

Een geïntrigeerde theorie over het menselijk redeneren in de suppression task: Het achterhalen van één correct model



Universiteit Utrecht

Bacheloreindwerkstuk

Naam: Anne Vos
Opleiding: bachelor Kunstmatige intelligentie UU
ECTS: 7,5 EC

Begeleidster: Rosalie Iemhoff
Tweede beoordeelster: Rianne van Lambalgen

Datum: 7 november 2019

Abstract

Door twee modellen die de uitkomsten van de suppression task voorspellen met elkaar te vergelijken draagt deze studie bij aan de integratie van theorieën over het menselijk redeneren. Door de suppression task op die condities waar de modellen van elkaar verschillen uit te voeren kan deze vergelijking worden gemaakt en de onderzoeksvraag: “*Lijkt het model van Stenning & van Lambalgen (2012) of dat van Iemhoff & Janssen het beste in overeenstemming te zijn met de resultaten van de suppression task?*” worden beantwoord. De uitkomsten van het experiment worden, gebruikmakend van een threshold, vergeleken met beide modellen en aan de hand van een binomiaal toets wordt beredeneerd of deze vergelijking terecht mag worden gemaakt. Uit deze analyse volgt dat er een voorkeur kan worden ontleend aan het model van Iemhoff & Janssen.

In de discussie wordt ingegaan op enige tekortkomingen van het onderzoek en wordt de vooraf vastgestelde threshold in twijfel getrokken. Vervolgens wordt beredeneerd dat geen enkele andere threshold de gevonden conclusie zal ontkrachten. Toch is er reden om aan te nemen dat het model van Iemhoff & Janssen niet het ideale model is om de suppression task in kaart te brengen. Vervolgonderzoek zou zich kunnen focussen op het ontwikkelen van een alternatief model.

Inhoudsopgave

Hoofdstuk 1 Inleiding

Hoofdstuk 2 Theoretisch kader

2.1 Suppression task

2.2 Modellen

-Stenning & van Lambalgen

-Iemhoff & Janssen

2.3 Threshold

Hoofdstuk 3 Methodologie

3.1 Proefpersonen

3.2 Experiment

3.3 Procedure

3.4 Data-analyse

-Chi-kwadraattoets ter vergelijking onderzoeksresultaten en resultaten van Byrne en Dietz et al. (2012)

- Binomiaal toets threshold

Hoofdstuk 4 Onderzoeksresultaten

4.1 Resultaten suppression task

4.2 Vergelijking onderzoeksresultaten en resultaten van Byrne (1989) en Dietz et al. (2012)

4.3 Resultaten na toepassing van de threshold

4.4 Resultaten binominaal toets

Hoofdstuk 5 Conclusie

Hoofdstuk 6 Discussie

6.1 Nuance conclusie

6.2 Meerdere thresholds

6.3 Aantal proefpersonen

6.4 relevantie kunstmatige intelligentie

6.5 Vervolgonderzoek

Hoofdstuk 1 Inleiding

Menselijk redeneren is iets wat elk moment van de dag gebeurt. Het is hoe beslissingen tot stand komen, met alle gevolgen van dien. In het dagelijks leven worden vaak alleen de uitkomsten van redeneringen opgemerkt. Toch is het van belang om ook het redeneringsproces zelf onder de loep te nemen. Dit belang wordt door menig wetenschapper erkent, al jarenlang weet het menselijk redeneren hen te boeien. Er zijn tal van onderzoeken gedaan naar de manier waarop beslissingen en conclusies tot stand komen en de mentale processen die hieraan ten grondslag liggen.

Vele logici hebben een uitspraak gedaan over deze onderliggende mentale processen. Zo beweren enkele wetenschappers dat er sprake is van bepaalde inferentieregels die afhankelijk zijn van de inhoud van de variabelen (Copi, 1982).

Tal van andere wetenschappers sluiten zich aan bij de theorie dat een redenering afhangt van inferentieregels die op een syntactische manier op de logische vorm van premissen opereren (Braine, 1978; Reiser & Romain, 1984; Otherson, 1975; Risp, 1983; Sperber & Wilson, 1986). En weer een andere theorie gaat ervan uit dat bepaalde gevolgtrekkingen tot stand kunnen komen indien voorgaande proposities al zijn vastgesteld (Braine, 1978).

Elk van deze theorieën probeert een psychologisch plausibele onderbouwing te geven voor het menselijke redeneringsproces.

Eén manier om dit proces in kaart te brengen, is door participanten een logistische test te laten uitvoeren, om vervolgens te onderzoeken hoe de uitkomsten van zo een taak tot stand komen. De 'suppression task', ontworpen en voor het eerst uitgevoerd door Byrne (1989), is zo een taak (Byrne, 1989). De suppression task kent vier verschillende premissen van de vorm "als A dan B", waar afhankelijk van de vraag bepaalde conclusies dienen te worden getrokken. Uit de resultaten van de suppression task blijkt dat men niet via de klassieke logica lijkt te redeneren. Dit leidt tot de vraag op welke manier men dan wel tot een bepaalde gevolgstrekking komt.

Byrne (1989) heeft in haar onderzoek de suppression task uitgevoerd en interpreteert de afwijkende resultaten als volgt: "The idea is that people represent all possible states of affairs which are consistent with their world knowledge and the content of the sentences. A conclusion is only endorsed if there is no model which contradicts it (Byrne, 1989, p.61).

Volgens Byrne (1989) leidt het introduceren van verschillende premissen tot het construeren van verschillende modellen, elk voorzien van hun eigen tegenvoorbeelden.

Met haar theorieën introduceert Byrne (1989) het zogeheten 'suppression effect'. Dit effect speelt een belangrijke rol in psychologische theorieën over conditioneel redeneren.

Toch lijken deze theorieën niet als algemeen geldende waarheid te worden aangenomen.

Stenning en van Lambalgen (2012) komen terug op het idee van de klassieke logica als basis van menselijk redeneren en ontwikkelen een model in overeenstemming met de bevindingen van onder andere Byrne (1989). Dit model zal het uitgangspunt worden voor dit onderzoek en wordt vergeleken met een variant van dat model, dat van Rosalie Iemhoff en Chris Janssen. Beide modellen bieden verklaringen voor redeneerstappen die worden gemaakt in de suppression task. Met als uitkomst bepaalde conclusies die dienen te worden getrokken tijdens deze taak. De uitkomsten van de twee modellen verschillen echter van elkaar op twee condities.

Het is van essentieel belang om te onderzoeken welk van deze modellen het best in staat is het menselijk redeneren in kaart te brengen om een zo accuraat mogelijke theorie over conditioneel redeneren te kunnen ontwikkelen. Gebleken is dat er momenteel veel verschillende (varianties van) theorieën zijn en worden toegepast. Evans (2002) beschrijft het belang van een meer geïntegreerde theorie over het menselijk redeneren. Eén correct model over de uitkomsten van de suppression task draagt bij aan de integratie van deze theorieën.

Om een uitspraak te doen over de twee modellen zal in dit onderzoek de suppression task opnieuw worden uitgevoerd op die punten waar de twee modellen van elkaar verschillen. De uitkomsten van dit experiment zullen worden vergeleken met de uitkomsten van de twee modellen. De volgende vraag zal in dit onderzoek beantwoord gaan worden:

Lijkt het model van Stenning & van Lambalgen (2012) of dat van Iemhoff & Janssen het beste in overeenstemming te zijn met de resultaten van de suppression task?

Een deel van deze onderzoeksvraag vraagt om theoretische onderbouwing. Hoofdstuk 1 richt zich op de bespreking van de suppression task en de twee modellen. Ook zal hier blijken dat de totstandkoming van een conclusie voor deze onderzoeksvraag minder rechtlijnig is dan gedacht. Om toch een nuttige uitspraak te doen over de twee modellen zullen er, onder andere in het theoretische deel, verschillende manieren en invalshoeken aan de orde komen om de onderzoeksvraag te beantwoorden.

De methodesectie wijdt zich aan het toelichten van de precieze uitvoering van de suppression task voor dit experiment en de statistische testen die aan de orde komen om de resultaten te analyseren. In de conclusie worden de uitkomsten van deze testen samengevat. De focus van de discussie ligt op de verschillende implicaties van het onderzoek en het nuanceren van de conclusie.

Modellen die uitkomsten van de suppression task voorspellen en verklaren gaan verder dan de taak zelf. Het in kaart brengen van menselijk redeneren kan in verschillende situaties inzicht geven. Door een correct en accuraat model te ontwikkelen kunnen menselijke denkstappen worden voorspeld. Zijn de uitkomsten van deze voorspellingen ook de gewenste uitkomst in bepaalde situaties? Indien dit niet het geval is, kan worden beargumenteerd dat bepaalde taken of beslissingen beter door een machine, die volkomen rationeel en logistisch kan redeneren, kan worden gemaakt in plaats van de mens. Dit onderzoek is dus interessant voor het vakgebied kunstmatig intelligentie omdat dit soort inzichten een argumentatieve rol kunnen hebben in de discussie of bepaalde beslissingen moeten worden gemaakt door mens of machine.

Hoofstuk 2 Theoretisch kader

2.1 Suppression task

De suppression task onderbouwt de bevinding van het ontbreken van logisch redeneren tijdens een conditionele redentietask. Tijdens deze taak dienen proefpersonen de juiste gevolgstrekking te maken, gegeven verschillende premissen.

De taak kent, in zijn simpelste vorm, de volgende premisse:

Zin 1: "If she has an essay to write then she will study in the library"

De taak breidt zich vervolgens uit door een additionele en een alternatieve premisse toe te voegen:

Zin 2: "If she has a textbook to read then she she will study in the library."

Zin 3: "If the library stays open then she will study late in the library."

Ter verduidelijking noemen we het eerste gedeelte van zin 1 P (She has an essay to write), het tweede gedeelte (she will study in the library) van zin 1 noemen we Q.

De proefpersonen ontvangen naast een combinatie van verschillende premissen vier soorten vraagstellingen:

- (a) Conditie modus ponens (MP): Bij deze conditie wordt P gegeven en daarbij gevraagd of Q het geval is.
- (b) Conditie modus tollens (MT): Bij deze conditie wordt de ontkenning van Q gegeven en daarbij gevraagd of de ontkenning van P het geval is.
- (c) Conditie denial of antecedent (DA): Bij deze conditie wordt de ontkenning van P gegeven en daarbij gevraagd of de ontkenning van Q het geval is.
- (d) Conditie affirmation of consequent (AC): Bij deze conditie wordt Q gegeven en daarbij gevraagd of P het geval is.

Proefpersonen van de suppression task worden onderverdeeld in drie groepen. Groep 1 ontvangt premisse 1, groep 2 ontvangt premisse 1 en 2 en groep 3 ontvangt premisse 1 en 3. Alle drie de groepen dienen voor elke conditie de vraag te beantwoorden. Tabel 1 en 2 tonen achtereenvolgens de uitkomsten van de suppression task uitgevoerd in het experiment van Byrne (1989) en Dietz et al. (2012). De getallen in deze tabel weergeven het percentage proefpersonen dat de gestelde vraag met "ja" heeft beantwoord. Uit de uitkomsten van de suppression task blijkt dat het toevoegen van de verschillende premissen van invloed is op de getrokken conclusies. Om die reden kan worden geconcludeerd dat er niet volkomen logisch wordt geredeneerd.

Tabel 1 toont de resultaten van de suppression task uitgevoerd door Byrne (1989)

Byrne (1989)	Conditie MP	Conditie MT	Conditie DA	Conditie AC
Premisse 1	96%	92%	46%	71%
Premissen 1,2	96%	96%	4%	13%
Premissen 1,3	38%	33%	63%	54%

Tabel 2 toont de resultaten van de suppression task uitgevoerd door Dietz et al. (2012)

Dietz et al. (2012)	Conditie MP	Conditie MT	Conditie DA	Conditie AC
Premisse 1	96%	69%	46%	53%
Premissen 1,2	96%	96%	4%	16%
Premissen 1,3	38%	44%	36%	55%

2.2 Modellen

Het ontbreken van redenering volgens de klassieke logica tijdens de suppression task duidt op een andere strategie die aan het redeneringsproces ten grondslag ligt. Om dit zo goed mogelijk in kaart te brengen hebben Stenning & van Lambalgen (2012) een model ontwikkeld dat de uitkomsten van de suppression task voorspelt en verklaart. Naar dit model zal in het vervolg worden gerefereerd als het SL-model. Het is voor het doel van dit experiment niet noodzakelijk de theorieën achter dit model te bespreken. Wat wel van belang is, zijn de uitkomsten van het model. Het model van Stenning & van Lambalgen is afgebeeld in tabel 3. Hierin is per conditie, voor elke verschillende combinatie van premissen, het verwachte antwoord op de vraag die wordt gesteld te zien.

Tabel 3 Model Stenning & van Lambalgen (2012).

	Conditie MP	Conditie MT	Conditie DA	Conditie AC
premissie 1	Ja	Nee	Ja	Ja
premissen 1,2	Ja	Ja	Nee	Nee
premissen 3,4	Nee	Nee	Ja	Ja

Rosalie Iemhoff en Chris Janssen zetten hun vraagtekens bij enkele details van het model van Stenning & van Lambalgen (2012) en de theorieën die hieraan ten grondslag liggen en hebben als aanvulling op het oorspronkelijke model een eigen model ontwikkeld. Dit zal worden aangeduid als het nieuwe model. Ook bij dit model zal er niet worden ingegaan op de theorieën en verklaringen. De uitkomsten van dit nieuwe model zijn op dezelfde manier als het model van Stenning & van Lambalgen (2012) weergegeven in tabel 4.

Tabel 4 Model Iemhoff & Janssen

	Conditie MP	Conditie MT	Conditie DA	Conditie AC
premissie 1	Ja	Nee	Ja	Ja
premissen 1,2	Ja	Ja	Nee	Nee
premissen 3,4	Nee	Nee	Nee	Nee

Te zien is dat de twee modellen op conditie DA en AC van elkaar verschillen. Waar Stenning & van Lambalgen (2012) een “ja” verwachten op de vraag gesteld in deze condities, hebben Iemhoff & Janssen hun motivaties om hier een “nee” te verwachten.

2.3 Threshold

Gezien het feit dat de modellen een andere uitkomstmaat hebben dan de resultaten van de suppression task zelf is het van belang een mogelijkheid te vinden om dit toch met elkaar te vergelijken. Voor de suppression task wordt voor elke conditie het percentage vastgesteld dat “ja” antwoordt op de gestelde vraag. Terwijl de modellen geen percentage maar een binominaal antwoord, namelijk een “ja” of een “nee” tonen. Eén manier om de uitkomsten van het experiment en de twee modellen toch met elkaar te kunnen vergelijken, is gebruik te maken van een zogeheten ‘threshold’. Het vaststellen van een threshold geeft het verschil aan tussen het gewenste en ongewenste antwoord. Een threshold is het percentage participanten dat antwoord X moet hebben gegeven, om X het gewenste antwoord te laten zijn (De Jong, 2019). Op deze manier kan worden vergeleken of dit gewenste antwoord overeenkomt met de conclusies in het model. Een op het eerste gezicht voor de hand liggende threshold is 50%. Echter, het blijkt dat dit percentage, toegepast op de uitkomsten van Byrne (1989) en Dietz et al. (2012), niet overeenkomt met de conclusies van de twee modellen (De Jong, 2019). Wanneer de threshold wordt verhoogd naar 75% voor de resultaten van Byrne (1989) en 65% voor de resultaten van Dietz et al. (2012) blijkt dat de gewenste antwoorden voor een groot deel wel overeenkomen met de antwoorden van het model.

Voor de vergelijking van de experimentele resultaten en de twee modellen zal in dit onderzoek een threshold van 70% worden gehanteerd. Gezien de verwachting dat de uitkomsten van de suppression task in deze studie in lijn zijn met die van Dietz et al. (2012) en Byrne (1989) zal de threshold die in dit onderzoek wordt gehanteerd het gemiddelde zijn van de twee eerdere gevonden thresholds door de Jong (2019). Dit betekent dus dat 70% van de participanten X als antwoord moet geven, wil X het gewenste antwoord zijn. Vervolgens kan de vergelijking worden gemaakt of dit antwoord ook is terug te vinden in één van de twee modellen.

Hoofdstuk 3 Methodologie

3.1 Proefpersonen

Aan het onderzoek hebben 48 studenten deelgenomen (27 vrouwen, 21 mannen, leeftijd: 19-25). Geen van deze studenten had enige ervaring met logica.

De studenten werden willekeurig onderverdeeld in drie verschillende groepen, elk bestaande uit 16 proefpersonen.

Groep 1 beantwoordt voor elke conditie de vragen behorend bij premisse 1.

Groep 2 beantwoordt voor elke conditie de vragen behorend bij premisse 1 en 2.

Groep 3 beantwoordt voor elke conditie de vragen behorend bij premisse 1 en 3.

3.2 Experiment

Voor dit onderzoek is een deel van de originele suppression task, beschreven in het theoretische deel, uitgevoerd. Dit werd enkel uitgevoerd op de condities waar het SL-model en het nieuwe model van elkaar verschillen, namelijk conditie DA en AC.

De premissen en de bijbehorende vragen zijn hieronder per groep opgesteld.

Groep 1:

Vraag 1. Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek.

Stel: Marie moet geen essay schrijven.

Marie gaat niet naar de bibliotheek. Juist/Onjuist

Vraag 2. Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek.

Stel: Marie gaat naar de bibliotheek.

Marie moet een essay schrijven. Waar/Onwaar

Groep 2:

Vraag 1. Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek. Als Marie een boek moet lezen, gaat ze naar de bibliotheek.

Stel: Marie moet geen essay schrijven.

Marie gaat niet naar de bibliotheek. Waar/Onwaar

Vraag 2. Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek. Als Marie een boek moet lezen, gaat ze naar de bibliotheek.

Stel: Marie gaat naar de bibliotheek.

Marie moet een essay schrijven. Waar/Onwaar

Groep 3:

Vraag 1. Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek. Als de bibliotheek open is, gaat ze naar de bibliotheek.

Stel: Marie moet geen essay schrijven.

Marie gaat niet naar de bibliotheek. Waar/Onwaar

Vraag 2. Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek. Als de bibliotheek open is, gaat ze naar de bibliotheek.

Stel: Marie gaat naar de bibliotheek.

Marie moet een essay schrijven. Waar/Onwaar

3.3 Procedure

De vragen werden aan de proefpersonen aangeboden op papier. De eerste pagina bevatte uitleg van het experiment. Elke proefpersoon beantwoordde de vragen in stilte en zonder te overleggen. Er is geen restrictie op de tijd toegepast waarbinnen te proefpersonen de vragen moesten hebben beantwoord.

Na het lezen van de premisse(n) en vragen diende elke proefpersoon het gewenste antwoord “Waar” of “Onwaar” te omcirkelen.

3.4 Data-analyse

Aan de hand van de uitkomsten van de taak zijn per conditie percentages vastgesteld. Een percentage X betekent dat X procent van het totaal aantal roefpersonen voor antwoord optie “Waar” heeft gekozen.

Om de betrouwbaarheid van het onderzoek te waarborgen zijn verschillende statistische testen uitgevoerd. Allereerst zijn de percentuele uitkomsten van het experiment door middel van een Chi-kwadraattoets geanalyseerd en vergeleken met de uitkomsten van de suppression task in het experiment van Byrne (1989) en Dietz et al. (2012). Deze parameter vrije toets is gekozen om na te gaan of deze twee verdelingen van elkaar verschillen.

Aan de hand van de vooraf vastgestelde threshold van 70 % zijn deze resultaten op eenzelfde manier weergegeven als de twee modellen. Dit om de vergelijking met een model mogelijk te maken. Gezien het aantal proefpersonen is er met behulp van SPSS een binomiaal toets uitgevoerd om na te gaan of deze vergelijking terecht mag worden gemaakt. In deze test werd onderzocht of het percentage daadwerkelijk significant verschilt van de threshold van 70%, of dat dit toeval kan zijn geweest.

Hoofstuk 4 Onderzoeksresultaten

4.1 Resultaten suppression task

De resultaten van de suppression task, getest op conditie DA en AC zijn te zien in tabel 5. Een hoog percentage houdt in dat veel participanten de gevraagde conclusie met “Waar” hebben beantwoord.

Tabel 5: De resultaten van de suppression task, getoond in percentages

	DA	AC
Premisse 1	12,5%	31,25%
Premissen 1,2	6,25%	18,75%
Premissen 1,3	56,25%	18,75%

4.2 Vergelijking onderzoeksresultaten en resultaten van Byrne (1989) en Dietz et al. (2012)

Aan de hand van een Chi-kwadraattoets is gekeken naar de verschillen tussen de resultaten uit tabel 3 en de resultaten van de suppression task gevonden door Byrne (1989).

Voor premisse 1, conditie DA is de p-waarde waarde 0.0748, voor conditie AC is deze 0.0706. Voor premisse 1 en 2, conditie DA is de p-waarde 0.8234, voor conditie AC is deze 0.7283. Voor premisse 1 en 3, conditie DA is de p-waarde 0.7568, voor conditie AC is deze 0.0835.

Eenzelfde test is uitgevoerd om de verschillen tussen de resultaten uit tabel 5 en die van de suppression task uitgevoerd door Dietz et al. (2012) te berekenen.

Deze test leverde de volgende p-waardes op: Voor premisse 1, conditie DA is de p-waarde 0.0748, voor conditie AC is deze 0.3122. Voor premisse 1 en 2, conditie DA is de p-waarde 0.8234, voor conditie AC is deze 0.8710. Voor premisse 1 en 3 is de p-waarde voor conditie DA 0.3598, voor conditie AC is deze 0.0760

In het onderzoek van Byrne (1989) en Dietz et al. (2012) werden 8 proefpersonen per conditie gebruikt.

4.3 Resultaten na toepassing van de threshold

Tabel 6 laat op eenzelfde manier als de twee modellen de verwachte antwoorden zien op de suppression task. In deze tabel betekent “Nee” dat minder dan 70% de gevraagde conclusie met “Waar” heeft beantwoord.

Tabel 6: Resultaten na het toepassen van een threshold van 70%, getoond in binaire antwoorden.

	DA	AC
Premisse 1	Nee	Nee
Premissen 1,2	Nee	Nee
Premissen 1,3	Nee	Nee

4.4 Resultaten binomiaal toets

De uitkomsten van de binomiaal toets tonen per groep de volgende p-waardes (1-zijdig), indien vergeleken met een testproportie van 0,7:

Groep 1: conditie DA: 0.000, conditie AC: 0.002

Groep 2: conditie DA: 0.000, conditie AC: 0.000

Groep 3: conditie DA: 0.175, conditie AC: 0.000

Hoofdstuk 5 Conclusie

Op basis van de gevonden resultaten kan er een voorkeur worden ontleend aan het nieuwe model. Echter toont tabel 3 een te optimistisch resultaat. Uit de uitkomsten van de binomiaal toets blijkt dat, indien er een betrouwbaarheidsinterval van 95% wordt gehanteerd, op conditie DA voor groep 3 geen significant verschil is gevonden met de threshold van 70%. Dit houdt in dat het percentage 56,65% per toeval lager kan zijn dan de gestelde threshold. Voor deze conditie kan dus niet met zekerheid een “Nee” worden geconcludeerd. Voor de overige condities en groepen kan dit wel, wat resulteert in de volgende tabel.

Tabel 7: Resultaten na toepassen van threshold van 70% en binominale test met een testproportie van 0.7. “-“ betekent dat er op basis van de gevonden p-waarde geen uitspraak mag worden gedaan op deze conditie.

	DA	AC
Premisse 1	Nee	Nee
Premissen 1,2	Nee	Nee
Premissen 1,3	-	Nee

De vergelijking tussen tabel 7 en de twee modellen is ter verduidelijking visueel weergegeven in tabel 8. Gezien het feit dat het SL-model en het nieuwe model van elkaar verschillen op beide condities voor groep 3 zal hier de focus op liggen. Te zien in de tabel is dat de uitkomsten van het experiment op conditie AC voor groep 3 overeenkomst met het nieuwe model en verschilt met het model van Stenning & van Lambalgen (2012). Om deze reden kan als antwoord op de onderzoeksvraag worden gesteld dat het model van Iemhoff & Janssen beter in overeenstemming lijkt te zijn met de resultaten van de suppression task dan het model van Stenning & van Lambalgen (2012). Deze conclusie vraagt echter wel om wat nuance. Dit zal in de discussie aan bod komen.

Tabel 8: Vergelijking verkregen resultaten met het model van Stenning & van Lambalgen (2012) en Iemhoff & Janssen.

		DA	AC			DA	AC
SL-model & Nieuw model	Premisse 1	Ja	Ja	Vos	Premisse 1	Nee	Nee
SL-model & Nieuwe model	Premissen 1,2	Nee	Nee	Vos	Premissen 1,2	Nee	Nee
SL-Model	Premissen 1,3	Ja	Ja	Vos	Premissen 1,3	-	Nee
Nieuwe model	Premissen 1,3	Nee	Nee	Vos	Premissen 1,3	-	Nee

Hoofdstuk 6 Discussie

6.1 Nuance conclusie

De conclusie dat het nieuwe model meer in lijn is met de resultaten van de suppression task dan het SL-model wordt door het feit dat er geen significant antwoord kon worden gevonden voor conditie DA op groep 3 verzwakt. Deze studie kan helaas geen uitspraak doen over één belangrijk punt waarop de modellen met elkaar vergeleken kunnen worden. Daarnaast is te zien in tabel 6 dat zowel het nieuwe model als het SL-Model voor groep 1 op beide condities niet overeenkomt met de gevonden antwoorden. Dit heeft geen effect op welk model meer overeenkomt met de gevonden resultaten, aangezien beide modellen hierop verschillen. Het roept echter wel de vraag op of één van de twee modellen überhaupt de antwoorden op de suppression task juist weet weer te geven.

6.2 Meerdere thresholds

Daarnaast is er reden om de vooraf gekozen threshold van 70% in twijfel te trekken. De verwachting was dat de resultaten van de suppression task uitgevoerd in deze studie in lijn zouden zijn met die van Dietz et al. (2012) en Byrne (1989). Om die reden is voor een threshold van 70% gekozen. Echter blijken de resultaten toch op verschillende punten sterk af te wijken van de originele resultaten van Dietz et al. (2012) en Byrne (1989). Uit de resultaten van de Chi-kwadraattoets blijkt dat indien er een betrouwbaarheidsinterval van 90% zou worden gehanteerd, dit verschil op enkele condities significant is. Gezien de kleine steekproefgrootte voor alle condities kunnen we concluderen dat de gevonden p-waardes duiden op een groot verschil tussen de suppression task uitgevoerd in dit experiment en die van Byrne (1989) en Dietz et al. (2012).

Tabel 9: Weergave van de resultaten van de suppression task uitgevoerd door Vos (2019), na het toepassen van een binomiaal toets met als testproportie de verschillende thresholds.

	Threshold 50%	Threshold 55%	Threshold 60%	Threshold 65%	Threshold 75%	Threshold 80%	Threshold 85%	Threshold 90%																																
Premisse 1	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>-</td></tr> </table>	DA	AC	Nee	-	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee
DA	AC																																							
Nee	-																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
Premissen 1,2	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
Premissen 1,3	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>-</td><td>Nee</td></tr> </table>	DA	AC	-	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>-</td><td>Nee</td></tr> </table>	DA	AC	-	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>-</td><td>Nee</td></tr> </table>	DA	AC	-	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>-</td><td>Nee</td></tr> </table>	DA	AC	-	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>-</td><td>Nee</td></tr> </table>	DA	AC	-	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee	<table border="1"> <tr><td>DA</td><td>AC</td></tr> <tr><td>Nee</td><td>Nee</td></tr> </table>	DA	AC	Nee	Nee
DA	AC																																							
-	Nee																																							
DA	AC																																							
-	Nee																																							
DA	AC																																							
-	Nee																																							
DA	AC																																							
-	Nee																																							
DA	AC																																							
-	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							
DA	AC																																							
Nee	Nee																																							

Een andere threshold zou kunnen resulteren in andere uitkomsten. Een hogere threshold kan betekenen dat er op conditie DA voor groep 3 wel een significant verschil kan worden gevonden. Om een idee te geven wat verschillende thresholds zouden doen voor de gevonden resultaten is een overzicht hiervan weergegeven in tabel 9. Voor de totstandkoming van deze resultaten is een alfa-level van 0.05 gehanteerd, de statistische berekeningen zijn uitgevoerd in SPSS gebruikmakend van een binominaal toets. Gezien het feit dat een threshold het minimale percentage is dat antwoord X moet hebben gegeven om antwoord X het gewenste antwoord te laten zijn, is er geen reden om dit te weergeven voor een threshold lager dan 50%. Tabel 9 laat zien dat zelfs wanneer de threshold wordt verlaagd naar 50%, dit niet resulteert in een “ja” voor conditie DA in groep 3. Op conditie AC voor groep 3 wordt een significant verschil gevonden tussen de experimentele uitkomsten en elke mogelijke threshold. Deze twee bevindingen maken dat ondanks het feit dat de gekozen threshold van 70% misschien niet de juiste was, een verandering in de thresholds niet zal resulteren in een model wat de voorkeur ontleend aan dat van Stenning & van Lambalgen (2012).

6.3 Aantal proefpersonen

Nadat er verschillende punten aan bod zijn gekomen die de conclusie dan wel versterken of verzwakken is het belangrijk om aandacht te schenken aan het laatste punt wat invloed kan hebben op de gevonden conclusie. De keuze voor het aantal proefpersonen was gebaseerd op het originele experiment van Byrne (1989), waar per conditie 8 proefpersonen zijn gebruikt. Een verdubbeling van dit aantal leek voldoende. Echter is de kans dat dat er per toeval een groter of kleiner percentage wordt gevonden dan de gekozen threshold voor dit aantal vrij waarschijnlijk. Indien er meer proefpersonen zouden zijn kunnen dezelfde percentages toch een significant verschil opleveren. Dit kan belangrijk zijn voor conditie DA voor groep 3, zowel in het voordeel als in het nadeel van het nieuwe model.

6.4 Relevatie kunstmatige intelligentie

Zoals kort besproken in de inleiding kunnen bevindingen over het menselijk redeneren interessant zijn voor kunstmatige intelligentie. Het maken van een beslissing is sterk gerelateerd aan de manier waarop er, door een mens of machine, wordt geredeneerd (Pomerol, 1997). Voordat men een beslissing maakt zullen zij eerst redeneren. Zowel het SL-model als het nieuwe model proberen dit redeneringsproces in kaart te brengen, maar uit beide modellen blijkt dat er niet altijd logisch wordt geredeneerd. Dit betekent dat er ook niet altijd de meest logische beslissing zal worden gemaakt, iets wat in sommige gevallen wel de gewenste uitkomst is. Bijvoorbeeld in het geval van medische beslissingen, hier wordt een logische benadering als het meest duidelijk bevonden (Bury, 1973-2019).

Mensen trekken logische conclusies in twijfel wanneer er extra informatie wordt toegevoegd, ondanks dat deze informatie de conclusie die getrokken dient te worden niet beïnvloedt (Byrne, 1989). Gezien intelligente systemen op zo'n manier kunnen worden geprogrammeerd dat er puur logistisch naar bepaalde problemen gekeken kan worden, zonder assumpties over de context te maken, overstijgt het machinaal redeneren hier dat van de mens. Door onderzoek te doen naar het menselijk redeneren kan worden onderzocht wat de exacte punten zijn waarop machinaal redeneren dat van de mens overstijgt. Door inzichten als deze kunnen in situaties waar logisch naar een vraagstuk moet worden gekeken intelligente systemen worden ingezet. Indien de mens hier niet in staat is puur logisch te redeneren.

Eén van de doelen van een kunstmatig intelligent systeem is de manier waarop mensen redeneren te begrijpen (Gardner, 1984). Wanneer een systeem dit beheerst kan het op eenzelfde manier als de mens interacteren met zijn omgeving en hedendaagse problemen oplossen. Om dit te doen moet een intelligent systeem weten wat de redenatiestappen zijn die

in het menselijk brein worden gemaakt en tot wat voor een uitkomst dit vervolgens leidt. Door middel van soortgelijke modellen als getoond in deze studie kan een intelligent systeem hierin getraind worden (Le, Baydin, Zinkov, & Wood, 2017). Om deze reden is het van belang één correct model te ontwikkelen, iets waar aan deze studie zo goed mogelijk heeft proberen bij te dragen.

6.5 vervolgonderzoek

De resultaten van de Chi-kwadraattoets maken dat de vooraf vastgestelde threshold van 70% in twijfel kan worden getrokken. Ondanks dat een andere threshold niet tot andere resultaten zal lijden kan vervolgonderzoek zich richten op het vinden van de juiste threshold.

De focus van het huidige experiment lag slechts op de vergelijking tussen de twee modellen. De onderzoeksvraag is beantwoord en er kan een voorkeur worden ontleend aan het model van Iemhoff & Janssen. Desalniettemin volgt daar niet uit dat dit model het “ideale” model is. Zoals eerder beschreven verschilt het nieuwe model op twee punten met de uitkomsten van het experiment. Voor het beantwoorden van de onderzoeksvraag was dit niet van belang, maar het roept wel de vraag op of beide modellen in staat waren de uitkomsten van de suppression task op een juiste manier te weergeven. Voor het doel van het onderzoek is de suppression task slechts uitgevoerd op condities DA en AC. Eventueel vervolgonderzoek zou de suppression task kunnen herhalen, ditmaal volledig uitgevoerd en met meer proefpersonen per conditie. Een grotere proefpersoon omvang zou kunnen zorgen voor een significant resultaat op conditie DA. Deze bevinding zal de voorkeur voor het nieuwe model rechtvaardigen.

Echter dient er ook rekening mee gehouden te worden dat de bevinding van dit vervolgonderzoek is dat het nieuwe model de antwoorden op suppression task niet juist weet te weergeven. Indien dat het geval is biedt dit ruimte voor de ontwikkeling van een geheel nieuw model. De totstandkoming van beide modellen en achterliggende theorieën zijn in deze studie niet aan bod gekomen. De Jong (2019) geeft een duidelijk overzicht van zowel de totstandkoming van het model van Stenning & van Lambalgen (2012) als dat van Iemhoff & Janssen. Dit kan verduidelijking bieden bij het ontwerpen van een alternatief model.

Bronnen

Braine, M. D. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological review*, 85(1), 1-21

Bury, J. F. (1973-2019). Quantitative and Qualitative Approaches to Reasoning Under Uncertainty in Medical Decision Making. In *Lecture Notes in Computer Scienc.* Springer.

Copi, I.M. (1982). *Introduction to logic*. (6th ed.). London: MacMillan.

[Dietz et al., 2012] Dietz, E.-A., Ho'ldobler, S., and Ragni, M. (2012). A computational logic approach to the suppression task. In Proceedings of the Annual Meeting of the Cognitive Science Society, volume 34.

Evans, J. S. (1991). Theories of human reasoning: The fragmented state of the art. *Theory & Psychology*, 1(1), 83-105.

Evans, J. S (2002). St. BT (1982) *The psychology of deductive reasoning*. London: Routledge & Kegan Paul.

Gardner, A. V. (1984). An artificial intelligence approach to legal reasoning. *Doctoral Dissertation*.

Jong, C. d. (2019). Human Reasoning In the Suppression Task: The Relevance Of Commonsense Knowledge.

Le, T. A., Baydin, A. G., Zinkov, R., & Wood, F. (2017). *Using synthetic data to train neural networks is model-based reasoning* Publisher: IEEE.

Osherson, D.N. (1975). *Logical abilities in children. Volume 3*. Hillsdale. NJ: Erlbaum.

Pomerol, J.-C. (1997). Artificial intelligence and human decision making. *European Journal of Operational Research*, 3-25.

Rips, L. J. (1983a). Reasoning as a central intellectual ability. In R.J. Sternberg (Ed.), *Advances in the study of human intelligence*. Hillsdale: NJ: Erlbaum.

Rips, L. J. (1983b). Cognitive processes in propositional reasoning. *Psychological review*, 90, 38-71. Rips, L. J. (1988). Deduction. In R.J. Sternberg and E.E. Smith (Eds.), *The psychology of human thought*. Cambridge: Cambridge University Press.

Rumain, B., Connell, J., & Braine, M. D. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. *Developmental psychology*, 19(4), 471-481.

Sperber, D. & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge: Cambridge University Press

[Stenning and Van Lambalgen, 2012] Stenning, K. and Van Lambalgen, M. (2012). *Human reasoning and cognitive science*. MIT Press.

Bijlage

Leeftijd:

Geslacht:

Op de volgende bladzijde staat een aantal aannames, met een daarbij behorende vraag. Lees eerst de aannames en beoordeel of er een juiste conclusie wordt getrokken, door het gewenste antwoord (waar of onwaar) te omcirkelen.

Vraag 1 en 2 staan los van elkaar, beschouw dit dus als twee individuele vragen.

Vraag 1 Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek.

Stel: Marie moet geen essay schrijven.

Marie gaat niet naar de bibliotheek. Waar/Onwaar

Vraag 2 Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek.

Stel: Marie gaat naar de bibliotheek.

Marie moet een essay schrijven. Waar/Onwaar

Leeftijd:

Geslacht:

Op de volgende bladzijde staat een aantal aannames, met een daarbij behorende vraag. Lees eerst de aannames en beoordeel of er een juiste conclusie wordt getrokken, door het gewenste antwoord (onjuist of juist) te omcirkelen.

Vraag 1 en 2 staan los van elkaar, beschouw dit dus als twee individuele vragen.

Vraag 1 Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek. Als Marie een boek moet lezen gaat ze naar de bibliotheek.

Stel: Marie moet geen essay schrijven.

Marie gaat niet naar de bibliotheek. Waar/Onwaar

Vraag 2 Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek. Als Marie een boek moet lezen, gaat ze naar de bibliotheek.

Stel: Marie gaat naar de bibliotheek.

Marie moet een essay schrijven. Waar/Onwaar

Leeftijd:

Geslacht:

Op de volgende bladzijde staat een aantal aannames, met een daarbij behorende vraag. Lees eerst de aannames en beoordeel of er een juiste conclusie wordt getrokken, door het gewenste antwoord (onjuist of juist) te omcirkelen.

Vraag 1 en 2 staan los van elkaar, beschouw dit dus als twee individuele vragen.

Vraag 1 Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek. Als de bibliotheek open is, gaat ze naar de bibliotheek.

Stel: Marie moet geen essay schrijven.

Marie gaat niet naar de bibliotheek. Waar/Onwaar

Vraag 2 Stel: Als Marie een essay moet schrijven, gaat ze naar de bibliotheek. Als de bibliotheek open is, gaat ze naar de bibliotheek.

Stel: Marie gaat naar de bibliotheek.

Marie moet een essay schrijven. Waar/Onwaar