiten. Deze twee soorten uitleg vormen in dit onderzoek dan ook de basis voor uitlegbaarheid van algoritmische besluiten in de publieke sector.

Uitlegbaarheid in dit onderzoek

In dit onderzoek wordt als uitgangspunt genomen dat uitleg over algoritmische beslissingen op twee manieren vormgegeven kan worden. In beiden gevallen betreft het een uitleg na afloop van het nemen van de beslissing:

1. *Procedurele uitleg*: uitleg over de procedure (beslisbomen, classificatiestructuren, etc.) waarin de beslissing tot stand gekomen is.
2. *Inhoudelijke uitleg*: uitleg over de redenering (logica, individuele omstandigheden, etc.) van een beslissing.

Deze twee soorten uitleg worden zowel benoemd in literatuur over algoritmische besluiten in het algemeen als in onderzoek over overheidsbesluiten. Dat maakt dat deze soorten uitleg geschikt zijn om als handvatten te gebruiken bij het onderzoeken van de causale relatie tussen uitlegbaarheid en vertrouwen in algoritmische besluiten in de publieke sector.

Conclusie

Allereerst is duidelijk geworden dat uitlegbaarheid als middel ingezet kan worden om transparantie over algoritmische besluitvormingsprocessen te realiseren. Verder is de tweedeling met betrekking tot wat er in een uitleg kan staan een belangrijke opbrengst van deze paragraaf. Hieruit blijkt dat er twee soorten uitleg gegeven kunnen worden na afloop van het nemen van een algoritmisch besluit in de publieke sector. Ten eerste een *procedurele uitleg*, waarin informatie staat over de procedure waarin de beslissing tot stand is gekomen. Ten tweede een *inhoudelijke uitleg*, waarin informatie wordt gegeven over de motivering van een specifieke beslissing. Hiermee is de tweede deelvraag "*Hoe kunnen algoritmische besluiten uitgelegd worden?*" beantwoord. In paragraaf 3.4 zullen deze twee vormen van uitleg in verband gebracht worden met het concept 'vertrouwen'. Hiervoor is het van belang dat dit concept eerst verder uitgediept wordt. Dit gebeurt in de volgende paragraaf.

3.3 VERTROUWEN

De derde deelvraag “Hoe kan ‘vertrouwen’ gedefinieerd worden in de context van algoritmische besluiten?” zal hieronder beantwoord worden door eerst te kijken naar vertrouwen in het algemeen, gevolgd door een specificatie van vertrouwen van burgers in algoritmische besluiten.

3.3.1 Vertrouwen

In vele disciplines is onderzoek gedaan naar vertrouwen. Onderzoek naar waarom en hoe mensen vertrouwen hebben in iets of iemand is gedaan in de psychologie, sociologie, politicologie, economie en organisatiewetenschappen (Grimmelikhuijsen, 2012a, p. 29). Ook is ‘vertrouwen’ geanalyseerd op verschillende niveaus: 1) vertrouwen tussen mensen, 2) vertrouwen in teams, 3) vertrouwen tussen organisaties, en 4) vertrouwen in systemen in instituties (Grimmelikhuijsen, 2012a, p. 29). Door de veelheid aan wetenschappelijk onderzoek naar vertrouwen en de verscheidenheid aan definities, is sprake van “*a confusing potpourri of definitions*” (Shapiro, 1987, p. 625). Rousseau, Sitkin, Burt en Camerer (1998, p. 395) hebben dit ‘allegaatje’ gevat in een invloedrijke multidisciplinaire definitie: “*a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another*”.

Uit bovenstaande definitie komen vier elementen van vertrouwen naar voren: 1) een psychologische staat, 2) een acceptatie van kwetsbaarheid, 3) positieve verwachtingen en 4) de intentie of het gedrag van een andere partij. Uitgaande van de definitie die Grimmelikhuijsen (2012a, p. 36) in zijn dissertatie gebruikt, verwijst de *psychologische staat* naar een individuele beslissing (beïnvloed door de sociale context) om de overheid te vertrouwen. Het is een mix van cognitieve (rationele) en affectieve (gevoels-) overwegingen en betreft zowel bestaande als nieuwe relaties met de overheid. Het tweede element, de *acceptatie van kwetsbaarheid* refereert naar een actor (partij A) die afhankelijk is van een andere actor (partij B) (Grimmelikhuijsen, 2012a, p. 31). *Positieve verwachtingen*, ten derde, verwijst naar het feit dat er vaak sprake is van onzekerheid bij vertrouwensrelaties. Om deze reden creëren mensen vaak verwachtingen gebaseerd op het beeld dat ze hebben van de ander. Indien ze de intenties en het gedrag van de ander als betrouwbaar ervaren, creëren ze positieve verwachtingen (Grimmelikhuijsen, 2012a, p. 37). Het vierde element, *intentie of gedrag van een andere partij*, gaat tot slot in op ‘de ander’ (Grimmelikhuijsen, 2012a, p. 31). De eigenschappen van de ander (bijvoorbeeld de overheid) zijn essentieel in de definitie van vertrouwen, omdat de overheid andere intenties heeft of gedrag vertoont dan bijvoorbeeld een private onderneming.

De definitie van Rousseau et al. (1998) beschrijft de basiselementen waaruit vertrouwen bestaat, maar geeft nog geen concrete aanknopingspunten hoe vertrouwen gemeten

kan worden in dit onderzoek. Er mist nog een specificering van de verschillende partijen van vertrouwen. Dit onderzoek focust zich op het vertrouwen van burgers (partij A) in algoritmische besluiten (partij B). De volgende paragraaf zal hier verder op ingaan door te definiëren wat vertrouwen in algoritmische besluiten inhoudt.

3.3.2 Vertrouwen in algoritmische besluiten

Definitie

Steeds vaker worden overheidsbesluiten genomen door geautomatiseerde systemen of door intelligente beslissystemen (Bovens & Zouridis, 2002; Madsen & Gregor, 2000). Bij algoritmische besluitsystemen is extra aandacht vereist voor het vertrouwen van burgers in deze beslissingen. De 'algoritmisering' van de overheid roept namelijk nieuwe vragen op hoe vertrouwen gewaarborgd kan worden (Grimmelikhuisen, 2012b; Meijer, 2009). Het vertrouwen dat burgers dan hebben, is niet meer in een menselijke beslisser. Het gaat hier om vertrouwen in (beslissingen van) een intelligente beslissysteem. Verschillende auteurs refereren hiervoor naar *human-computer trust* (HCT) (Madsen & Gregor, 2000; Nothdurft, Lang, Klepsch & Minker, 2013; Sousa, Lamas & Diaz, 2014). HCT gaat over *"the extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent decision aid"* (Madsen & Gregor, 2000, p. 1).

Vertrouwen als houding en op vertrouwen gebaseerd gedrag

In deze definitie komen twee elementen duidelijk naar voren. Ten eerste het *vertrouwen (confidence)* dat de gebruiker heeft in de aanbevelingen, acties en beslissingen van het beslissysteem. Ten tweede de *bereidheid (willingness)* van de gebruiker om te handelen naar die adviezen, acties en besluiten van het beslissysteem.

In 3.3.1 kwam al naar voren dat 'vertrouwen' een psychische staat betreft. Vertrouwen is hier als het ware een overtuiging die iemand heeft. Uit de definitie van vertrouwen in algoritmische besluiten blijkt echter dat vertrouwen meer is dan alleen een overtuiging. Het gaat zowel om de overtuiging om een beslissing te vertrouwen, als de bereidheid om te handelen naar die beslissing. Uit onderzoek van McKnight, Choudhury & Kacmar (2002) komt naar voren dat dit twee verschillende theoretische concepten zijn. Ze onderscheiden een *vertrouwende houding* en *op vertrouwen gebaseerd gedrag*. Indien er sprake is van een vertrouwende houding, dan zal (de intentie tot) gedrag hierop volgen (Matook, Brown & Rolf, 2015; McKnight, Choudhury & Kacmar, 2002; Moody, Galletta & Lowry, 2014; Schlosser, White & Lloyd, 2006).

Een *vertrouwende houding* gaat in dit onderzoek dus over het hebben van vertrouwen in de besluiten van een intelligent beslissysteem. *Op vertrouwen gebaseerd gedrag* gaat over de bereidheid om hiernaar te handelen. Het is dus een intentie om te handelen gebaseerd op het vertrouwen dat een burger heeft in het beslissysteem. Dit komt

overeen met hoe Madsen en Gregor (2000) het vertrouwen in besluiten van een intelligent beslissysteem definiëren.

Cognitieve en affectieve overwegingen

Madsen en Gregor (2000) stellen, net zoals in de algemene definitie van vertrouwen naar voren komt, dat HCT voorkomt uit cognitieve en affectieve overwegingen van de gebruiker. Op cognitieve gebaseerde overwegingen van HCT gaan over de intellectuele perceptie van de gebruiker over de eigenschappen van het beslissysteem, zoals de begrijpelijkheid en de betrouwbaarheid van de algoritmische beslissingen. De op affectieve gebaseerde overwegingen van HCT gaan over de emotionele reactie van de gebruiker op het beslissysteem, zoals de mate waarin een gebruiker het fijn vindt om het systeem te gebruiken (Madsen & Gregor, 2000).

Nothdurft, Richter en Minker (2014, p. 52) beschrijven dat bij eenmalige interacties tussen mens en computer of bij interacties die weinig plaatsvinden, het op cognitieve gebaseerd vertrouwen het meest relevant is. De auteurs stellen dat het op affectieve gebaseerd vertrouwen niet beïnvloedbaar is in een korte tijdspanne. In dit onderzoek wordt gekeken naar het vertrouwen van burgers in besluiten van de slimme keuzehulp. Dit betreft een intelligent beslissysteem waar burgers nog geen of weinig ervaring mee hebben. Om deze reden wordt bij vertrouwen in dit onderzoek gekeken naar soorten uitleg die bijdragen aan het op cognitieve gebaseerd vertrouwen.

Conclusie

Waar de algemene definitie van vertrouwen nog met name ging over een psychologische staat, gaat de specifieke definitie van vertrouwen in besluiten van een intelligent beslissysteem (HCT) een stap verder. Niet alleen heeft een burger vertrouwen in de aanbevelingen, acties en beslissingen van zo'n algoritmisch beslissysteem. Ook heeft de burger de bereidheid om te handelen naar die aanbevelingen, acties en beslissingen. Het vertrouwen in algoritmische besluiten bestaat dus uit een *vertrouwende houding en op vertrouwen gebaseerd gedrag* van de burger. Dit beantwoordt de derde theoretische deelvraag "*Hoe kan 'vertrouwen' gedefinieerd worden in de context van algoritmische besluiten?*".

3.4 RELATIE UITLEGBAARHEID EN VERTROUWEN

De laatste theoretische deelvraag “*Wat zijn mogelijke relaties tussen uitlegbaarheid en vertrouwen van burgers in algoritmische besluiten?*” zal hieronder uitgewerkt worden. Eerst worden mogelijke relaties tussen uitlegbaarheid en vertrouwen in algoritmische besluiten verkend. Dit gebeurt door kort in te gaan op wat voor effect uitleg op overheidsbesluiten in het algemeen heeft, gevolgd door een bespreking van de huidige stand van zaken in de wetenschap met betrekking tot de relatie van uitlegbaarheid en vertrouwen in algoritmische besluiten in de publieke sector. Daarna volgt een theoretisch model, waarin de voor dit onderzoek verwachte causale relaties worden weergegeven. Dit zijn vier hypothesen, die tevens worden toegelicht, waarmee een antwoord wordt gevormd op de laatste theoretische deelvraag.

3.4.1 Relatie uitlegbaarheid en vertrouwen

In dit onderzoek staat transparantie over besluitvormingsprocessen centraal. Grimmeliikhuijsen (2012a, p. 232) constateert in zijn dissertatie dat transparantie over besluitvormingsprocessen een positief effect heeft op het vertrouwen in de overheid. Zodra burgers weten hoe een beslissing tot stand is gekomen, neemt het vertrouwen in de overheid toe. Hieruit kan worden afgeleid dat er een positieve relatie bestaat tussen uitlegbaarheid en vertrouwen bij overheidsbesluiten in het algemeen. Of deze relatie ook bestaat bij algoritmische overheidsbesluiten wordt in de wetenschap bediscussieerd.

Wetenschappelijk debat

Waar in de wetenschap consensus over is, is dat uitlegbare AI belangrijk is. In een (normatief) academisch debat wordt geschreven over de grote belofte van het effect van transparantie op het vertrouwen in algoritmische beslissingen (Arrieta et al., 2020; Gunning, 2017; Larsson & Heintz, 2020; Ryan & Stahl, 2020). Dit thema staat zelfs centraal in de ethische richtsnoeren die opgesteld zijn door de Europese Unie om betrouwbare AI te realiseren (AI HLEG, 2019). Er moet transparantie zijn, maar ook weer niet te veel, want anders gaan mensen het systeem ‘bespelen’ (*gaming the system*), waardoor onwenselijk strategisch gedrag, zoals belastingontduiking, mogelijk is (De Fine Licht & De Fine Licht, 2020; Kroll et al., 2017). Er bestaat dus een dunne scheidslijn waarop beleidsmakers zich begeven met betrekking tot het verschaffen van voldoende juiste en begrijpelijke uitleg.

Het is moeilijk om te bepalen wat er nou in zo’n uitleg moet staan om het vertrouwen in een algoritmisch besluit te waarborgen. Het is namelijk nog vrij onduidelijk wat voor effect transparantie precies heeft op het vertrouwen van gebruikers in een algoritmisch beslissysteem. In de wetenschap zijn hierover verschillende geluiden te horen. Enkele onderzoeken tonen een positief effect (Kizilcec, 2016; Nothdurft et al., 2014; Springer & Whittaker, 2019; Wang & Benbasat, 2007), terwijl andere onderzoeken geen of een

negatief resultaat vonden (Cramer et al., 2008; Tintarev & Masthoff, 2012). De onderzoeken waaruit een positief effect naar voren komt, laten zien dat een vorm van uitleg over het algoritmisch beslissysteem het vertrouwen van de gebruiker vergroot in die beslissing. Wat voor uitleg dat dan precies zou moeten zijn, verschilt per onderzoek. Sommige onderzoeken pleiten voor een combinatie van verschillende soorten uitleg om het vertrouwen in de algoritmische beslissing te vergroten (Nothdurft et al., 2014; Wang & Benbasat, 2007), terwijl andere onderzoeken juist vinden dat één bepaalde vorm van uitleg het vertrouwen beïnvloedt (Kizilcec, 2016; Springer & Whittaker, 2019). De onderzoeken met een negatief of geen resultaat tonen daarentegen aan dat uitleg over hoe een bepaalde beslissing of advies tot stand komt, geen toename in vertrouwen realiseert. Een uitleg zou hoogstens kunnen leiden tot een hogere mate van acceptatie van het besluit, maar het vertrouwen in het besluit neemt hiermee niet toe (Cramer et al., 2008).

Bovengenoemde onderzoeken richtten zich echter grotendeels op zogenoemde *recommender systems* (aanbevelingssystemen die de voorkeur van iemand voorspellen) en vonden veelal plaats in gecontroleerde (laboratorium)omgevingen. Dit maakt het lastig deze resultaten te vertalen naar de publieke context. Er is nog geen empirisch experimenteel onderzoek gedaan in de publieke sector op dit specifieke thema waar op voortgebouwd kan worden. Op dit moment is daarom sprake van wetenschappelijke onduidelijkheid over de causale relatie van uitlegbaarheid over algoritmische besluiten in de publieke sector. Dat maakt de wetenschappelijke relevantie van dit onderzoek zo groot. In de volgende paragraaf zal een theoretisch model ontwikkeld worden dat voortbouwt op bestaande kennis over uitlegbaarheid en vertrouwen in algoritmische besluiten. Zo kan later getoetst worden of deze relatie ook aanwezig is in de context van publieke algoritmische besluitvormingsprocessen.

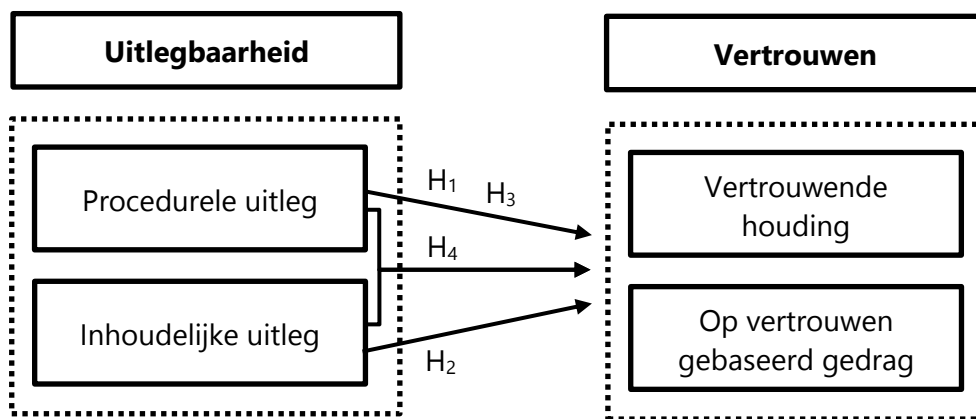
3.4.2 Theoretisch model

In dit onderzoek is gekozen om voor het vertrouwen in algoritmische besluiten gebruik te maken van het concept van HCT (Madsen en Gregor; 2000). Om deze reden is het interessant om te kijken naar onderzoek dat gedaan is naar het effect van verschillende soorten uitleg op HCT. Het onderzoek van Nothdurft et al. (2014) is hiervoor geschikt. In hun onderzoek wordt gekeken naar het effect van twee soorten uitleg op het vertrouwen in algoritmische beslissingen. De auteurs kijken enerzijds naar *transparency explanations*, die gaan over het vergroten van het begrip van hoe het beslissysteem werkt. Dit komt overeen met de procedurele uitleg, die in paragraaf 3.2.4 uitgewerkt is. Anderzijds kijken ze naar *justifications*, die gaan over de motivering van een specifiek besluit. Deze vorm van uitleg komt overeen met de inhoudelijke uitleg (zie paragraaf 3.2.4). Nothdurft et al. (2014) hebben dus het effect onderzocht van een procedurele en een inhoudelijke uitleg op het vertrouwen in de besluiten van een intelligent beslissysteem. De auteurs

concluderen dat deze twee vormen van uitleg een positief effect hebben op het op cognitie gebaseerd vertrouwen van de gebruiker.

Volgens Nothdurft et al. (2014) hebben dus zowel een procedurele uitleg als een inhoudelijke uitleg een positief effect op de *vertrouwende houding* en op het *op vertrouwen gebaseerd gedrag* van de gebruiker. Voor dit onderzoek betekent dit dat uitgegaan kan worden dat er een positieve relatie bestaat tussen uitlegbaarheid en vertrouwen bij algoritmische besluiten. De relatie tussen uitlegbaarheid en vertrouwen in algoritmische besluiten is weergegeven in figuur 3.2. Ook zijn in de figuur de verwachte causale relaties in de vorm van hypothesen schematisch weergegeven. Deze worden onder de figuur toegelicht.

Figuur 3.2: Relatie uitlegbaarheid en vertrouwen bij algoritmische besluiten



Zowel een procedurele uitleg als een inhoudelijke uitleg hebben een positief effect op het vertrouwen in een algoritmisch besluit. Deze positieve relaties zijn niet alleen door Nothdurft et al. (2014) gevonden, maar worden ook in andere onderzoeken erkend. Zo beargumenteren Hind (2019) en Kim en Routledge (2017) dat een inhoudelijke uitleg inderdaad zorgt voor een toename in vertrouwen in algoritmische besluiten. De onderzoeken van Kizilcec (2016) en Rader et al. (2018) ondersteunen daarnaast de positieve relatie tussen een procedurele uitleg en het vertrouwen in een algoritmisch besluit.

Dit geeft aanleiding tot het opstellen van hypothesen, die later in dit onderzoek empirisch getoetst zullen worden. Ten eerste kan een positieve relatie verwacht worden tussen een procedurele uitleg en het vertrouwen van burgers in een algoritmische beslissing. Het vertrouwen zal hoger zijn bij een procedurele uitleg dan wanneer er geen uitleg gegeven wordt over de beslissing. Een tweede positieve relatie wordt verwacht tussen een inhoudelijke uitleg en het vertrouwen van burgers in een algoritmische beslissing. Ook hier zal de mate van vertrouwen hoger zijn met een inhoudelijke uitleg dan wanneer er geen uitleg gegeven wordt over de beslissing.

Hypothese 1: Een *procedurele uitleg* heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.

Hypothese 2: Een *inhoudelijke uitleg* heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.

Verschillende auteurs leggen bij het effect van uitlegbaarheid op vertrouwen in algoritmische besluiten een koppeling met procedurele rechtvaardigheid (Kizilcec, 2016; Lee, Jain, Cha, Ojha & Kusbit, 2019; Meijer et al., 2019). Het concept van procedurele rechtvaardigheid is door Thibaut et al. (1974) ontwikkeld en in 1990 door Tyler uitgewerkt. Dit concept houdt in dat als een proces eerlijk en rechtvaardig is, mensen sneller de uitkomst accepteren.

In het kader van dit onderzoek wordt rekening gehouden met deze bevinding in zoverre dat als een individu verwacht een positief advies te krijgen van de slimme keuzehulp, maar deze verwachting geschonden wordt (er komt een afwijzend advies), het vertrouwen in de algoritmische beslissing lager is. Kizilcec (2016, pp. 2392-2393) vond in zijn onderzoek dat transparantie bij zo'n geschonden verwachting het vertrouwen kan vergroten. Als de verwachting van een individu niet geschonden wordt, dan maakt de mate van transparantie niet uit. Transparantie is dus alleen effectief bij individuen waarbij de verwachting niet bevestigd wordt (Kizilcec, 2016, pp. 2392-2393; Rader et al., 2018, p. 10). Kizilcec (2016) concludeert hierbij dat een *procedurele uitleg* een sterker positief effect heeft op het vertrouwen dan een *inhoudelijke uitleg*. Dit komt door een toename van het gevoel van procedurele rechtvaardigheid (en dus vertrouwen in de beslissing) door een uitleg over het proces. Dit houdt dus in dat wanneer mensen het gevoel hebben dat het besluitvormingsproces eerlijk is, zij zich eerder zullen conformeren naar het besluit omdat zij hier vertrouwen in hebben (Kim & Mauborgne, 1998; Lee et al., 2019). Het concept van procedurele rechtvaardigheid is dus inherent onderdeel van de kracht van een procedurele uitleg.

Dit leidt tot het formuleren van de derde hypothese, die gaat over het verschil in grootte van het effect tussen een procedurele en een inhoudelijke uitleg. Verwacht wordt dat beiden een positief effect hebben op het vertrouwen van burgers in algoritmische beslissingen, maar dat een procedurele uitleg -vanwege een toename in het gevoel van procedurele rechtvaardigheid- een sterker positief effect heeft op het vertrouwen dan een inhoudelijke uitleg. Procedurele rechtvaardigheid versterkt dus het gevoel van vertrouwen bij een afwijzend advies van de slimme keuzehulp ('geen aangifte doen'), waarbij de verwachting van de burger wordt geschonden.

Hypothese 3: Een *procedurele uitleg* heeft een sterker positief effect op vertrouwen van burgers in algoritmische beslissingen dan een *inhoudelijke uitleg*.

Wanneer meerdere soorten uitleg gecombineerd worden, bestaat het risico dat men zoveel informatie krijgt, dat de boodschap niet meer goed gelezen wordt en men ontevreden is met de beslissing. Teveel transparantie doet dan afbreuk aan het vertrouwen. Dit wordt ook wel een *information overload* genoemd (Birkinshaw & Cable, 2017; Hosseini, Shariri, Phalp & Ali, 2015). Kizilcec (2016) laat in zijn onderzoek zien dat meer informatie dan de *procedurele uitleg* de aandacht voor de procedurele rechtvaardigheid van het beslisproces wegneemt, omdat er sprake is van een *information overload*. Hierdoor wordt de gebruiker gevoed in zijn gevoel van een ontorechte uitkomst ten opzichte van zijn initiële verwachting (Kizilcec, 2016, pp. 2392).

Om het gevoel van procedurele rechtvaardigheid bij een beslissing te waarborgen is het dus van belang dat er niet te veel uitleg gegeven wordt. In het onderzoek van Kizilcec (2016) wordt vastgesteld dat een procedurele uitleg gecombineerd met een inhoudelijke uitleg voor zo'n *information overload* zorgt. Het is aannemelijk dat het toevoegen van andere vormen van uitleg aan een procedurele uitleg ook voor zo'n *overload* zorgen, maar dit is niet zo nauwkeurig onderzocht als de combinatie die Kizilcec hanteert.

In dit onderzoek wordt dan ook aangenomen dat een combinatie van een *inhoudelijke* en een *procedurele* uitleg op basis van deze bevinding zorgt voor een negatief effect op de mate van vertrouwen in een algoritmische beslissing. Dit leidt tot het opstellen van de vierde en laatste hypothese, waarin een negatieve relatie wordt verwacht tussen een combinatie van een procedurele en een inhoudelijke uitleg op het vertrouwen dat burgers hebben in een algoritmische beslissing. Dit komt doordat een *information overload* het gevoel van procedurele rechtvaardigheid doet afnemen.

Hypothese 4: Een *procedurele uitleg gecombineerd met een inhoudelijke uitleg* heeft een negatief effect ten opzichte van enkel een procedurele of een inhoudelijke uitleg op vertrouwen van burgers in algoritmische beslissingen.

Conclusie

De vier hypothesen schetsen het antwoord op de laatste deelvraag *“Wat zijn mogelijke relaties tussen uitlegbaarheid en vertrouwen van burgers in algoritmische besluiten?”*. Zowel een procedurele als een inhoudelijke uitleg hebben een positief effect op het vertrouwen van burgers in algoritmische beslissingen ten opzichte van geen uitleg. Hierbij wordt verwacht dat het effect van een procedurele uitleg sterker is dan dat van een inhoudelijke uitleg, vanwege het gevoel van procedurele rechtvaardigheid dat met de eerste uitleg gewaarborgd wordt. Tot slot lijkt een combinatie-uitleg (procedureel en inhoudelijk) minder vertrouwen te genereren dan enkel een procedurele of een inhoudelijke uitleg, omdat een *information overload* afbreuk doet aan het gevoel van procedurele rechtvaardigheid. In het volgende hoofdstuk worden de methoden besproken die gebruikt zijn bij het onderzoeken van de vier hypothesen en bij de beantwoording van de empirische deelvragen.

4 Methoden

Voortbouwend op de theoretische bevindingen uit het vorige hoofdstuk over uitlegbaarheid van en vertrouwen in algoritmische beslissingen (en de relatie daartussen), wordt in dit hoofdstuk uitgelicht welke methoden gehanteerd worden bij de beantwoording van de empirische deelvragen. Allereerst wordt aandacht besteed aan de wetenschapsfilosofische benadering. Vervolgens komen de onderzoeksmethode en operationalisatie aan bod. Daarna wordt ingegaan op de selectie van respondenten en hoe de verkregen data zijn geanalyseerd. Hierbij is het van belang dat de betrouwbaarheid en validiteit van het onderzoek besproken worden. Het hoofdstuk sluit af met een ethische verantwoording.

4.1 WETENSCHAPSFILOSOFISCHE BENADERING

Bryman (2016, pp. 16-37) maakt een onderscheid in de wetenschapsfilosofie naar:

- Ontologische kenmerken: hoe/wat *is* het?
- Epistemologische kenmerken: hoe is het te *kennen*?
- Methodologische kenmerken: hoe is het te *onderzoeken*?

In dit onderzoek wordt wat betreft *ontologische* afwegingen gekozen voor een objectivistische benadering. Dit houdt in dat er vanuit wordt gegaan dat een sociaal fenomeen en haar betekenis onafhankelijk van sociale actoren bestaan (Bryman, 2016, p. 29). Op *epistemologisch* vlak neemt dit onderzoek een positivistische benadering in. Dit houdt in dat kennis alleen verworven kan worden door empirisch onderzoek (Bryman, 2016, p. 24). In dit onderzoek zijn vanuit de theorie vier hypothesen opgesteld die getest zijn op basis van de analyse van empirische data. De onderzoeksconstructen en relaties zijn op basis van de theorie geoperationaliseerd. Dit onderzoek is daarmee deductief van aard (Bryman, 2016, p. 21). De bevindingen uit dit onderzoek kunnen aan de hand van de theorie beschreven en verklaard worden. Op het gebied van *methodologie*, tot slot, wordt in dit onderzoek dan ook gekozen voor een kwantitatieve onderzoeksbenadering (Bryman, 2016, pp. 31-32). Middels een kwantitatief surveyexperiment zijn de hypothesen getoetst. De volgende paragraaf zal uitgebreid ingaan op deze onderzoeksmethode.

4.2 ONDERZOEKSMETHODEN

In deze paragraaf wordt allereerst het onderzoeksdesign toegelicht. Vervolgens wordt aandacht besteed aan de casus die in dit onderzoek centraal staat. Tot slot wordt uitgewerkt hoe de experimentele vignetten zijn vormgegeven en hoe de begrijpelijkheid en realiteit van de vignetten zijn gewaarborgd.

4.2.1 Surveyexperiment als onderzoeksdesign

In de hoofdvraag van dit onderzoek zit een effect van een onafhankelijke variabele op een afhankelijke variabele verwerkt:

“Wat is het effect van uitlegbaarheid van algoritmische besluiten op het vertrouwen van burgers in deze besluiten?”

Hier is de uitlegbaarheid van besluiten van de slimme keuzehulp de onafhankelijke variabele en het vertrouwen van burgers in deze besluiten de afhankelijke. In dit onderzoek wordt dus gezocht naar een causaal verband: de invloed van variabele X op Y. Een experiment is een geschikte onderzoeksmethode om zo'n causale relatie vast te kunnen stellen (Bryman, 2016, p. 44). Bij een experiment kan de onderzoeker namelijk variabelen manipuleren, zodat verschillende scenario's met elkaar vergeleken kunnen worden. Het doel van een experiment is dan ook om veranderingen in een groep toe te kennen aan een bepaalde behandeling.

In dit onderzoek wordt een *2x2 factorieel experimenteel design* gehanteerd, waarbij twee onafhankelijke variabelen worden getoetst. Dit is dus een *two-way design*, omdat er twee onafhankelijke variabelen onderzocht worden. Alle mogelijke combinaties (4) worden hierbij getoetst: wel/geen procedurele uitleg en wel/geen inhoudelijke uitleg. Dit leidt ertoe dat er drie treatmentgroepen vergeleken kunnen worden met een controlegroep (Bouwman & Grimmelikhuisen, 2016, p. 113). Hierbij worden de resultaten achteraf geobserveerd. Verder wordt hier gebruik gemaakt van een *between-subjects* experimenteel design. Dit houdt in dat de participanten willekeurig verdeeld worden over de controle en de treatmentgroepen (Morton & Williams, 2010, p. 86). In tabel 4.1 is het experiment schematisch weergegeven.

Tabel 4.1: Experimenteel design

		Inhoudelijke uitleg	
		Niet	Wel
Procedurele uitleg	Niet	<i>Controlegroep</i> Geen uitleg	<i>Treatment 2</i> Inhoudelijke uitleg
	Wel	<i>Treatment 1</i> Procedurele uitleg	<i>Treatment 3</i> Procedurele en inhoudelijke uitleg

In dit onderzoek is specifiek gekozen voor een surveyexperiment. Hierbij wordt het experiment uitgevoerd middels een vragenlijst en verschillende vignetten. In een vignet wordt een situatie kort beschreven (Bryman, 2016, p. 259). In een surveyexperiment vindt

een experimentele behandeling plaats op basis van een (hypothetische) casus. Hierna volgt een vragenlijst. De respondenten weten niet in welke groep zij zitten en zullen willekeurig één van de vier typen uitleg voorgelegd krijgen.

Een voordeel van surveyexperimenten is dat een grote groep respondenten getest kan worden, waardoor de resultaten van het experiment representatiever zijn voor de populatie (Mutz, 2011, p. 15). Een ander voordeel is dat surveyexperimenten goedkoop zijn (in tegenstelling tot bijvoorbeeld sommige labexperimenten) en makkelijk te implementeren zijn (Blom-Hansen, Morton & Serritzlew, 2015, p. 161).

Aan een surveyexperiment kleeft ook een nadeel. Een mogelijke beperking van een surveyexperiment is dat de treatment een lage intensiteit heeft (Blom-Hansen et al., 2015, pp. 161-162). Respondenten zitten niet in een laboratoriumomgeving van waaruit ze gestimuleerd worden om zich te concentreren op het experiment. Bij een surveyexperiment kunnen deelnemers tijdens het experiment de vragenlijst open laten staan en iets anders gaan doen. In dit onderzoek is geprobeerd om dit zo veel mogelijk te ondervangen.

Door een specifieke vraag op te nemen om te checken of respondenten hun aandacht bij het experiment hebben, kan dit nadeel ondervangen worden. Een dergelijke vraag wordt ook wel een *attention check* genoemd (Abbey & Meloy, 2017). In het surveyexperiment is de stelling *“Let op: vul bij deze vraag geen antwoord in.”* (schaal: 1-7) opgenomen. 10 van de 536 respondenten hebben bij de attentiecheck wél een antwoord ingevuld. Deze respondenten zijn meegenomen in de analyse, maar in bijlage 8.5 staat dezelfde analyse beschreven, maar dan uitgevoerd met enkel de respondenten die de geslaagd waren voor de attentiecheck ($N = 526$). Deze analyse verandert inhoudelijk niks aan de resultaten.

Een andere manier om te controleren of respondenten hun aandacht bij het experiment hadden en niet tussendoor iets anders zijn gaan doen, is door een filter toe te passen op hoe lang (in seconden) respondenten over het experiment deden ($M = 1148,12$; $SD = 3939,79$). In de analyse in het resultatenhoofdstuk zijn alle respondenten meegenomen. In bijlage 8.5 zijn dezelfde analyses uitgevoerd waarin de respondenten die langer dan twee standaarddeviaties afweken van het gemiddelde, eruit gefilterd zijn. Ook dit heeft geen inhoudelijke impact op de resultaten van het onderzoek. De volgende paragraaf zal ingaan op de casus die in dit onderzoek gebruikt is om de causale relatie te kunnen onderzoeken.

4.2.2 Casus: de slimme keuzehulp

De casus die in dit onderzoek onderzocht wordt met betrekking tot de relatie tussen uitlegbaarheid en vertrouwen in algoritmische besluiten is de slimme keuzehulp van de

politie. Bryman (2016, pp.62-63) onderscheidt verschillende soorten casussen die een onderzoeker kan gebruiken. In dit onderzoek is gekozen voor een zogenaamde *"representative or typical case"*, ook wel een voorbeeldgeval genoemd (Bryman, 2016, pp. 62-63). Dit houdt in dat de casus een voorbeeld is van een situatie die veel voorkomt. De casus wordt dan niet gekozen omdat deze extreem of ongebruikelijk is, maar omdat ze een brede categorie van gevallen (situaties) representeren. Zo'n casus biedt een geschikte context om de brede onderzoeksvraag te beantwoorden (Bryman, 2016, p. 62).

In dit geval is de casus een situatie waarin het voortkomt dat uitleg gegeven moet worden over een algoritmisch besluit of advies. De casus van de slimme keuzehulp voorziet dit onderzoek van een context waarbinnen de uitlegbaarheid van algoritmische besluiten en het vertrouwen van burgers daarin getest kan worden. Het gekozen voorbeeld sluit aan op het eerder besproken debat rondom 'uitlegbare AI' en is dus geschikt om de hoofdvraag van dit onderzoek te beantwoorden.

Wat houdt de casus precies in? Bij het doen van aangifte tegen internetoplichting kijkt de slimme keuzehulp mee of alle benodigde informatie is ingevuld en stelt zo nodig extra vragen om tot een advies te kunnen komen over of de zaak al dan niet over een strafbaar feit gaat. De politie gebruikt dus een intelligent beslissysteem bij de beoordeling van aangiftes tegen internetoplichting. Oplichting- en fraudezaken kunnen een grote impact hebben op slachtoffers. Door het verlies van (veel) geld kunnen ze bijvoorbeeld een negatief zelfbeeld ontwikkelen: ze vinden zichzelf "dom" (Kunst & Van Dijk, 2009, p. 51). Daarnaast beoordeelt 90% van de mensen die via de slimme keuzehulp aangifte willen doen hun zaak van internetoplichting als 'ernstig' of 'zeer ernstig' (Jorritsma, Cremers & De Jong, 2020). Dit onderstreept de impact die het besluit van de slimme keuzehulp kan hebben op het leven van het slachtoffer. Hierdoor is aannemelijk dat het advies van de slimme keuzehulp over de fraudezaak impact heeft op het slachtoffer en vergroot de noodzaak van uitlegbare beslissingen.

Op dit moment kunnen slachtoffers ondanks een negatief advies toch nog aangifte doen (Politie, 2019a). Deze mogelijkheid zal op termijn verdwijnen, zodra de slimme keuzehulp geoptimaliseerd is. Hiervoor is, zoals al eerder vastgesteld, van belang dat burgers het advies van de slimme keuzehulp vertrouwen en hier ook naar handelen. Volgens Soares (2020) kiest 72% van de mensen dat een negatief advies krijgt er toch voor om door te klikken en alsnog aangifte te doen. Als voorbeeld: in de periode van 15 mei tot en met 15 juni 2020 besloten 830 van de 1150 mensen die een afwijzing kregen van de slimme keuzehulp ('geen aangifte doen') ervoor om toch aangifte te doen. Dit houdt in dat bijna driekwart van de mensen dat een negatief advies van de slimme keuzehulp krijgt, dit advies niet vertrouwt en hier niet naar handelt. Dat maakt deze casus erg geschikt om het effect van uitlegbaarheid op algoritmische beslissingen te onderzoeken.

4.2.3 Experimentele vignetten

Om het surveyexperiment uit te voeren voor de casus van de slimme keuzehulp zijn verschillende experimentele vignetten opgesteld. Eerst kregen alle respondenten dezelfde situatie te lezen (zie bijlage 8.4). Hier zat geen verschil in voor de controlegroep en de treatmentgroepen. Op basis van deze casus werden de respondenten gevraagd om een aantal vragen in te vullen (zie bijlage 8.4) die vergelijkbaar zijn aan de vragen die slimme keuzehulp stelt bij het doen van aangifte. Vervolgens kregen ze, op basis van hun groepsindeling, het bijbehorende vignet te zien waarin het advies van de slimme keuzehulp werd getoond, gevolgd door de uitleg (zie bijlage 8.4). De experimentele vignetten in dit onderzoek zijn gebaseerd op de werkelijke uitleg die burgers nu ontvangen als zij met behulp van de slimme keuzehulp aangifte willen doen van internetoplichting. In de volgende alinea (operationalisatie) wordt dieper ingegaan op de elementen waaruit de verschillende soorten uitleg in de vignetten zijn opgebouwd.

4.2.4 Begrijpelijkheid van de experimentele situaties

Het is belangrijk voor een vignettenonderzoek dat respondenten zich tot op zekere hoogte in kunnen leven in de situaties die zij voorgelegd krijgen (Bryman, 2016, p. 260). Hiervoor moeten respondenten de experimentele situatie begrijpen. In het surveyexperiment worden respondenten namelijk gevraagd om zich in te leven in een casus en op basis hiervan de vragen te beantwoorden. In de vragenlijst is een schaal (schaal: 1-7) opgenomen om te checken of de respondenten zich goed in konden leven in de aan hen voorgelegde casus. Hier kwam uit dat respondenten zich goed konden inleven in de situatie (zie paragraaf 4.6.4). Daarnaast is het experiment getest met twaalf testrespondenten. Na het testen is feedback gevraagd aan de respondenten, die vervolgens verwerkt is in de vignetten en de vragenlijst. Deze feedback heeft bijgedragen aan de begrijpelijkheid van de experimentele situaties, omdat respondenten ($N = 536$) gemiddeld hoog scoren op de mate waarin ze zich in konden leven in de situatie ($M = 5,76$, $SD = 1,19$).

4.2.5 Realiteit van de experimentele situaties

Naast het feit dat de experimentele situatie begrijpelijk moeten zijn, is het voor de maatschappelijke relevantie van dit onderzoek belangrijk dat de gebruikte vignetten realistisch zijn. De deelnemers aan het experiment zijn namelijk willekeurige burgers die niet per se zelf een ervaring hebben gehad met internetoplichting. De casus en de vignetten moeten daarom begrijpelijk en realistisch zijn. Hoe realistischer deze zijn, hoe relevanter de resultaten van het onderzoek zijn voor de politie. Om de realiteit van de vignetten te waarborgen zijn de experimentele situaties voorgelegd aan twee medewerkers van het team Pre Development van de Politie-ICT in Utrecht. Dit heeft ervoor gezorgd dat de uitleg over de werking van de slimme keuzehulp (hoe beslissingen

tot stand komen en de motivering achter de specifieke beslissing) overeenkomt met de werkelijkheid.

4.3 OPERATIONALISATIE

In deze paragraaf worden eerst de twee niveaus van vertrouwen geoperationaliseerd die gebruikt worden bij het meten van vertrouwen van burgers in algoritmische besluiten: *vertrouwende houding* en *op vertrouwen gebaseerd gedrag*. De *vertrouwende houding* wordt geoperationaliseerd aan de hand van het construct 'vertrouwen in advies'. Bij de operationalisatie van dit construct is gebruik gemaakt van vragenlijsten van andere wetenschappers. Dit heeft als voordeel dat ze al in de praktijk getoetst zijn, waardoor de betrouwbaarheid en validiteit van het onderzoek vergroot worden (zie paragraaf 4.7) (Bryman, 2016, p. 261).

Het *op vertrouwen gebaseerd gedrag* wordt gemeten aan de hand van twee variabelen. Ten eerste het construct 'bereidheid om naar het advies te handelen', dat de intentie meet van de gebruiker van de slimme keuzehulp om het advies op te volgen. Ten tweede aan de hand van het 'daadwerkelijke gedrag van de gebruiker', dat inzicht geeft in hoe respondenten handelen op basis van het advies dat zij krijgen.

Na de operationalisatie van de afhankelijke variabele (vertrouwen) volgt die van de onafhankelijke variabele (uitlegbaarheid). Hierbij wordt toegelicht wat in dit onderzoek wordt verstaan onder een 'procedurele uitleg' en een 'inhoudelijke uitleg'. Op basis van deze operationalisatie zijn de experimentele vignetten tot stand gekomen. Tot slot wordt aandacht besteed aan de manipulatiecheck en enkele overige vragen die gesteld zijn in de vragenlijst.

4.3.1 Vertrouwende houding

Het construct 'vertrouwen in advies' is afgeleid van Fogg (1999). Deze schaal gaat namelijk over het vertrouwen van de mens in de computer. Fogg (1999) beschrijft hoe de items zowel voor het algemeen vertrouwen in de computer gebruikt kunnen worden, als voor het vertrouwen in specifieke onderdelen van een computer, zoals een interface. Dit maakt de schaal relevant voor de slimme keuzehulp, omdat dit een systeem die de communicatie tussen mens en computer faciliteert.

Er zijn enkele aanpassingen gedaan aan de originele schaal. Zo zijn twee van de zes items uit de originele schaal weggelaten. Deze items gingen over of het desbetreffende product of onderdeel "goed" (*good*) is en "goedbedoeld" (*well-intentioned*) (Fogg, 1999, p. 86). In de context van de slimme keuzehulp passen deze items minder goed, omdat ze een normatief element bevatten. Als burgers het advies krijgen om geen aangifte te doen terwijl ze bijvoorbeeld geld zijn verloren aan een product dat ze nooit ontvangen hebben, dan kan verwacht worden dat ze het advies niet goed vinden. Of ze het advies

vertrouwen is namelijk iets anders. Het kan namelijk zo zijn dat ze het advies dat de slimme keuzehulp geeft wel vertrouwen, maar niet goed vinden omdat ze liever een ander advies gezien hadden. Door de twee normatieve items te verwijderen uit de schaal, blijven er vier items over.

Deze stellingen zijn vertaald naar het Nederlands, omdat de vragenlijst ook in het Nederlands wordt afgenomen. Verder is de formulering van de stellingen aangepast om ze beter aan te sluiten op de voorgelegde situatie, zonder de betekenis van de stellingen te veranderen. Dit houdt onder andere in dat er wordt gesproken over een 'advies' in plaats van een besluit. Dit komt omdat de politie deze term gebruikt bij de uitleg van de beslissing van de slimme keuzehulp.

Construct	Vragen
Vertrouwen in advies	Ik vind het advies van de slimme keuzehulp betrouwbaar. Ik vind het advies van de slimme keuzehulp onbevooroordeeld. Ik vind het advies van de slimme keuzehulp waarheidsgetrouw. Ik vertrouw op het advies van de slimme keuzehulp.

4.3.2 Op vertrouwen gebaseerd gedrag

Bereidheid om naar advies te handelen

Het construct 'bereidheid om naar advies te handelen' is tot stand gekomen aan de hand van de schaal van McKnight et al. (2002). De originele schaal meet aan de hand van vijf items de intentie van een koper om het advies op te volgen van een online verkoper. Deze schaal hebben de auteurs zelf opgesteld voor hun specifieke casus en moet daarom aangepast worden aan de context van de slimme keuzehulp. Eén item valt hierbij weg, omdat dit item gaat over een veilig gevoel dat handelen op basis van het advies teweeg zou brengen. Dit gaat echter over de specifieke casus waarin persoonlijke informatie gedeeld moet worden voor een juridisch advies.

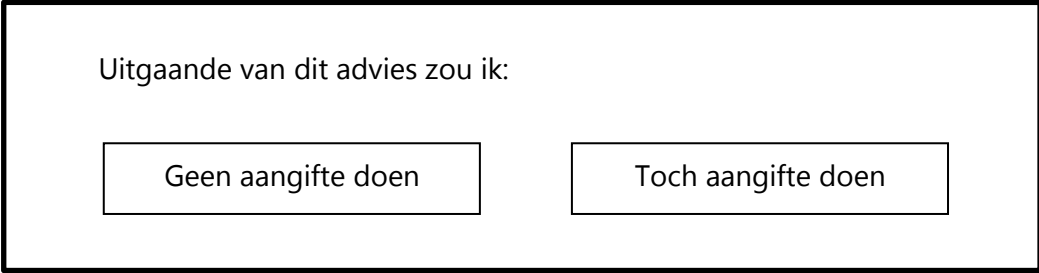
Dit houdt in dat er vier items overblijven. Net zoals bij het vorige construct zijn hier ook de stellingen vertaald naar het Nederlands. Verder is de formulering van de stellingen aangepast om ze beter aan te sluiten op de voorgelegde situatie, zonder de betekenis van de stellingen te veranderen. Zo wordt hier ook gesproken over een 'advies'.

Construct	Vragen
Bereidheid om naar advies te handelen	Het is zeer waarschijnlijk dat ik luister naar het advies van de slimme keuzehulp. Ik leg me neer bij het advies van de slimme keuzehulp. Ik vind het geen probleem om te handelen naar het advies van de slimme keuzehulp. Ik ben bereid het advies van de slimme keuzehulp op te volgen.

Daadwerkelijk gedrag gebruiker

Een voordeel van de casus die in dit onderzoek gebruikt wordt bij het surveyexperiment is dat het daadwerkelijke gedrag van de gebruiker vastgelegd kan worden. Naast de bereidheid die de respondent heeft om te handelen naar het advies dat hij krijgt, wordt de respondent gevraagd of hij wel of geen aangifte zou doen (zie figuur 4.1). Dit geeft inzicht in hoe de respondent werkelijk handelt op basis van het advies. Hierdoor kan het op vertrouwen gebaseerd gedrag vollediger gemeten worden.

Figuur 4.1: Operationalisatie 'daadwerkelijk gedrag gebruiker'



Uitgaande van dit advies zou ik:

Geen aangifte doen	Toch aangifte doen
--------------------	--------------------

Na het zien van het advies om geen aangifte te doen en de uitleg daarbij, kunnen burgers nu (in de originele vorm van de slimme keuzehulp) alsnog ervoor kiezen om aangifte te doen ('doorgaan met aangifte'). Uit gesprekken bij de Politie-ICT signaleerde de onderzoeker dat de knop 'doorgaan met aangifte' voor verwarring kan zorgen. De knop is daarom voor dit onderzoek aangepast naar 'toch aangifte doen' om deze duidelijker te maken voor de gebruikers van de slimme keuzehulp. De opties die de respondenten te zien kregen, zijn inhoudelijk dus wel vergelijkbaar met de opties die nu gegeven worden door de slimme keuzehulp en geven inzicht in het daadwerkelijke gedrag van de respondenten naar aanleiding van het advies dat zij kregen van de slimme keuzehulp.

4.3.3 Factoranalyse

Voor de acht variabelen tezamen van 'vertrouwen in advies' en 'bereidheid om naar advies te handelen' is een factoranalyse uitgevoerd. Zo'n factoranalyse legt onderlinge correlaties tussen de verschillende variabelen bloot, waardoor deze onderverdeeld kunnen worden in meerdere factoren. In de analyse is gekozen voor de optie om te zoeken naar twee factoren, omdat dit op basis van de theorie te verwachten viel.

Om te kijken in hoeverre de dataset geschikt is voor een factoranalyse, zijn twee testen gedaan. Ten eerste is de Bartlett's Test of Sphericity uitgevoerd. Deze bleek significant $\chi^2(28) = 3982,88; p < 0,001$. Kleine p -waarden, zoals hier het geval is, tonen aan dat de dataset geschikt is voor een factoranalyse (Field, 2009, p. 660). Ten tweede is een Kaiser-Meyer-Olkin (KMO) test gedaan om te kijken wel deel van de variantie in de variabelen veroorzaakt zou kunnen worden door onderliggende factoren. Hierbij geldt dat een

factoranalyse relevant is bij een hoge KMO-waarde (dichtbij 1,0). Waarden tussen de 0,9 en 1,0 worden gezien als “*superb*” (Field, 2009, p. 647). Dat was hier het geval: KMO = 0,92.

In deze analyse wordt er op basis van de theorie vanuit gegaan dat de variabelen onderling correleren, omdat ze zijn afgeleid van gevalideerde schalen. Hierdoor is gekozen voor een Oblique-rotatie: de Direct Oblimin rotatie met een standaard schuinheid (Δ) van 0. In bijlage 8.2 is de pattern matrix van de factoranalyse weergegeven.

Uit de analyse komen twee factoren naar voren. De twee schalen van ‘vertrouwen in advies’ en ‘bereidheid om naar advies te handelen’ die eerder besproken zijn, kunnen onderscheiden worden in de resultaten van de factoranalyse. Eén variabele laadt echter op beide factoren: ‘Ik vertrouw op het advies van de slimme keuzehulp’. Dit kan verklaard worden door het feit dat ‘bereidheid om naar advies te handelen’ in de theorie ook wordt gezien als onderdeel van vertrouwen in algoritmische besluiten. Het is dus niet verwonderlijk dat deze variabele op beide factoren laadt. Vanwege grondige theoretische onderbouwing wordt in dit onderzoek gekozen om de hierboven besproken indeling aan te houden. Hiervoor moeten de Cronbach’s alpha van de twee schalen berekend worden. Tabel 4.2 geeft de resultaten hiervan weer.

Tabel 4.2: Cronbach’s alpha van de gehanteerde constructen

Construct	Cronbach’s alpha	Items
Vertrouwen in advies	0,89	4
Bereidheid om naar advies te handelen	0,93	4

De Cronbach’s alpha van beide constructen is hoog genoeg om de schalen te gebruiken. Uitgegaan wordt namelijk van een Cronbach’s alpha $> 0,7$ voor een geslaagde betrouwbaarheidsanalyse (Field, 2009, p. 675). Bij het tweede construct werd de Cronbach’s alpha niet hoger na het verwijderen van een item. Bij het eerste construct zou de Cronbach’s alpha 0,91 worden als het item ‘Ik vind het advies van de slimme keuzehulp onbevooroordeeld.’ verwijderd zou worden. In dit onderzoek is ervoor gekozen om dit item niet te verwijderen. Ten eerste omdat de huidige Cronbach’s alpha al hoog is en het verwijderen van het item maar voor een geringe verhoging zou zorgen. Ten tweede omdat uit de factoranalyse niet naar voren kwam dat er reden zou zijn om aan te nemen dat dit item niet bij het construct zou passen.

4.3.4 Procedurele uitleg

Uit het theoretisch kader kwam naar voren dat een procedurele uitleg verschillende elementen kan bevatten, zoals uitleg over de gehanteerde *beslisbomen en -technieken*. In onderstaande figuur is de procedurele uitleg weergegeven die gebruikt is in dit onderzoek. Het complete experimentele vignet is te zien in bijlage 8.4. Onder het vignet

wordt uitgelegd hoe de uitleg tot stand is gekomen. Hierbij wordt steeds gerefereerd aan de theoretische bevindingen over een procedurele uitleg, zoals die in het theoretisch kader al aan bod kwamen. De sterke concrete koppeling van het experimentele vignet aan de theorie zorgt voor een hoge constructvaliditeit. Hier wordt in paragraaf 4.6.5. nog dieper op ingegaan.

Figuur 4.2: Procedurele uitleg

Waarom krijg ik het advies om geen aangifte te doen?

De webwinkel www.sneakerwinkel.nl staat bij de politie bekend als een betrouwbare webwinkel en bij dergelijke webwinkels is er vrijwel nooit sprake van online fraude.

Hoe komen we tot dit advies?

- De slimme keuzehulp analyseert uw tekst door gebruik te maken van verschillende **algoritmes**.
- Er wordt o.a. gebruik gemaakt van een **tekstanalyse** om te zoeken naar bepaalde termen en combinaties van woorden die kunnen wijzen op een fraudezaak.
- Op basis van deze tekstanalyse is gekeken of er nog **relevante vragen** zijn.
- Uit uw antwoorden op deze vragen bleek dat u een aankoop heeft gedaan bij www.sneakerwinkel.nl.
- Deze winkel is in het databestand van de slimme keuzehulp aangemerkt als "**betrouwbaar**" en bij dergelijke webwinkels is er doorgaans **geen sprake van fraude**.

Deze procedurele uitleg bevat informatie over de *beslistechneik*. Rader et al. (2018, pp. 9-10) en Kizilcec (2016, p. 2393) benadrukken dat het bij een procedurele uitleg van belang is dat het duidelijk wordt hoe het beslissysteem functioneert. Informatie over de beslistechneik is namelijk een belangrijk kenmerk van een procedurele uitleg (De Fine Licht et al., 2014, p.127). In de procedurele uitleg wordt duidelijk gemaakt hoe de slimme keuzehulp tot zijn advies 'geen aangifte doen' is gekomen. De respondent wordt verteld dat de slimme keuzehulp aan de hand van een tekstanalyse (met behulp van algoritmen) kijkt of er al geconcludeerd kan worden dat er mogelijk sprake is van fraude, of dat er nog naar ontbrekende informatie gevraagd moet worden (Bex, 2019).

Daarnaast wordt in de uitleg aan de respondent kenbaar gemaakt dat er een *beslisboom* wordt gehanteerd. Zoals Guidotti et al. (2018, p. 8) en Kim & Routledge (2018, pp. 3-4) aantonen, maken regels, een beslisboom en/of een als-dan-redenering de uitleg van een

algoritmische beslissing begrijpelijk voor de gebruiker. Zo wordt uit de procedurele uitleg duidelijk dat als de slimme keuzehulp zoekt naar bepaalde (combinaties van) woorden die kunnen wijzen op een fraudezaak. Dit zijn de regels (in dit geval gebaseerd op de wet) die het voor de burger inzichtelijk maken hoe de slimme keuzehulp tot zijn uitleg komt. Verder zit in deze procedurele uitleg ook het element van een als-dan-redenatie. Als er nog niet voldoende informatie is om een conclusie te trekken, dan stelt de slimme keuzehulp extra vragen.

4.3.5 Inhoudelijke uitleg

Vanuit de theorie werd duidelijk dat een inhoudelijke uitleg een aantal elementen kan bevatten, namelijk uitleg over de *redenering* van een specifieke beslissing en uitleg over eventuele *individuele omstandigheden* die meegewogen worden bij het maken van de beslissing. Figuur 4.3 laat de inhoudelijke uitleg zien die in dit onderzoek gebruikt is. Hierbij geldt weer dat het gehele vignet terug te vinden is in de bijlage (zie 8.4). Uit welke elementen de inhoudelijke uitleg is opgebouwd wordt onder de figuur toegelicht. Net zoals bij de procedurele uitleg is hier ook sprake van een sterke concrete koppeling van het experimentele vignet aan de theorie.

Figuur 4.3: Inhoudelijke uitleg

Waarom krijg ik het advies om geen aangifte te doen?

De webwinkel www.sneakerwinkel.nl staat bij de politie bekend als een betrouwbare webwinkel en bij dergelijke webwinkels is er vrijwel nooit sprake van online fraude.

Waarom is dit een betrouwbare webwinkel?

- De webwinkel heeft het **keurmerk** 'Webshop Keurmerk'.
- Een webwinkel die aangesloten is bij dit keurmerk is een **uitgebreide screeningsprocedure** doorgegaan.
- Het keurmerk bemiddelt bij een geschil met een webwinkel.
- Als deze bemiddeling niet tot het gewenste resultaat leidt, kunt u het geschil voorleggen aan een **onafhankelijke geschillencommissie**.
- Kijk op de website van de webwinkel voor wat u moet doen om uw geld terug te krijgen.

Uit de theorie kwam naar voren dat het bij een inhoudelijke uitleg van belang is dat de achterliggende argumenten voor een bepaalde beslissing duidelijk worden (Kizilcec, 2016). Dit kan gedaan worden door uitleg te geven over de individuele omstandigheden die ervoor gezorgd hebben dat iemand een bepaalde beslissing te zien krijgt (Kim & Routledge, 2018, pp. 3-4).

In deze inhoudelijke uitleg wordt de respondent duidelijk gemaakt waarom in zijn of haar specifieke geval het advies 'geen aangifte doen' gegeven wordt. De uitleg bevat een *redenering* waarom in het specifieke geval van www.sneakerwinkel.nl sprake is van een betrouwbare webwinkel, wat van groot belang is bij een inhoudelijke uitleg (Kim & Routledge, 2018, pp. 3-4). In de uitleg wordt de respondent verteld dat er rekening gehouden wordt met bepaalde eigenschappen of kenmerken (bij een webshop met zo'n keurmerk komt doorgaans geen internetfraude voor) die zijn afgeleid uit zijn individuele casus, wat volgens Kim en Routledge (2018) een sterk kenmerk is van een inhoudelijke uitleg.

4.3.6 Manipulatiecheck

Om te controleren of de manipulatie in het experiment succesvol plaats heeft gevonden zijn een vijftal stellingen (schaal: 1-7) toegevoegd:

Geef aan in hoeverre de volgende stelling van toepassing is op de door u gelezen situatie:

1. De slimme keuzehulp gaf **geen** goede uitleg over het advies.
2. De slimme keuzehulp gaf uitleg over de manier waarop het advies tot stand is gekomen.
3. De slimme keuzehulp maakte duidelijk welke stappen zijn gezet om tot het advies te komen.
4. Ik begrijp waarom het in mijn geval niet zinvol was om aangifte te doen.
5. De slimme keuzehulp gaf mij duidelijke uitleg over waarom ik het advies kreeg om geen aangifte te doen.

De experimentele groepen zijn met elkaar vergeleken voor deze vijf vragen. Hiervoor is vijf keer een one-way ANOVA uitgevoerd. Tot slot wordt een korte conclusie gegeven over de resultaten van de manipulatiecheck.

De slimme keuzehulp gaf geen goede uitleg over het advies.

De ANOVA was statistisch significant, wat betekent dat de mate waarin respondenten ervaren dat ze geen goede uitleg kregen over het advies, beïnvloed werd door de experimentele groep waarin zij zaten, $F(3, 532) = 173,45, p < 0,001, \eta^2 = 0,50$.

Uit de post hoc analyse met Games-Howell (bij $\alpha = 0,05$) kwam naar voren dat respondenten in de controlegroep zonder uitleg ($M = 6,12, SD = 1,54$) significant hoger scoorden dan respondenten in een van de drie treatmentgroepen met uitleg. Respondenten die een procedurele uitleg ontvingen ($M = 3,23, SD = 1,69$) scoorden significant hoger op deze stelling dan respondenten met een inhoudelijke uitleg ($M = 2,63, SD = 1,46$) en met een combinatie van een procedurele en inhoudelijke uitleg ($M = 2,39, SD = 1,33$). Er waren geen significante verschillen tussen respondenten met een

inhoudelijke uitleg en respondenten die een combinatie van inhoudelijke en procedurele uitleg kregen.

De slimme keuzehulp gaf uitleg over de manier waarop het advies tot stand is gekomen.

De ANOVA was statistisch significant, wat betekent dat de mate waarin respondenten ervaren dat ze geen goede uitleg kregen over het advies, beïnvloed werd door de experimentele groep waarin zij zaten, $F(3, 532) = 372,30$, $p < 0,001$, $\eta^2 = 0,68$.

Uit de post hoc analyse met Games-Howell (bij $\alpha = 0,05$) kwam naar voren dat respondenten in de controlegroep zonder uitleg ($M = 1,62$, $SD = 1,21$) significant lager scoorden dan respondenten in een van de drie treatmentgroepen met uitleg. Echter, er waren geen significante verschillen tussen respondenten die een procedurele uitleg ($M = 5,55$, $SD = 1,09$), een inhoudelijke uitleg ($M = 5,45$, $SD = 1,30$) en een combinatie van een procedurele en inhoudelijke uitleg ontvingen ($M = 5,75$, $SD = 1,14$).

De slimme keuzehulp maakte duidelijk welke stappen zijn gezet om tot het advies te komen.

De ANOVA was statistisch significant, wat betekent dat de mate waarin respondenten ervaren dat de slimme keuzehulp duidelijk maakte welke stappen zijn gezet om tot het advies te komen, beïnvloed werd door de experimentele groep waarin zij zaten, $F(3, 532) = 265,93$, $p < 0,001$, $\eta^2 = 0,60$.

Uit de post hoc analyse met Games-Howell (bij $\alpha = 0,05$) kwam naar voren dat respondenten in de controlegroep zonder uitleg ($M = 1,53$, $SD = 1,05$) significant lager scoorden dan respondenten in een van de drie treatmentgroepen met uitleg. Verder scoorden respondenten die een combinatie-uitleg ontvingen ($M = 5,54$, $SD = 1,18$) significant hoger dan respondenten met een procedurele uitleg ($M = 4,98$, $SD = 1,48$) en met een inhoudelijke uitleg ($M = 5,10$, $SD = 1,48$). Er waren geen significante verschillen tussen de procedurele en inhoudelijke groep.

Ik begrijp waarom het in mijn geval niet zinvol was om aangifte te doen.

De ANOVA was statistisch significant, wat betekent dat de mate waarin respondenten begrepen waarom het in hun geval niet zinvol was om aangifte te doen, beïnvloed werd door de experimentele groep waarin zij zaten, $F(3, 532) = 119,59$, $p < 0,001$, $\eta^2 = 0,40$.

Uit de post hoc analyse met Games-Howell (bij $\alpha = 0,05$) kwam naar voren dat respondenten in de controlegroep zonder uitleg ($M = 2,05$, $SD = 1,67$) significant lager scoorden dan respondenten in een van de drie treatmentgroepen met uitleg. Respondenten die een procedurele uitleg kregen ($M = 4,39$, $SD = 1,82$), scoorden ook significant lager dan respondenten met een inhoudelijke uitleg ($M = 5,44$, $SD = 1,48$), of

een combinatie-uitleg ($M = 5,30$, $SD = 1,64$). Er waren geen significante verschillen tussen de respondenten met een inhoudelijke uitleg of een combinatie-uitleg.

De slimme keuzehulp gaf mij duidelijke uitleg over waarom ik het advies kreeg om geen aangifte te doen.

De ANOVA was statistisch significant, wat betekent dat de mate waarin respondenten ervaren dat ze duidelijke uitleg ontvingen over waarom zij het advies kregen om geen aangifte te doen, beïnvloed werd door de experimentele groep waarin zij zaten, $F(3, 532) = 309,81$, $p < 0,001$, $\eta^2 = 0,64$.

Uit de post hoc analyse met Games-Howell (bij $\alpha = 0,05$) kwam naar voren dat respondenten in de controlegroep zonder uitleg ($M = 1,58$, $SD = 1,12$) significant lager scoorden dan respondenten in een van de drie treatmentgroepen met uitleg. Respondenten die een procedurele uitleg kregen ($M = 5,04$, $SD = 1,49$), scoorden ook significant lager dan respondenten met een inhoudelijke uitleg ($M = 5,70$, $SD = 1,05$), of een combinatie-uitleg ($M = 5,53$, $SD = 1,40$). Er waren geen significante verschillen tussen de respondenten met een inhoudelijke uitleg of een combinatie-uitleg.

Conclusie manipulatiecheck

Uit de manipulatiecheck kan opgemaakt worden dat de experimentele behandeling van de controlegroep geslaagd is. Zij ervoeren duidelijk dat zij geen of minder goede uitleg ontvingen dan de andere drie groepen. De verschillen tussen de andere drie treatmentgroepen is minder groot. Dit kan verklaard worden doordat er maar hele subtiele verschillen in bewoording tussen de laatste vier stellingen zaten. In de discussie wordt gereflecteerd op de uitkomsten van de manipulatiecheck in het licht van de gevonden resultaten.

4.3.7 Overige vragen

Naast de concepten die hierboven geoperationaliseerd zijn en de manipulatiecheck, bevatte de vragenlijst nog een aantal overige vragen: controlevragen, demografische vragen en feedbackvragen. Deze zullen hieronder toegelicht worden.

Demografische vragen

In de vragenlijst zijn een aantal demografische variabelen opgenomen, zodat een weergave kan worden gegeven van de demografische opbouw van de respondentengroep. Voor dit onderzoek waren de volgende drie vragen relevant:

- Wat is uw geslacht?
- Wat is uw geboortjaar?
- Wat is uw hoogst behaalde onderwijsniveau?

Controlevragen

Aan het einde van de vragenlijst volgden nog twee controlevragen. Deze vragen gingen over het vertrouwen van de respondenten in de politie en in technologie in het algemeen. Deze variabelen kunnen mogelijk het effect van uitlegbaarheid van algoritmische besluiten op het vertrouwen van burgers in deze besluiten beïnvloeden. Dit zijn dus variabelen die een invloed kunnen hebben op de relatie tussen de afhankelijke en de onafhankelijke variabele (Seltman, 2012, p. 12). In dit onderzoek kan aangenomen worden dat in hoeverre een respondent vertrouwen heeft in de politie en technologie in het algemeen van invloed kunnen zijn op het effect van uitlegbaarheid op vertrouwen in algoritmische besluiten. Om dit mogelijke effect te toetsen zijn de volgende vragen (schaal: 1-7) toegevoegd aan de vragenlijst:

Geef aan hoeveel vertrouwen u heeft in de volgende zaken:

- Vertrouwen in de politie in het algemeen
- Vertrouwen in technologie in het algemeen

Feedbackvragen

Tot slot zijn een aantal feedbackvragen (schaal: 1-7) voorgelegd aan de respondenten. Deze hadden als doel om te toetsen in hoeverre mensen zich in konden leven in de voorgelegde experimentele situaties:

Geef aan in hoeverre u het eens bent met de volgende stellingen:

- Ik kon mij inleven in de aan mij voorgelegde situatie.
- Ik heb een soortgelijke situatie van internetoplichting wel eens meegemaakt.

4.4 SELECTIE RESPONDENTEN

Een surveyexperiment is een onderzoek dat wordt afgenomen bij een voor de populatie representatieve steekproef (Mutz, 2011, p. 14). In surveyexperimenten wordt geprobeerd om een verzameling aan respondenten (*sample*) te creëren die representatief is voor de populatie (in dit geval Nederland). Een voordeel van een surveyexperiment is dat de experimentele omstandigheden op grote schaal getoetst kunnen worden (Mutz, 2011, p. 16). Bij een laboratoriumexperiment is dit vaak moeilijker, door de beperkte tijd, middelen en mensen. In dit onderzoek zijn de respondenten geworven via sociale mediakanalen LinkedIn en Facebook, maar ook door verspreiding van het surveyexperiment binnen het netwerk van de onderzoeker. Hierbij is dus gebruik gemaakt van *convenience sampling*, wat inhoudt dat de respondenten geworven zijn op basis van toegankelijkheid voor de onderzoeker (Bryman, 2016, p. 187). Dit kan bijgedragen hebben aan het feit dat de steekproef niet geheel representatief is voor de populatie. In de discussie zal hier verder aandacht aan worden besteed.

Bij de indeling van respondenten in de experimentele groepen is randomisatie toegepast. Hierbij is gebruik gemaakt van een *random assignment* (Bryman, 2016, p. 45). De deelnemers aan het experiment werden willekeurig in één van de vier groepen ingedeeld. De kans voor een respondent om in een bepaalde groep te belanden was dus 25%. In totaal namen $N = 536$ mensen deel aan het experiment. Tabel 4.3 geeft een kort overzicht weer van hoe die totale steekproef eruit zag. Voor een compleet overzicht van de beschrijvende statistieken van de steekproef, zie bijlage 8.1.

Tabel 4.3: Overzicht steekproef

Variabele		Verdeling
Geslacht	Man	42,7%
	Vrouw	56,2%
		Gemiddelde
Geboortjaar		1980
Opleidingsniveau		HBO- of WO-bachelor

4.4.1 Beschrijving van de experimentele groepen

In dit onderzoek zijn de respondenten evenredig over vier experimentele groepen verdeeld. Uitval tijdens het surveyexperiment kan ervoor zorgen dat de groepen verschillen in grootte. Voor het experiment is het aanbevolen om groepen te hebben die dicht bij elkaar liggen wat betreft het aantal respondenten. Zoals in paragraaf 4.7.4. verder toegelicht wordt, is de gewenste aantal grootte van de totale steekproef minimaal 492 respondenten. In elke groep zouden dus minimaal 123 respondenten zitten. De daadwerkelijke *sample size* van dit experiment was $N = 536$, dus dat zou betekenen dat er in elke groep ongeveer 134 respondenten zouden moeten zijn. Hieronder is een verdeling per groep te zien.

Tabel 4.4: Steekproefgrootte experimentele groepen

Experimentele groep	Steekproefgrootte
Controle: geen uitleg	$n = 133$
Treatment 1: procedurele uitleg	$n = 134$
Treatment 2: inhoudelijke uitleg	$n = 138$
Treatment 3: procedurele en inhoudelijke uitleg	$n = 131$

Zoals uit de tabel naar voren komt, is de verdeling nagenoeg evenredig. Dit houdt in dat de verdeling van respondenten met succes is gelukt. Vervolgens is een randomisatiecheck uitgevoerd. Dit is gedaan om te controleren of de gevonden verschillen in resultaten daadwerkelijk toe te schrijven zijn aan de experimentele behandeling. Als de randomisatie geslaagd is, is namelijk de experimentele behandeling

het enige verschil tussen de verschillende groepen (Bryman, 2016, p. 45). Tabel 4.5 laat de resultaten van deze randomisatiecheck zien.

Tabel 4.5: Randomisatiecheck

Variabele	Statistiek	Significantie (p)
Geslacht	$\chi^2 = 2,82$	0,42
Opleidingsniveau*	$\chi^2 = 1,72$	0,63
Geboortjaar	$F = 2,63$	0,05
Vertrouwen in de politie in het algemeen	$F = 2,04$	0,11
Vertrouwen in technologie in het algemeen	$F = 0,05$	0,99
Inleven in de situatie	$F = 0,67$	0,57
Soortgelijke situatie eerder meegemaakt	$F = 0,18$	0,91

**Hierbij zijn enkel de categorieën 'midden' en 'hoog' meegenomen, omdat er te weinig respondenten in de categorie 'laag' zaten om de analyse toe te passen.*

Uit de tabel komt naar voren dat er voor alle variabelen, op één na, geen significante verschillen ($\alpha = 0,05$) zijn tussen de groepen. Alleen bij de variabele 'geboortjaar' waren de resultaten significant. Een post-hoc analyse met Sidak wees echter uit dat er geen significante verschillen in leeftijd tussen de verschillende experimentele groepen waren vast te stellen. Hieruit kan geconcludeerd worden dat de randomisatie geslaagd is en dat verschillen in resultaten enkel te wijten zijn aan de experimentele behandeling. Een enkele keer was er geen sprake van een normaalverdeling of homogeniteit van varianties, wat voorwaarden zijn voor het uitvoeren van een one-way ANOVA-toets. In een grote steekproef is dit niet ongebruikelijk en de ANOVA wordt gezien als een robuuste toets, wat inhoudt dat de resultaten van de toets nog steeds interpreteerbaar zijn als enkele voorwaarden geschonden worden (Field, 2009, pp. 155, 359). Om deze reden is de toets toch gebruikt bij de randomisatiecheck.

4.5 DATA-ANALYSE

Wat betreft de data-analyse is het van belang om toe te lichten welke toetsen zijn uitgevoerd. Om te toetsen of er een verschil bestaat tussen de gemiddelden van de controle- en treatmentgroepen wordt een 2x2 factoriale ANOVA-toets gebruikt. Dit is gedaan voor zowel de variabele 'vertrouwen in het advies van de slimme keuzehulp' als voor de variabele 'bereidheid om te handelen naar dit advies'. Aan de hand van contrasten, gespecificeerd naar de vier hypothesen (zie bijlage 8.3), is gekeken welke groepen significant van elkaar verschilden. Verder is een binaire logistische regressie uitgevoerd om te toetsen of het soort uitleg dat respondenten kregen een voorspellend effect heeft op of respondenten al dan niet aangifte zouden doen.

4.6 BETROUWBAARHEID EN VALIDITEIT

Allereerst zal kort ingegaan worden op de betrouwbaarheid van dit onderzoek. Daarna worden vier soorten validiteit behandeld: interne, externe, ecologische, statistische en constructvaliditeit.

4.6.1 Betrouwbaarheid

De betrouwbaarheid van een onderzoek gaat over in hoeverre de gevonden resultaten te repliceren zijn (Bryman, 2016, p. 158). Allereerst is de betrouwbaarheid van dit onderzoek gewaarborgd door de meetschalen deductief op te stellen en zo min mogelijk aan te passen. De meetschalen zijn vervolgens getest en zijn allen betrouwbaar (Cronbach's $\alpha > 0,7$). Dit houdt in dat de indicatoren waaruit de schalen zijn opgemaakt, consistent zijn (Bryman, 2016, p. 157). Ten tweede zijn de respondenten in dit onderzoek willekeurig verdeeld over de experimentele groepen. Hierbij is onderzocht of de groepen vergelijkbaar met elkaar waren en is er een randomisatiecheck uitgevoerd. Hier kwamen geen significante resultaten uit die aanleiding kunnen geven om aan te nemen dat de respondenten niet willekeurig verdeeld waren.

4.6.2 Interne validiteit

Interne validiteit gaat over de mate waarin vastgesteld kan worden dat er sprake is van een causaal verband (Bryman, 2016, p. 41). Bij een experimenteel design is het gebruikelijk dat de interne validiteit hoog is (Bryman, 2016, p. 44). In dit onderzoek is randomisatie toegepast, waardoor de respondenten eerlijk over de groepen verdeeld zijn. Hierdoor blijven de eigenschappen van de verschillende groepen gelijk. Enkel wat betreft de experimentele behandeling verschillen de groepen, waardoor een gevonden effect direct toe te schrijven valt aan een verandering van de onafhankelijke variabele (Bryman, 2016, p. 42). Met betrekking tot de interne validiteit treden vier mogelijke risico's op: selectie-effecten, maturatie, attritie en testeffecten (Bryman, 2016, p. 47).

Mogelijke selectie-effecten (systematische verschillen van kenmerken van respondenten tussen de experimentele groepen) zijn voorkomen door de willekeurige plaatsing van respondenten in de controle- of behandelgroepen middels het programma *Qualtrics*. Ook is het risico op maturatie voorkomen. Dit houdt in dat natuurlijke veranderingen door de tijd heen een invloed hebben op de resultaten. Het experiment heeft maar twee weken geduurd, waardoor maturatie van de resultaten geminimaliseerd is. Verder is de respons in alle groepen nagenoeg even groot. Dit houdt in dat er niet in één van de groepen een groter gedeelte van de respondenten is uitgevallen, waardoor attritie zou optreden. Deze beperking houdt in dat in één groep meer uitval is van respondenten dan in de andere groepen, waardoor de resultaten minder betrouwbaar zijn. Tot slot is de kans op testeffecten erg klein bij een surveyexperiment, omdat de respondenten thuis

de enquête ingevuld hebben. Ze zaten niet in een testruimte waarbij een onderzoeker aanwezig was, waardoor ze mogelijk anders reageren.

4.6.3 Externe validiteit

De externe validiteit van een onderzoek betreft de mate waarin de resultaten van een onderzoek generaliseerbaar zijn buiten de onderzochte context (Bryman, 2016, p. 47). Het is hierbij dus van belang dat de resultaten te generaliseren zijn naar de hele populatie. Een veelgehoorde kritiek op experimenten is dat deze een hoge interne validiteit, maar een lage externe validiteit hebben (Seltman, 2015). Dit komt voort uit het feit dat de resultaten van een experiment alleen geldig zijn voor de specifieke onderzochte situatie en hierdoor moeilijk te generaliseren zijn. In dit onderzoek is gebruik gemaakt van een op de populatie-gebaseerd surveyexperiment. Een voordeel hiervan is dat de resultaten, mits de steekproef representatief is voor de populatie, betrekking hebben op de gehele populatie (Mutz, 2012, p. 153). Dit draagt bij aan de generaliseerbaarheid van de resultaten van dit onderzoek. In de discussie wordt nog dieper ingegaan op de externe validiteit van dit onderzoek.

4.6.4 Ecologische validiteit

Een ander mogelijk probleem met betrekking tot de generaliseerbaarheid van de resultaten zou kunnen voortkomen uit het feit dat de respondenten in dit onderzoek geen 'echte' slachtoffers van internetfraude zijn. Dit heeft betrekking op de ecologische validiteit van het onderzoek, waarbij het van belang is dat bevindingen uit dit onderzoek generaliseerbaar zijn naar andere alledaagse situaties (Bryman, 2016, p. 47). Het risico dat dat niet het geval zou zijn is zoveel mogelijk ondervangen door de casus en de experimentele vignetten in samenspraak met de politie op te stellen, waardoor deze erg reëel zijn. Zoals al eerder benoemd, konden respondenten zich goed inleven in de situatie ($M = 5,77$; $SD = 1,19$). Dit kan verklaard worden door het feit dat het tegenwoordig voor veel mensen steeds normaler is om via internet iets te bestellen. Voor respondenten is het daarom makkelijk om zich in te leven in een situatie waarbij zij geen pakketje ontvangen, maar er wel voor betaald hebben. Dit draagt bij aan de ecologische validiteit van dit onderzoek.

4.6.5 Statistische conclusievaliditeit

Statistische conclusievaliditeit gaat over of het onderzoek in statistisch opzicht zodanig is opgezet dat er daadwerkelijk een verband gedetecteerd kan worden en dus een kwantitatieve conclusie mogelijk is. Met behulp van het programma *G*Power 3.1* is de gewenste grootte van de steekproef berekend. Voor de *a priori power* ($1-\beta$) is uitgegaan van 0,80. Omdat de effectgrootte vooraf nog niet bekend was, is een inschatting gemaakt van 0,15. Dit houdt een relatief klein effect van het soort uitleg op de mate van vertrouwen (Seltman, 2012). Bij een foutmarge van 0,05 komt bij vier groepen de

steekproefgrootte uit op 492. Aan het experiment deden uiteindelijk 536 respondenten mee. Dit houdt in dat het minimale aantal gehaald is en het experiment voldoende krachtig is om mogelijke effecten aan te tonen.

4.6.6 Constructvaliditeit

Tot slot wordt bij de constructvaliditeit nagegaan of er een voldoende sterke link is tussen de experimentele behandelingen, metingen en theoretische concepten. In dit onderzoek zijn hypothesen deductief ontwikkeld, wat bijdraagt aan de constructvaliditeit (Bryman, 2016, p. 159). Er zijn concepten gebruikt die gebaseerd zijn op onderzoeken van Fogg (1999) en McKnight et al. (2002). Beide onderzoeken zijn in de wetenschap veelvuldig gebruikt. Daarnaast is voor beide constructen de Cronbach's alpha berekend. In beide gevallen was er sprake van een hoge Cronbach's alpha, wat duidt op een goede constructvaliditeit. Ook liet de factoranalyse zien dat de items uit de constructen laadden op twee factoren die overeenkomen met de twee constructen.

4.7 ETHISCHE VERANTWOORDING

Bij een experimenteel design is het van belang om na te denken over de rol van ethiek in het onderzoek. Hierbij zijn volgens (Bryman 2016, p. 125) vier principes die in acht genomen moeten worden: het voorkomen van schade aan deelnemers, het informeren van deelnemers over het experiment, het waarborgen van de privacy van deelnemers en het voorkomen van misleiding van deelnemers.

4.7.1 Schade

Allereerst is het van belang dat deelnemers aan het experiment niet geschaad worden. Dit kan bijvoorbeeld fysieke schade zijn, een toename in stress, een afname in zelfverzekerdheid, etc. (Bryman, 2016, p. 126). In dit onderzoek is er geen reden om aan te nemen dat de deelnemers aan het experiment geschaad zijn. De enige manipulatie die zij immers ondervonden hebben is een verschil in uitleg en het niet weten dat andere respondenten mogelijk een ander vignet voorgelegd kregen. Aan het eind van het onderzoek is een open vraag toegevoegd waarin respondenten opmerkingen over het onderzoek konden plaatsen. Hier zijn geen negatieve antwoorden gegeven die betrekking hebben op ethische aspecten van dit onderzoek, wat een extra bevestiging is dat respondenten geen schade hebben ondervonden door deelname aan dit experiment.

4.7.2 Informed consent

Een tweede principe gaat over *informed consent*: deelnemers aan het experiment dienen geïnformeerd te worden (Bryman, 2016, pp. 129-131). Voor deelname aan het experiment zijn respondenten in de introductie van de vragenlijst geïnformeerd over het doel en de aard van het onderzoek. Ook werd in deze tekst aangegeven dat hun

gegevens anoniem werden verwerkt en dat ze op elk moment mochten stoppen met het experiment. In de introductietekst werd tevens het e-mailadres van de onderzoeker getoond, zodat deelnemers bij verdere vragen contact konden opnemen met de onderzoeker. Aan het eind van de introductie werd de respondent medegedeeld dat door op de knop 'verder' te klikken, akkoord gegaan werd met het gebruik van hun antwoorden in dit onderzoek.

4.7.3 Privacy

Ten derde noemt Bryman (2016, pp. 131-133) dat de privacy van deelnemers niet in gevaar mag worden gebracht. De onderzoeker heeft alle gegevens anoniem verwerkt door de respondenten te anonimiseren en hun IP-adres niet op te slaan. Verder wordt het programma *Qualtrics*, dat gebruikt is voor de verzameling van data in dit onderzoek, als veilig beschouwd voor het verwerken van onderzoeksgegevens. De resultaten van het onderzoek zijn opgeslagen op *Yoda*. In het kader van dit onderzoek zullen de data minimaal 6 weken bewaard blijven.

4.7.4 Misleiding

Het vierde principe houdt in dat deelnemers niet misleid mogen worden. In dit onderzoek hebben respondenten een experimentele behandeling ondergaan (Bryman, 2016, pp. 133-134). Dit betrof echter enkel het lezen van een uitleg over het advies van de slimme keuzehulp, die ten behoeve van het experiment gemanipuleerd was. De respondenten zijn dus zelf niet misleid. Verder werd vooraf duidelijk gemaakt aan de respondent dat het niet ging om een echte aangifte bij de politie, maar dat dit voor het doel van het onderzoek nageemaakt was.

Ook kregen respondenten de gegevens die ze in moesten vullen over de aangifte door de onderzoeker aangeleverd. Ze hoefden de slimme keuzehulp niet te doorlopen aan de hand van een eigen ervaring met internetoplichting. Respondenten waren dus geen 'echte' slachtoffers van internetfraude die de slimme keuzehulp gebruikten om aangifte te doen. Hierdoor was er geen sprake van misleiding van respondenten. Tot slot werd op het eind van de vragenlijst een *debriefing* gegeven. Hierdoor zijn respondenten op de hoogte gebracht van de werkwijze en het doel van het onderzoek. Ook werd na afloop het emailadres van de onderzoeker getoond, zodat bij vragen of opmerkingen over de werkwijze van het onderzoek contact opgenomen kon worden. Er zijn geen reacties binnengekomen over het onderzoek, wat bijdraagt aan de bevestiging dat respondenten niet misleid zijn in dit onderzoek.

5 Resultaten

De causale relatie tussen de uitlegbaarheid van algoritmische besluiten en het vertrouwen dat burgers in deze besluiten hebben, staat in dit hoofdstuk centraal. Aan de hand van drie empirische deelvragen zal gekeken worden of deze causale relatie bestaat:

1. Wat is de invloed van uitlegbaarheid op het vertrouwen van gebruikers in het advies van de slimme keuzehulp?
2. Wat is de invloed van uitlegbaarheid op de bereidheid van gebruikers om naar het advies van de slimme keuzehulp te handelen?
3. Wat is de invloed van uitlegbaarheid op het daadwerkelijke gedrag van gebruikers van de slimme keuzehulp?

Dit hoofdstuk is opgedeeld in drie onderdelen. In het eerste onderdeel wordt gekeken naar de *vertrouwende houding* van de respondent. De eerste empirische deelvraag past binnen dit niveau van vertrouwen. Via een analyse van de samengestelde schaal 'vertrouwen in advies', wordt de vertrouwende houding van de respondent vastgesteld.

In het tweede onderdeel van dit hoofdstuk wordt aandacht besteed aan *op vertrouwen gebaseerd gedrag*. In lijn met het methodenhoofdstuk valt dit niveau van vertrouwen uiteen in twee onderdelen. Zo wordt enerzijds via de samengestelde schaal 'bereidheid om naar advies te handelen', de intentie om naar het advies te handelen, gemeten. Anderzijds wordt het daadwerkelijke gedrag van de respondent op basis van het advies vastgesteld door hen de keuze te laten maken om al dan niet aangifte te doen. In dit onderdeel worden de tweede en derde empirische deelvraag behandeld. Bij de behandeling van elk van de drie deelvragen volgt een tussentijds overzicht van de toetsing van de hypothesen en een deelconclusie met de beantwoording van de desbetreffende deelvraag.

Het derde laatste onderdeel van dit hoofdstuk combineert de resultaten van de drie deelvragen en kijkt of er verschillen of overeenkomsten zijn tussen de resultaten van de verschillende concepten van vertrouwen. In deze slotsom volgt een eindconclusie over de toetsing van de vier hypothesen.

5.1 VERTROUWENDE HOUDING

Een 2x2 factoriaal ANOVA is uitgevoerd om de mate van vertrouwen in het advies van de slimme keuzehulp te vergelijken met respondenten die wel of geen procedurele uitleg kregen, gecombineerd met wel of geen inhoudelijke uitleg.

Beschrijvende statistiek

De beschrijvende statistiek ($N = 536$) van de afhankelijke variabele 'vertrouwen in advies' is weergegeven in tabel 5.1.¹

Tabel 5.1: Beschrijvende statistiek 'vertrouwen in advies'

Experimentele groep	<i>M</i>	<i>SD</i>
Controlegroep (geen uitleg)	3,60	1,31
Treatment 1 (procedurele uitleg)	4,50	1,32
Treatment 2 (inhoudelijke uitleg)	5,18	1,21
Treatment 3 (procedurele en inhoudelijke uitleg)	5,05	1,25
Totaal	4,59	1,41

Toetsende statistiek

Het effect van een procedurele uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was significant, $F(1, 532) = 12,46$; $p < 0,001$; $\eta_p^2 = 0,02$. Respondenten die geen procedurele uitleg kregen ($M = 4,40$; $SD = 1,49$) hadden significant minder vertrouwen in het advies van de slimme keuzehulp dan respondenten die wel een procedurele uitleg ontvingen ($M = 4,77$; $SD = 1,31$).

Het effect van een inhoudelijke uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was ook significant, $F(1, 532) = 94,24$; $p < 0,001$; $\eta_p^2 = 0,15$. Respondenten die geen inhoudelijke uitleg kregen ($M = 4,05$; $SD = 1,39$) hadden significant minder vertrouwen in het advies van de slimme keuzehulp dan respondenten die wel een inhoudelijke uitleg ontvingen ($M = 5,11$; $SD = 1,23$).

Zoals opgemerkt kan worden uit de toetsende statistiek, is de effectgrootte van een inhoudelijke uitleg veel groter (0,15) dan van een procedurele uitleg (0,02). Met een inhoudelijke uitleg kan dus 15% van de verschillen in de mate van vertrouwen tussen de groepen verklaard worden, terwijl dat bij een procedurele uitleg maar 2% is.

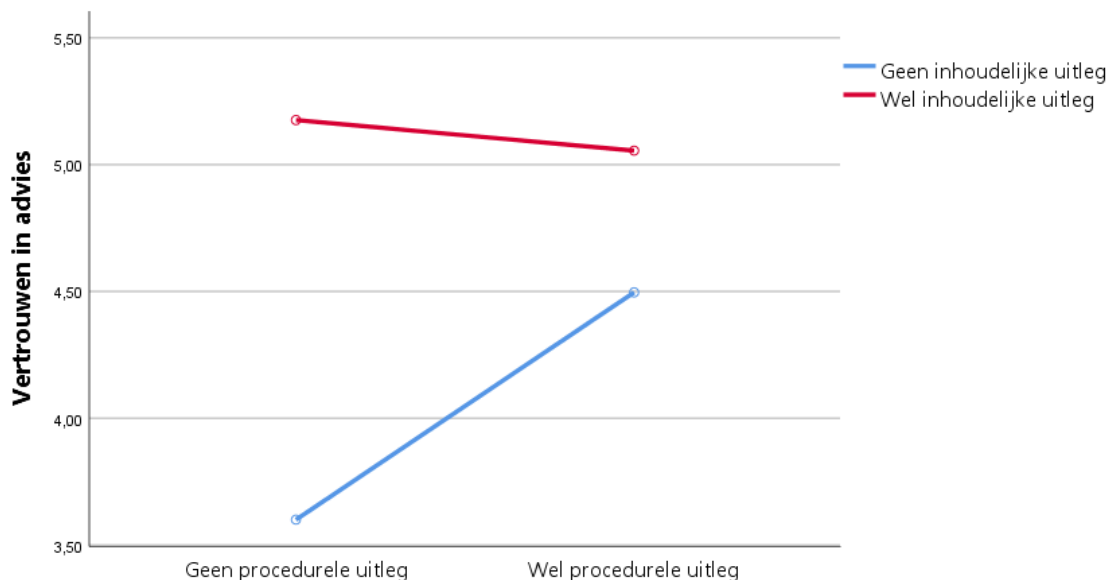
Verder kwam uit de analyse naar voren dat er sprake is van een significant interactie-effect. Dit houdt in dat het effect van een onafhankelijke variabele (procedurele uitleg) op de afhankelijke variabele (vertrouwen in advies) afhangt van een andere onafhankelijke variabele (inhoudelijke uitleg) (Field, 2009, p. 788).

¹ De gemiddelden in de beschrijvende statistiek en toetsende statistiek wijken af van elkaar omdat ze via verschillende toetsen berekend zijn. Voor de beschrijvende statistiek is een one-way ANOVA uitgevoerd, waarbij de gemiddelden per groep berekend worden. Bij de toetsende statistiek is gebruik gemaakt van 2x2 factoriaal ANOVA, waarbij de gemiddelden per factor worden berekend.

Het effect van een procedurele uitleg op het vertrouwen van respondenten in het advies van de slimme keuzehulp hangt af van of ze een inhoudelijke uitleg gehad hebben, $F(1, 532) = 21,38$; $p < 0,001$; $\eta_p^2 = 0,04$. In figuur 5.1 is deze interactie weergegeven.

De figuur laat zien dat een inhoudelijke uitleg een enorm groot effect heeft op vertrouwen. Als daar een procedurele uitleg aan toegevoegd wordt, zorgt dit niet voor meer vertrouwen. Indien een inhoudelijke uitleg ontbreekt, dan zorgt de aanwezigheid van een procedurele uitleg wel voor een hogere mate van vertrouwen in het advies van de slimme keuzehulp dan wanneer er geen uitleg is. Verder is te zien in de figuur dat bij een procedurele uitleg het vertrouwen in het advies verhoogd kan worden door een inhoudelijke uitleg toe te voegen.

Figuur 5.1: Het effect van een procedurele uitleg en een inhoudelijke uitleg op het vertrouwen in het advies van de slimme keuzehulp.



Om de hypothesen te toetsen, is een vervolganalyse uitgevoerd aan de hand van contrasten. Op basis van de vier hypothesen zijn contrasten gespecificeerd (zie bijlage 8.3). Hieruit komt ten eerste naar voren dat respondenten met een procedurele uitleg een significant hogere mate van vertrouwen in het advies van de slimme keuzehulp hebben dan respondenten die geen uitleg ontvingen, $t(532) = 5,76$; $p < 0,001$. Dit ondersteunt hypothese 1.

Ten tweede toonden de contrasten aan dat respondenten die een inhoudelijke uitleg ontvingen een significant hogere mate van vertrouwen in het advies van de slimme keuzehulp hebben dan respondenten die geen uitleg ontvingen, $t(532) = 10,19$; $p < 0,001$. Dit resultaat ondersteunt hypothese 2.

Ten derde kwam naar voren dat respondenten met een procedurele uitleg significant minder vertrouwen hadden in het advies van de slimme keuzehulp dan respondenten met een inhoudelijke uitleg, $t(532) = -4,40$; $p < 0,001$. Dit resultaat komt niet overeen met hypothese 3, waarbij juist verwacht werd dat een procedurele uitleg meer effect heeft op het vertrouwen in een besluit dan een inhoudelijke uitleg. Hypothese 3 wordt daarom verworpen.

Tot slot lieten de contrasten zien dat er geen significante verschillen zijn tussen respondenten die een combinatie-uitleg ontvingen (procedureel en inhoudelijk) en respondenten die ofwel een procedurele ofwel een inhoudelijke uitleg ontvingen, $t(532) = -1,62$; $p = 0,11$. Ook hypothese 4 wordt op basis van deze analyse verworpen. Tabel 5.2 vat de uitkomsten van de toetsing van de hypothesen middels contrasten samen.

Tabel 5.2: Toetsing hypothesen voor 'vertrouwen in advies'

Hypothese	Resultaat
H ₁ : Een procedurele uitleg heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.	Aangenomen
H ₂ : Een inhoudelijke uitleg heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.	Aangenomen
H ₃ : Een procedurele uitleg heeft een sterker positief effect op vertrouwen van burgers in algoritmische beslissingen dan een inhoudelijke uitleg.	Verworpen
H ₄ : Een procedurele uitleg gecombineerd met een inhoudelijke uitleg heeft een negatief effect ten opzichte van enkel een procedurele of een inhoudelijke uitleg op vertrouwen van burgers in algoritmische beslissingen.	Verworpen

Deelconclusie

Het antwoord op de eerste deelvraag "Wat is de invloed van uitlegbaarheid op het vertrouwen van gebruikers in het advies van de slimme keuzehulp?" is driedig. Ten eerste is vastgesteld dat 'een uitleg' meer vertrouwen genereert dan 'geen uitleg'. Ten tweede is uit de resultaten op te maken dat de aanwezigheid van een inhoudelijke uitleg zorgt voor een groter vertrouwen in het advies van de slimme keuzehulp dan een procedurele uitleg. Ten derde wijst het interactie-effect uit dat als er een inhoudelijke uitleg gegeven wordt, de aanwezigheid van een procedurele uitleg niet zorgt voor een toename aan vertrouwen. Het geven van een procedurele uitleg zorgt echter wel voor een toename in vertrouwen als er geen uitleg gegeven wordt. Het vertrouwen in het advies van de slimme keuzehulp door een procedurele uitleg kan tot slot vergroot worden door daar een inhoudelijke uitleg aan toe te voegen.

5.2 OP VERTROUWEN GEBASEERD GEDRAG

5.2.1 Bereidheid om naar advies te handelen

Een 2x2 factoriaal ANOVA is uitgevoerd om de mate van bereidheid om te handelen naar het advies van de slimme keuzehulp te vergelijken met respondenten die wel of geen procedurele uitleg kregen, gecombineerd met wel of geen inhoudelijke uitleg.

Beschrijvende statistiek

De beschrijvende statistiek ($N = 536$) van de afhankelijke variabele 'bereidheid om naar advies te handelen' is weergegeven in tabel 5.3.²

Tabel 5.3: Beschrijvende statistiek 'bereidheid om naar advies te handelen'

Experimentele groep	<i>M</i>	<i>SD</i>
Controlegroep (geen uitleg)	3,33	1,49
Treatment 1 (procedurele uitleg)	4,56	1,47
Treatment 2 (inhoudelijke uitleg)	5,23	1,22
Treatment 3 (procedurele en inhoudelijke uitleg)	5,10	1,31
Totaal	4,56	1,56

Toetsende statistiek

Het effect van een procedurele uitleg op de mate van bereidheid om te handelen naar het advies van de slimme keuzehulp was significant, $F(1, 532) = 21,62$; $p < 0,001$; $\eta_p^2 = 0,04$. Respondenten die geen procedurele uitleg kregen ($M = 4,40$; $SD = 1,66$) hadden een significant lagere gerapporteerde bereidheid om te handelen naar het advies van de slimme keuzehulp dan respondenten die wel een procedurele uitleg ontvingen ($M = 4,83$; $SD = 1,42$).

Ook het effect van een inhoudelijke uitleg op de mate van bereidheid om te handelen naar het advies van de slimme keuzehulp was significant, $F(1, 532) = 105,27$; $p < 0,001$; $\eta_p^2 = 0,17$. Respondenten die geen inhoudelijke uitleg kregen ($M = 3,95$; $SD = 1,60$) hadden significant minder bereidheid om te handelen naar het advies van de slimme keuzehulp dan respondenten die wel een inhoudelijke uitleg ontvingen ($M = 5,17$; $SD = 1,26$).

Hier kan, net zoals bij de *vertrouwende houding* het geval was, opgemerkt worden dat de effectgrootte van een inhoudelijke uitleg veel groter is (0,17) dan van een procedurele uitleg (0,04). Met een inhoudelijke uitleg kan dus 17% van de verschillen tussen groepen

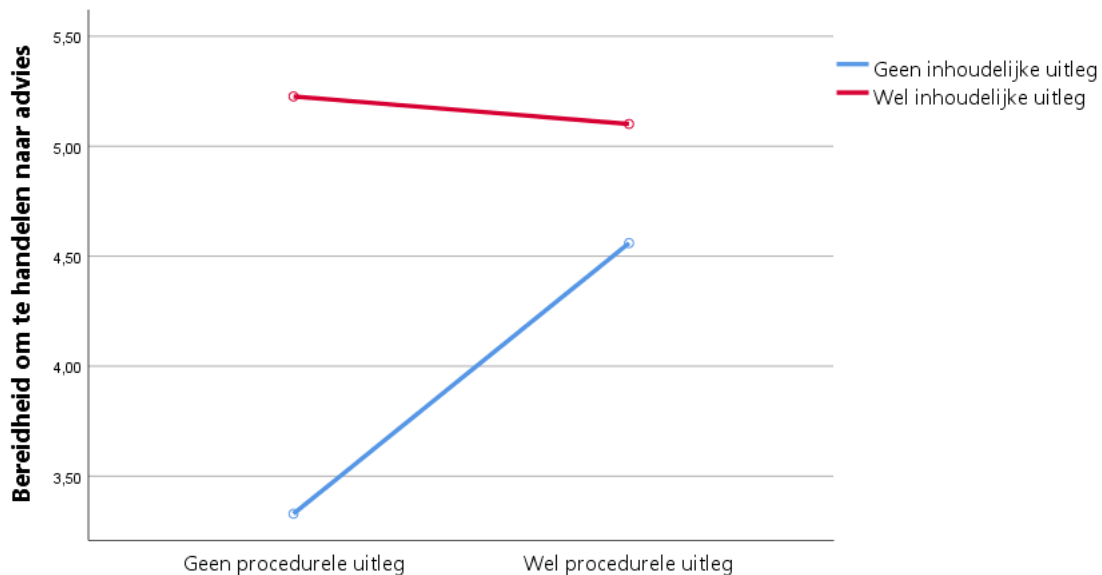
² De gemiddelden in de beschrijvende statistiek en toetsende statistiek wijken af van elkaar omdat ze via verschillende toetsen berekend zijn. Voor de beschrijvende statistiek is een one-way ANOVA uitgevoerd, waarbij de gemiddelden per groep berekend worden. Bij de toetsende statistiek is gebruik gemaakt van 2x2 factoriaal ANOVA, waarbij de gemiddelden per factor worden berekend.

in de mate van bereidheid om te handelen verklaard worden, terwijl dat bij een procedurele uitleg maar 4% is.

Tot slot kwam uit de analyse ook naar voren dat er sprake is van een significant interactie-effect. Het effect van een procedurele uitleg op de bereidheid van respondenten om te handelen naar het advies van de slimme keuzehulp hangt af van of ze een inhoudelijke uitleg gehad hebben, $F(1, 532) = 32,54$; $p < 0,001$; $\eta_p^2 = 0,06$. In figuur 5.2 is deze interactie weergegeven.

Het interactie-effect is vergelijkbaar met het effect dat gevonden was bij de *vertrouwende houding*. Uit de figuur komt weer naar voren dat een inhoudelijke uitleg een enorm groot effect heeft op de bereidheid van respondenten om te handelen naar het advies dat zij kregen. Als daar een procedurele uitleg aan toegevoegd wordt, zorgt dit niet voor meer bereidheid. Indien een inhoudelijke uitleg echter ontbreekt, dan zorgt de aanwezigheid van een procedurele uitleg wel voor een grotere mate van bereidheid om te handelen naar het advies van de slimme keuzehulp dan wanneer er geen uitleg is. Verder is te zien in de figuur dat bij een procedurele uitleg de bereidheid om te handelen naar het advies verhoogd kan worden door een inhoudelijke uitleg toe te voegen.

Figuur 5.2: Het effect van een procedurele uitleg en een inhoudelijke uitleg op de bereidheid om te handelen naar het advies van de slimme keuzehulp.



Om de hypothesen te kunnen toetsen, is weer een vervolganalyse uitgevoerd met behulp van contrasten die gespecificeerd zijn op basis van de vier hypothesen. Allereerst wijst deze analyse uit dat respondenten die een procedurele uitleg ontvingen, significant meer bereidheid hebben om te handelen naar het advies van de slimme keuzehulp dan

respondenten uit de controlegroep zonder uitleg, $t(264,87) = 6,80$; $p < 0,001$. Hiermee is ondersteuning gevonden voor hypothese 1.

Daarnaast laten de contrasten zien dat het geven van een inhoudelijke uitleg een significant hogere bereidheid tot stand brengt om te handelen naar het advies van de slimme keuzehulp dan de afwezigheid van een uitleg, $t(254,89) = 11,45$; $p < 0,001$. Dit resultaat ondersteunt hypothese 2.

Verder komt, in lijn met de bevindingen bij het concept 'vertrouwen in advies', naar voren dat respondenten die een inhoudelijke uitleg ontvangen significant meer bereid zijn om te handelen naar het advies dan respondenten die een procedurele uitleg ontvingen, $t(258,12) = -4,07$; $p < 0,001$. Dit strookt niet met de verwachting die opgesteld is in hypothese 3, waarbij juist uitgegaan werd van een sterker positief effect van een procedurele uitleg. Dit zorgt voor de verwerping van hypothese 3.

Ten slotte werd uit de vervolganalyse duidelijk dat er geen significante verschillen zijn tussen respondenten die een combinatie-uitleg ontvingen en respondenten die enkel een procedurele of inhoudelijke uitleg kregen, $t(262,68) = -1,48$; $p = 0,14$. Hypothese 4 wordt op basis van deze resultaten verworpen. In tabel 5.4 zijn de resultaten van de toetsing van de vier hypothesen samengevat.

Tabel 5.4: Toetsing hypothesen voor 'bereidheid om naar advies te handelen'

Hypothese	Resultaat
H ₁ : Een procedurele uitleg heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.	Aangenomen
H ₂ : Een inhoudelijke uitleg heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.	Aangenomen
H ₃ : Een procedurele uitleg heeft een sterker positief effect op vertrouwen van burgers in algoritmische beslissingen dan een inhoudelijke uitleg.	Verworpen
H ₄ : Een procedurele uitleg gecombineerd met een inhoudelijke uitleg heeft een negatief effect ten opzichte van enkel een procedurele of een inhoudelijke uitleg op vertrouwen van burgers in algoritmische beslissingen.	Verworpen

Deelconclusie

Het antwoord op de tweede deelvraag "Wat is de invloed van uitlegbaarheid op de bereidheid van gebruikers om naar het advies van de slimme keuzehulp te handelen?" toont grote gelijkennis met het antwoord op de eerste vraag. Ook hier is te zien dat 'een

uitleg' zorgt voor een grotere bereidheid om naar het advies van de slimme keuzehulp te handelen dan 'geen uitleg'. Wederom tonen de resultaten aan dat de aanwezigheid van een inhoudelijke uitleg een sterker positief effect heeft dan een procedurele uitleg. In lijn met het gevonden interactie-effect bij de *vertrouwende houding* kan hier ook geconcludeerd worden dat wanneer een inhoudelijke uitleg gegeven wordt, de toevoeging van een procedurele uitleg niet zorgt voor een toename aan bereidheid om naar het advies te handelen. Het geven van een procedurele uitleg zorgt echter wel voor een toename in bereidheid als er geen uitleg gegeven wordt. Tot slot is hier ook sprake van een verhoging in bereidheid om te handelen naar het advies van de slimme keuzehulp door aan een procedurele uitleg een inhoudelijke uitleg toe te voegen.

5.2.2 Daadwerkelijk gedrag gebruiker

Een binaire logistische regressie is uitgevoerd om te kijken of het soort uitleg dat respondenten kregen een voorspellend effect heeft op of respondenten al dan niet aangifte zouden doen. Deze analyse komt het meest in de buurt bij het meten van het daadwerkelijke gedrag van de gebruiker van de slimme keuzehulp op basis van het advies dat zij kregen.

Beschrijvende statistiek

De beschrijvende statistiek ($N = 536$) van de afhankelijke variabele 'daadwerkelijk gedrag gebruiker' is weergegeven in tabel 5.5.

Tabel 5.5: Beschrijvende statistiek 'daadwerkelijk gedrag gebruiker'

Experimentele groep	Geen aangifte (%)	Wel aangifte (%)
Controlegroep (geen uitleg)	36,1	63,9
Treatment 1 (procedurele uitleg)	67,9	32,1
Treatment 2 (inhoudelijke uitleg)	90,6	9,4
Treatment 3 (procedurele en inhoudelijke uitleg)	88,5	11,5
Totaal	70,9	29,1

Op dit moment kiest 72% van de gebruikers van de 'echte' slimme keuzehulp bij een afwijzend advies ('geen aangifte doen') ervoor om alsnog aangifte te doen (Soares, 2020). Deze beschrijvende statistiek laat zien dat bij alle drie de treatmentgroepen het percentage respondenten dat ervoor kiest om wel aangifte te doen bij het negatieve advies dat ze kregen, lager ligt dan 72%. Met de aanwezigheid van een inhoudelijke uitleg (bij treatmentgroepen 2 en 3) ligt dit percentage zelfs aanzienlijk lager, respectievelijk 9,4% en 11,5%. Dit geeft al een indicatie over de invloed van de verschillende soorten uitleg op het daadwerkelijke gedrag van de gebruiker van de slimme keuzehulp. In het volgende onderdeel zal dit statistisch getoetst worden.

Toetsende statistiek

Het omnibus model voor de logistische regressie analyse was significant, χ^2 ($df = 2$, $N = 536$) = 114,61; $p < 0,001$, Cox en Snell $R^2 = 0,19$; Nagelkerke $R^2 = 0,28$. Het model is 78% nauwkeurig in haar voorspellingen of gebruikers al dan niet aangifte gaan doen. Tabel 5.6 geeft de coëfficiënten voor de voorspellers van het model weer.

Tabel 5.6: Voorspellende coëfficiënten voor het model dat 'geen aangifte doen' voorspelt ($N = 536$)

	<i>b</i>	<i>SE (b)</i>	<i>p</i>	<i>Exp(B)</i> [95% BI]*
Constante	-0,36			
Procedurele uitleg	0,89	0,22	<0,001	2,44 [1,59, 3,73]
Inhoudelijke uitleg	2,16	0,24	<0,001	8,64 [5,39, 13,84]

*BI = betrouwbaarheidsinterval.

Zoals uit tabel 5.6 naar voren komt, zijn zowel een procedurele uitleg als een inhoudelijke uitleg significante voorspellers die het voorspellend vermogen van het model aanzienlijk verbeteren. De procedurele uitleg laat een oddsratio zien van 2,4. De odds op geen aangifte doen is dus 2,4 keer zo hoog voor respondenten die wel een procedurele uitleg hebben gehad dan voor respondenten zonder een procedurele uitleg. Dit geeft aanleiding tot het aannemen van hypothese 1. Ook de beschrijvende statistiek ondersteunt deze bevinding. Hierin is te zien dat 67,9% van de respondenten met een procedurele uitleg ervoor kiest om geen aangifte te doen, terwijl dat bij de controlegroep zonder uitleg maar 36,1% van de respondenten is.

De inhoudelijke uitleg laat een oddsratio zien van 8,6. Dit betekent dat de odds op het doen van geen aangifte 8,6 keer hoger is voor respondenten die wel een inhoudelijke uitleg hebben gehad dan voor respondenten zonder een inhoudelijke uitleg. Dit resultaat ondersteunt hypothese 2. De beschrijvende statistiek komt ook overeen met deze bevinding. 90,6% van de respondenten dat een inhoudelijke uitleg krijgt, kiest ervoor om geen aangifte te doen, ten opzichte van slechts 36,1% van de respondenten in de controlegroep.

Ook geeft het resultaat van de logistische regressieanalyse aan dat de voorspelde stijging van de kans dat iemand geen aangifte doet groter is bij een inhoudelijke uitleg dan bij een procedurele uitleg. Dit is niet in lijn met hypothese 3 en deze wordt daarom verworpen op basis van deze resultaten. Dit wordt ondersteund door de beschrijvende statistiek, waaruit blijkt dat 90,6% van de respondenten met een inhoudelijke uitleg ervoor kiest om geen aangifte te doen, tegenover 67,9% van de respondenten met een procedurele uitleg.

Aan de hand van de resultaten van de binaire logistische regressieanalyse kan geen uitspraak gedaan worden over hypothese 4. In de toets wordt geen interactie-effect onderzocht tussen de twee soorten uitleg, waardoor geen effecten van de combinatie-uitleg getoetst zijn. Uit de beschrijvende statistiek kan wel worden opgemaakt dat een vergelijkbare conclusie getrokken kan worden met de eerdere bevindingen over hypothese 4. Hieruit is af te leiden dat 88,5% van de respondenten met een combinatie-uitleg besloot om geen aangifte te doen op basis van het advies van de slimme keuzehulp. Dit is meer dan bij respondenten met een procedurele uitleg (67,9%), maar minder dan bij respondenten die een inhoudelijke uitleg kregen (90,6%). Tabel 5.7 vat de tussentijdse bevindingen van de toetsing van de hypothesen samen.

Tabel 5.7: Toetsing hypothesen voor 'daadwerkelijk gedrag gebruiker'

Hypothese	Resultaat
H ₁ : Een procedurele uitleg heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.	Aangenomen
H ₂ : Een inhoudelijke uitleg heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.	Aangenomen
H ₃ : Een procedurele uitleg heeft een sterker positief effect op vertrouwen van burgers in algoritmische beslissingen dan een inhoudelijke uitleg.	Verworpen
H ₄ : Een procedurele uitleg gecombineerd met een inhoudelijke uitleg heeft een negatief effect ten opzichte van enkel een procedurele of een inhoudelijke uitleg op vertrouwen van burgers in algoritmische beslissingen.	n.v.t.

Deelconclusie

Het antwoord op derde en laatste empirische deelvraag *“Wat is de invloed van uitlegbaarheid op het daadwerkelijke gedrag van gebruikers van de slimme keuzehulp?”* ligt ook in lijn met de antwoorden op de eerste twee deelvragen. Uit de binaire logistische regressie komt naar voren dat zowel een procedurele als een inhoudelijke uitleg een voorspellend vermogen hebben bij de vraag of een respondent al dan niet aangifte gaat doen. De aanwezigheid van een inhoudelijke uitleg voorspelt een significant grotere kans (8,6 keer zo groot) dat een respondent geen aangifte gaat doen dan de aanwezigheid van een procedurele uitleg (2,4 keer zo groot).

5.3 SLOTSOM

Indien de resultaten van de analyses van de onderdelen *vertrouwende houding* en *op vertrouwen gebaseerd gedrag* gecombineerd worden, ontstaat een samenhangend geheel. Resultaten van beide analyses tonen aan dat er een causale relatie bestaat tussen de uitlegbaarheid van algoritmische besluiten en het vertrouwen dat burgers in deze besluiten hebben. Hieronder worden de resultaten van de verschillende onderdelen gecombineerd bij de bespreking van de vier hypothesen.

H₁: Een procedurele uitleg heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.

Uit de twee factoriale ANOVA-toetsen van 'vertrouwen in advies' enerzijds en 'bereidheid om naar advies te handelen' anderzijds komt naar voren dat een procedurele uitleg meer vertrouwen genereert dan de afwezigheid van een uitleg. Ook de logistische regressie ondersteunt deze bevinding. De aanwezigheid van een procedurele uitleg voorspelt een hogere kans dat een gebruiker van de slimme keuzehulp geen aangifte gaat doen dan bij de afwezigheid van een procedurele uitleg. Hiermee is vastgesteld dat hypothese 1 aangenomen kan worden.

H₂: Een inhoudelijke uitleg heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.

Ook bij een inhoudelijke uitleg lieten de twee factoriale ANOVA-toetsen zien dat de aanwezigheid ervan meer vertrouwen genereert dan de afwezigheid van een uitleg. De resultaten van de logistische regressie onderschrijven deze bevinding. De aanwezigheid van een inhoudelijke uitleg voorspelt een hogere kans dat een gebruiker van de slimme keuzehulp geen aangifte gaat doen dan bij de afwezigheid van een inhoudelijke uitleg. Hypothese 2 wordt tevens aangenomen.

H₃: Een procedurele uitleg heeft een sterker positief effect op vertrouwen van burgers in algoritmische beslissingen dan een inhoudelijke uitleg.

Beide factoriale ANOVA-toetsen toonden aan dat een inhoudelijke uitleg zorgt voor een grotere mate van vertrouwen in het advies van de slimme keuzehulp dan een procedurele uitleg. Dit resultaat wordt ook ondersteund door de logistische regressie, waaruit naar voren komt dat een inhoudelijke uitleg een hogere kans voorspelt dat een gebruiker geen aangifte gaat doen dan een procedurele uitleg. Deze bevindingen zijn tegenovergesteld aan de verwachting die is opgesteld. Hypothese 3 wordt daarom verworpen.

H₄: Een procedurele uitleg gecombineerd met een inhoudelijke uitleg heeft een negatief effect ten opzichte van enkel een procedurele of een inhoudelijke uitleg op vertrouwen van burgers in algoritmische beslissingen.

De factoriale ANOVA-toetsen lieten geen significante resultaten zien tussen de mate van vertrouwen dat respondenten met een combinatie-uitleg hadden, en de respondenten die ofwel een procedurele ofwel een inhoudelijke uitleg ontvingen. Dit leidt tot het verwerpen van hypothese 4. Zoals eerder toegelicht, kon met de binaire logistische regressie geen uitspraak gedaan worden over hypothese 4. In tabel 5.8 wordt een overzicht gegeven van de toetsing van de vier hypothesen per vertrouwensdimensie.

Met betrekking tot de vierde hypothese kunnen uit de factoriale ANOVA-toetsen nog wel andere relevante bevindingen afgeleid worden (zie ook figuur 5.1 en 5.2). Deze hebben te maken met de eerder besproken interactie-effecten die optreden. Als de respondent een inhoudelijke uitleg ontvangt, dan voegt een procedurele uitleg niets meer toe. Het vertrouwen in het advies van de slimme keuzehulp wordt dan niet groter. Als de respondent echter geen uitleg ontvangt, dan heeft de toevoeging van een procedurele uitleg wel een positief effect op de mate van vertrouwen in het advies van de slimme keuzehulp. Tot slot laten de interactie-effecten ook zien dat wanneer een respondent een procedurele uitleg ontvangt, de toevoeging van een inhoudelijke uitleg ook zorgt voor een toename in vertrouwen. Dit grotere positieve effect van een inhoudelijke uitleg ten opzichte van een procedurele uitleg komt ook naar voren in de beantwoording van de drie deelvragen.

Tabel 5.8: Toetsing hypothesen

Hypothese	<i>Vertrouwende houding</i>	<i>Op vertrouwen gebaseerd gedrag</i>	
	Vertrouwen	Bereidheid	Gedrag
H ₁ : Procedurele uitleg > geen uitleg	Aangenomen	Aangenomen	Aangenomen
H ₂ : Inhoudelijke uitleg > geen uitleg	Aangenomen	Aangenomen	Aangenomen
H ₃ : Procedurele uitleg > inhoudelijke uitleg	Verworpen	Verworpen	Verworpen
H ₄ : Combinatie-uitleg < inhoudelijke of procedurele uitleg	Verworpen	Verworpen	n.v.t.

6 Conclusie en discussie

In het resultatenhoofdstuk zijn de bevindingen over het effect van verschillende soorten uitleg op het vertrouwen van burgers in het advies van de slimme keuzehulp op een rij gezet. In dit hoofdstuk zal aan de hand van de gevonden resultaten een antwoord geformuleerd worden op de vraag: *“Wat is het effect van uitlegbaarheid van algoritmische besluiten op het vertrouwen van burgers in deze besluiten?”*. Daarna zullen de wetenschappelijke en maatschappelijke implicaties van dit onderzoek worden besproken.

6.1 AI, IS DAT WEL TE VERTROUWEN?

In de context van ‘mensgerichte AI’, moet verantwoord omgegaan worden met het gebruik van algoritmen bij publieke besluitvormingsprocessen. Om vertrouwen in algoritmische besluiten te waarborgen, moeten deze besluiten transparant zijn (AI HLEG, 2019, p. 22). Begrijpelijke uitleg zou een oplossing zijn om die transparantie te bewerkstelligen. Dit sluit aan bij de roep om ‘uitlegbare AI’, waarbij steeds meer mensen bepleiten dat begrijpelijke uitleg noodzakelijk is om verantwoord door te kunnen gaan met het gebruik van algoritmische besluitvormingsprocessen (Diakopoulos, 2016; Kizilcec, 2016; Kroll et al., 2017).

In de publieke sector was het effect van uitlegbaarheid op het vertrouwen in algoritmische besluiten echter nog niet eerder empirisch onderzocht. Dit onderzoek bekijkt aan de hand van een surveyexperiment wat voor effect verschillende soorten uitleg hebben op het vertrouwen van burgers in het advies van de slimme keuzehulp. De resultaten zijn opvallend consistent. Wat voor uitleg iemand krijgt over een algoritmische beslissing heeft een groot effect op hoeveel vertrouwen deze persoon heeft in de beslissing.

Procedurele uitleg: klein, maar significant positief effect op vertrouwen

Een procedurele uitleg heeft van de drie treatmentgroepen het kleinste effect op het vertrouwen in het algoritmische besluit. Een uitleg over hoe de beslissing tot stand is gekomen, zorgt voor enige mate van vertrouwen in het algoritmische besluit. Dit vertrouwen kan verhoogd worden door aan zo’n procedurele uitleg nog een inhoudelijke uitleg toe te voegen. Wanneer geen inhoudelijke uitleg gegeven kan worden, om wat voor reden dan ook, laten de resultaten zien dat het alsnog relevant is om een procedurele uitleg te geven. Het is dan voor het waarborgen van vertrouwen in een algoritmische beslissing wenselijk om wel toe te lichten hoe het besluitvormingsproces in elkaar zit. Een uitleg over hoe de algoritmische beslissing tot stand is gekomen, levert namelijk meer vertrouwen op in zo’n besluit dan wanneer er geen uitleg is.

Inhoudelijke uitleg: groot, positief effect op vertrouwen

Het grootste effect op de mate van vertrouwen in een algoritmische beslissing is gevonden bij een inhoudelijke uitleg. Wanneer mensen snappen waarom zij in hun specifieke geval het advies krijgen om geen aangifte te doen, hebben zij het meeste vertrouwen in het advies van de slimme keuzehulp. Bij iemand met een inhoudelijke uitleg, kan het vertrouwen in de algoritmische beslissing niet verhoogd worden door het toevoegen van een procedurele uitleg. Het toevoegen van extra uitleg doet juist afbreuk aan het vertrouwen in een besluit.

Inhoudelijke uitleg maakt procedurele uitleg overbodig

Een uitleg waarbij zowel wordt ingegaan op hoe de beslissing tot stand is gekomen (procedure) en waarom er in dat specifieke geval zo besloten wordt (inhoud), levert ook een hoge mate van vertrouwen op in een algoritmische beslissing. Dit is niet zo hoog als bij enkel een inhoudelijke uitleg, maar wel hoger dan bij een procedurele of geen uitleg. Een inhoudelijke uitleg overschaduwet dus een procedurele uitleg. Indien een inhoudelijke uitleg gegeven kan worden, is een procedurele uitleg overbodig in het verhogen van het vertrouwen dat mensen hebben in een algoritmische beslissing.

AI is wel te vertrouwen!

Dit onderzoek heeft bevestigd dat transparantie bijdraagt aan het vertrouwen in een algoritmische beslissing. Een uitleg bij een beslissing leidt altijd tot meer vertrouwen in een besluit dan geen uitleg. Het vertrouwen wordt vergroot wanneer men uitleg krijgt over de redenering achter de beslissing, via een inhoudelijke uitleg. Voor het vertrouwen in de beslissing is het minder van belang om te weten hoe het besluitvormingssysteem werkt. Mensen willen weten waarom zij in hun specifieke geval het advies krijgen om geen aangifte te doen. Kortom, om verantwoord door te kunnen gaan met het gebruik van algoritmen bij publieke besluitvormingsprocessen is het van belang dat er -waar mogelijk- altijd een inhoudelijke uitleg wordt gegeven over de beslissing die de burger te zien krijgt. Immers: het effect van een inhoudelijke uitleg op het vertrouwen van burgers in algoritmische besluiten is erg groot. Een inhoudelijke uitleg haalt dus de meeste twijfel weg.

6.2 ACADEMISCHE IMPLICATIES

Eerst wordt de wetenschappelijke waarde van dit onderzoek toegelicht. Daarna is er aandacht voor methodologische beperkingen, gevolgd door enkele aanbevelingen voor vervolgonderzoek.

6.2.1 Wetenschappelijke waarde

Uitlegbare AI in de publieke sector

De belofte van AI in de publieke sector is groot. Niet alleen in Nederland, maar wereldwijd gebruiken overheden steeds meer AI-toepassingen bij de uitvoering van verschillende taken (Mols, 2019). Met de toenemende voordelen die AI-toepassingen hebben, zoals een effectievere en efficiëntere dienstverlening (Lepri et al., 2018), groeit ook de aandacht naar verantwoord gebruik van deze toepassingen. Omdat empirisch onderzoek naar uitlegbare AI in de publieke sector ontbrak, is in dit onderzoek voortgebouwd op algemene kennis over de relatie tussen de uitlegbaarheid van algoritmische beslissingen en het vertrouwen in deze beslissingen. Zoals uit de theorie naar voren kwam, wordt deze specifieke relatie in het onderzoek van Kizilcec (2016) empirisch getoetst. Zijn bevinding dat een procedurele uitleg het meeste vertrouwen zou genereren, moet genuanceerd worden in de context van de publieke sector.

Dit onderzoek toont namelijk aan dat niet een procedurele, maar juist een inhoudelijke uitleg zorgt voor het meeste vertrouwen. Deze constatering is in lijn met het werk van De Fine Licht en De Fine Licht (2020), die op basis van hun literatuuronderzoek stellen dat een inhoudelijke motivering van een algoritmische beslissing leidt tot een toename aan in hoeverre burgers die beslissing legitiem en betrouwbaar achten. Zij beargumenteren dat niet een procedurele uitleg, maar juist een inhoudelijke uitleg de burger het gevoel geeft dat de overheid goede intenties heeft en de burger respecteert door specifiek in te gaan op de motivering van een bepaalde beslissing.

Aan de hand van de resultaten van dit onderzoek kan nu ook empirisch met zekerheid worden gezegd dat een inhoudelijke uitleg het meeste effect heeft op het vertrouwen in een algoritmische beslissing in de publieke sector. Door theorieën uit de computerwetenschappen samen te voegen met literatuur uit de bestuurskunde, kon dit empirische onderzoek inzicht creëren in uitlegbare AI in de publieke sector. Voor de 'digitale overheid van de toekomst' blijft het relevant om inzichten uit deze vakgebieden te combineren om verantwoord gebruik van AI-toepassingen binnen de publieke sector te waarborgen (Kankanhalli, Charalabidis & Mellouli, 2019; Mergel, Rethemeyer & Isett, 2016).

De kracht van experimenteel onderzoek

In dit onderzoek is voor het eerst aan de hand van experimentele methoden gekeken naar het effect van uitlegbaarheid op vertrouwen in algoritmische beslissingen in de publieke sector. Dit heeft op twee manieren bijgedragen aan de wetenschappelijke waarde van dit onderzoek. Ten eerste zorgt de gehanteerde methode voor nieuwe inzichten in het vraagstuk van uitlegbare AI in de publieke sector. Experimenteel onderzoek is relatief nieuw in de bestuurskunde (Margretts, 2011). Aan de hand van experimenten kunnen causale relaties onderzocht worden waardoor theorieën verder ontwikkeld kunnen worden (Bouwman & Grimmelikhuijsen, 2016). Dat was hier ook het geval. Voor dit onderzoek is de slimme keuzehulp nagemaakt in het surveyexperiment. Respondenten doorliepen alle stappen van de 'echte' slimme keuzehulp. Vervolgens werd gesimuleerd dat de slimme keuzehulp op basis van algoritmen tot een advies kwam. Deze realistische manipulatie is een groot pluspunt van dit experiment, omdat inzicht kon worden opgedaan in de causale relatie tussen uitlegbaarheid en vertrouwen, waardoor theorieën omtrent uitlegbare AI in de publieke sector verder ontwikkeld kunnen worden.

Ten tweede is het doen van experimenteel onderzoek erg geschikt voor het onderzoeken van het gedrag van individuen, maar het wordt nog weinig gebruikt in de bestuurskunde (Bozeman & Scott, 1992; Grimmelikhuijsen, Jilke, Olsen & Tummers, 2017). Bestuurskundig onderzoek spitst zich voornamelijk toe op het niveau van organisaties of specifieke groepen mensen en niet op het niveau van het individu (Margretts, 2011). Grimmelikhuijsen, Jilke, Olsen en Tummers (2017) laten juist het belang zien van de gedragsbestuurskunde, waarin psychologie en bestuurskunde gecombineerd worden. Hiermee kunnen interessante inzichten opgedaan worden over de houding en het gedrag van individuen (Grimmelikhuijsen et al., 2017, p.46). In dit onderzoek is het op vertrouwen gebaseerd gedrag van de respondent onderzocht aan de hand van een surveyexperiment. Door de respondent te laten kiezen om wel of geen aangifte te doen op basis van het advies van de slimme keuzehulp, kon zijn of haar gedrag onderzocht worden. De resultaten tonen aan dat mensen die een specifieke uitleg krijgen bij het algoritmische besluit, minder snel aangifte zullen doen bij een afwijzend advies van de slimme keuzehulp. Aan de hand van experimenteel onderzoek kon dus het daadwerkelijke gedrag van de gebruiker van de slimme keuzehulp (wel of geen aangifte doen) gemeten worden, waardoor nieuwe inzichten zijn opgedaan in de gedragsbestuurskunde.

Reflectie op de hypothesen

Uit de literatuur kwamen vier hypothesen naar voren, die getoetst zijn in het surveyexperiment. Uit de resultaten bleek dat de twee verwachtingen die gebaseerd waren op het concept van procedurele rechtvaardigheid geen stand hielden. Zo werd vanuit de theorie verwacht dat een procedurele uitleg een sterker effect zou hebben op

het vertrouwen in een algoritmische beslissing, vanwege de elementen in de uitleg die ingingen op hoe het beslissingsproces werkt. Hierdoor zouden mensen het gevoel moeten hebben dat het proces eerlijk verlopen is, waardoor ze vertrouwen zouden hebben in het besluit en zich neer zouden leggen bij de uitkomst. Hoe kan het dat die verwachting niet opging in dit onderzoek?

Een eerste mogelijke verklaring hiervoor kan worden gevonden in de resultaten van de manipulatiecheck. Hieruit kan afgeleid worden dat respondenten die een procedurele uitleg ontvingen, vonden dat ze minder (goede) uitleg kregen dan respondenten uit de inhoudelijke of de combinatiegroep. Zij ervoeren over het algemeen dus minder uitleg. Het is mogelijk dat de lagere mate van vertrouwen in algoritmische beslissingen bij een procedurele uitleg dus voortkomt uit het feit dat deze uitleg beschouwd werd als 'minder uitleg' dan de uitleg van de twee andere behandelgroepen. Dit komt overeen met het op cognitie gebaseerd vertrouwen, dat besproken is in het theoretisch kader. Dit wordt namelijk beïnvloed door de mate waarin een burger ervaart dat de uitkomsten van een intelligent beslissysteem gemotiveerd zijn (Madsen & Gregor, 2000; Nothdurft et al., 2014). Een afname in op cognitie gebaseerd vertrouwen zorgt voor een afname in het vertrouwen dat iemand heeft in de besluiten van een intelligent beslissysteem. In dit onderzoek is dit ook waargenomen: als de respondent ervaart dat hij of zij minder uitleg krijgt (en de motivering van het besluit minder sterk vindt), daalt zijn of haar vertrouwen in het advies van de slimme keuzehulp.

Een tweede mogelijke verklaring heeft betrekking op het al dan niet verschaffen van een handelingsperspectief in een uitleg. In een handelingsperspectief worden concrete instrumenten genoemd die mensen sturing kunnen geven (Ministerie van Algemene Zaken, 2015, p. 16). In de inhoudelijke uitleg in dit onderzoek werd aan mensen een handelingsperspectief geboden. Men werd namelijk verteld dat ze naar de onafhankelijke geschillencommissie konden stappen van het Keurmerk waaraan de webshop verbonden was en dat ze op de website van de webshop konden kijken om erachter te komen hoe ze hun geld terug konden krijgen. Er werd daarentegen geen handelingsperspectief gegeven bij de procedurele uitleg. Hierin stond hoe de uitleg van de slimme keuzehulp tot stand is gekomen, maar geen vervolgstappen die mensen konden ondernemen. Op de open vragen in de vragenlijst gaven enkele respondenten met een procedurele uitleg aan dat ze ervoor kozen om alsnog aangifte te doen omdat ze niet wisten wat ze anders moesten doen. Het ontbreken van een handelingsperspectief bij de procedurele uitleg heeft dus afbreuk gedaan aan de kracht van het gevoel van procedurele rechtvaardigheid. In de volgende paragraaf wordt hier nog meer aandacht aan besteed door in te gaan op hoe dit in vervolgonderzoek anders aangepakt zou kunnen worden.

De andere hypothese die niet aangenomen kon worden, stelde dat een combinatie-uitleg minder vertrouwen zou realiseren dan enkel een procedurele of een inhoudelijke

uitleg, vanwege het risico op een *information overload*. De resultaten laten zien dat de aanwezigheid van een inhoudelijke uitleg altijd voor meer vertrouwen zorgt dan bij de afwezigheid hiervan. Een combinatie-uitleg zorgt dus voor meer vertrouwen dan een procedurele uitleg. Aangenomen kan worden, omdat het concept van procedurele rechtvaardigheid niet het verwachte effect had zoals hierboven beschreven is, dat de vierde hypothese die hier op voortbouwt om dezelfde reden geen standhoudt. Wel is het zo dat enkel een inhoudelijke uitleg meer vertrouwen geeft dan een combinatie-uitleg (waarin ook een inhoudelijke uitleg verwerkt zit). Dit zou verklaard kunnen worden aan de hand van het feit dat meer transparantie (en dus uitleg) niet altijd voor meer vertrouwen zorgt, zoals ook uit de theorie naar voren kwam (Birkinshaw & Cable, 2017; Hosseini, Shariri, Phalp & Ali, 2015). Teveel transparantie kan namelijk afbreuk doen aan het vertrouwen dat men heeft in een algoritmische beslissing (De Fine Licht & De Fine Licht, 2020; Kizilcec, 2016). Dat was ook te zien in dit onderzoek, waarbij meer uitleg dan enkel een inhoudelijke motivering zorgde voor een afname in vertrouwen.

6.2.2 Beperkingen en aanbevelingen voor vervolgonderzoek

Externe validiteit

De belangrijkste methodologische beperking van dit onderzoek heeft te maken met de externe validiteit. Dit gaat over de generaliseerbaarheid van de resultaten buiten de onderzoekscontext (Seltman, 2012, p. 201). De resultaten van dit onderzoek zouden generaliseerbaar moeten zijn naar de hele populatie, in dit geval alle inwoners van Nederland. Dat kan door de samenstelling van de steekproef echter niet gedaan worden. De steekproef is namelijk niet representatief voor de populatie.

De man/vrouw-verhouding en de gemiddelde leeftijd van de steekproef wijken maar enigszins af van de verhoudingen in de werkelijke populatie. Van de respondenten in dit onderzoek was 42,7% man, 56,2% vrouw en was de gemiddelde leeftijd 40 jaar. In Nederland is 49,7% man, 50,4% vrouw en is de gemiddelde leeftijd 42 jaar (CBS, 2019b). De verdeling naar hoogst behaald opleidingsniveau in dit onderzoek is daarentegen zorgelijker voor de representativiteit van de steekproef. Van de respondenten in dit onderzoek is 2,8% laag, 16,6% middelbaar en 80,4% hoogopgeleid. Dit is niet vergelijkbaar met de verdeling in de populatie. In Nederland is namelijk 27,9% laag, 38,1% middelbaar en 32,5% hoogopgeleid (Onderwijs in Cijfers, z.d.).

Een verklaring voor dit grote verschil is te vinden in het feit dat in dit onderzoek gebruik is gemaakt van *convenience sampling*. Een groot deel van de respondenten komt uit het netwerk van (vrienden en familie van) de onderzoeker, die veelal uit een hoogopgeleid milieu komen. De vraag is of het verschil tussen de steekproef en populatie in hoogst behaald opleidingsniveau problematisch is. Om dit uit te zoeken zijn de analyses opnieuw uitgevoerd waarbij enkel de laag en middelbaar opgeleide respondenten zijn meegenomen (zie bijlage 8.5.4). De resultaten van deze analyses vertonen vergelijkbare

patronen met de bevindingen in het resultatenhoofdstuk. Hierdoor kan aangenomen worden dat het verschil niet problematisch is.

Ondanks dat het verschil niet problematisch blijkt, wordt toch aanbevolen voor vervolgonderzoek om meer aandacht te besteden aan de representativiteit van de steekproef, zodat de resultaten beter generaliseerbaar zijn naar de populatie. Dit kan door gebruik te maken van een 'respondenten *pool*' waaruit naar verhouding de juiste aantallen respondenten per hoogst behaald opleidingsniveau gehaald kunnen worden.

Ecologische validiteit

Een andere methodologische beperking heeft betrekking op de ecologische validiteit van dit onderzoek. Ecologische validiteit gaat over in hoeverre de bevindingen uit dit onderzoek generaliseerbaar zijn naar andere alledaagse situaties (Bryman, 2016, p. 47). In dit geval waren de respondenten geen echte slachtoffers van internetfraude. De vraag doet zich dan voor in hoeverre hun beleving van het doorlopen van de slimme keuzehulp overeenkomt met die van 'echte' slachtoffers. Het is voor te stellen dat bij echte slachtoffers meer emoties meespelen dan bij de respondenten in dit onderzoek.

Uit het onderzoek bleek echter dat respondenten zich erg goed in konden leven in de hypothetische situatie. Gemiddeld scoorden zij een 5,8 op een schaal van 7. Dit geeft aan dat ze zich goed konden voorstellen hoe het voelt om slachtoffer te zijn van internetfraude. Dit is te verklaren aan de hand van het feit dat het steeds normaler is geworden om via internet aankopen te doen. In 2019 kocht namelijk 79,4% van de Nederlanders online één of meerdere aankopen (CBS, 2019a). Dit is sinds de coronacrisis toegenomen, waarmee ook een enorme groei te zien was in internetfraude (Voskuil, 2020). Mensen mochten de deur niet meer uit, waardoor het aantrekkelijker werd om aankopen online te doen. Kortom, die hypothetische situatie waarin respondenten zich moesten inleven, was geen ver-van-hun-bed-show.

Ondanks dat het aannemelijk is dat respondenten zich goed konden voorstellen hoe het is om een 'echt' slachtoffer te zijn van internetfraude, is het interessant om te kijken hoe daadwerkelijke slachtoffers de verschillende soorten uitleg ervaren. In de tijdspanne van dit onderzoek was het niet mogelijk om het experiment uit te voeren bij 'echte' slachtoffers. In vervolgonderzoek is het daarom raadzaam om genoeg tijd uit te trekken om het onderzoek via de politie uit te zetten bij mensen die de slimme keuzehulp invullen omdat zij menen slachtoffer te zijn van internetfraude.

Handelingsperspectief

Een derde beperking gaat over het ontbreken van een handelingsperspectief bij de procedurele uitleg. Zoals hierboven al toegelicht is, was de inhoudelijke uitleg (en daarmee ook de combinatie-uitleg) voorzien van alternatieve stappen die respondenten konden ondernemen, anders dan aangifte doen. Dit kan invloed gehad hebben op het feit dat respondenten de procedurele uitleg als 'minder uitleg' ervoeren. Hen werd

namelijk niet verteld welke andere handelingsalternatieven zij hadden. Het is daarom relevant om in vervolgonderzoek te kijken naar de effecten van zo'n handelingsperspectief. Daarnaast is het voor vervolgonderzoek van belang dat de experimentele groepen gelijk worden getrokken, dus als één groep handelingsalternatieven in de uitleg te zien krijgt, moet dit ook voor de andere groepen gelden.

Geschonden verwachtingen

In dit onderzoek is enkel onderzocht hoe respondenten waarvan de verwachting geschonden is, reageren op verschillende soorten uitleg. Hun verwachting, dat zij het advies van de slimme keuzehulp zouden krijgen om wél aangifte te doen, werd niet waargemaakt. Zij kregen namelijk het advies om geen aangifte te doen. Dit is gebaseerd op het eerder besproken concept van *procedurele rechtvaardigheid* (Tyler, 1990). Als een besluit niet uitpakt zoals iemand verwacht, dan draagt een uitleg over het besluitproces bij aan het gevoel dat het proces in ieder geval rechtvaardig is geweest. Mensen accepteren dan sneller het besluit dat zij niet hadden zien aankomen. Voortbouwend op dit concept was in onderzoeken over het vertrouwen in algoritmische besluiten te zien dat enkel wanneer een verwachting van iemand geschonden wordt, het uitmaakt wat voor soort uitleg gegeven wordt (Kizilcec, 2016; Rader et al., 2018). Als de verwachting van een individu niet wordt geschonden, dan zou de mate van transparantie niet uitmaken. In vervolgonderzoek is het interessant om te kijken of dat ook het geval is bij besluiten in de publieke sector, die direct impact (kunnen) hebben op het leven van individuen. Hiervoor moet naast een casus waarin de verwachting van de respondent wordt geschonden (zoals dat in dit onderzoek het geval was), ook een casus opgenomen waarin de verwachting van de respondent overeenkomt met het daadwerkelijke advies van de slimme keuzehulp. Dan kan onderzocht worden of een uitleg überhaupt effect heeft op het vertrouwen van respondenten waarvan de verwachting overeenkomt met het advies, en zo ja, of die effecten vergelijkbaar zijn met de in dit onderzoek gevonden resultaten.

Onderzoeksmethode

Omdat de onderzoeksvraag een causale relatie bevat, zijn in dit onderzoek enkel experimentele methoden gebruikt. Uit de kwantitatieve data komen interessante bevindingen naar voren met betrekking tot het effect van uitlegbaarheid op vertrouwen. Om meer diepgaande kennis op te doen over waarom respondenten een inhoudelijke uitleg betrouwbaarder ervaren dan een procedurele uitleg, is het interessant om kwalitatieve data te verzamelen. Dit onderzoek kan daarom als kapstok dienen voor niet-experimenteel vervolgonderzoek. Uit interviews kunnen overwegingen van respondenten onderzocht worden met betrekking tot de mate van vertrouwen dat zij hebben in de uitleg van algoritmische beslissingen. Dat zou waardevol zijn voor de wetenschap, omdat zo meer inzicht verkregen wordt in uitlegbare AI in de publieke sector.

6.3 MAATSCHAPPELIJKE IMPLICATIES

Dit onderzoek sluit af door de maatschappelijke waarde van dit onderzoek te bespreken en enkele aanbevelingen te doen voor de werkpraktijk.

6.3.1 Maatschappelijke waarde

De toenemende impact van AI-toepassingen in de publieke sector is enorm. Stolze (2018) spreekt van 'algoritmisering', omdat algoritmen niet meer weg te denken zijn uit onze huidige samenleving. Zowel op Europees als op nationaal niveau wordt gepleit voor 'verantwoorde mensgerichte AI' om het vertrouwen van de burger in de digitale overheid te vergroten (Rijksoverheid, 2019). Transparantie over het gebruik en de werking van AI staan daarbij centraal. Dit onderzoek onderstreept het belang van mensgerichte AI voor het vertrouwen van burgers in algoritmische beslissingen van de overheid.

Het blijkt namelijk dat mensen graag een motivering willen over waarom in hun specifieke geval een bepaald besluit genomen wordt. De uitleg die gegeven moet worden over algoritmische beslissingen moet daarom mensgericht zijn. Dit geeft reden tot nadenken over beslissingen waarbij geen individuele motivering gegeven kan worden. Als de motivering van een beslissing zo complex is dat het niet uitlegbaar is op individueel niveau, moet er goed nagedacht worden of deze AI-toepassing wel wenselijk is.

Hier kan bijvoorbeeld sprake van zijn bij het gebruik van zogenoemde *black box* modellen, waarbij de input en de output van een algoritme bekend zijn, maar het onduidelijk is hoe het beslissingsproces verloopt. Hierdoor is de beslissing zowel moeilijk te voorspellen als uit te leggen. Deze ondoorzichtigheid wordt vergroot door het feit dat algoritmen vaak geheim worden gehouden. Vanwege commerciële of veiligheidsredenen maken overheden de algoritmen die ten grondslag liggen aan het besluitvormingsproces niet openbaar (Vetzo & Gerards, 2019, p. 12).

De resultaten van dit onderzoek laten zien dat dit schadelijk kan zijn voor het vertrouwen van burgers in algoritmische besluiten. Als we verder willen met het gebruik van algoritmen bij besluitvormingsprocessen, is het van belang dat er goed wordt nagedacht over wat soort algoritmen we willen gebruiken in de publieke sector. Sterke kunstmatige intelligentie, zoals *black box* modellen, lijkt minder geschikt voor een sector die zijn besluiten moet kunnen motiveren aan burgers. Zwakke kunstmatige intelligentie, zoals de slimme keuzehulp die tekstanalyses uitvoert, is daarom wenselijker. Besluiten die hier uit voortvloeien zijn immers op individueel niveau uitlegbaar en dragen daarmee bij aan een verantwoorde mensgerichte AI-aanpak.

6.3.2 Aanbevelingen voor de werkwijze

Ten slotte worden op drie niveaus aanbevelingen gedaan voor de werkwijze: aan beleidsmakers die zich bezighouden met AI-toepassingen, aan ontwikkelaars die de AI-toepassingen ontwerpen en aan de politie.

Beleidsmakers: Heiligt het doel de middelen?

Voor beleidsmakers in de gehele publieke sector die zich bezighouden met AI-toepassingen wordt aanbevolen om steeds weer kritisch te blijven op het gebruik van algoritmen bij beslissingen die burgers individueel raken. Hierbij kunnen ze zichzelf afvragen of het doel wel de middelen heiligt. Algoritmische beslissystemen hebben het imago dat ze nauwkeurig, effectief en efficiënt zijn, waarbij ze de capaciteiten van een ambtenaar voorbijstreven. Om deze reden wordt steeds vaker gekozen om beslissingen door algoritmen te laten nemen in plaats van door een menselijke beslisser. Algoritmen zijn hip in de publieke sector. De resultaten van dit onderzoek laten echter zien dat het van groot belang is om een inhoudelijke motivering te kunnen geven over de genomen beslissing. Als een beslisproces zo complex is dat de motivering van een besluit niet aan burgers uitgelegd kan worden, heiligt het doel de middelen niet.

Ontwikkelaars: Massamaatwerk door AI.

Op allerlei terreinen in de publieke sector werken ontwikkelaars die AI-toepassingen ontwerpen. Aan hen is de aanbeveling om bij de ontwikkeling van zo'n toepassing extra aandacht te besteden aan de uitleg die ze ontwerpen voor de beslissingen die uit zo'n geautomatiseerd beslissysteem rollen. Dit onderzoek leert hen dat ze hierbij zo specifiek mogelijk moeten proberen te zijn. Mensgerichte AI vraagt om een inhoudelijke motivering waarom een bepaalde beslissing wordt genomen. De uitleg moet zo specifiek mogelijk zijn. Het is dus aan ontwikkelaars de taak om massamaatwerk te realiseren bij AI-toepassingen.

Politie: De kost gaat voor de baat uit.

Op dit moment wordt in de uitleg van de slimme keuzehulp nog weinig toegespitst op specifieke inhoudelijke argumenten waarom iemand het advies krijgt om geen aangifte te doen. De gebruiker krijgt een standaard uitleg zoals *“De webwinkel staat bij de politie bekend als een betrouwbare webwinkel”*. Hierin wordt bijvoorbeeld niet verteld waarom de webwinkel bekend staat als betrouwbaar, terwijl dit onderzoek juist aantoonde dat dat het vertrouwen in het advies van de slimme keuzehulp vergroot. Uit cijfers van de politie blijkt dat 72% van de mensen dat een afwijzend advies krijgt ('geen aangifte doen'), alsnog aangifte doet (Soares, 2020). Dit onderzoek laat zien dat dit percentage verlaagd kan worden naar 9,4% door een inhoudelijke uitleg te geven over de beslissing van de slimme keuzehulp. Indien de politie gebruik wil maken van AI-toepassingen, moet dit op een verantwoorde wijze gedaan worden. Dit houdt in dat burgers vertrouwen hebben in

de beslissingen die volgen uit de AI-toepassingen. De politie wordt daarom aanbevolen om de uitleg van de slimme keuzehulp zo specifiek mogelijk te maken.

Het aanpassen van de uitleg van de slimme keuzehulp is niet alleen maatschappelijk gezien wenselijk, maar kan ook financieel gezien voordelig zijn. Zo kost het beoordelen van een aangifte van internetfraude circa €2,-. Afgelopen maand (15 mei - 15 juni 2020) kreeg de politie ruim 800 aangiften binnen van mensen die een afwijzing kregen van de slimme keuzehulp en er toch voor kozen om aangiften te doen (Soares, 2020). Een simpele rekensom leert ons dat het ruim €1600,- per maand kost om deze aangiften te beoordelen. Op jaarbasis gaat dit om bijna €20.000,-. Als het concept van de slimme keuzehulp in de toekomst ook toegepast gaat worden op andere aangifteterreinen, kan dit bedrag nog hoger oplopen. Het is dus lonend om de uitleg zo specifiek mogelijk te maken, zodat het vertrouwen in het advies van de slimme keuzehulp toeneemt en minder mensen ervoor kiezen om alsnog aangifte te doen in het geval van een afwijzing. Dit betekent dus extra investeren aan de voorkant, maar dit zal op de lange termijn voordelig zijn voor de politie.

7 Literatuur

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, 53(1), 63-70.

AI HLEG (8 april 2019). *Ethische richtsnoeren voor betrouwbare KI*. Geraadpleegd via <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (16 juni 2020)

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.

Berbers, Y., Bonte, L., Man, H. de, Eynikel, J., Heene, A., Overschee, W. van, Vandewalle, J. & Verstraeten, P. (2018). Verantwoordelijk omgaan met digitalisering: een oproep naar overheden en bedrijfsleven, waartoe ook de burger toe kan/moet bijdragen. *KVAB Standpunt*, 61.

Berendsen, L. (2007). *Bureaucratische drama's* (Dissertatie). University of Utrecht.

Bex, F. (2019). Artificiële intelligentie in de praktijk: Lessen uit onderzoek bij de politie. *Rechtstreeks*, 16(2), 26-34.

Binns, R. (2018). Algorithmic accountability and public reason. *Philosophy & technology*, 31(4), 543-556.

Birkinshaw, J., & Cable, D. (2017). The dark side of transparency. *McKinsey Quarterly*, 88-95.

Blom-Hansen, J., Morton, R., & Serritzlew, S. (2015). Experiments in public management research. *International Public Management Journal*, 18(2), 151-170.

Boer, M. G. W. den (1998). Steamy windows: Transparency and openness in justice and home affairs. *Openness and transparency in the European Union*, 91-105.

Bouwman, R., & Grimmelikhuijsen, S. G. (2016). Experimental public administration from 1992 to 2014: A systematic literature review and ways forward. *International Journal of Public Sector Management*, 29(2), 110-131.

Bovens, M. A. P., & Zouridis, S. (2002). From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. *Public administration review*, 62(2), 174-184.

Bovens, M. A. P., Zouridis, S., & Eck, M. van (2018). Algoritmische rechtstoepassing in de democratische rechtsstaat. *Nederlands juristenblad*, 93(40), 3008-3017.

Bozeman, B., & Scott, P. (1992). Laboratory experiments in public policy and management. *Journal of Public Administration Research and Theory*, 2(3), 293-313.

Bryman, A. (2016). *Social Research Methods* (5e ed.). Oxford: Oxford University Press.

Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies*. New York: W.W. Norton & Company.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1-12.

Busch, P. A., & Henriksen, H. Z. (2018). Digital discretion: A systematic literature review of ICT and street-level discretion. *Information Polity*, 23(1), 3-28.

Campbell, M., Hoane Jr, A. J., & Hsu, F. H. (2002). Deep blue. *Artificial intelligence*, 134(1-2), 57-83.

CBS (2019a). *Online winkelen; kenmerken aankoop, persoonskenmerken*. Geraadpleegd via <https://www.cbs.nl/nl-nl/cijfers/detail/83430NED?q=online%20shoppen> (21 juni 2020)

CBS (2019b). *Bevolking; kerncijfers*. Geraadpleegd via <https://www.cbs.nl/nl-nl/cijfers/detail/37296ned> (22 juni 2020)

Cook, F. L., Jacobs, L. R., & Kim, D. (2010). Trusting what you know: Information, knowledge, and confidence in social security. *The Journal of Politics*, 72(2), 397-412.

Cramer, H., Evers, V., Ramlal, S., Someren, M. van, Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, 18(5), 455-496.

Davis, J. (1998). Access to and Transmission of Information: Position of the Media, in V. Deckmyn and I. Thomson (eds) *Openness and Transparency in the European Union*. Maastricht: European Institute of Public Administration, 121-126.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56-62.

Dignum, M. V., & Hoven, J. van den (2016). Reflecties op het verantwoord gebruik van kunstmatige intelligentie. *Justitiële Verkenningen*, 42(3), 108-115.

Doran, D., Schulz, S., & Besold, T. R. (2017). What does explainable AI really mean? A new conceptualization of perspectives. *ArXiv preprint arXiv:1710.00794 2071*, 2017.

Eck, B. M. A. van (2018). Computerbesluiten bij de overheid; recht en rekenen. *Ars Aequi*, 67(11), 962-967.

Ferrucci, D. A. (2012). Introduction to "this is watson". *IBM Journal of Research and Development*, 56(3), 1.

Field, A. P. (2009). *Discovering statistics with SPSS* (3^e ed.). Londen: Sage.

Fine Licht, K., de & Fine Licht, J. de (2020). Artificial intelligence, transparency, and public decision-making. *AI & SOCIETY*, 1-10.

Fine Licht, J. de, Naurin, D., Esaiasson, P., & Gilljam, M. (2014). When does transparency generate legitimacy? Experimenting on a context-bound relationship. *Governance*, 27(1), 111-134.

Fogg, B.J., & Tseng, H. (1999). The Elements of Computer Credibility. In *Proceedings of ACM CHI 99 Conference on Human Factors in Computing Systems*, 80-87. New York: ACM Press.

Fry, H. (2019). *Algoritmes aan de macht: Hoe blijf je menselijk in een geautomatiseerde wereld?*. Amsterdam: De Geus.

Grimmelikhuijsen, S. G. (2012a). *Transparency and trust. An experimental study of online disclosure and trust in government* (Dissertatie). Universiteit Utrecht.

Grimmelikhuijsen, S. G. (2012b). Linking transparency, knowledge and citizen trust in government: An experiment. *International Review of Administrative Sciences*, 78(1), 50-73.

Grimmelikhuijsen, S. G., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review*, 77(1), 45-56.

Grimmelikhuijsen, S. G., Porumbescu, G., Hong, B., & Im, T. (2013). The effect of transparency on trust in government: A cross-national comparative experiment. *Public Administration Review*, 73(4), 575-586.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), 1-42.

Gunning, D. (2017). *Explainable Artificial Intelligence (XAI)*. Geraadpleegd via <https://www.darpa.mil/attachments/XAIProgramUpdate.pdf> (25 mei 2020)

Harbers, M. (2018). *Verstand erbij: verantwoord ontwerp van toepassingen met kunstmatige intelligentie: openbare les*. Rotterdam: Hogeschool Rotterdam Uitgeverij.

Hind, M. (2019). Explaining explainable AI. *XRDS: Crossroads, The ACM Magazine for Students*, 25(3), 16-19.

Hoffmann, A. L. (2019). Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, 22(7), 900-915.

Hogan, B. (2015). From invisible algorithms to interactive affordances: Data after the ideology of machine learning. E. Bertone & S.A. Matei (eds.), *Roles, trust, and reputation in social media knowledge markets - theory and methods*. Heidelberg: Springer, 103-117.

Hosseini, M., Shahri, A., Phalp, K. T., & Ali, R. (2015). Transparency as a requirement. In *Joint proceedings of REFSQ 2015 workshops, research method track, and poster track, co-located with the 21st international conference on requirements engineering: foundation for software quality (REFSQ)*, 222-223.

Hutink, H., & Wouda, F. (2019). Verantwoorde zorg vereist verantwoorde AI. *Zorgvisie ICT*, 20(2), 16-17.

ICAI (z.d.). *Police Lab AI*. Geraadpleegd via <https://icai.ai/police-lab-ai/> (16 januari 2020)

Jorritsma, W., Cremers, A., & Jong, J. de (2020, 2 juli). *Onderzoek keuzehulp internetaangifte aankoopfraude*. Soesterberg: TNO.

Kankanhalli, A., Charalabidis, Y., & Mellouli, S. (2019). IoT and AI for smart government: A research agenda. *Government Information Quarterly*, 36(2), 304-309.

Kim, W. C., & Mauborgne, R. (1998). Procedural justice, strategic decision making, and the knowledge economy. *Strategic management journal*, 19(4), 323-338.

Kim, T. W., & Routledge, B. R. (2018). Informational Privacy, A Right to Explanation, and Interpretable AI. In *2018 IEEE Symposium on Privacy-Aware Computing (PAC)*, 64-74.

Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2390-2395.

Kool, L., Timmer, J., & Est, R. van (2015). *De datagedreven samenleving. Achtergrondstudie*. Den Haag: Rathenau Instituut.

Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633-706.

Kunst, M. & Dijk, J.J.M. van (2009). *Slachtoffers van fraude; een explorerend onderzoek naar de impact van diverse vormen van financieel-economische criminaliteit*. Tilburg: INTERVICT.

Laat, P. B. de (2018). Algorithmic decision-making based on machine learning from Big Data: Can transparency restore accountability?. *Philosophy & Technology*, 31(4), 525-541.

Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet Policy Review*, 9(2), 1-16.

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. In *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1-26.

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4), 611-627.

Lind, E. A., & Tyler, T. R. (1988). *The Social Psychology of Procedural Justice*. New York: Springer Science & Business Media.

Lipton, Z. C. (2018). The mythos of model interpretability. *Queue*, 16(3), 31-57.

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. In *Proceedings of the 11th Australasian Conference on Information Systems*, 6-8.

Mansbridge, J. (2009). A "selection model" of political representation. *Journal of Political Philosophy*, 17(4), 369-398.

Margetts, H. Z. (2011). Experiments for public management research. *Public Management Review*, 13(2), 189-208.

Matook, S., Brown, S. A., & Rolf, J. (2015). Forming an intention to act on recommendations given via online social networks. *European Journal of Information Systems*, 24(1), 76-92.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The journal of strategic information systems*, 11(3-4), 297-323.

Meijer, A. (2009). Understanding modern transparency. *International Review of Administrative Sciences*, 75(2), 255-269.

Meijer, A. J., Schäfer, M. T., & Branderhorst, M. (2019). Principes voor goed lokaal bestuur in de digitale samenleving: Een aanzet tot een normatief kader. *Bestuurswetenschappen*, 73(4), 8-23.

Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, 76(6), 928-937.

Ministerie van Algemene Zaken (2015). *Gereedschap voor gedragsverandering*. Geraadpleegd via <https://www.communicatierijk.nl/documenten/publicaties/2015/12/11/gereedschap-voor-gedragsverandering> (19 juni 2020)

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2009). *Nederlandse code voor goed openbaar bestuur. Beginselen van deugdelijk overheidsbestuur*. Den Haag: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21.

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency*, 279-288.

Mols, B. (2019). *Internationaal AI-beleid*. Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.

Moody, G. D., Galletta, D. F., & Lowry, P. B. (2014). When trust and distrust collide online: The engenderment and role of consumer ambivalence in online consumer behavior. *Electronic Commerce Research and Applications*, 13(4), 266-282.

Morton, R. B., & Williams, K. C. (2010). *Experimental political science and the study of causality: From nature to the lab*. New York: Cambridge University Press.

Mutz, D. C. (2011). *Population-based survey experiments*. Princeton: Princeton University Press.

Nagtegaal (2019). *Kunstmatige intelligentie. De bakboter van de toekomst*. Geraadpleegd via <https://dv2030.nl/sites/default/files/2019-10/Veilig%20en%20duurzaam-gebundelde%20essays%20sep%202019.pdf#page=38> (1 juni 2020)

Nothdurft, F., Lang, H., Klepsch, M., & Minker, W. (2013). Bases of human-computer trust and explanations. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 1737-1742.

Nothdurft, F., Richter, F., & Minker, W. (2014). Probabilistic human-computer trust handling. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 51-59.

Odekerken, D. (2020, 4 februari). *Keuzehulp Aangifte Internetoplichting. Toelichting voor ontwikkelaars*. Utrecht: Politie.

Onderwijs in Cijfers (z.d.). *Hoogst behaald onderwijsniveau*. Geraadpleegd via <https://www.onderwijsincijfers.nl/kengetallen/onderwijs-algemeen/hoogst-behaald-opleidingsniveau> (21 juni 2020)

Politie (2019a, 26 september). *Sneller duidelijkheid bij aangifte internetoplichting*. Geraadpleegd via <https://www.politie.nl/nieuws/2019/september/26/sneller-duidelijkheid-bij-aangifte-internetoplichting.html> (16 februari 2020)

Politie (2019b, 16 januari). *'Kunstmatige intelligentie vergroot onze slagkracht'*. Geraadpleegd via <https://www.politie.nl/nieuws/2019/januari/16/%E2%80%98kunstmatige-intelligentie-vergroot-onze-slagkracht.html> (16 februari 2020)

Politie (2020). *Aangifte of melding doen*. Geraadpleegd via <https://aangifte.politie.nl/iaai-preintake/#/> (21 juni 2020)

Poorter, J. de, & Goossens, J. (2019). Effectieve rechtsbescherming bij algoritmische besluitvorming in het bestuursrecht. *Nederlands Juristenblad*, 2019(44), 3303-3312.

Prins, C. (2016). De (digitale) scholen zijn weer begonnen. *Nederlands Juristenblad*, 91(30), 2161.

Rachlin, H. (2012). Making IBM's computer, Watson, human. *The Behavior Analyst*, 35(1), 1-16.

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1-13.

Rijksoverheid (2019). *Strategisch Actieplan Artificiële Intelligentie*. Den Haag: Ministerie van Economische Zaken en Klimaat.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3), 393-404.

Ryan, M., & Stahl, B. C. (2020). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*, in print.

Schlosser, A. E., White, T. B., & Lloyd, S. M. (2006). Converting web site visitors into buyers: how web site investment increases consumer trusting beliefs and online purchase intentions. *Journal of marketing*, 70(2), 133-148.

Schouwenburg, G. van, & Smits, J. (2019). Datagedreven rechtspraak? Het kan. Maar hoe? *Rechtstreeks*, 16(2), 61-69.

Seltman, H.J. (2012). *Experimental Design and Analysis*. Pittsburgh: Carnegie Mellon University.

Shapiro, S. P. (1987). The social control of impersonal trust. *American journal of Sociology*, 93(3), 623-658.

Soares, C. A. (2020, 15 juni). *Evaluatierapportage Keuzehulp Q1/Q2 2020*. Utrecht: Politie-ICT.

Sousa, S., Lamas, D., & Dias, P. (2014). A model for Human-computer Trust. In *International Conference on Learning and Collaboration Technologies*, 128-137.

Springer, A., & Whittaker, S. (2019, March). Progressive disclosure: empirically motivated approaches to designing effective transparency. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 107-120.

Stolze, J. (2018). *Algoritmisering, wen er maar aan!*. Amsterdam: Boom.

Thibaut, J., Walker, L., LaTour, S., & Houlden, P. (1974). Procedural justice as fairness. *Stanford Law Review*, 26(6), 1271-1290.

Tintarev, N., & Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5), 399-439.

Tyler, T. R. (1990). *Why people obey the law: Procedural justice, legitimacy, and compliance*. New Haven: Yale University Press.

US Public Policy Council (2017). *Statement on Algorithmic Transparency and Accountability*. Geraadpleegd via https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf (1 juni 2020)

Vetzo, M., & Gerards, J. H. (2019). Algoritme-gedreven technologieën en grondrechten. *Computerrecht*, 3(1), 10-19.

Voskuil, K. (2020, 6 juni). Enorme groei bij onlinefraude. *Algemeen Dagblad*. Geraadpleegd via https://digitalekrant.ad.nl/algemeendagblad/18681/article/1149384/1/1/render/?token=b90fd75f84cba3bac6b5f0ffd47a7fc6&vl_platform=ios&vl_app_id=be.persgroep.algemeendagblad&vl_app_version=3.4.9

Wachter, S., Mittelstadt, B., & Floridi, L. (2017a). Transparent, explainable, and accountable AI for robotics. *Science Robotics*, 2(6), ean6080.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017b). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2), 76-99.

Walle, S. van der, & Bouckaert, G. (2003). Public service performance and trust in government: the problem of causality. *International Journal of Public Administration*, 26(8-9), 891-913.

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems*, 23(4), 217-246.

Yeung, K. (2018). Algorithmic regulation: a critical interrogation. *Regulation & Governance*, 12(4), 505-523.

Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, 41(1), 118-132.

8 Bijlagen

8.1 BESCHRIJVENDE STATISTIEK RESPONDENTEN

Hieronder volgt een overzicht van de beschrijvende statistiek ($N=536$).

Tabel 8.1: Geslacht respondenten

Geslacht	Aantal	Percentage (%)
Man	229	42,7
Vrouw	301	56,2
Wil ik niet zeggen	6	1,1

Tabel 8.2: Geboortejaar respondenten

Geboortejaar	Aantal	Percentage (%)
≤1949	8	1,5
1950-1959	42	7,8
1960-1969	124	23,1
1970-1979	72	13,4
1980-1989	40	7,5
1990-1999	234	43,7
≥2000	8	1,5
<i>Missing</i>	8	1,5

Tabel 8.3: Opleidingsniveau respondenten

Geslacht	Aantal	Percentage (%)
Basisonderwijs	1	0,2
VMBO, HAVO-, VWO-onderbouw, MBO-1	13	2,4
HAVO, VWO, MBO-2, -3, of -4	89	16,6
HBO- of WO-bachelor	216	40,3
HBO-, WO-master, doctor	215	40,1
Onbekend	2	0,4

8.2 FACTORANALYSE

Tabel 8.4: Factoranalyse

Variabele	Factor 1	Factor 2
Ik vind het advies van de slimme keuzehulp betrouwbaar.		,719
Ik vind het advies van de slimme keuzehulp onbevooroordeeld.		,650
Ik vind het advies van de slimme keuzehulp waarheidsgetrouw.		,920
Ik vertrouw op het advies van de slimme keuzehulp.	,528	,439
Het is zeer waarschijnlijk dat ik luister naar het advies van de slimme keuzehulp.	,895	
Ik leg me neer bij het advies van de slimme keuzehulp.	,920	
Ik vind het geen probleem om te handelen naar het advies van de slimme keuzehulp.	,665	
Ik ben bereid het advies van de slimme keuzehulp op te volgen.	,922	
Verklaarde variantie (%)	68,73	5,12

8.3 CONTRASTEN

Tabel 8.5: Contrasten bij de 2x2 factoriale ANOVA's

Contrast	Geen uitleg	Procedurele uitleg	Inhoudelijke uitleg	Procedurele en inhoudelijke uitleg
1	-1	1	0	0
2	-1	0	1	0
3	0	1	-1	0
4	0	1	1	-2

Gebaseerd op de vier hypothesen:

- **Hypothese 1:** Een *procedurele uitleg* heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.
- **Hypothese 2:** Een *inhoudelijke uitleg* heeft een positief effect ten opzichte van geen uitleg op vertrouwen van burgers in algoritmische beslissingen.
- **Hypothese 3:** Een *procedurele uitleg* heeft een sterker positief effect op vertrouwen van burgers in algoritmische beslissingen dan een *inhoudelijke uitleg*.
- **Hypothese 4:** Een *procedurele uitleg* gecombineerd met een *inhoudelijke uitleg* heeft een negatief effect ten opzichte van enkel een procedurele of een inhoudelijke uitleg op vertrouwen van burgers in algoritmische beslissingen.

8.4 EXPERIMENT

[Intro]

Hartelijk dank voor uw deelname aan dit onderzoek over het aangiftesysteem van de politie. Dit onderzoek wordt uitgevoerd in samenwerking tussen de Universiteit Utrecht en de Nationale Politie.

Dit onderzoek kijkt naar de werking van 'de slimme keuzehulp'. De slimme keuzehulp is ontwikkeld door de politie om u een advies te geven over of het voor u zinvol is om aangifte te doen. De slimme keuzehulp maakt daarbij gebruik van door u ingevoerde gegevens. De resultaten van dit onderzoek worden gebruikt om het aangiftesysteem van de politie te verbeteren. Deelname aan dit onderzoek zal ongeveer 8 tot 10 minuten duren.

Privacyverklaring:

Deelname aan dit onderzoek is geheel vrijwillig. U kunt op elk moment stoppen met het onderzoek. Door rechtsonder op 'verder' te klikken gaat u er mee akkoord dat uw antwoorden gebruikt worden voor analyse in dit onderzoek. De antwoorden die u geeft, zullen vertrouwelijk en anoniem behandeld worden. Ook zullen uw antwoorden alleen worden gebruikt voor analyses op groepsniveau. Bij vragen over het onderzoek kunt u contact opnemen met e.n.nieuwenhuizen@students.uu.nl.

----- pagina split -----

[Informatie vooraf casus]

U krijgt zo een situatie voorgelegd. Hierbij wordt u gevraagd om u in te leven in de betreffende situatie. Na het lezen van de situatie wordt u gevraagd om aan de hand van de slimme keuzehulp aangifte te doen. **Let op:** u doet niet echt aangifte bij de politie. Dit is nagebootst voor het doel van het onderzoek.

Over uw ervaring met de slimme keuzehulp zal een korte vragenlijst aan u voorgelegd worden. Probeer u zich zo goed mogelijk in te leven in de betreffende situatie.

----- pagina split -----

[Casus]

Stelt u zich de volgende situatie voor:

Drie weken geleden heeft u nieuwe schoenen gekocht via de webshop www.sneakerwinkel.nl. U heeft hiervoor € 150,- betaald. De website gaf aan dat de sneakers binnen 5 werkdagen bij u thuis afgeleverd zouden worden.

Na drie weken heeft u nog steeds geen pakket ontvangen en ook geen bevestiging van verzending toegestuurd gekregen.

Omdat u denkt dat u opgelicht bent, gaat u naar de website van de politie om aangifte te doen van internetoplichting. Om aangifte te doen, moet u eerst enkele gegevens invullen.

Heeft u bovenstaande tekst gelezen? Klik dan op "verder".

----- pagina split -----

[Doorlopen slimme keuzehulp]

Uitleg over de slimme keuzehulp

U wilt aangifte doen van internetoplichting.

De politie wilt u graag advies geven of aangifte doen in uw geval zinvol is. Daarom doorloopt u eerst de slimme keuzehulp:

- U begint met vertellen wat er is gebeurd.
- De slimme keuzehulp zal uw verhaal bekijken en als dat nodig is enkele vragen stellen.
- Ten slotte adviseert de slimme keuzehulp of u aangifte kunt doen of dat u beter andere stappen kunt nemen.

----- pagina split -----

Stap 1: Wat is er gebeurd?

Hier kunt u vertellen wat er is gebeurd.

Het is belangrijk dat u zo volledig mogelijk bent. Denk bij het beschrijven aan:

- of u betaald heeft voor wat u wilde kopen

- of er door de verkoper iets geleverd is
- hoe lang u gewacht heeft voor u aangifte ging doen
- bij welke (web)winkel u een aankoop heeft gedaan

Beschrijf hieronder **in eigen woorden** wat er gebeurd is.

Indien u de details van de situatie vergeten bent, gaat u met uw muis op deze zin staan. De situatiebeschrijving wordt weergegeven.

----- pagina split -----

Stap 2: Bevindingen van de slimme keuzehulp

Op basis van uw input komt de keuzehulp met de volgende bevindingen. Vink de juiste bevinding(en) aan:

Indien u de details van de situatie vergeten bent, gaat u met uw muis op deze zin staan. De situatiebeschrijving wordt weergegeven.

- U heeft een product gekocht.
- U heeft een redelijke tijd gewacht voor u aangifte wilde doen.
- U verwachtte een pakket.
- Er is een product geleverd.

----- pagina split -----

Stap 3: Vragen over wat er is gebeurd

Naar aanleiding van uw verhaal heeft de slimme keuzehulp nog enkele vragen.

Indien u de details van de situatie vergeten bent, gaat u met uw muis op deze zin staan. De situatiebeschrijving wordt weergegeven.

Heeft u de verkoper betaald voor het product of een aanbetaling gedaan?

- Ja
- Nee

Heeft u besteld bij een webwinkel?

- Ja
- Nee

----- pagina split -----

Vul hier de naam of url (internetadres) van de webshop in:

----- pagina split -----

[Randomizer: random assignment in groep; n per groep = gelijk]

[Vignet 1: Controlegroep – geen uitleg]

Stap 4: Ons advies

Ons advies: Geen aangifte doen.

Op basis van uw verhaal lijkt het niet zinvol om aangifte te doen.

Uitgaande van dit advies zou ik:

Geen aangifte doen

Toch aangifte doen

----- pagina split -----

[Vignette 2: Treatment 1 – procedurele uitleg]

Stap 4: Ons advies

Ons advies: Geen aangifte doen.

Op basis van uw verhaal lijkt het niet zinvol om aangifte te doen.

Waarom krijg ik het advies om geen aangifte te doen?

De webwinkel www.sneakerwinkel.nl staat bij de politie bekend als een betrouwbare webwinkel en bij dergelijke webwinkels is er vrijwel nooit sprake van online fraude.

Hoe komen we tot dit advies?

- De slimme keuzehulp analyseert uw tekst door gebruik te maken van verschillende **algoritmes**.
- Er wordt o.a. gebruik gemaakt van een **tekstanalyse** om te zoeken naar bepaalde termen en combinaties van woorden die kunnen wijzen op een fraudezaak.
- Op basis van deze tekstanalyse is gekeken of er nog **relevante vragen** zijn.
- Uit uw antwoorden op deze vragen bleek dat u een aankoop heeft gedaan bij www.sneakerwinkel.nl.
- Deze winkel is in het databestand van de slimme keuzehulp aangemerkt als "**betrouwbaar**" en bij dergelijke webwinkels is er doorgaans **geen sprake van fraude**.

Uitgaande van dit advies zou ik:

Geen aangifte doen

Toch aangifte doen

----- pagina split -----

[Vignet 3: Treatment 2 – inhoudelijke uitleg]

Stap 4: Ons advies

Ons advies: Geen aangifte doen.

Op basis van uw verhaal lijkt het niet zinvol om aangifte te doen.

Waarom krijg ik het advies om geen aangifte te doen?

De webwinkel www.sneakerwinkel.nl staat bij de politie bekend als een betrouwbare webwinkel en bij dergelijke webwinkels is er vrijwel nooit sprake van online fraude.

Waarom is dit een betrouwbare webwinkel?

- De webwinkel heeft het **keurmerk** 'Webshop Keurmerk'.
- Een webwinkel die aangesloten is bij dit keurmerk is een **uitgebreide screeningsprocedure** doorgegaan.
- Het keurmerk bemiddelt bij een geschil met een webwinkel.
- Als deze bemiddeling niet tot het gewenste resultaat leidt, kunt u het geschil voorleggen aan een **onafhankelijke geschillencommissie**.
- Kijk op de website van de webwinkel voor wat u moet doen om uw geld terug te krijgen.

Uitgaande van dit advies zou ik:

Geen aangifte doen

Toch aangifte doen

----- pagina split -----

[Vignette 4: Treatment 3 – procedurele en inhoudelijke uitleg]

Stap 4: Ons advies

Ons advies: Geen aangifte doen.

Op basis van uw verhaal lijkt het niet zinvol om aangifte te doen.

Waarom krijg ik het advies om geen aangifte te doen?

De webwinkel www.sneakerwinkel.nl staat bij de politie bekend als een betrouwbare webwinkel en bij dergelijke webwinkels is er vrijwel nooit sprake van online fraude.

Hoe komen we tot dit advies?

- De slimme keuzehulp analyseert uw tekst door gebruik te maken van verschillende **algoritmes**.
- Er wordt o.a. gebruik gemaakt van een **tekstanalyse** om te zoeken naar bepaalde termen en combinaties van woorden die kunnen wijzen op een fraudezaak.
- Op basis van deze tekstanalyse is gekeken of er nog **relevante vragen** zijn.
- Uit uw antwoorden op deze vragen bleek dat u een aankoop heeft gedaan bij www.sneakerwinkel.nl.
- Deze winkel is in het databestand van de slimme keuzehulp aangemerkt als "**betrouwbaar**" en bij dergelijke webwinkels is er doorgaans **geen sprake van fraude**.

Waarom is dit een betrouwbare webwinkel?

- De webwinkel heeft het **keurmerk** 'Webshop Keurmerk'.
- Een webwinkel die aangesloten is bij dit keurmerk is een **uitgebreide screeningsprocedure** doorgegaan.
- Het keurmerk bemiddelt bij een geschil met een webwinkel.
- Als deze bemiddeling niet tot het gewenste resultaat leidt, kunt u het geschil voorleggen aan een **onafhankelijke geschillencommissie**.
- Kijk op de website van de webwinkel voor wat u moet doen om uw geld terug te krijgen.

Uitgaande van dit advies zou ik:

Geen aangifte doen

Toch aangifte doen

----- pagina split -----

[Open vraag na uitleg]

Enkel stellen In het geval van 'toch aangifte doen'

- Waarom kiest u er op basis van het advies van de slimme keuzehulp toch voor om aangifte te doen?

Enkel stellen In het geval van 'geen aangifte doen'

- Waarom kiest u er op basis van het advies van de slimme keuzehulp geen aangifte te doen?

----- pagina split -----

[Overige vragen]

Naar aanleiding van de situatie die u net heeft gelezen, geef antwoord op de volgende vragen:

[Vragen vertrouwen in advies] (Likertschaal: 1 -7)

Geef aan in hoeverre u het eens bent met de volgende stellingen. Er zijn geen goede of foute antwoorden.

- Ik vind het advies van de slimme keuzehulp betrouwbaar.
- Ik vind het advies van de slimme keuzehulp onbevooroordeeld.
- Ik vind het advies van de slimme keuzehulp waarheidsgetrouw.
- Ik vertrouw op het advies van de slimme keuzehulp.

----- pagina split -----

[Vragen bereidheid om naar advies te handelen] (Likertschaal: 1 -7)

Geef aan in hoeverre u het eens bent met de volgende stellingen. Er zijn geen goede of foute antwoorden.

- Het is zeer waarschijnlijk dat ik luister naar het advies van de slimme keuzehulp.
- Ik leg me neer bij het advies van de slimme keuzehulp.
- Ik vind het geen probleem om te handelen naar het advies van de slimme keuzehulp.
- Ik ben bereid het advies van de slimme keuzehulp op te volgen.

----- pagina split -----

[Manipulatiecheck] (Likertschaal 1-7)

Geef aan in hoeverre de volgende stelling van toepassing is op de door u gelezen situatie:

- De slimme keuzehulp gaf **geen** goede uitleg over het advies.
- De slimme keuzehulp gaf uitleg over de manier waarop het advies tot stand is gekomen.
- De slimme keuzehulp maakte duidelijk welke stappen zijn gezet om tot het advies te komen.
- Ik begrijp waarom het in mijn geval niet zinvol was om aangifte te doen.
- De slimme keuzehulp gaf mij duidelijke uitleg over waarom ik het advies kreeg om geen aangifte te doen.

----- pagina split -----

[Demografische vragen]

Wat is uw geslacht?

- Man
- Vrouw
- Anders
- Wil ik niet zeggen

Wat is uw geboortjaar?

Wat is uw hoogst behaalde onderwijsniveau?

- Basisonderwijs
- VMBO, HAVO-, VWO-onderbouw, MBO-1
- HAVO, VWO, MBO-2, -3 of -4
- HBO-, WO-bachelor
- HBO-, WO-master, doctor Ja
- Onbekend

----- pagina split -----

[Controlevragen] (Likertschaal: 1 -7)

Geef hier aan hoeveel vertrouwen u heeft in de volgende zaken:

- Vertrouwen in de politie in het algemeen
- Vertrouwen in technologie in het algemeen

----- pagina split -----

[Feedbackvragen] (Likertschaal: 1 -7)

Geef aan in hoeverre u het eens bent met de volgende stellingen:

- Ik kon mij inleven in de aan mij voorgelegde situatie.
- Ik heb een soortgelijke situatie van internetoplichting wel eens meegemaakt.

Eventuele extra toelichting:

----- pagina split -----

[Eerdere ervaring met slimme keuzehulp]

Heeft u -los van dit onderzoek- de slimme keuzehulp eerder gebruikt bij het doen van aangifte tegen internetoplichting?

- Ja
- Nee

----- pagina split -----

[Extra opmerkingen]

Indien u nog opmerkingen heeft over dit onderzoek, vult u deze dan hieronder in.

----- pagina split -----

[Debriefing]

Hartelijk bedankt voor uw deelname aan het onderzoek van de Universiteit Utrecht voor de Nationale Politie. Sinds september 2019 gebruikt de politie de slimme keuzehulp bij aangiftes van internetoplichting. Uw hulp aan dit onderzoek draagt bij aan de verbetering van het aangiftesysteem van de politie. De resultaten van dit onderzoek kunt u vanaf medio juli 2020 inzien via de [website](#) van het Nationaal Politielab Artificial Intelligence.

In dit onderzoek is getest wat de invloed is van het soort uitleg dat u krijgt over het advies van de slimme keuzehulp op uw vertrouwen in dit advies. Er is aan u een situatie voorgelegd en u heeft met behulp van de slimme keuzehulp aangifte willen doen. Hierna werd u gevraagd om een vragenlijst in te vullen over het vertrouwen in het advies van de slimme keuzehulp. De antwoorden die u gegeven heeft, zullen vertrouwelijk en anoniem behandeld worden. Bij verdere vragen over het onderzoek kunt u contact opnemen met e.n.nieuwenhuizen@students.uu.nl.

Uw antwoorden zijn opgeslagen dus u kunt dit venster sluiten.

8.5 AANVULLENDE ANALYSES

8.5.1 Filtering respondentent

De analyses zijn opnieuw uitgevoerd, maar hierbij zijn de antwoorden van acht respondentent eruit gefilterd. Ten eerste van respondentent die langer dan twee keer de standaarddeviatie van het gemiddelde deden over het invullen van de vragenlijst, $M = 1148,12$; $SD = 3939,79$. Bij deze respondentent bestaat namelijk het risico dat zij afgeleid zijn geraakt tijdens het experiment, waardoor de manipulatie niet het gewenste effect zou kunnen hebben. In totaal waren er zeven respondentent die langer dan 9027,20 seconden deden over het surveyexperiment. Deze zijn uit de analyse gehaald.

Ten tweede is gekeken naar respondentent die bij 'stap 1' van de slimme keuzehulp geen inhoudelijk antwoord gaven (bijvoorbeeld "asldfdkfdkjdkdsjf"), omdat het advies van de slimme keuzehulp dan niet overeenkomt met hun antwoord. Dit kan een negatief effect hebben op het vertrouwen dat respondentent hebben in het advies. Er was één respondent die dit gedaan had.

Onderstaande analyses zijn dus op basis van $N = 528$. De resultaten van deze analyses komen overeen met de bevindingen in het resultatenhoofdstuk.

Vertrouwen in het advies

Het effect van een procedurele uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was significant, $F(1, 524) = 13,11$; $p < 0,001$; $\eta_p^2 = 0,02$. Respondentent die geen procedurele uitleg kregen ($M = 4,39$; $SD = 1,49$) hadden significant minder vertrouwen in het advies van de slimme keuzehulp dan respondentent die wel een procedurele uitleg ontvingen ($M = 4,78$; $SD = 1,32$).

Het effect van een inhoudelijke uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was ook significant, $F(1, 524) = 92,32$; $p < 0,001$; $\eta_p^2 = 0,15$. Respondentent die geen inhoudelijke uitleg kregen ($M = 4,04$; $SD = 1,40$) hadden significant minder vertrouwen in het advies van de slimme keuzehulp dan respondentent die wel een inhoudelijke uitleg ontvingen ($M = 5,11$; $SD = 1,23$).

Hier is een vergelijkbaar interactie-effect als bij de originele analyses gevonden. Het effect van een procedurele uitleg op het vertrouwen van respondentent in het advies van de slimme keuzehulp hangt af van of ze een inhoudelijke uitleg gehad hebben, $F(1, 524) = 20,96$; $p < 0,001$; $\eta_p^2 = 0,04$.

Bereidheid om te handelen naar advies

Het effect van een procedurele uitleg op de mate van bereidheid om te handelen naar het advies van de slimme keuzehulp was significant, $F(1, 524) = 21,78$; $p < 0,001$; $\eta_p^2 =$

0,04. Respondenten die geen procedurele uitleg kregen ($M = 4,29$; $SD = 1,67$) hadden een significant lagere gerapporteerde bereidheid om te handelen naar het advies van de slimme keuzehulp dan respondenten die wel een procedurele uitleg ontvingen ($M = 4,83$; $SD = 1,42$).

Ook het effect van een inhoudelijke uitleg op de mate van bereidheid om te handelen naar het advies van de slimme keuzehulp was significant, $F(1, 524) = 103,69$; $p < 0,001$; $\eta_p^2 = 0,17$. Respondenten die geen inhoudelijke uitleg kregen ($M = 3,94$; $SD = 1,61$) hadden significant minder bereidheid om te handelen naar het advies van de slimme keuzehulp dan respondenten die wel een inhoudelijke uitleg ontvingen ($M = 5,16$; $SD = 1,27$).

Vergelijkbaar met de originele analyse is hier ook sprake van een interactie-effect. Het effect van een procedurele uitleg op de bereidheid van respondenten om te handelen naar het advies van de slimme keuzehulp hangt af van of ze een inhoudelijke uitleg gehad hebben, $F(1, 524) = 32,04$; $p < 0,001$; $\eta_p^2 = 0,06$.

Daadwerkelijk gedrag gebruiker

Het omnibus model voor de logistische regressie analyse was significant, $\chi^2 (df = 2, N = 528) = 114,64$; $p < 0,001$, Cox en Snell $R^2 = 0,20$; Nagelkerke $R^2 = 0,28$. Het model is 78% nauwkeurig in haar voorspellingen of gebruikers al dan niet aangifte gaan doen. Tabel 8.6 geeft de coëfficiënten voor de voorspellers van het model weer.

Tabel 8.6: Voorspellende coëfficiënten voor het model dat 'geen aangifte doen' voorspelt ($N = 528$)

	<i>b</i>	<i>SE (b)</i>	<i>p</i>	<i>Exp(B)</i> [95% BI]*
Constante	-0,38			
Procedurele uitleg	0,91	0,22	<0,001	2,49 [1,62, 3,82]
Inhoudelijke uitleg	2,16	0,24	<0,001	8,69 [5,42, 13,95]

*BI = betrouwbaarheidsinterval.

Zoals uit tabel 8.6 naar voren komt, zijn zowel een procedurele uitleg als een inhoudelijke uitleg significante voorspellers die het voorspellend vermogen van het model aanzienlijk verbeteren. De procedurele uitleg laat een oddsratio zien van 2,5. De odds op geen aangifte doen is dus 2,5 keer zo hoog voor respondenten die wel een procedurele uitleg hebben gehad dan voor respondenten zonder een procedurele uitleg. De inhoudelijke uitleg laat een oddsratio zien van 8,7. Dit betekent dat de odds op het doen van geen aangifte 8,7 keer hoger is voor respondenten die wel een inhoudelijke uitleg hebben gehad dan voor respondenten zonder een inhoudelijke uitleg.

8.5.2 Attentiecheck

Tenslotte zijn de respondenten die 'gezakt' zijn voor de attentiecheck uit de analyse gefilterd. Dit ging om tien respondenten. Bij deze respondenten bestaat namelijk het risico dat ze andere vragen ook niet goed gelezen hebben, waardoor het experimentele effect niet gemeten kan worden. Omdat de resultaten vergelijkbaar zijn met die in het resultatenhoofdstuk, zijn die hieronder verkort weergegeven. Hierbij is $N = 526$.

Vertrouwen in het advies

Het effect van een procedurele uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was significant, $F(1, 522) = 13,13$; $p < 0,001$; $\eta_p^2 = 0,03$. Respondenten die geen procedurele uitleg kregen hadden significant minder vertrouwen in het advies van de slimme keuzehulp dan respondenten die wel een procedurele uitleg ontvingen.

Het effect van een inhoudelijke uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was ook significant, $F(1, 522) = 101,76$; $p < 0,001$; $\eta_p^2 = 0,16$. Respondenten die geen inhoudelijke uitleg kregen hadden significant minder vertrouwen in het advies van de slimme keuzehulp dan respondenten die wel een inhoudelijke uitleg ontvingen.

Hier is een vergelijkbaar interactie-effect als bij de originele analyses gevonden. Het effect van een procedurele uitleg op het vertrouwen van respondenten in het advies van de slimme keuzehulp hangt af van of ze een inhoudelijke uitleg gehad hebben, $F(1, 522) = 22,56$; $p < 0,001$; $\eta_p^2 = 0,04$.

Bereidheid om te handelen naar advies

Het effect van een procedurele uitleg op de mate van bereidheid om te handelen naar het advies van de slimme keuzehulp was significant, $F(1, 522) = 22,66$ $p < 0,001$; $\eta_p^2 = 0,04$. Respondenten die geen procedurele uitleg kregen hadden een significant lagere gerapporteerde bereidheid om te handelen naar het advies van de slimme keuzehulp dan respondenten die wel een procedurele uitleg ontvingen.

Ook het effect van een inhoudelijke uitleg op de mate van bereidheid om te handelen naar het advies van de slimme keuzehulp was significant, $F(1, 522) = 109,67$; $p < 0,001$; $\eta_p^2 = 0,17$. Respondenten die geen inhoudelijke uitleg kregen hadden significant minder bereidheid om te handelen naar het advies van de slimme keuzehulp dan respondenten die wel een inhoudelijke uitleg ontvingen.

Vergelijkbaar met de originele analyse is hier ook sprake van een interactie-effect. Het effect van een procedurele uitleg op de bereidheid van respondenten om te handelen naar het advies van de slimme keuzehulp hangt af van of ze een inhoudelijke uitleg gehad hebben, $F(1, 522) = 32,56$; $p < 0,001$; $\eta_p^2 = 0,06$.

Daadwerkelijk gedrag gebruiker

Het omnibus model voor de logistische regressie analyse was significant, $\chi^2 (df = 2, N = 526) = 118,16$; $p < 0,001$, Cox en Snell $R^2 = 0,20$; Nagelkerke $R^2 = 0,29$. Het model is 78% nauwkeurig in haar voorspellingen of gebruikers al dan niet aangifte gaan doen. Tabel 8.7 geeft de coëfficiënten voor de voorspellers van het model weer.

Tabel 8.7: Voorspellende coëfficiënten voor het model dat 'geen aangifte doen' voorspelt ($N = 526$)

	<i>b</i>	<i>SE (b)</i>	<i>p</i>	<i>Exp(B)</i> [95% BI]*
Constante	-0,39			
Procedurele uitleg	0,91	0,22	<0,001	2,50 [1,62, 3,85]
Inhoudelijke uitleg	2,24	0,25	<0,001	9,40 [5,79, 15,27]

*BI = betrouwbaarheidsinterval.

Zoals uit tabel 8.7 naar voren komt, zijn zowel een procedurele uitleg als een inhoudelijke uitleg significante voorspellers die het voorspellend vermogen van het model aanzienlijk verbeteren. De procedurele uitleg laat een oddsratio zien van 2,5. De odds op geen aangifte doen is dus 2,5 keer zo hoog voor respondenten die wel een procedurele uitleg hebben gehad dan voor respondenten zonder een procedurele uitleg. De inhoudelijke uitleg laat een oddsratio zien van 9,4. Dit betekent dat de odds op het doen van geen aangifte 9,4 keer hoger is voor respondenten die wel een inhoudelijke uitleg hebben gehad dan voor respondenten zonder een inhoudelijke uitleg.

8.5.3 Samengestelde schaal

De Cronbach's alpha van de items uit de twee constructen 'vertrouwen in advies' en 'bereidheid om naar advies te handelen' samengenomen is zeer hoog (0,94). Hierom is opnieuw een 2x2 factorial ANOVA uitgevoerd met de samengestelde schaal waarin deze acht variabelen van de twee constructen samengenomen worden. Hieronder wordt naar die nieuwe variabele gerefereerd als 'vertrouwen in het advies'. Ook hier zijn de resultaten ($N = 536$) vergelijkbaar met de originele bevindingen.

Het effect van een procedurele uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was significant, $F(1, 532) = 19,19$; $p < 0,001$; $\eta_p^2 = 0,04$. Respondenten die geen procedurele uitleg kregen ($M = 4,35$; $SD = 1,49$) hadden significant minder vertrouwen in het advies van de slimme keuzehulp dan respondenten die wel een procedurele uitleg ontvingen ($M = 4,80$; $SD = 1,30$).

Het effect van een inhoudelijke uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was ook significant, $F(1, 532) = 113,40$; $p < 0,001$; $\eta_p^2 = 0,18$. Respondenten die geen inhoudelijke uitleg kregen ($M = 4,00$; $SD = 1,42$) hadden

significant minder vertrouwen in het advies van de slimme keuzehulp dan respondenten die wel een inhoudelijke uitleg ontvingen ($M = 5,14$; $SD = 1,17$).

Hier is een vergelijkbaar interactie-effect als bij de originele analyses gevonden. Het effect van een procedurele uitleg op het vertrouwen van respondenten in het advies van de slimme keuzehulp hangt af van of ze een inhoudelijke uitleg gehad hebben, $F(1, 532) = 30,52$; $p < 0,001$; $\eta_p^2 = 0,05$.

8.5.4 Laag en middelbaar opgeleid

Omdat in de steekproef relatief veel hoogopgeleide respondenten zaten, zijn de analyses opnieuw uitgevoerd met enkel de laag en middelbaar opgeleide respondenten. Onderstaande analyses zijn op basis van $N = 103$. De resultaten van deze analyses vertonen dezelfde patronen als de bevindingen in het resultatenhoofdstuk wat betreft het op vertrouwen gebaseerd gedrag. De resultaten van de vertrouwende houding wijken af van de bevindingen in het resultatenhoofdstuk.

Vertrouwen in het advies

Het effect van een procedurele uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was niet significant, $F(1, 99) = 2,23$; $p = 0,14$. Er bestond geen significant verschil tussen respondenten die geen procedurele uitleg kregen ($M = 4,30$; $SD = 1,47$) en respondenten die wel een procedurele uitleg ontvingen ($M = 4,63$; $SD = 1,45$) in hun vertrouwen in het advies van de slimme keuzehulp.

Het effect van een inhoudelijke uitleg op de mate van vertrouwen in het advies van de slimme keuzehulp was wel significant, $F(1, 99) = 15,86$; $p < 0,001$; $\eta_p^2 = 0,14$. Respondenten die geen inhoudelijke uitleg kregen ($M = 3,99$; $SD = 1,16$) hadden significant minder vertrouwen in het advies van de slimme keuzehulp dan respondenten die wel een inhoudelijke uitleg ontvingen ($M = 5,02$; $SD = 1,46$).

Hier is geen vergelijkbaar interactie-effect als bij de originele analyses gevonden. Er is geen significant interactie-effect, $F(1, 99) = 0,76$; $p = 0,39$.

Bereidheid om te handelen naar advies

Het effect van een procedurele uitleg op de mate van bereidheid om te handelen naar het advies van de slimme keuzehulp was significant, $F(1, 99) = 6,29$; $p = 0,01$; $\eta_p^2 = 0,06$. Respondenten die geen procedurele uitleg kregen ($M = 4,06$; $SD = 1,58$) hadden een significant lagere gerapporteerde bereidheid om te handelen naar het advies van de slimme keuzehulp dan respondenten die wel een procedurele uitleg ontvingen ($M = 4,72$; $SD = 1,60$).

Ook het effect van een inhoudelijke uitleg op de mate van bereidheid om te handelen naar het advies van de slimme keuzehulp was significant, $F(1, 99) = 17,88$; $p < 0,001$; $\eta_p^2 = 0,15$. Respondenten die geen inhoudelijke uitleg kregen ($M = 3,88$; $SD = 1,73$) hadden significant minder bereidheid om te handelen naar het advies van de slimme keuzehulp dan respondenten die wel een inhoudelijke uitleg ontvingen ($M = 5,02$; $SD = 1,25$).

Vergelijkbaar met de originele analyse is hier ook sprake van een interactie-effect. Het effect van een procedurele uitleg op de bereidheid van respondenten om te handelen naar het advies van de slimme keuzehulp hangt af van of ze een inhoudelijke uitleg gehad hebben, $F(1, 99) = 4,14$; $p = 0,04$; $\eta_p^2 = 0,04$.

Daadwerkelijk gedrag gebruiker

Het omnibus model voor de logistische regressie analyse was significant, $\chi^2 (df = 2, N = 103) = 30,09$; $p < 0,001$, Cox en Snell $R^2 = 0,25$; Nagelkerke $R^2 = 0,35$. Het model is 75% nauwkeurig in haar voorspellingen of gebruikers al dan niet aangifte gaan doen. Tabel 8.8 geeft de coëfficiënten voor de voorspellers van het model weer.

Tabel 8.8: Voorspellende coëfficiënten voor het model dat 'geen aangifte doen' voorspelt ($N = 103$)

	<i>b</i>	<i>SE (b)</i>	<i>p</i>	<i>Exp(B)</i> [95% BI]*
Constante	-1,00			
Procedurele uitleg	1,13	0,50	0,025	3,09 [1,15, 8,26]
Inhoudelijke uitleg	2,52	0,56	<0,001	12,39 [4,19, 36,62]

*BI = betrouwbaarheidsinterval.

Zoals uit tabel 8.8 naar voren komt, zijn zowel een procedurele uitleg als een inhoudelijke uitleg significante voorspellers die het voorspellend vermogen van het model aanzienlijk verbeteren. De procedurele uitleg laat een oddsratio zien van 3,1. De odds op geen aangifte doen is dus 3,1 keer zo hoog voor respondenten die wel een procedurele uitleg hebben gehad dan voor respondenten zonder een procedurele uitleg. De inhoudelijke uitleg laat een oddsratio zien van 12,4. Dit betekent dat de odds op het doen van geen aangifte 12,4 keer hoger is voor respondenten die wel een inhoudelijke uitleg hebben gehad dan voor respondenten zonder een inhoudelijke uitleg.