Rathenau Instituut

Online ontspoord

Een verkenning van schadelijk en immoreel gedrag op het internet in Nederland

Rapport

Auteurs
Mariëtte van Huijstee, Wouter Nieuwenhuizen, Mathilde Sanders, Eef Masson en Pieter van Boheemen

Foto omslag
Shutterstock

Bij voorkeur citeren als:
Rathenau Instituut (2021). *Online ontspoord – Een verkenning van schadelijk en immoreel gedrag op het internet in Nederland*. Den Haag (auteurs: Huijstee, M. van, W. Nieuwenhuizen, M. Sanders, E. Masson en P. van Boheemen)

Voorwoord

Pedojagen. Phishing. Cyberverslaving. Wraakporno. Desinformatie. Zomaar wat voorbeelden van schadelijk en immoreel gedrag op het internet. Door specifieke eigenschappen van het internet, zoals het grote bereik van online berichten, de (gevoelsmatige) anonimiteit van mensen achter hun scherm, of de onmiddellijkheid waarmee een video wereldwijd te zien is, worden dit soort gedragsfenomenen gefaciliteerd. Dat richt grote schade aan, voor individu en samenleving.

Maar over wat voor gedrag hebben we het eigenlijk? Hoe vaak komt het voor? En wat kan de overheid hiertegen doen? Op verzoek van het WODC deed het Rathenau Instituut onderzoek naar schadelijk en immoreel gedrag online. We hielden interviews en voerden gesprekken met deskundigen uit beleid, wetenschap en praktijk, en deden een literatuurstudie van wetenschappelijke en journalistieke bronnen en beleidsstukken. Op basis hiervan en met de expertise uit eerder onderzoek van het Rathenau Instituut, introduceren we een taxonomie waarin we 22 fenomenen van online schadelijk en immoreel gedrag categoriseren: van online informatiemanipulatie tot haat en zelfbeschadiging.

Uit ons onderzoek blijkt dat alle internetgebruikers in Nederland vroeg of laat te maken kunnen krijgen met schadelijk en immoreel gedrag online. Burgers zijn op het internet onvoldoende beschermd; online komen hun grondrechten in het geding. Lange tijd leek het internet een domein van zelfregulering en zelfredzaamheid in de samenleving. Maar om schadelijk en immoreel gedrag tegen te gaan, is een actieve overheid nodig. Een overheid die niet alleen reageert wanneer gedrag al is ontspoord, maar die ook proactief ingrijpt in de online omgeving, zodat schade wordt voorkomen en grondrechten van burgers worden beschermd. We presenteren een strategische agenda waarmee de overheid, in samenwerking met bedrijven, maatschappelijke organisaties en burgers, grip kan krijgen op online schadelijk en immoreel gedrag.

Het Rathenau Instituut doet al 35 jaar onderzoek naar de impact van technologie op de samenleving. Met dit rapport wil het Rathenau Instituut bijdragen aan de maatschappelijke dialoog over welk gedrag online wenselijk en toelaatbaar is. Online doen zich steeds weer nieuwe fenomenen voor. En morele normen zijn aan verandering onderhevig. Een actieve overheid en publiek debat over online schadelijk en immoreel gedrag zijn daarom noodzakelijk.

Dr. ir. Melanie Peters
Directeur Rathenau Instituut

Samenvatting

Inleiding
Het internet heeft bepaalde kenmerken waardoor online gedrag makkelijk ontspoort. Een persoon die op straat nooit een voorbijganger zou uitschelden, kan daar op Twitter geen moeite mee hebben. Iemand die nooit uit de buurtsuper zou stelen, kan online een lagere drempel ervaren en overgaan tot creditcarddiefstal. In het boek *Evil Online*, geschreven door Dean Cocking en Jeroen van den Hoven in 2018, wordt het internet geduid als een omgeving waarin schadelijk en immoreel gedrag geïnspireerd, gefaciliteerd en aangejaagd wordt. Voor het Ministerie van Justitie en Veiligheid vormde dit boek de aanleiding om de vraag te stellen wat de status is van dergelijke 'ontsporingen' in Nederland.

Het WODC verzocht het Rathenau Instituut om de volgende centrale onderzoeksvraag te beantwoorden: *Wat zijn de aard en de omvang van online schadelijk en immoreel gedrag in Nederland, wat zijn de onderliggende mechanismen en oorzaken, en welke handelingsperspectieven zijn er voor het ministerie en de overheid als geheel voor het beperken van schadelijk en immoreel gedrag op internet?*

Met dit rapport zet het Rathenau Instituut een schijnwerper op online gedrag dat zich in een moreel schemergebied bevindt, en waar de overheid nu nog handelingsverlegen is. Het gaat om online gedrag dat als schadelijk en/of immoreel kan worden geduid. Dat gedrag kan schadelijk zijn voor individuen, maar ook voor grotere groepen of de samenleving als geheel. Een deel van het gedrag dat we in dit onderzoek bespreken is in strijd met bepaalde grondrechten en wetten, en daarmee onrechtmatig of strafbaar. Toch blijkt het voor internetgebruikers online een stuk lastiger om te beoordelen wanneer iets door de beugel kan. De online omgeving is niet de facto wettelozer of grenzelozer dan de offline wereld, maar wordt wel sneller zo ervaren.

Het Rathenau Instituut introduceert met dit rapport voor het eerst een overzicht van schadelijk en immoreel gedrag online in Nederland door middel van een taxonomie. Deze taxonomie kan als kapstok dienen voor een gecoördineerde aanpak door de Rijksoverheid, in samenwerking met bedrijfsleven en maatschappelijke actoren. Het Rathenau Instituut wil met dit onderzoek ook bijdragen aan de maatschappelijke discussie over welk gedrag online wenselijk en toelaatbaar is. We weten dat morele normen aan verandering onderhevig zijn en publiek debat hierover noodzakelijk is.

Aanpak

Het rapport behandelt de volgende deelvragen:

1. Wat is de taxonomie van de online gedragingen en online fenomenen die schadelijk kunnen zijn voor individuen of groepen, en daarmee van invloed kunnen zijn op de morele infrastructuur van de samenleving?
2. Wat is in Nederland de aard van deze problematische gedragingen en fenomenen?
3. Wat is in Nederland de schaal waarop de problematische gedragingen en fenomenen zich voordoen, in termen van actoren, slachtoffers en maatschappelijke schade?
4. Hoe zijn deze problematische gedragingen en fenomenen en de daaruit voortvloeiende maatschappelijke schade, verbonden aan de werking, onderliggende mechanismen en inrichting van de online omgeving? Met andere woorden: hoe is de online wereld een facilitator en katalysator voor schadelijke uitingen en gedragingen op internet en sociale media?
5. Welke handelingsopties zijn er, nationaal en internationaal, al ontwikkeld voor het beperken van schadelijk en immoreel gedrag online en de maatschappelijke schade die daaruit voortvloeit, en welke lessen zijn daar uit te trekken?
6. Welke handelingsopties heeft de Nederlandse overheid?

Bij het beantwoorden van elk van de deelvragen is gebruik gemaakt van een combinatie van methoden, bestaande uit literatuuronderzoek, interviews, workshops en bijeenkomsten met deskundigen uit beleid, praktijk en wetenschap. In totaal hebben 56 deskundigen uit wetenschap, beleid en praktijk aan het onderzoek bijgedragen.

Aard en omvang van schadelijk en immoreel gedrag online

Dit onderzoek brengt voor het eerst schadelijk en immoreel gedrag online in Nederland in beeld in al zijn facetten. Het Rathenau Instituut ontwikkelde een taxonomie met zes categorieën van schadelijk en immoreel gedrag online met daaronder 22 verschillende fenomenen waar alle internetgebruikers in Nederland vroeg of laat mee te maken kunnen krijgen.

Bron: Rathenau Instituut

Figuur 1 Taxonomie van schadelijk en immoreel gedrag online

Taxonomie van immoreel en schadelijk gedrag online[1]

De schade die het gedrag in deze taxonomie veroorzaakt, kan ernstig zijn voor individuen, groepen en de samenleving als geheel. Deze kan variëren van een tienermeisje dat zichzelf uithongert omdat ze online in een extreme challenge terecht komt met leeftijdsgenoten, en vrouwelijke journalisten en wetenschappers die zichzelf niet meer durven uitspreken, tot maatschappelijke ontwrichting door de verspreiding van complottheorieën en desinformatie.

Uit het beeld dat de experts en literatuur schetsen over de aard en omvang van deze fenomenen in de taxonomie wordt duidelijk, dat alle Nederlanders het risico lopen om als slachtoffer, dader of omstander betrokken te raken bij dit gedrag. Iedereen kan te maken krijgen met schadelijk en immoreel gedrag online. Toch lopen bij bepaalde fenomenen sommige groepen meer risico dan anderen, afhankelijk van hun leeftijd, geslacht, etniciteit, seksuele voorkeur,

1 Zie begrippenlijst in dit rapport voor definities van de fenomenen.

geloofsovertuiging of opleidingsniveau. Het is op basis van de beschikbare gegevens lastig om hierover algemene uitspraken te doen.

Het onderzoek laat zien dat voor diverse fenomenen tot op heden nauwkeurige definities en systematische metingen ontbreken. Het is niet zinvol om te proberen vast te stellen welk fenomeen het meest zorgelijk is, omdat dit afhangt van de gekozen criteria: het aantal slachtoffers, de ernst van de schade, of de mogelijke schade in de toekomst. We concluderen dat alle fenomenen op hun eigen wijze zorgelijk zijn, voor de maatschappij als geheel, voor individuen of groepen individuen.

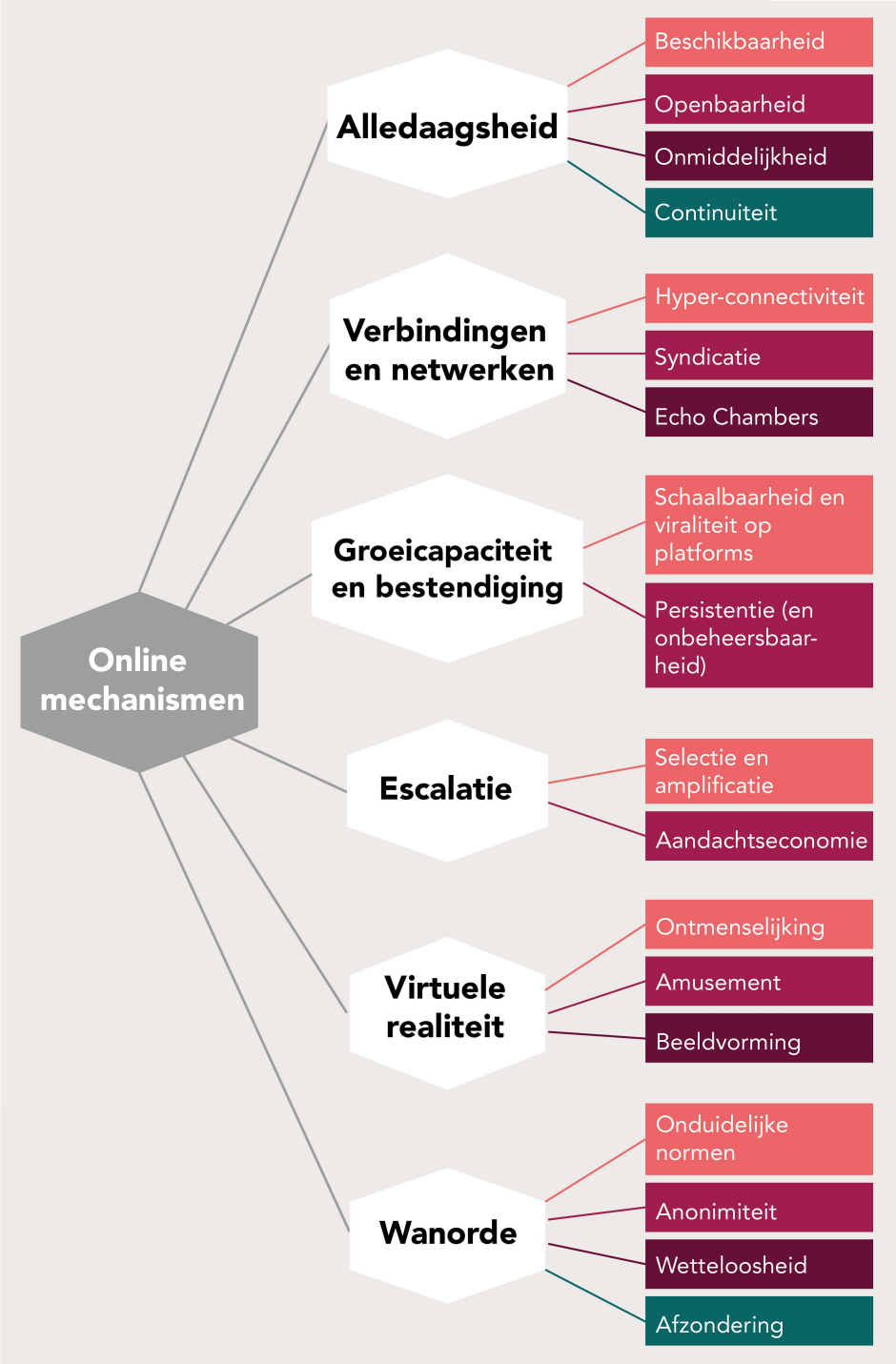
Mechanismen

Het internet als omgeving kent mechanismen die van invloed zijn op menselijk gedrag. Mensen kunnen door die online mechanismen bijvoorbeeld anders met normen en regels omgaan. Naast de mechanismen van het internet zijn er nog veel meer factoren van invloed op menselijk gedrag, zoals sociale, psychologische, culturele en economische oorzaken. Al deze factoren spelen een rol bij de ontwikkeling van schadelijk en immoreel gedrag online. Dit rapport richt zich op de mechanismen die kenmerkend zijn voor het internet.

Het onderzoek heeft in totaal 18 online eigenschappen en mechanismen geïdentificeerd die een rol spelen in het inspireren, faciliteren en aanjagen van schadelijk en immoreel gedrag online: 1) beschikbaarheid, 2) openbaarheid, 3) onmiddellijkheid, 4) continuïteit, 5) hyper-connectiviteit, 6) syndicatie, 7) echo chambers, 8) schaalbaarheid en viraliteit op platformen, 9) persistentie (en onbeheersbaarheid), 10) selectie en amplificatie, 11) aandachtseconomie, 12) ontmenselijking, 13) amusement, 14) beeldvorming, 15) onduidelijke normen, 16) anonimiteit, 17) (schijnbare) wetteloosheid, 18) afzondering. Deze mechanismen zijn samengebracht onder zes beschrijvende kenmerken van het internet:

1. Alledaagsheid
2. Verbindingen en netwerken
3. Groeicapaciteit en bestendiging
4. Escalatie
5. Virtuele realiteit
6. Wanorde

Een overzicht van alle mechanismen en hun indeling is te vinden in de figuur hieronder.



Bron: Rathenau Instituut

Figuur 2 Overzicht van online mechanismen

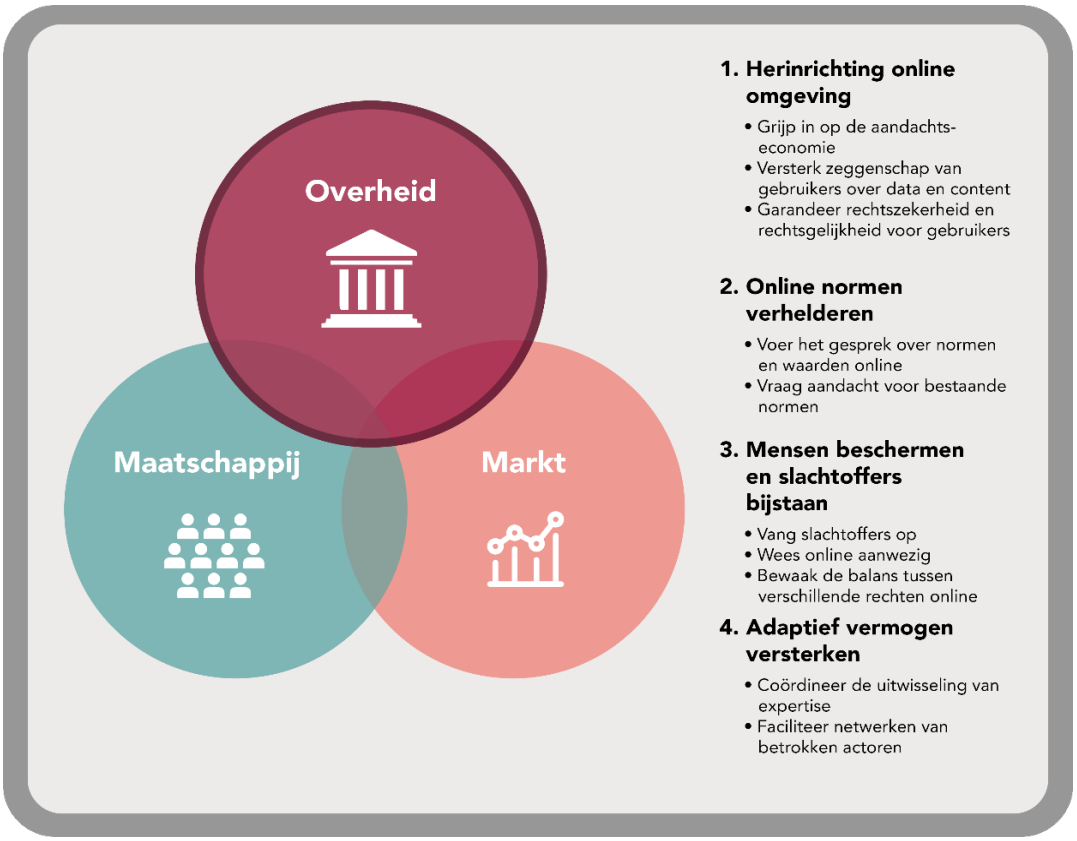
De casuïstiek in het rapport laat zien dat bij zeer verschillende fenomenen dezelfde mechanismen een rol kunnen spelen, en dat de mechanismen in combinatie voorkomen. Zo spelen syndicatie (het gemakkelijk vinden van gelijkgestemden online) en viraliteit (snelle, oncontroleerbare verspreiding van content online) zowel in de Casus online shaming, de Casus desinformatie en de Casus verstoord eetgedrag een rol. Ingrijpen in de mechanismen, zoals het vereisen van transparantie van de aanbevelingsmechanismen van online content of het opheffen van anonimiteit van internetgebruikers in bepaalde omgevingen, heeft zin bij het voorkomen of beperken van schadelijk en immoreel gedrag online. Maar dergelijke interventies vereisen zorgvuldige afweging en maatschappelijk debat. Die mechanismen van het internet kunnen namelijk ook leiden tot sociaal wenselijk gedrag en maatschappelijke verdiensten. Zo maakt anonimiteit online het voor klokkenluiders mogelijk om maatschappelijke misstanden aan te kaarten. Met ingrijpen in deze mechanismen worden mogelijk ook de positieve effecten beperkt of teniet gedaan.

Handelingsperspectieven
Het internet leek altijd een domein van zelfregulering en zelfredzaamheid van de samenleving, waar de overheid geen rol had en gebruikers zichzelf wel zouden redden. Uit dit onderzoek blijkt echter dat grondrechten in het geding zijn; burgers zijn op het internet onvoldoende beschermd. Bedrijven, maatschappelijke organisaties en burgers hebben een actieve overheid nodig om schadelijk en immoreel gedrag online tegen te gaan en sociaal wenselijk gedrag online te bevorderen.

Het rapport bevat een overzicht van bestaande maatregelen die overheden, bedrijven, hulpverleners en anderen al genomen hebben om schadelijk gedrag online aan te pakken. Dit overzicht van bestaande initiatieven geeft inzicht in de interventies die al werken en veelbelovend zijn om schadelijk en immoreel gedrag online te beperken of te voorkomen. Maar het laat ook zien waar in de aanpak nog gaten vallen en dus ruimte is voor extra interventies. De belangrijkste observatie is dat veel van de huidige initiatieven vooral *reactief* zijn van aard. Ze zijn met name gericht op de bestrijding van symptomen van schadelijk en immoreel gedrag, en nauwelijks op de onderliggende mechanismen. Daarbij zien we wel verschillen tussen de diverse actoren. Met name overheden en platformbedrijven zijn vooralsnog weinig proactief bezig. Bij platformbedrijven is dat niet zo verwonderlijk. Sleutelen aan mechanismen betekent immers dat de keuze gemaakt moet worden voor een alternatieve vorm van platformontwerp. Maar dit brengt onzekerheden mee ten aanzien van verdienmodellen – en bedrijven bewegen zich nu eenmaal binnen een competitieve markt. In de praktijk zijn het dan ook vooral andere, kleinschaliger partijen die met alternatieve vormen van ontwerp experimenteren.

De analyse van bestaande maatregelen laat zien dat overheden vooral in actie komen als gedragingen uit de hand lopen en dus in toom gehouden moeten worden. Hun interventies zijn tot nu toe vooral reactief. Het overzicht in dit rapport van de mechanismen die online schadelijk gedrag in de hand werken, kan overheden en andere partijen helpen om pro-actiever op te treden.

Op basis van interviews en gesprekken met deskundigen uit beleid, wetenschap en praktijk, vele wetenschappelijke, journalistieke bronnen en beleidsstukken, en de expertise uit eerder onderzoek en analyse van het Rathenau Instituut, introduceren we een strategische agenda voor de Rijksoverheid. Hierin identificeren we vier thema's waarbinnen de Rijksoverheid een sturende, coördinerende en faciliterende rol kan vervullen om in samenwerking met actoren uit de markt en maatschappij, schadelijk en immoreel gedrag online aan te pakken en een veilige online omgeving kan bevorderen.



Bron: Rathenau Instituut

Figuur 3 Strategische agenda voor de aanpak van schadelijk en immoreel gedrag online

Het eerste thema – *Herinrichting van de online omgevin*g – bevat handvatten voor de Rijksoverheid om de online mechanismen die schadelijk en immoreel gedrag online mede veroorzaken, ten goede te keren. Zo doet het rapport een aantal suggesties om in te grijpen in de aandachtseconomie online. Het tweede thema – *Online normen verhelderen* – gaat in op de rol van de Rijksoverheid, markt en maatschappij bij het vernieuwen van de maatschappelijke afspraken over normen en waarden online. De handelingsopties bij dit thema beogen een bredere bewustwording en groter begrip van schadelijk en immoreel gedrag in de samenleving te bewerkstelligen. Het derde thema – *Mensen beschermen en slachtoffers bijstaan* – bevat suggesties voor de Rijksoverheid, handhaving en uitvoeringsorganisaties om beter te reageren op de fenomenen van schadelijk en immoreel gedrag online en de schade hiervan. Zo doen we een aantal voorstellen voor de overheid om meer zichtbaar en aanwezig te zijn online. Het vierde thema – *Adaptief vermogen versterken* – bevat suggesties voor de Rijksoverheid om grip te krijgen en te houden op schadelijk en immoreel gedrag online, dat continu in beweging is. Deze handelingsopties zijn gericht op het toekomstbestendig maken van de strategische agenda.

Inhoud

Begrippenlijst

Onderstaande definities zijn gebaseerd op wetenschappelijk onderzoek en journalistieke bronnen. Bronverwijzing vindt u terug in hoofdstuk 3, waarin we de door het Rathenau Instituut ontwikkelde taxonomie van schadelijk en immoreel gedrag online presenteren.

Cancellen: online shaming waarbij wordt opgeroepen om iemand uit te sluiten uit diens community bij wijze van sociale straf.

Catfishing: het opzettelijk misleiden van anderen door delen van je eigen identiteit te verbergen of te veranderen, meestal in de context van online dating en soms zonder de intentie om iemand ooit in het echt te ontmoeten.

Complotdenken: de overtuiging dat bepaalde gebeurtenissen of situaties geen toeval zijn, maar in het geheim zijn gemanipuleerd door machtige groepen met verkeerde bedoelingen.

Cyberverslaving: excessieve en ongecontroleerde online activiteit met langdurig gebruik van internet in het algemeen en sociale media, gaming en pornosites in het bijzonder.

Cyberchondrie: excessief of herhaaldelijk zoeken naar informatie over gezondheid leidend tot onnodige paniek of zorgen over de gezondheid.

Cryptofraude: oplichting waarbij mensen worden aangemoedigd cryptomunten te (ver)kopen, soms in de vorm van een piramidespel, bijvoorbeeld om de beurswaarde van cryptomunten te manipuleren.

Cyberbullying: herhaaldelijk online pesten door een groep of individu van een slachtoffer dat zich niet gemakkelijk kan verdedigen.

Challenges: het aanmoedigen van mensen om bepaalde (gevaarlijke) opdrachten uit te voeren en vervolgens online te delen.

Desinformatie: het verspreiden van informatie die ‘onjuist’ of ‘misleidend’ is met de intentie om kwaadwillend te handelen of schade te berokkenen.

Digitaal vigilantisme: een vorm van collectieve actie, morele afkeuring of terechtwijzing (bijvoorbeeld via online shaming, intimidatie of doxing) gericht aan personen die ongewenst sociaal gedrag vertonen.

Doxing: het openbaar maken van iemands persoonlijke, sensitieve en privé-informatie zoals adres, telefoonnummer, paspoort, werkgever, gegevens van familie en foto's van iemands kinderen.

Extreme pranks: een vorm van vernedering en voor de gek houden (plagerige grap) die plaatsvindt tussen dader, slachtoffer en omstanders. Online nemen pranks vaak de vorm aan van gefilmde "offline" pranks, waarbij de reactie van slachtoffers (verwarring, verbijstering, ongemak of schaamte) breed wordt uitgemeten.

Griefing: doelbewust irriteren van andere spelers in online games door bepaalde spelelementen zo te manipuleren dat andere spelers daar last van hebben.

Grooming: een proces waarbij een volwassene een relatie van seksueel misbruik ontwikkelt door middel van technologie, zoals op sociale media. In dit verband wordt ook gesproken van 'digitale kinderlokkerij'.

Haatzaaien: alle uitingsvormen die zorgen voor de verspreiding, aanstichting, aanmoediging of legitimering van raciale haat, xenofobie, antisemitisme of andere vormen van haat gebaseerd op intolerantie.

Hacking: alle soorten activiteiten rondom het ongeautoriseerd toegang (proberen te) verkrijgen tot computersystemen.

Kwakzalverij: onbevoegd uitoefenen van de geneeskunst door iemand die beweert ziekten te kunnen genezen met een nutteloos of zelfs schadelijk geneesmiddel.

Pedojagen: vorm van digitaal vigilantisme waarbij burgers zich voordoen als kind om pedofielen 'in de val' te lokken en ze vervolgens zelf te straffen of ze aan te geven bij de politie.

Phishing: manier om via het internet allerlei soorten informatie te ontfutselen over personen, bijvoorbeeld via valse e-mails of websites.

Pro-ana coach: persoon die jonge meisjes met een eetstoornis aanmoedigt om verder af te vallen, veelal met als doel seksueel expliciet materiaal van (minderjarige) meisjes te bemachtigen.

Sexting: verspreiden of delen van seksueel getinte berichten, foto's of video's van zichzelf via mobiele telefoons of andere media.

Sextortion: een vorm van afpersing waarbij een dader dreigt om zonder toestemming seksueel beeldmateriaal te openbaren van een slachtoffer om deze te

dwingen om meer van dit soort foto’s te sturen, geld te betalen of om (seksueel getinte) opdrachten uit te voeren.

Shaming: vorm van digitaal vigilantisme waarbij publieke morele kritiek online wordt geuit als reactie op het overschrijden van sociale normen.

Shame sexting: het zonder toestemming maken en/of verspreiden van seksueel getinte beelden of video’s.

Stalking: het herhaaldelijk intimideren, lastigvallen en soms bedreigen van slachtoffers.

Sock puppet: een valse identiteit (sokpop) die wordt aangenomen om anderen te misleiden. Een sock puppet kan op sociale media bijvoorbeeld worden ingezet voor catfishing of trolling.

Trolling: het opzettelijk dwars zitten van mensen in online gemeenschappen met gedrag dat niet als acceptabel wordt gezien, zoals mensen uitschelden, ruzie zoeken of zich negatief uitlaten over anderen. Daarnaast bestaat tegenwoordig ook een bredere interpretatie van het begrip trolling, namelijk het gebruik van nepaccounts om desinformatie te verspreiden en het publieke debat te beïnvloeden.

Wraakporno: zonder toestemming bezitten, openbaar maken en verspreiden van (gestolen) seksueel beeldmateriaal door bijvoorbeeld hackers, (ex)partners, kindermisbruikers, verkrachters en mensenhandelaren.

Verstoord eetgedrag (eetstoornissen): psychische stoornissen die worden gekenmerkt door verstoord eetgedrag en/of inadequaat compensatiegedrag (braken, laxeren). Mensen met een eetstoornis hebben een verstoord lichaamsbeeld, zijn veel bezig met hun gewicht of lichaamsvorm en zijn erg bang om aan te komen.

1 Inleiding

Het internet heeft bepaalde kenmerken waardoor online gedrag gemakkelijk ontspoort. Een persoon die op straat nooit een voorbijganger zou uitschelden, kan daar op Twitter geen moeite mee hebben. Iemand die nooit uit de buurtsuper zou stelen, kan online een lagere drempel ervaren en overgaan tot creditcarddiefstal. In het boek *Evil Online* (Cocking & van den Hoven, 2018), wordt het internet geduid als een omgeving waar schadelijk en immoreel gedrag geïnspireerd, gefaciliteerd en aangejaagd wordt. Voor het Ministerie van Justitie en Veiligheid vormde dit boek de aanleiding om de vraag te stellen wat de status is van dergelijke 'ontsporingen' in Nederland. Het Ministerie stelde de volgende centrale onderzoeksvraag aan WODC: *Wat is de aard en de omvang van online schadelijk en immoreel gedrag in Nederland, wat zijn de onderliggende mechanismen en oorzaken, en welke handelingsperspectieven zijn er voor het Ministerie en de overheid als geheel voor het beperken van online schadelijk en immoreel gedrag op internet?*

Het WODC verzocht het Rathenau Instituut dit onderzoek uit te voeren. Kerntaak van het Rathenau Instituut is het in kaart brengen van de maatschappelijke effecten van technologieën en het aanreiken van handelingsperspectieven voor het beschermen van het publieke belang. Vertrekpunt voor al ons onderzoek is het beschermen van publieke waarden in relatie tot technologische ontwikkeling: welke waarden zijn in het geding, en wat is de rol van overheid, burgers en bedrijfsleven om deze waarden te beschermen? Vanuit deze kerntaak heeft het Rathenau Instituut veel ervaring met onderzoek naar de schadelijke effecten van internettechnologieën, en het aanreiken van handelingsperspectieven.

De centrale vraag valt in een aantal deelvragen uiteen, die in de opeenvolgende hoofdstukken in dit rapport behandeld worden.

1. Wat is in algemene zin de taxonomie van de online gedragingen en online fenomenen die schadelijk kunnen zijn voor individuen of groepen en daarmee van invloed kunnen zijn op de morele infrastructuur van de samenleving?
2. Wat is in Nederland de aard van deze problematische gedragingen en fenomenen?
3. Wat is in Nederland de schaal waarop de problematische gedragingen en fenomenen zich voordoen, in termen van actoren, slachtoffers en maatschappelijke schade?

4. Hoe zijn deze problematische gedragingen en fenomenen en de daaruit voortvloeiende maatschappelijke schade verbonden aan de werking, onderliggende mechanismen en inrichting van de online omgeving? Met andere woorden: hoe is de online wereld een facilitator en katalysator voor schadelijke uitingen en gedragingen op internet en sociale media?
5. Welke handelingsopties zijn er nationaal en internationaal al ontwikkeld voor het beperken van online schadelijk en immoreel gedrag en de maatschappelijke schade die eruit voortvloeit, en welke lessen zijn daar uit te trekken?
6. Welke handelingsopties lijken geschikt voor de Nederlandse overheid?

De concepten ‘online’, ‘schadelijk en immoreel’ in de onderzoeksvragen verdienen nadere toelichting.

1.1 Online

Ons dagelijks leven is in tal van aspecten verbonden met het internet. Daardoor valt het onderscheid tussen online en offline soms lastig te maken. Wat we online doen, heeft ook offline impact, en andersom. Toch valt er een onderscheid te maken tussen handelingen die niet kunnen bestaan zonder het internet (zoals een bedrijfswebsite beheren, foto’s posten en chatten op sociale media, of online gamen) en gedrag waarvoor geen internet nodig is (zoals een wandeling maken in het bos). Ook is er gedrag dat voorheen al bestond (zoals pesten), maar dat sinds het internet een online variant gekregen heeft (cyberpesten). In dit onderzoek kijken we naar schadelijk en immoreel gedrag online. Dus cyberpesten in tegenstelling tot alle vormen van pesten en online discriminatie in plaats van discriminatie in algemene zin. Het onderzoek gaat ook in op het verschil tussen offline en online gedragingen. We kijken naar onderliggende mechanismen die kenmerkend zijn voor het internet en online gedrag beïnvloeden.

1.2 Schadelijk en immoreel

In dit onderzoek kijken we naar online gedrag dat als schadelijk en/of immoreel kan worden geduid. Het gaat om gedrag waarvan internetgebruikers niet altijd helder voor ogen hebben waar de sociale en morele grenzen liggen. Dat gedrag kan schadelijk zijn voor individuen, maar ook voor grotere groepen of de samenleving als geheel. Als samenleving hebben we morele normen vastgelegd in wet- en regelgeving, in mensenrechtenverdragen en in impliciete onderlinge sociale afspraken. Rechten, waaronder grondrechten, gelden zowel online als offline. Bijvoorbeeld de vrijheid van meningsuiting van de persoon die een uiting doet op

internet (zoals een pro-ana blog of tweet) en de rechten en belangen van de ander, (zoals iemand die zich laat meeslepen door pro-ana content en wiens gezondheid wordt geschaad). Niet elke uiting wordt beschermd door de vrijheid van meningsuiting: haatzaaien wordt bijvoorbeeld niet beschermd.

Sommige gedragingen die we in dit onderzoek bespreken, kunnen in strijd zijn met bepaalde grondrechten en wetten en daarmee onrechtmatig of strafbaar zijn. Waar voor een rechter de grenzen van het toelaatbare online wellicht helder zijn, blijkt het voor mensen online een stuk lastiger om te beoordelen wanneer iets door de beugel kan. De online omgeving is dus niet de facto wettelozer of grenzelozer dan de offline wereld, maar wordt wel sneller zo ervaren. Verschillen in morele opvattingen en vaagheid van morele grenzen op het internet, oftewel morele mist online, kunnen in de online omgeving allerlei schadelijke gevolgen hebben. Dit onderzoek richt zich op gedrag dat zich in dit morele schemergebied bevindt. De overheid is hier nu nog handelingsverlegen, en is zoekende naar een passende manier om grondrechten online te beschermen. Het Rathenau Instituut wil met dit onderzoek bijdragen aan het ontwikkelen van een overheidsinstrumentarium en aan de maatschappelijke discussie over wat we online wenselijk vinden. Morele normen zijn aan verandering onderhevig en publiek debat hierover is noodzakelijk.

1.3 Afbakening

Het gaat in dit onderzoek dus over schadelijk en immoreel gedrag in relatie tot het internet, waarbij het internet een faciliterende, aanjagende of inspirerende rol speelt. We onderzoeken niet al het gedrag dat mogelijk schadelijk of immoreel is. We onderzoeken ook niet al het mogelijke morele, positieve, altruïstische gedrag dat online plaatsvindt. En we kijken niet naar alle mogelijke mechanismen die een rol spelen bij schadelijk en immoreel gedrag online, zoals sociale, psychologische of economische factoren, maar alleen naar mechanismen die kenmerkend zijn voor het internet. De fenomenen die binnen het bereik van het onderzoek vallen, zijn ondergebracht in een taxonomie met zes categorieën en in totaal 22 fenomenen van schadelijk en immoreel gedrag online, 17 onderliggende mechanismen en een agenda voor de overheid aan de hand van vier strategische thema’s.

Bij het lezen van het rapport is het belangrijk de volgende aspecten in het achterhoofd te houden:

1. Het gaat hier om een verkennend, exploratief onderzoek met een breed onderzoeksveld. De inzichten moeten dan ook niet als uitputtend worden beschouwd, maar als een stap in de bewustwording over de schaduwkanten van het internet, en hoe daar als mens en samenleving mee om te gaan.

2. Dat dit onderzoek zich richt op schadelijk online gedrag, wil niet zeggen dat het internet alleen maar schadelijk gedrag faciliteert. Online connectiviteit heeft ook veel positieve kanten, zoals de potentie om een grote sociale groep te bereiken en meer directe interactie tussen bestuur en burgers te creëren. Deze positieve kanten van het internet vallen buiten het bereik van dit onderzoek, maar moeten wel worden meegewogen bij het ontwerpen van maatregelen tegen de schadelijke effecten.

Er komen wekelijks nieuwe termen en fenomenen rondom online schadelijk en immoreel gedrag bij. Dat hebben we gedurende het onderzoek ook ervaren. Zo stond ‘pedojagen’ bij de start van dit onderzoek in de publieke schijnwerpers, maar is de aandacht hiervoor bij de afronding weer geluwd. Bij de afronding van dit onderzoek kwam cryptospeculatie op. Toch verwachten we dat onze taxonomie, overzicht van onderliggende mechanismen en handelingsopties wel voor langere tijd houvast bieden.

1.4 Leeswijzer

Dit rapport bestaat uit de volgende onderdelen.

- Een methodologische verantwoording, met toelichting bij elke fase van het onderzoek (hoofdstuk 2).
- Enkele casussen van schadelijk en immoreel gedrag online (tussen de hoofdstukken door). Deze illustreren diverse fenomenen en demonstreren de werking van de achterliggende mechanismen. Daarnaast verhelderen de casussen de rol van verschillende actoren die bij online gedrag betrokken zijn.
- Een taxonomie van schadelijk en immoreel gedrag online, waarin de aard en (waar mogelijk) omvang van de besproken fenomenen in Nederland uiteengezet worden (hoofdstuk 3). Het gaat om 22 fenomenen, onderverdeeld in zes categorieën. In dit hoofdstuk worden onderzoeksvragen 1 tot en met 3 beantwoord.
- Een overzicht en bespreking van de online mechanismen achter schadelijk en immoreel gedrag op het internet (hoofdstuk 4). Hierin ligt de focus op de mechanismen die geduid worden als specifiek voor de online omgeving. In dit hoofdstuk wordt onderzoeksvraag 4 beantwoord.
- Een overzicht van initiatieven die nu al genomen worden om schadelijk en immoreel gedrag online te voorkomen of tegen te gaan (hoofdstuk 5). De initiatieven worden besproken op het niveau van type aanpak en vanuit het perspectief van de verantwoordelijke actor of partij (zoals overheid, bedrijf,

maatschappelijke organisatie of hulpverleningsinstantie). In dit hoofdstuk wordt onderzoeksvraag 5 beantwoord.

- Een strategische agenda waarbij handelingsopties voor de Nederlandse overheid in samenwerking met actoren in markt en maatschappij worden aangereikt om schadelijk en immoreel gedrag online te voorkomen, te beperken en/of te repareren (hoofdstuk 6). In dit hoofdstuk wordt onderzoeksvraag 6 beantwoord.

2 Aanpak

Dit onderzoek behandelt verschillende deelvragen. Bij het beantwoorden van elk van de deelvragen is gebruik gemaakt van een combinatie van methoden. Nadat het Rathenau Instituut de onderzoeksopzet had bepaald, is het onderzoeksobject afgebakend. Daarbij hebben de begeleidingscommissie en ambtenaren van verschillende ministeries input gegeven.

Vervolgens is literatuuronderzoek gedaan naar de aard, omvang en oorzaken van schadelijk en immoreel gedrag online. Daarbij is gebruik gemaakt van wetenschappelijke literatuur, beleidsstukken en rapporten in opdracht van beleidsmakers, journalistieke artikelen en gedragscodes van platformen. Daarnaast zijn interviews gehouden met deskundigen uit wetenschap, beleid en praktijk. Een thematische analyse van al deze bronnen heeft geleid tot een taxonomie van schadelijk en immoreel gedrag. Tot slot is gewerkt aan een overzicht van handelingsopties en oplossingsrichtingen. Daarbij is eveneens gebruik gemaakt van literatuur, en van een verkennende sessie met vertegenwoordigers uit beleid, praktijk en wetenschap. Vervolgens zijn alle bevindingen voorgelegd aan een aantal deskundigen in een validatiebijeenkomst, wat de bevindingen verder heeft aangescherpt en verrijkt. In verschillende fasen van het onderzoek hebben de onderzoekers van het Rathenau Instituut overleg gehad met de door het WODC ingestelde begeleidingscommissie (zie Bijlage 1). In totaal hebben 56 deskundigen uit wetenschap, beleid en praktijk aan het onderzoek bijgedragen. Een nadere toelichting van de gebruikte methoden volgt hieronder.

2.1 Verkenning

Na de vaststelling van de onderzoeksopzet zijn we het onderzoek begonnen met een verkenningsworkshop met ambtenaren van ministeries, handhaving en hulporganisaties om beter grip te krijgen op de kennisbehoefte en ons te helpen ons onderzoeksonderwerp af te bakenen. De opzet en deelnemers zijn opgenomen in Bijlage 2.

2.2 Literatuuronderzoek

De casuïstiek uit het boek *Evil Online* is als startpunt gebruikt. Dit leverde een lijst met fenomenen en mechanismen op, die we hebben gebruikt als zoektermen in

databases van wetenschappelijke literatuur en nieuwsmedia. Via de sneeuwbalmethode zijn op basis van deze lijst meer relevante artikelen en rapporten gevonden. Een literatuurbron werd relevant geacht als die inzicht gaf in de aard van fenomenen, als die cijfers bevatte over de omvang van deze fenomenen in Nederland, als de literatuurbron inzicht gaf in de oorzaken en mechanismen, of ideeën bevatte voor handelingsopties. Ook een inventarisatie van de richtlijnen en gedragsvoorschriften van online platformen leverde tientallen begrippen op die onder de noemer ‘schadelijk en immoreel gedrag online’ vallen. De inzichten uit het literatuuronderzoek zijn in alle hoofdstukken verwerkt en de bronnen zijn te vinden in de literatuurlijst.

2.3 Interviews

Het literatuuronderzoek is aangevuld met 15 interviews met experts en deskundigen (zie Bijlage 3). De geïnterviewden zijn zodanig geselecteerd, dat de interviews gezamenlijk de breedte van de fenomenen en onderliggende oorzaken en mechanismen bestreken. De interviewleidraad is opgenomen in Bijlage 4. De inzichten uit de interviews zijn in de verschillende hoofdstukken verwerkt.

2.4 Werksessie

Op 13 april 2021 heeft het onderzoeksteam van het Rathenau Instituut een werksessie georganiseerd rond mogelijkheden voor de aanpak van schadelijk en immoreel gedrag online. Tijdens het literatuuronderzoek en de interviews waren vijf oplossingsrichtingen in beeld gekomen waarvoor veel animo was, maar die niet uitgewerkt waren in de vorm van concrete initiatieven. Het doel van de werksessie was om deze oplossingsrichtingen verder te concretiseren, in een dialoog tussen medewerkers van verschillende ministeries, handhavings- en hulporganisaties, en onderzoekers, vertegenwoordigers van maatschappelijke organisaties en anderen met relevante expertise. Bij de werksessie waren 22 deelnemers betrokken, verdeeld over vijf oplossingsrichtingen (zie Bijlage 5).

- online toezicht en hulp
- het gesprek over normen online
- waardengedreven inrichting van platformen
- technologische oplossingen
- handhaving van wetten en regels in de online omgeving

De opgedane inzichten zijn verwerkt in de strategische agenda die in hoofdstuk 6 gepresenteerd wordt.

2.5 Validatiebijeenkomst

Op 26 mei 2021 heeft het onderzoeksteam van het Rathenau Instituut ter validatie van de onderzoeksresultaten een expertmeeting georganiseerd. Daarbij waren onderzoekers betrokken en medewerkers van uitvoeringsorganisaties en maatschappelijke organisaties.

Voorafgaand aan de bijeenkomst kregen de deelnemers een samenvatting van het onderzoek ongestuurd (ca. 20 pagina's). Tijdens de bijeenkomst kregen ze de gelegenheid op de onderzoeksresultaten te reageren. De nadruk lag daarbij op de handelingsopties die voortkomen uit de analyse in het rapport. Het doel was om samen met de aanwezigen te komen tot een prioritering van handelingsopties en te reflecteren op de rol die verschillende partijen kunnen aannemen bij de aanpak van schadelijk en immoreel gedrag online. Bij de validatiebijeenkomst waren zeven personen betrokken (zie Bijlage 6). De inzichten zijn verwerkt in het gehele rapport.

2.6 Begeleidingscommissie

Dit rapport kwam tot stand op verzoek van het WODC. Bij de start van het onderzoek is een begeleidingscommissie ingesteld, bestaande uit een vertegenwoordiger van het Ministerie van Justitie & Veiligheid, het WODC en drie deskundigen (voor samenstelling zie Bijlage 1). De begeleidingscommissie had een adviserende rol en is op vier momenten tijdens verschillende fasen van het onderzoek met het onderzoeksteam bijeengekomen. Hun adviezen zijn naar inzicht van het Rathenau Instituut verwerkt. De verantwoordelijkheid voor de inhoud van het rapport ligt volledig bij het Rathenau Instituut.

Casus: online shaming

Deze casus gaat over online shaming, een uiting van digitaal vigilantisme. Het is een bewerking van een waargebeurde casus. We beginnen met een beschrijving van de gebeurtenissen, waarna we reflecteren op de rol die verschillende actoren daarin spelen. Online mechanismen die een rol spelen in deze casus zijn vetgedrukt en worden in hoofdstuk 4 verder uitgewerkt.

Casus
Manu is vorig jaar seksueel misbruikt door een man die nog steeds een belangrijke maatschappelijke functie vervult. De politie raadde wegens gebrek aan bewijs aan om geen aangifte te doen. Zelf heeft Manu aanwijzingen dat deze man nog meer slachtoffers heeft gemaakt, maar hij weet dat niet zeker. Zijn vertrouwen in de rechtstaat is geschaad. Na lang twijfelen besluit Manu in de openbaarheid te treden op sociale media, om anderen te waarschuwen en zo te voorkomen dat de dader nog meer slachtoffers kan maken.

Ondanks dat Manu weinig volgers heeft op sociale media, wordt zijn bericht gedeeld door iemand met **veel meer bereik**. Meer slachtoffers beginnen hun ervaringen te delen met dezelfde dader. Manu schrikt van de **media-aandacht** in kranten en op tv die opeens ontstaat rondom de zaak, maar voelt zich ook gesteund door de ervaringen van anderen. Het blijkt dat meer mensen naar de politie zijn gestapt met aantijgingen tegen dezelfde man, maar dat het nooit tot onderzoek en vervolging is gekomen. Mensen beginnen zich op sociale media tegen de man te keren en iemand publiceert ook zijn thuisadres online (doxing). Hij wordt bedreigd en voorlopig door zijn werkgever op non-actief gesteld (cancelling).

Ook Manu wordt online beschuldigd van liegen en het zoeken van aandacht, wat zorgt voor herbeleving van het trauma dat het misbruik zijn eigen schuld zou zijn. Het socialemediaplatform besluit berichten over deze zaak te **verbergen in de tijdlijn** van mensen, omdat die berichten de regels van het platform rondom smaad en laster zou overtreden. In een populaire talkshow vertelt een woordvoerder van de politie dat tegen deze man nooit aangiften zijn gedaan. Slachtoffers reageren woedend op sociale media, omdat aangifte doen hen juist werd afgeraden. Het vertrouwen van de slachtoffers in de rechtsstaat daalt nog verder.

Reflectie
In deze casus zijn diverse actoren betrokken. Niet alleen Manu en degene die Manu beschuldigt spelen een rol, maar ook omstanders, socialemediaplatformen en de politie hebben impact op de gevolgen van Manu’s gedrag. Het is in dit

voorbeeld ook moeilijk om over dader en slachtoffer te spreken, omdat Manu en de man die beschuldigd wordt, allebei als slachtoffer én dader gezien kunnen worden. Manu is slachtoffer van seksueel misbruik, maar kan zelf ook als ‘dader’ worden gezien vanwege zijn rol in het naar buiten brengen van de beschuldigingen. Manu is niet uit op het toebrengen van schade aan de beschuldigde, maar wil tegelijkertijd wel voorkomen dat hij nog meer slachtoffers maakt. Dat laat zien dat uitingen van online schadelijk gedrag niet zwart-wit hoeven te zijn en ook voor omstanders vaak moeilijk op waarde te schatten zijn.

Het voorbeeld maakt duidelijk hoe snel situaties op internet kunnen escaleren door de **onbeheersbaarheid** en **persistentie** als informatie eenmaal online staat. Daarin spelen verschillende online mechanismen een rol, zoals de **schaalbaarheid** en **viraliteit** van online platformen, de **openbaarheid** van het internet en de **onbeschaafde omgangsnormen** die vaak online gelden. Voor deze casus is het vooral belangrijk te onthouden dat mensen online vaak de controle over de situatie verliezen. Doordat het bericht viral gaat, verliest Manu de autonomie over de verspreiding van zijn boodschap. Andere mensen achterhalen de gegevens van de man die Manu beschuldigt en zij bedreigen hem, terwijl Manu zelf het trauma opnieuw herbeleeft door de vele reacties van ongeloof.

De beschuldigde heeft door de **onmiddellijkheid** van het internet, waardoor gedrag meteen effect heeft, geen tijd en mogelijkheid om zich te verweren. Voor hem vindt een versneld proces van veroordeling plaats, doordat zijn werkgever hem op non-actief stelt en ook de traditionele media over hem schrijven. Dat is een wezenlijk verschil met wanneer de politie en het OM een onderzoek hadden ingesteld naar zijn gedrag. Ook zorgt het online aspect ervoor dat verschillende actoren direct snel handelen. Omstanders scharen zich achter de kant van Manu of trekken zijn verhaal juist in twijfel, door berichten van anderen te verspreiden of zelf berichten te plaatsen. **Syndicatie,** het gemakkelijk vinden van gelijkgestemden online, speelt hierin een grote rol. Afhankelijk van de online sociale omgeving waarin mensen zich bevinden, wordt bepaald gedrag misschien van hen verwacht: je uitspreken tegen onrecht is een manier om te laten zien dat je zelf wel aan de moreel juiste kant staat.

Het socialemediaplatform heeft in dit voorbeeld besloten om de verspreiding van berichten over Manu’s zaak te verbergen, waardoor het platform ingrijpt in de **amplificatie** van berichten online. Dit soort afwegingen tussen vrijheid van expressie en mogelijke schade zijn voor platformen moeilijk te maken en komt ze regelmatig op kritiek te staan. Mede doordat de politie in haar gesprek met Manu besloot om geen verder onderzoek in te stellen, nam Manu uiteindelijk het heft in eigen hand. Ongenoegen en wantrouwen in de strafrechtketen om voor rechtvaardigheid te zorgen, speelt bij alle vormen van digitaal vigilantisme een rol.

3 Taxonomie van schadelijk en immoreel gedrag online

In dit hoofdstuk beschrijven we de aard en omvang van schadelijk en immoreel gedrag online. Dat doen we aan de hand van een door het Rathenau Instituut ontwikkelde taxonomie. Om grip te krijgen op het veranderlijke en veelzijdige onderwerp ‘schadelijk en immoreel gedrag online’, heeft het onderzoeksteam gedurende het hele onderzoeksproces gewerkt aan een taxonomie die de soorten schadelijk en immoreel gedrag categoriseert en specifieke fenomenen indeelt.

Eerst zijn sommige in het onderzoek geïdentificeerde fenomenen samengevoegd die slecht van elkaar te onderscheiden zijn. Vervolgens zijn de fenomenen thematisch gesorteerd en is verder rekening gehouden met de (belangrijkste) drijfveer achter het gedrag (zoals eigenrichting of sadisme) en kenmerken van de slachtoffers. Dit is een iteratief proces geweest waarin de onderzoekers tot het laatste moment aan de taxonomie hebben geschaafd.

In dit hoofdstuk lichten we de taxonomie eerst kort toe. Daarna presenteren we de zes categorieën van deze taxonomie, waarbij we de aard en omvang van verschillende vormen van ‘online ontsporing’ benoemen.

3.1 Taxonomie

De taxonomie in Figuur 3.1 brengt verschillende vormen van online ontsporing onder in zes hoofdcategorieën:

1. online informatiemanipulatie;
2. digitaal vigilantisme;
3. online haat;
4. online pesterij en geweld;
5. cyberbedrog; en
6. online zelfbeschadiging.

Binnen deze categorieën identificeren we 22 fenomenen (zie de begrippenlijst vooraan in dit rapport).

In de taxonomie staan gedragingen die niet altijd onrechtmatig zijn en vaak straffeloos gebeuren. Soms gaat het om online varianten van gedrag dat offline ook al bestond, zoals bedreiging of intimidatie, en soms gaat het om nieuwe vormen

van gedrag die mogelijk worden gemaakt door het internet, zoals doxing, phishing, hacking of sock puppeting.

Bron: Rathenau Instituut

Figuur 3.1 Taxonomie van immoreel en schadelijk gedrag online

Hierna bespreken we per categorie en per fenomeen de kenmerken van online schadelijk en immoreel gedrag. Vervolgens behandelen we ook de omvang van elk fenomeen, voor zover daarvoor cijfers beschikbaar zijn.

3.2 Online informatiemanipulatie

Onder online informatiemanipulatie verstaan we hier het verspreiden van allerlei vormen van online informatie (tekst, beeld, audio) die als feitelijk wordt gepresenteerd, maar feitelijke onjuistheden bevat. Dit kan bewust gebeuren (zoals in het geval van verspreiding van desinformatie) of onbewust (in het geval van complotdenken). De (bewuste) verspreiding van onjuiste informatie zorgt voor misleiding, onrust en verwarring, en zorgt ervoor dat informatie moeilijker op waarde te schatten wordt (Rathenau Instituut, 2018a, 2021b). Onder de categorie

‘informatiemanipulatie’ vallen: **complotdenken, desinformatie, sock puppeting en kwakzalverij**. Deze fenomenen zijn niet altijd strafbaar op zichzelf, al kunnen ze wel onderdeel zijn van strafbaar handelen. Informatiemanipulatie is niet per se gericht op specifieke individuen of groeperingen, en treft ook vaak de maatschappij als geheel.

Complotdenken

We spreken van ‘complottheorieën’ als mensen de overtuiging hebben dat bepaalde gebeurtenissen of situaties in het geheim en achter de schermen gemanipuleerd zijn door machtige groepen met verkeerde bedoelingen (COMPACT Education Group, 2020). Complotdenkers denken over het algemeen dat niets is wat het lijkt en dat veel gebeurtenissen niet door toeval zijn ontstaan maar in het kader van vooropgezette, kwade plannen. Daarmee verwerpen ze de mogelijkheid dat de werkelijkheid chaotisch en complex is. Ze hebben vaak de neiging om ‘daders’ aan te wijzen voor wat er niet goed gaat.

De term ‘complotdenken’ ligt gevoelig: hij is waardegeladen en kan veroordelend overkomen richting mensen die in complotten geloven. Wetenschappers geven aan dat het woord complotdenken vaak gebruikt wordt om de positie van mensen op een bepaald thema te delegitimeren (Husting & Orr, 2007). De Nederlandse socioloog Jaron Harambam stelt op basis van zijn promotieonderzoek dat het diskwalificeren van complotdenkers niet zonder risico is (Harambam, 2017, p. 75). Soms kan nieuw bewijs er namelijk ook voor zorgen dat complotdenkers in hun gelijk worden gesteld (Mortimer, 2017). Het Rathenau Instituut (2018a) is daarom terughoudend in het aanwijzen van mensen als complotdenkers en zal deze term dan ook terughoudend gebruiken en telkens kwalificeren.

Online verspreiden complottheorieën zich sneller, maar het fenomeen complotdenken is al eeuwenoud. Al in het jaar 68 geloofden sommige Romeinen niet dat keizer Nero echt zelfmoord had gepleegd en gingen ze ervan uit dat er een complot zat achter zijn dood (Champlin, 1998). Wat online complotdenken onderscheidt van de offline variant van dit fenomeen, is de mogelijkheid voor complotdenkers om zich online gemakkelijk te verenigen en informatie te delen. Hierdoor kunnen ze nog meer overtuigd raken van hun denkbeelden. Ook zijn er online eenvoudig andere complottheorieën te vinden die voor een bestaande groep complotdenkers aannemelijk overkomt (Bessi et al., 2015).

Mensen die online complottheorieën verspreiden, wantrouwen vaak traditionele media en wetenschap (Rathenau Instituut, 2018a, 2018b). Vaak voelen zij zich niet vertegenwoordigd door dit soort instituties, omdat ze elitair zouden zijn (Harambam, 2017). Voor hen is het dus belangrijk om hun overtuigingen naar buiten te brengen. Het (onbewust) verspreiden van complottheorieën verschilt van het (doelbewust)

zaaien van twijfel over bepaalde gebeurtenissen, omdat er niet noodzakelijk een kwade intentie achter zit.

De **schade** van complotdenken is potentieel groot, omdat het fenomeen kan bijdragen aan antidemocratisch gedachtegoed en kan leiden tot een afname van vertrouwen in publieke instituties (Sternisko et al., 2020). Ook kunnen complotdenkers schade toebrengen aan individuen door hen te beschuldigen van geheime kwade intenties, hen te bedreigen (Schildkamp & Rodenburg, 2021) of door strafbare feiten te plegen, geïnspireerd door complotten. In de Verenigde Staten zijn kidnappingen, achtervolgingen en een moord in verband te brengen met de complottheorie Qanon (The Guardian, 2020). De terrorist die in Christchurch het vuur opende op moskeegangers was geïnspireerd door online complottheorieën over witte suprematie (The Independent, 2019), net als een Duitse terrorist die zich in 2020 beriep op racistische complottheorieën (AP, februari 2020).
Complotdenken kan dus reële schade veroorzaken, zowel op individueel als op maatschappelijk niveau. Complotdenken *an sich* is niet strafbaar**,** maar bepaalde uitingen of gedragingen die eruit voortvloeien kunnen wel strafbaar zijn.

Omvang
Al sinds de opkomst van het internet circuleren in Nederland online allerlei soorten maatschappijkritische, niet-wetenschappelijk bewezen theorieën, bijvoorbeeld over de landing op de maan, de holocaust en de aanslagen van 9/11. Een bekende theorie biedt alternatieve verklaringen voor ‘chemtrails’: condenssporen van vliegtuigen in de lucht. Via onder andere een online petitie sturen de aanhangers hiervan al lange tijd aan op een Tweede-Kamerdebat over het onderwerp. De petitie is onlangs nog door meer dan 40.000 mensen ondertekend (Petities.nl, 2021). Naast deze klassieke theorieën heeft ook Covid-19 gezorgd voor nieuw wantrouwen. In Nederland manifesteert dit wantrouwen zich behalve online ook steeds vaker in de offline wereld, bijvoorbeeld bij demonstraties en het in brand steken van 5G-zendmasten (NCTV, 2021).

Het blijkt lastig te zijn om via peilingen te achterhalen hoe groot de aanhang van extremistische complottheorieën zoals de QAnon-beweging daadwerkelijk is. Aanhangers van de QAnon-complottheorie zijn uit op de omverwerping van de democratische rechtsstaat waarbij geweld mag worden gebruikt (Bellingcat, 2021). Inschattingen van de omvang van het totale aantal QAnon-aanhangers lopen in de VS uiteen van 3% tot 14% van de bevolking (Shanahan, 2021).

Bron: Rathenau Instituut

Figuur 3.2 Online informatiemanipulatie

Voor Nederland bestaan er geen schattingen, maar er zijn wel wat indicaties van de omvang van complotdenken in het algemeen te geven. Volgens de NOS zouden zo'n 40.000 berichten op Nederlandse Facebookgroepen over QAnon leiden tot ruim een half miljoen interacties (likes, shares, comments) (Bouma, 2020). Viruswaarheid, de actiegroep die het coronabeleid van de overheid bekritiseert en wantrouwt, heeft op Twitter ongeveer 10.000 volgers van de in totaal 2,9 miljoen Nederlandse twittergebruikers. Nadat hun accounts zijn verwijderd, verhuizen complotdenkers vaak naar andere platformen zoals Bitchute, Gab, Telegram of Parler.

Uit onderzoek van *NRC* en de UvA blijkt dat ondanks de maatregelen van de mainstream platformen Facebook, Instagram, Twitter en YouTube, het aantal berichten dat populaire complotdenkers en virussceptici plaatsen, nergens significant minder werd tussen juli 2020 en januari 2021 (Kist & Van den Bos, 2021; Motivaction, 2021). Complottheorieën zouden vooral worden geloofd door jongvolwassenen met weinig vertrouwen in nieuws afkomstig uit de reguliere kanalen (zoals kranten of nieuws- en actualiteitenprogramma's op radio of TV). Verschillende onderzoeken naar de aanhangers van bekende complotten geven aan dat deze groep circa 9% tot 15% van de bevolking uitmaakt (Ipsos, 2020; Motivaction, 2021). Jongeren, laagopgeleiden en aanhangers van extreemlinkse en -rechtse partijen zijn oververtegenwoordigd in de groep die gelooft dat het coronavirus een biologisch wapen is (Ipsos, 2020; Motivaction, 2021). Dit is in lijn met bevindingen over het vertrouwen in de wetenschap en andere instituties zoals de media (Rathenau Instituut, 2018b). Volgens onderzoeksbureau Kieskompas zou één op de tien Nederlanders geloven dat rond corona vuile spelletjes worden gespeeld (Visser, 2020).

Desinformatie

Desinformatie verschilt van complotdenken. Complotdenkers geloven vaak echt in de informatie die ze verspreiden, terwijl achter desinformatie een doelgericht schadelijke intentie zit (COMPACT Education Group, 2020). **Online** kan desinformatie gemakkelijk verspreid worden. Het is niet altijd duidelijk welke kwaliteitswaarborging online kanalen hanteren, zo concludeerde het Rathenau Instituut in eerder onderzoek naar de digitalisering van het nieuws (Rathenau Instituut, 2018a).

Vaak wordt met desinformatie gedoeld op het verspreiden van informatie die 'onjuist' of 'misleidend' is (Rathenau Instituut, 2018a). Dat maakt het een lastig te definiëren term. Binnen de wetenschap gaan de laatste jaren meer stemmen op om bij de beoordeling van iemands gedrag rekening te houden met zijn intentie (Gelfert, 2018). Iemand zou bewust kwaadwillend moeten handelen, voordat iets als desinformatie aangemerkt kan worden. Maar ook deze definitie is in de praktijk lastig hanteerbaar, omdat het niet altijd duidelijk is of iemands intenties

kwaadwillend zijn. In plaats daarvan maakt het (Rathenau Instituut, 2018a, pp. 33–34) gebruik van het volgende onderscheid tussen verschillende typen foute informatie (voorgesteld door de Raad van Europa):

1. **Desinformatie:** inaccurate informatie die opzettelijk wordt gecreëerd en verspreid om schade toe te brengen aan personen, groepen, organisaties of landen.
2. **Misinformatie:** inaccurate informatie die onbewust wordt gecreëerd en verspreid, bijvoorbeeld vlak na een heftige gebeurtenis zoals een aanslag.
3. **Laster:** waarheidsgetrouwe informatie die wordt verspreid om schade te berokkenen aan personen, groepen, organisaties of landen, om te treiteren of aan te zetten tot haat.

Uit onderzoek naar de verspreiding van nepnieuws op Facebook blijkt dat mensen boven de 65 meer dan zeven keer zo vaak nepnieuws delen als jongeren. Om de schade van online desinformatie te beperken moeten dus zeker niet alleen jongeren worden aangesproken (A. Guess et al., 2019).

Het Rathenau Instituut ziet bij desinformatie vooral risico's op **schade** aan het publieke debat en het democratische proces (Rathenau Instituut, 2020b). Achter het verspreiden van desinformatie kunnen verschillende motieven zitten. Voor statelijke actoren is desinformatie een middel om verwarring te zaaien en maatschappelijke onrust te creëren in eigen land of andere landen. Maar verspreiders kunnen ook meer opportunistische of economische motieven hebben, bijvoorbeeld als ze nepnieuws 'viral' laten gaan om eraan te verdienen. Het verspreiden van desinformatie is niet bij wet verboden, maar kan gepaard gaan met strafbare gedragingen zoals haatzaaien, laster en oplichting. Desinformatie staat de laatste jaren hoog op de agenda van zowel platformbedrijven als de politiek.

Omvang

Naar de omvang van het fenomeen desinformatie is in Nederland nog maar weinig empirisch onderzoek gedaan (Rathenau Instituut, 2018a; ROB, 2019). Er zijn nauwelijks cijfers over de omvang van het fenomeen, en er bestaan ook weinig studies naar wie erachter zit en wat de impact ervan is (Prij & Janssens, 2020). Daarbij komt nog dat bronnen lastig vergelijkbaar zijn, door onduidelijkheid (en onenigheid) over de definitie van desinformatie. Die zorgen ervoor dat de omvang van het fenomeen steeds anders gemeten wordt (Common & Kleis Nielsen, 2021).

Internationaal zijn er wel wetenschappelijke studies gedaan naar de omvang van desinformatie. Zo zijn er *peer-reviewed* studies die aantonen dat nepnieuws slechts 0,15% van de dagelijkse mediaconsumptie van Amerikanen uitmaakt (Allen et al., 2020). Ook bezochten drie op de vier Amerikanen geen enkele nepnieuwswebsite in de aanloop naar de Amerikaanse verkiezingen in 2016 deed een kwart van de Amerikanen dit tenminste eenmalig (Guess et al., 2020). In de VS bestaat de

belangrijkste bezoekersgroep van nepnieuwswebsites uit mensen die zeer intensief gebruik maken van het internet, maar tegelijkertijd ook zeer geëngageerde en loyale gebruikers zijn van gevestigde nieuwswebsites (Nelson & Taneja, 2018).

Vooralsnog lijken in Nederland minder desinformatie en minder polarisatie van het medialandschap te bestaan dan in de Verenigde Staten (Rathenau Instituut, 2018a). Volgens het *Digital News Report* kent Nederland een relatief hoog vertrouwen in de media en zijn er weinig zorgen over nepnieuws (Reuters, 2020). Dit wordt deels verklaard door de relatief sterke positie van de publieke omroep die de standaard zet voor de kwaliteit van andere nieuwsmedia (zie Figuur 3.2). Er zijn maar weinig Nederlanders die hun nieuws alleen via sociale media krijgen, omdat zij ook televisie, radio en kranten raadplegen. Traditionele nieuwsmedia zijn sterk vertegenwoordigd in de top twintig van de socialemediafeed van Nederlanders (Möller et al., 2019).

In internationaal opzicht bevindt Nederland zich ook op het laagste niveau van *cyber troop capacity*, waarbij het gaat om het aantal betrokken actoren, instrumenten, permanente teams en uitgaven aan disinformatiecampagnes (University of Oxford, 2020, p. 18). Er is vrijwel geen bewijs te vinden voor buitenlandse desinformatie, en er zijn ook maar weinig voorbeelden gedocumenteerd van Nederlandse spelers die bijvoorbeeld Russische desinformatietactieken toepassen (Rogers & Niederer, 2019). Het recente geval in de Tweede Kamer, waarbij een komiek zich voordeed als Leonid Volkov (een naaste medewerker van de Russische oppositieleider Aleksei Navalvy) bewijst echter dat misleiding vanuit het buitenland ook hier niet ondenkbaar is. Ook bestond er een Nederlandse Telegramgroep, met bijna achthonderd coronasceptici die desinformatie verspreidden (Pointer, 2021b).

Sock puppeting

Een sock puppet is letterlijk een handpop gemaakt van een sok. In de **online** context betekent sock puppeting het aannemen van een valse identiteit om anderen te misleiden (Oleshchuk, 2020), vaak in de vorm van nepaccounts op online platformen. In veel online communities zijn sock puppets niet welkom en vermelden de gebruikersvoorwaarden bijvoorbeeld dat het niet toegestaan is om je voor te doen als iemand anders. Het is een typisch online fenomeen, omdat het internet zich bijzonder goed leent voor het aannemen van andere identiteiten.

Sock puppeting wordt soms in verband gebracht met desinformatie, omdat misleidende informatie met behulp van andere identiteiten verspreid kan worden. In november 2020 meldden diverse media bijvoorbeeld dat een witte Republikeinse kandidaat voor het Amerikaanse Congres zich op Twitter voordeed als een zwarte man en supporter van Trump (Espinoza, 2020). Met een sock puppet zou hij het publieke debat hebben willen beïnvloeden. Het voorbeeld is bijzonder omdat de Republikeinse kandidaat schijnbaar per ongeluk een bericht deelde dat bedoeld

zou zijn geweest voor zijn sock puppet account, waardoor het gebruik aan het licht kwam. Op Twitter weersprak hij deze beschuldigingen. Dit voorbeeld laat zien dat het niet moeilijk is om met sock puppets misleidende informatie te verspreiden en dat het tegelijkertijd lastig te bewijzen is.

Sock puppeting is een vorm van online informatiemanipulatie die ingezet kan worden voor veel andere vormen van schadelijk gedrag, zoals *phishing* (zie 3.6, cyberbedrog) en *trolling* (zie 3.5, pesterij en geweld). De **schade** die sock puppeting aanricht, hangt af van iemands doelen. Ze kan variëren van persoonlijke economische schade of imagoschade tot maatschappelijke schade. Dat laatste gebeurt wanneer het publieke debat met misleidende informatie wordt beïnvloed, net als bij desinformatie. Sock puppeting kan strafbaar zijn als het fenomeen wordt ingezet voor strafbare zaken als smaad, laster en haatzaaien. Ook wanneer iemands identiteit gestolen wordt en gebruikt wordt als sock puppet, kan dat strafbaar zijn. Voor sock puppeting wordt echter niet altijd identiteitsfraude gepleegd. Vaak worden ook nepprofielen met bijvoorbeeld computergegenereerde foto's gebruikt.

Omvang

Over de omvang van het fenomeen sock puppeting in Nederland zijn vooralsnog geen cijfers gevonden. Ook lijkt er vrijwel geen literatuur over het onderwerp te bestaan. Daarom verwijzen we hier naar wat bekend is over de omvang van de fenomenen *phishing*, *catfishing* en *trolling* (zie verder).

Kwakzalverij

Kwakzalverij is volgens Van Dales *Groot woordenboek* het "onbevoegd uitoefenen van de geneeskunst" door iemand die beweert een ziekte te kunnen genezen met een nutteloos geneesmiddel (Geerts & Den Boon, 1999).

Patiënten gaan vaak **online** om informatie te vinden over hun ziekte en eventuele (alternatieve) behandelmethoden (Delgado-López & Corrales-García, 2018). Zowel de pluriformiteit van informatie als de mogelijkheid tot lotgenotencontact maken van het internet een waardevol hulpmiddel voor patiënten. Maar kwakzalverij ligt er wel op de loer, omdat regulering ontbreekt en online informatie er soms moeilijk op waarde te schatten is – veel moeilijker dan offline. Uit onderzoek naar online informatie over kankerbehandelingen blijkt bijvoorbeeld dat het internet een veelgebruikte manier is om onbewezen en gevaarlijke behandelmethoden voor kanker aan te prijzen (Delgado-López & Corrales-García, 2018).

Toch is bij gebruik van de term 'kwakzalverij' dezelfde terughoudendheid geboden als bij 'complotdenken'. Het woord wordt namelijk makkelijk geassocieerd met alternatieve geneeswijzen. Alternatieve geneeswijzen worden echter pas kwakzalverij als ze 1) gangbare geneeswijzen afkeuren, 2) schadelijke therapieën

zonder waarschuwing aanraden, 3) veel geld kosten en 4) zich beroepen op bovennatuurlijke genezing (Offit, 2013).

De **schade** van kwakzalverij voor individuen kan groot zijn, bijvoorbeeld als mensen op basis van dit soort informatie onbetrouwbare therapieën ondergaan of zich van zorg onthouden op advies van kwakzalvers. Zo zorgde COVID-19 in 2020 voor het verspreiden van gevaarlijke informatie over niet-goedgekeurde, potentieel zeer schadelijke medicijnen (Freckelton QC, 2020). Bovendien kan het op obsessieve wijze zoeken naar informatie over gezondheid ook leiden tot cyberchondrie: angst veroorzaakt door alarmerende informatie over ziekten en kwalen die mensen op het spoor komen tijdens het online zoeken naar medische informatie (Eindhovens Dagblad, 2019) (Aiken, 2016).

In 2019 werd een homeopathische arts uit Eindhoven beboet na het maken van online reclame voor een ongeregistreerd medicijn tegen griep (Eindhovens Dagblad, 2019). Deze boete werd na een bestuursrechtelijke procedure opgelegd door de Inspectie voor de Gezondheidszorg, omdat de reclame voor het middel misleidend was. Kwakzalverij kan ook strafbaar zijn onder artikel 96 van de Wet op de Beroepen in de Individuele Gezondheidszorg, wanneer schade wordt toegebracht aan iemands gezondheid bij het uitoefenen van zorg.

Omvang
Over de omvang van kwakzalverij in Nederland zijn geen cijfers gevonden. Het Rathenau Instituut (2018a, p. 40) signaleerde dat van alle online hoaxen of onterechte waarschuwingen op het internet, bijna een derde betrekking had op de ‘gezondheidsrisico’s’ van bijvoorbeeld voedsel, gebruiksvoorwerpen of insecten.

Volgens de World Health Organisation verspreidt foutieve informatie over COVID-19 zich snel via sociale media en hindert dit de strijd tegen het coronavirus (Laato et al., 2020). Een recent onderzoek in de VS van het Center for Countering Digital Hate wees uit dat slechts 12 personen (accounts) verantwoordelijk waren voor de verspreiding via sociale media van circa 65% van de misinformatie met een anti-vaccinatie-boodschap (Bond, 2021). Sommige van deze accounts zijn op meerdere platformen actief, promoten natuurlijke gezondheidszorg en verkopen onder meer supplementen, workshops en boeken. De coronacrisis is een gat in de markt voor deze ondernemers, aldus de directeur van het Center (Brumfiel, 2021).

Door de lage kosten, gemakkelijke toegang tot het internet en de anonimiteit die men er ervaart, wordt zoeken naar medische informatie online steeds populairder (Zheng et al., 2020). Nederland staat op nummer twee in Europa qua percentage van de bevolking dat op internet naar informatie over gezondheid zoekt (Eurostat, 2021). In 2020 zocht 76% van de Nederlandse bevolking online naar dergelijke informatie, terwijl dit in 2011 nog 53% was (Eurostat, 2021).

3.3 Digitaal vigilantisme

Digitaal vigilantisme is een vorm van collectieve actie tegen personen die ongewenst sociaal gedrag vertonen. De drijfveer is **morele afkeuring en eigenrichting**. Online uit zich dit bijvoorbeeld in naming en shaming en doxing. Eigenrichting is op zichzelf niet strafbaar gesteld, maar gedrag dat daaraan vaak gekoppeld is (zoals het toepassen van geweld) wel.

Shaming wordt ingezet als tactiek om sociaal geldende normen te expliciteren, bijvoorbeeld door iemand online te wijzen op het racistische karakter van zijn of haar uitspraken. Dat maakt het ook lastig om te beoordelen of shaming 'gerechtvaardigd' is, en wanneer morele grenzen overschreden worden. Vigilantisme op het internet kan het karakter krijgen van burger-geïnitieerd politiewerk (*do-it-yourself policing*), als burgers het gevoel hebben dat de rechtsstaat niet voorziet in het terechtwijzen van bepaalde individuen of groepen. Het is een vorm van surveillance door burgers die het als hun plicht zien om anderen te wijzen op iets dat in hun ogen moreel verwerpelijk is. Dergelijke initiatieven kunnen in de online omgeving snel en spontaan ontstaan (zie hoofdstuk 4). Door de beschikbaarheid van grote hoeveelheid openbare persoonlijke informatie is het mogelijk om anderen continu te monitoren en op te sporen.

We onderscheiden drie vormen van digitaal vigilantisme: **pedojagen**, **doxing** en **shaming en cancellen**.

Pedojagen

Pedojagen is online 'jagen' op personen van wie het vermoeden bestaat dat zij pedofiel zijn. Pedojagers gebruiken **online** anonimiteit om zich voor te doen als kind en daarmee pedofielen 'in de val' te lokken (Hadjimatheou, 2019). Daarbij gebruiken ze soms ook fysiek geweld.

Criminologe Katerina Hadjimatheou wijst erop dat de term 'pedojagen' ethisch problematisch is, omdat hij mensen dehumaniseert. Het woord 'jagen' suggereert dat het geoorloofd is om op mensen te jagen, zoals dat ook bij dieren gebeurt (Hadjimatheou, 2019). Bij groepen die (online) geweld gebruiken tegen pedofielen is deze term populair, juist *omdat* hij dehumaniseert en haat aanwakkert, en tegelijkertijd hun eigen acties moreel legitimeert.

Pedojagers vinden dat hun doelen overeenkomen met die van de staat. Ze menen dat ze bijdragen aan het naleven van wetten en regels. In tegenstelling tot sommige andere mensen die overgaan tot eigenrichting, hebben pedojagers wel vertrouwen in de rechtsstaat: ze leveren mensen over aan de politie, met het idee dat traditionele instituties voor rechtvaardige berechting zullen zorgen. Veel pedojagers zien zichzelf dan ook niet als mensen die het recht in eigen hand nemen, maar als

onderzoeksjournalisten. Sommige groepen zijn sterk geprofessionaliseerd, trainen hun eigen mensen en accepteren donaties (Hadjimatheou, 2019).

Pedojagen biedt niet alleen bescherming aan (potentiële) slachtoffers, maar kan ook veel **schade** berokkenen – zowel aan individuen als aan de samenleving. Mensen van wie vermoed wordt dat ze pedofiel zijn en daarom online 'opgejaagd' worden, kunnen al 'door de massa veroordeeld' worden, voordat ze enige vorm van proces hebben gehad vanwege een verdenking van pedofiel gedrag. Dit kan gevolgen hebben voor hun sociale status, relaties en baan – ook als ze onschuldig zijn. Bovendien kunnen ze slachtoffer worden van fysiek geweld. Daarnaast scheppen acties van pedojagers precedenten voor eigenrichting van burgers. Ze ondergraven het functioneren van de rechtsstaat en richten dus ook grotere maatschappelijke schade aan. Pedojagen is op zichzelf niet strafbaar, behalve als bijvoorbeeld geweld wordt gebruikt. In mei 2021 veroordeelde de rechtbank in Arnhem vijf verdachten tot maximaal vijf jaar cel voor het zwaar mishandelen van een vermeende pedofiel (RTL Nieuws, 2021).

Omvang

Participatie van burgers binnen de opsporing – ook wel '*do-it-yourself policing'* genoemd – is een internationale trend, waarin Nederland tot de koplopers behoort (Denef et al., 2017). Het aantal burgers dat zelf start met opsporen, lijkt gestaag toe te nemen, onder meer door de democratisering van informatie, onderzoeksmiddelen en kennis (de Vries, 2018).

Pedojagers zijn in ons land al zeker tien jaar actief. Empirisch onderzoek naar pedojagen in Nederland ontbreekt. Desondanks lijkt het fenomeen zich uit te breiden (Herweijer & Ververs, 2020). Volgens Arnout de Vries van TNO komt dit enerzijds door de zichtbaarheid van online kindermisbruik voor volwassenen. Ouders kijken in tijden van pandemie vaker mee met het online gedrag van hun kinderen, waardoor er meer oog is voor mogelijke misstanden. Anderzijds interpreteert De Vries de toename ook als een gevolg van gebrek aan online handhaving (Herweijer & Ververs, 2020).

Ongeveer 90% van de burgers die helpt met opsporen, heeft goede intenties (de Vries, 2018). Maar soms gaat het fout. In november 2020 bijvoorbeeld, overleed een 73-jarige man in Arnhem bij de jacht op vermeende kindermisbruikers die volledig uit de hand liep (Sjoukes, 2020). Volgens de politie waren er tussen juli en november van 2020 ongeveer 250 gevallen waarbij pedojagers over de schreef gingen (Veldhuis & Ingabire, 2021).

Bron: Rathenau Instituut

Figuur 3.3 Digitaal vigilantisme

Een van de grootste accounts op Instagram, waar informatie en filmpjes van confrontaties worden uitgewisseld, is @pedohunterznl, dat begin 2021 ruim 41.000 volgers had (Instagram, z.d.). De Facebookgroep van een pedojager uit Deventer had 44.000 leden, tot die eind oktober werd verwijderd (Kraak, 2020). De toename van het aantal pedojagers valt samen met een jaarlijkse toename van het aantal

beelden van kindermisbruik dat online circuleert en dat onlangs ook werd gevonden binnen het grootste mainstream platform Pornhub (Grant, 2020a, 2020b).

Doxing

Doxing is het openbaar maken van iemands persoonlijke, gevoelige of privé-informatie. Bijvoorbeeld zijn of haar adres, telefoonnummer, paspoort of werkgever, of gegevens van familieleden of foto's van kinderen (MacAllister, 2016). De term doxing is waarschijnlijk afkomstig van de online hackergroep Anonymous en verwijst naar 'docs' of 'documenten'.

Net als veel andere vormen van immoreel en schadelijk gedrag, staat doxing vaak niet op zichzelf, maar is het een strategie die onderdeel vormt van andere vormen van **online** lastigvallen, zoals bedreiging en wraakporno. Doordat persoonlijke informatie op straat komt te liggen, lopen slachtoffers na doxing ook offline risico op schadelijk gedrag.

De informatie die wordt gebruikt om iemand te doxen, is vaak openbaar online beschikbaar, zonder dat daar hacking voor nodig is. Stel bijvoorbeeld dat iemand zijn werkgever op het professionele sociaalnetwerk LinkedIn heeft vermeld, dan kan dit in combinatie met andere publieke bronnen gebruikt worden als middel in een haatcampagne.

De redenen van mensen om aan doxing te doen, zijn niet eenduidig. Het kan louter voor het plezier zijn, maar het gedrag kan ook politieke doeleinden dienen of een middel zijn om aan 'zelfregulering' te doen in bepaalde gemeenschappen. Een voorbeeld van dit laatste is de haatcampagne #gamergate, die gericht was op vrouwen in de gaming community. De reden hiervoor was dat mannelijke leden van de gemeenschap zich verzetten tegen meer diversiteit onder gamers. Doxing vanuit seksistische motieven werd er veelvuldig gebruikt om vrouwen te bedreigen en te intimideren (Wingfield, 2014).

Doxing is **schadelijk** omdat privé-informatie van slachtoffers online gemakkelijk misbruikt kan worden, bijvoorbeeld om iemand te bedreigen of iemands werkgever lastig te vallen. Mensen die zich mengen in het publieke debat, kunnen bovendien bang worden om zelf slachtoffer te worden van doxing. Dat kan leiden tot online zelfcensuur.

Doxing is juridisch complex omdat er in veel landen, waaronder Nederland, geen juridische grond bestaat om het gedrag te vervolgen (MacAllister, 2016). Demissionair minister Ferdinand Grapperhaus heeft wel aangekondigd nog vóór het zomerreces van 2021 te willen komen met wetgeving om het fenomeen aan te pakken. Ook onderzoekt het Openbaar Ministerie of concrete gevallen van doxing als strafbaar aangemerkt kunnen worden (Redactie ScienceGuide, 2021).

Omvang
In de wetenschappelijke literatuur zijn er weinig internationale, empirische studies gevonden over doxing – laat staan over de omvang van het fenomeen in Nederland. Doxing wordt wel al sinds begin jaren 2000 gesignaleerd. En tijdens de coronapandemie is het fenomeen een stuk zichtbaarder geworden.

Zo meldde de Nationaal Coördinator Terrorismebestrijding en Veiligheid (NCTV) in het *Dreigingsbeeld Terrorime Nederland*, dat ieder kwartaal gepubliceerd wordt, dat boosheid en ongenoegen over de coronamaatregelen onder meer leiden tot doxing door activisten die persoonsgegevens van mensen die werken bij de politie en politici online zetten, als intimidatietactiek (Von Piekartz, 2020) In januari 2021 zijn de gegevens van tientallen undercoveragenten (zogenoemde ‘romeo’s’) gedoxed, en na de avondklokrellen is de online jacht op hen opgevoerd via Telegram- en Facebookgroepen, meldde *de Volkskrant* (Von Piekartz & Bahara, 2021). De Facebookgroep Steungroep Boeren en Burgers, die 165.000 leden telt, zette doxing in als intimidatiemiddel door te dreigen met de publicatie van gegevens van tientallen romeo’s (Von Piekartz & Bahara, 2021). In radicaal-rechtse Telegramgroepen, zoals De Bataafse Republiek (5.000 leden), gaat ook al een tijd een lijst rond met huisadressen van ‘linkse’ journalisten en ministers, soms met de oproep om een ‘burgerwacht’ te vormen die de genoemde personen ‘geweldloos op non-actief kan stellen’ (Von Piekartz & Bahara, 2021). En in maart 2020 kregen Twitteraars bij hun voordeur stikkers geplakt van Vizier op Links, een anoniem account met circa 16.000 volgers, dat linkse opiniemakers, activisten en politici het leven probeert onmogelijk te maken.

In de pers komen ook voorbeelden uit de VS langs, zoals het Twitteraccount @YesYoureRacist ('ja, je bent racist'), dat de identiteit van racisten probeert te achterhalen. Toen in augustus 2017 via dit account werd opgeroepen tot hulp bij het opsporen van de betogers in Charlottesville, groeide het aantal volgers binnen een paar dagen van 65.000 naar bijna 400.000 (van Houwelingen, 2017). Dit geeft een indicatie van hoe snel het fenomeen kan groeien.

Shaming en cancellen

Online shaming is het publiekelijk uiten van morele kritiek als reactie op het overschrijden van sociale normen (Billingham & Parr, 2020). Mensen die aan online shaming doen, hebben niet altijd tot doel dat iemand zich gaat schamen. Vaker willen ze aandacht voor een sociale gewoonte of norm vragen, die ter discussie stellen en anderen mobiliseren voor hun doel.

Publieke shaming kan enerzijds bijdragen aan de instandhouding en bekrachtiging van bestaande sociale normen – ook wanneer daar geen wet- en regelgeving voor bestaat (Billingham & Parr, 2020). Zo kan online shaming een effectieve manier zijn om mensen op misbruik te attenderen of racisme en seksisme te ‘ontmaskeren’

(Billingham & Parr, 2019). Maar anderzijds kan dit soort gedrag ook ingezet worden om vrouwen hun stem te ontnemen in het publieke debat door hun publiekelijk aan de schandpaal te nagelen en jonge meisjes te shamen voor hun seksualiteit (Levey, 2018). Shame-sexting (het zonder toestemming maken en verspreiden van seksueel getinte beelden of video's) beschrijven we in paragraaf 3.5 en komt hier dus niet aan bod.

Het internet maakt shaming gemakkelijker doordat groepen mensen zich eenvoudig kunnen verenigen en doordat gemeenschappen **online** gemakkelijker aan zelfregulering kunnen doen. Stel dat iemand in een community op Facebook iets zegt dat als racistisch wordt ervaren, dan kan een veroordelende opmerking van één persoon ervoor zorgen dat alle leden zich tegen hem of haar keren. Het internet zorgt er bovendien voor dat shaming erg moeilijk te controleren en in de hand te houden is.

In een recent wetenschappelijk artikel in de *European Journal of Philosophy* stellen de auteurs vijf morele grenzen aan shaming (Billingham & Parr, 2020). Als deze grenzen niet overschreden worden, zou shaming gerechtvaardigd *kunnen* zijn. Dat geeft nog steeds geen 'recht' op shaming, maar biedt wel basis voor morele rechtvaardiging. Shaming kan dus soms een 'nobel doel' dienen, maar het middel zelf is lastig te rechtvaardigen.

Tabel 1 Vijf morele grenzen aan online shaming

Moreel criterium	**Toelichting**
Proportionaliteit	Shaming is alleen proportioneel als de negatieve gevolgen kleiner zijn dan de positieve.
Noodzakelijkheid	Shaming is alleen geoorloofd als er geen andere manier is om hetzelfde te bereiken.
Respect voor privacy	Shaming moet met respect voor privacy gebeuren. Dat betekent dat irrelevante informatie uit het verleden buiten beschouwing moet blijven en dat zeer gevoelige informatie het liefst privé moet blijven.
Geen haat	Shaming mag niet gepaard gaan met bedreiging, seksisme, racisme en andere vormen van haat.
Reïntegratie	Shaming moet niet onmogelijk maken dat iemand terug kan keren in een bepaalde gemeenschap en shamers moeten zich bewust zijn van het risico op uitsluiting.

Bron: Billingham & Parr, 2020

Billingham & Parr's onderzoek (zie Tabel 1) laat zien dat de morele rechtvaardiging van shaming complex is. De intenties van mensen die online shamen zijn belangrijk. Geven ze anderen de kans om te leren van hun gedrag en dit bij te stellen? Is het voor slachtoffers mogelijk om na het shamen weer terug te keren in een bepaalde gemeenschap?

Omdat het online veel lastiger is om de gevolgen van een actie in de hand te houden (bijvoorbeeld omdat het bereik ervan niet te controleren valt), is het in deze omgeving ook veel moeilijker om te voldoen aan de vijf criteria. Bij shaming zijn omstanders bovendien cruciaal; er is geen sprake van een traditionele slachtoffer-dader-relatie, maar van collectieve actie. Mensen kunnen online ook moeilijker verantwoordelijk worden gehouden voor de negatieve gevolgen van shaming. Het is immers een vorm van collectieve actie met gedeelde verantwoordelijkheid.

Een bekend voorbeeld van online shaming is de #Metoo-beweging. Deze gaf een stem aan vrouwen die te maken hadden gehad met seksueel overschrijdend gedrag en die niet gehoord werden door bestaande instituties (Mendes et al., 2018). Door het internet te gebruiken, probeerden ze alsnog een vorm van rechtvaardigheid te krijgen (Powell, 2015). De beweging heeft ervoor gezorgd dat organisaties zich meer bewust geworden zijn van seksueel grensoverschrijdend gedrag en beter luisteren naar aantijgingen van slachtoffers (Leopold et al., 2021). Daar staat echter wel tegenover dat gebruikers van de hashtag ook onterechte aantijgingen maken, en de mensen die ze aanklagen zodoende schade berokkenen.

Voor slachtoffers kan de **schade** van online shaming groot zijn. Als gevolg van een publieke aanklacht kunnen ze bijvoorbeeld uit een bepaalde gemeenschap worden gezet, of verhinderd worden hun werk uit te voeren. Shaming is op zich niet strafbaar, maar smaad en laster zijn dat in Nederland wel. Smaad is het gericht verspreiden van negatieve uitingen over een ander, met als doel dat anderen dit horen. Smaad verandert in laster wanneer de dader weet dat de uitingen niet kloppen.

Cancellen is een variant van online shaming waarbij wordt opgeroepen tot een vorm van sociaal straffen waarbij iemand wordt uitgesloten uit een gemeenschap. Cancellen gaat nog een stap verder dan shaming, omdat het nadrukkelijk gericht is op het ontnemen van iemands machtspositie. De grens is wel fluïde: in het kader van #MeToo, bijvoorbeeld, zijn er ook heel wat publieke figuren teruggetreden.

Een bekend voorbeeld van shaming en cancellen uit Nederland is de situatie rond D66-Kamerlid Sidney Smeets, die zich uit de Tweede Kamer terugtrok na aantijgingen van jongens en mannen van seksueel grensoverschrijdend gedrag richting minderjarigen. Sidney Smeets beriep zich op het feit dat hij nooit de wet zou hebben overtreden, terwijl zijn "shamers" zich juist op bepaalde morele normen

beriepen die niet per se in het strafrecht verankerd zijn. Op deze manier probeerden zij normoverschrijding aan te kaarten bij iemand met een belangrijke machtspositie.

In de VS riepen medewerkers van Apple in een online petitie op tot het aftreden van een nieuw aangenomen medewerker (Ghaffary, 2021). Hij werd beschuldigd van seksistische opmerkingen in het verleden. Na interne druk is de medewerker vertrokken. Omdat online uitingen vaak nog terug te vinden zijn jaren nadat ze geplaatst zijn, maakt het internet het gemakkelijker om iemand aan te spreken op uitspraken uit het verleden.

Omvang

Eén van de bekendste voorbeelden van online shaming is de #MeToo-beweging die op 24 oktober 2017 op Twitter begon. In de eerste 24 uur werd de hashtag 12 miljoen keer gebruikt (CBS, 2017). Over hoe groot deze beweging in Nederland is, zijn geen cijfers gevonden. We bespreken hier niet de omvang van shame-sexting en andere seksueel getinte vormen van shaming, omdat we dat scharen onder ‘pesterij en geweld’ en niet onder digitaal vigilantisme.

In Nederland gaat de Stichting Online Shaming (SOS) de strijd aan tegen online shaming. Naar aanleiding van een zaak aangespannen door SOS, bijvoorbeeld, verbood de rechter in januari 2021 de website ZwarteLijstArtsen.nl die al een decennium lang een soort publieke ‘schandpaal’ is voor artsen (NOS, 2021a). Op deze particuliere website stonden bijna 900 artsen en zorgverleners, veelal met foto, die werden neergezet als plegers van ‘medische misdrijven’ en als ‘falende zorgverleners’ (SOS, 2021).

Volgens NRC lijkt online shamen door de coronapandemie in hevigheid en omvang te zijn toegenomen. De krant noemt geen cijfers, maar duidt shaming als een veelgebruikt middel om anderen ertoe aan te sporen zich aan de coronamaatregelen te houden, bijvoorbeeld door foto’s online te verspreiden (Van Noort, 2020).

SOS heeft evenmin cijfers over de totale omvang van shaming en cancellen in Nederland. De term shaming is namelijk zeer breed en slachtoffers hebben vaak niet door dat dit is wat zij meemaken. Afhankelijk van het soort shaming melden zij zich wellicht wel bij andere meldpunten zoals de kindertelefoon, Slachtofferhulp Nederland, het Expertisebureau Online Kindermisbruik (EOKM) of Helpwanted.nl.

Bij EOKM en helpwanted.nl komen met ‘mondjesmaat’ meldingen binnen van moslimmeisjes die online terecht worden gewezen, omdat zij zonder hoofddoek of met een diepe decolleté op straat lopen. EOKM denkt dat het aantal meldingen slechts het topje van de ijsberg is, omdat mensen die met shaming te maken

hebben bang zijn voor *victim blaming*. Ze schamen zich zelf ook vaak voor beelden en durven daardoor niet om hulp te vragen.

3.4 Online haat

Online haat omvat **haatzaaien**, **discriminatie**, **bedreiging** en **intimidatie**. Online haat richt zich op individuen, maar is vaak ook bedoeld om achterliggende groepen te schaden. Bij online haat is een vorm van ***xenofobie*** (afkeer van alles wat vreemd is) de belangrijkste drijfveer. Online haat komt voort uit afkeer tegen bepaalde groepen mensen, ook als het zich richt op individuen. Slachtoffers van online haat ervaren dikwijls de gevolgen van verschillende vormen van immoreel en schadelijk gedrag. Zo kunnen ze te maken krijgen met racisme, stalking en doxing.

Slachtoffers van online haat worden veroordeeld op basis van verschillende onderdelen van hun identiteit. Iemand die bijvoorbeeld zowel zwart, vrouw, als lesbisch is, kan online haat ervaren die zowel racistisch, seksistisch als homofoob van aard is. Dit wordt in de sociologie, genderstudies en rechtsgeleerdheid ‘intersectionaliteit’ of ‘kruispuntdenken’ genoemd: als ongelijkheid zich voordoet langs verschillende assen die elkaar snijden. Dit speelt een grote rol bij online haat. Mannen kunnen te maken krijgen met bedreiging en intimidatie, maar ervaren veel minder vaak dan vrouwen seksuele intimidatie (Vogels, 2021).

Haatzaaien

Een uniforme definitie van haatzaaien (*hate speech* in het Engels) ontbreekt. De Raad van Europa definieert het als volgt: ‘haatzaaien omvat alle uitingsvormen die zorgen voor verspreiding, aanstichting, aanmoediging of legitimering van raciale haat, xenofobie, antisemitisme of andere vormen van haat gebaseerd op intolerantie’ (Council of Europe, 2021).[2] Online haatzaaien is zowel **schadelijk** voor slachtoffers als voor de maatschappij in de breedte, omdat het zorgt voor een onveilige omgeving voor iedereen. Mensen kunnen voorzichtiger worden in hun online uitspraken uit angst om slachtoffer te worden van online haat.

Haatzaaien op internet verschilt van haatzaaien op straat, zo blijkt uit onderzoek van UNESCO (2015). Ten eerste kan haatzaaiende content **online** erg lang beschikbaar blijven op verschillende platformen. Ten tweede kan haatzaaiende content snel weer opduiken op een ander platform, ook als content elders verwijderd is. Ten derde zorgt anonimiteit op internet voor handhavingsproblemen, al heeft de politie vaak wel mogelijkheden tot identificatie maar niet genoeg capaciteit. Als laatste zorgt het internationale karakter van het internet ervoor dat

2 Vertaling door het Rathenau Instituut.

het moeilijk is om ‘nationaal’ op te treden tegen haatzaaien dat in andere landen gehost wordt (Gagliardone et al., 2015).

In het Wetboek van Strafrecht zijn drie artikelen opgenomen over haatzaaien: Artikel 137c, Artikel 137d en Artikel 137e. In Nederland vallen ‘daden’ zoals bedreigingen niet onder haatzaaien, omdat die apart strafbaar gesteld zijn in het Wetboek van Strafrecht (zie paragraaf 3.4). GroenLinks en de ChristenUnie hebben in juni 2020 een wetsvoorstel ingediend om *hate crimes*, misdrijven met een discriminatoir oogmerk, zwaarder te bestraffen (Bhikhie, 2020).

Omvang
Volgens de Raad van Europa zijn de gerapporteerde gevallen van online *hate speech* slechts ‘het topje van de ijsberg’ (ECRI, 2019). In Nederland zou ongeveer een tiende van alle tweets gericht aan vrouwelijke politici haat of agressie bevatten. Dit blijkt uit onderzoek van de Utrecht Data School en de *Groene Amsterdammer* (Saris & Van de Ven, 2021). Zij onderzochten 339.932 tweets die tussen 1 oktober 2020 en 26 februari 2021 zijn gestuurd naar alle vrouwen op Nederlandse kieslijsten. De conclusie was dat met name volksvertegenwoordigers die naast hun vrouw-zijn ook tot een minderheidsreligie behoren of van kleur zijn, veel haat over zich heen krijgen (Saris & Van de Ven, 2021).

Ook vrouwelijke journalisten en wetenschappers zijn doelwit van haatzaaien. Vrouwelijke columnisten hebben bijvoorbeeld meer last van haatzaaien via *hate speech* dan hun mannelijke tegenhangers, bleek uit een onderzoek van *de Volkskrant* in 2017. Van dertig vrouwelijke columnisten van vijf kranten en twee opiniebladen was twee derde een enkele keer of meerdere keren online bedreigd. De helft van de ondervraagde columnisten voelde zich soms tot vaak geïntimideerd door online reacties (Linnemann & Melchior, 2017). Wetenschappers hebben ook steeds meer last van online haatzaaien, *hate speech* of ‘vit-riool’, volgens KNAW-president Ineke Sluiter. Zij noemt bekende wetenschappers die doelwit zijn, zoals Marion Koopmans, wier Twittertijdlijn na elk media-optreden volstroomt met kritiek en bedreigingen (Digan, 2021).

Een ander bekend (en al iets ouder voorbeeld) van haatzaaien tegen vrouwelijke politici in Nederland is de ‘Uitzwaaidag’ van politica Sylvana Simons die op 6 december 2016 op Facebook werd georganiseerd als evenement. Hierbij toonden meteen al 39.000 mensen interesse en zeker 16.000 wilden komen (Wiegman, 2016). Tijdens de verkiezingen in maart 2021 bleef het haatzaaien tegen deze politica aanhouden. Traditionele media sluiten hierdoor soms de opties voor reacties onder berichten over politici (RTL, 2021). Bij politica en klimaatactiviste Kauthar Bouchallikht bevat bijna een derde van alle aan haar gerichte tweets seksistische haat of islamhaat. Op het dieptepunt kwamen bij haar een dag lang elke drie minuten haatberichten binnen (Saris & Van de Ven, 2021).

Discriminatie

We bespreken drie veelvoorkomende vormen van online discriminatie: seksisme, racisme en homofobie. Discriminatie is strafbaar onder Artikel 137c, 137d en 137e van het Wetboek van Strafrecht.

Seksisme

Seksisme omvat gedrag of houdingen die discrimineren puur op basis van geslacht (EIGE). **Online** is seksisme vaak verbonden met andere vormen van online immoreel en schadelijk gedrag, zoals bedreiging en cyberpesten. Het International Center For Research On Women (ICRW) gebruikt de term “technologisch gefaciliteerd gender-gerelateerd geweld” voor alle vormen van cyberpesten, intimidatie en (verbaal) geweld waarbij seksisme een rol speelt (Hinson et al., 2018). Vaak zijn vrouwen hiervan het slachtoffer.

Volgens Amnesty International (2018) zijn online seksisme en misogynie (vrouwenhaat) vaak bedoeld om vrouwen te intimideren of te kleineren. Circa 7,1% van de tweets die vrouwen ontvangen, is problematisch of schimpend (Amnesty International, 2017). Ook passieve en indirecte seksistische opmerkingen verpakt als grapje kunnen schadelijk zijn voor het welzijn van vrouwen, zo blijkt uit onderzoek van Harvard uit 2015 (Fox et al., 2015). Online seksisme is dus **schadelijk** omdat het een onveilige situatie oplevert voor directe slachtoffers en bijdraagt aan een minder veilige en vrije omgeving voor vrouwen in het algemeen (Plan International, 2020).

In Nederland is seksisme niet expliciet opgenomen in het Wetboek van Strafrecht, in tegenstelling tot bijvoorbeeld in België en Frankrijk. Wel is seksisme strafbaar onder discriminatie in Artikel 137c, 137d en 137e van het Wetboek van Strafrecht.

Racisme

Online racisme, ook wel cyber-racisme genoemd, onderscheidt zich van offline racisme doordat de racistische uitingen **online** plaatsvinden, en doordat racistische mensen elkaar online gemakkelijk kunnen vinden en zich zo kunnen verenigen (Bliuc et al., 2018). Mensen met racistische denkbeelden gebruiken het internet om hun denkbeelden te valideren en zichzelf het gevoel te geven dat ze ergens bij horen. Het internet “empowert” hen. Mensen die zich racistisch gedragen, kunnen er gemakkelijk hun denkbeelden delen, onder andere door de anonimiteit die het internet biedt. De motivatie achter online racisme ligt vaak in het schaden van mensen van kleur, het uitlokken van conflict en het normaliseren van racistisch gedachtegoed in het publieke debat (Bliuc et al., 2018).

De **schade** van online racisme voor slachtoffers is uitgebreid gedocumenteerd (Bliuc et al., 2018), maar specifiek Nederlands onderzoek blijkt schaars. Een onderzoek naar racisme in online gaming communities laat bijvoorbeeld zien hoe mannen van kleur proberen om te gaan met online racisme door stil te blijven en

zich niet uit te spreken uit angst voor meer haat (Ortiz, 2019). De rol van omstanders in het expliciet maken van sociale normen speelt ook bij online shaming een grote rol, zoals we hebben laten zien in paragraaf 3.3.

Homofobie
Homofobie is een angst voor, of haat tegenover, personen die zich emotioneel of seksueel aangetrokken voelen tot personen van dezelfde sekse, maar bij uitbreiding wordt de term ook gebruikt voor angst of haat ten aanzien van bijvoorbeeld mensen met andere dan cisgenderidentiteiten.[3] Er is sprake van discriminatie als iemands handelen door homofobie is ingegeven.

Een vorm van homofoob immoreel en schadelijk gedrag **online** is ‘outing’. Outing is het bekendmaken van iemands genderidentiteit of seksuele voorkeur zonder zijn of haar toestemming. De **schade** daarvan kan enorm zijn, bijvoorbeeld in landen waar lgbtq+-mensen vervolgd worden of in families waarin zij niet geaccepteerd worden. De Vlaamse publieke omroep VRT wijdde in augustus van 2020 een uitgebreid artikel aan online chatgroepen waarin wordt opgeroepen tot het “opsporen” van mensen uit de lgbtq+ gemeenschap (VRT, 2020). In sommige gevallen werd zelfs geld aangeboden om bepaalde mensen fysiek aan te vallen. Dat draagt bij aan het creëren van online én offline onveiligheid voor mensen die daarvan het slachtoffer zijn.

Omvang
Cijfers over de totale omvang van discriminatie in Nederland belichten slechts een deel van het totaal. Veel incidenten worden niet als discriminatie erkend of zijn niet gemeld bij de instanties (MiND, 2020).

Volgens het Meldpunt internetdiscriminatie (MiND) registreerde de politie over 2019 in Nederland in totaal 5.487 discriminatie-incidenten (zowel online als offline) (MiND, 2020). In 2016 waren dit nog 4.376 meldingen bij de politie (Mink & Van Bon, 2017). Er is hier dus sprake van een toename. Bij de Antidiscriminatiebureaus en voorzieningen (ADV’s) was er een kleine daling van 4.382 discriminatiemeldingen in 2019 ten opzichte van 4.761 in 2016. (Mink & Van Bon, 2017). Van al deze meldingen bestaat maar een klein deel (iets meer dan een tiende) uit online discriminatie (zie Figuur 3.4).

Herkomst was in 2019 de meest voorkomende discriminatiegrond (2.156 bij de politie en 1.922 keer bij ADV’s), daarna komen seksuele gerichtheid (1.603 registraties politie), antisemitisme (768 registraties bij de politie) handicap (552 registraties bij ADV’s) en geslacht (515 registraties bij ADV’s) (MiND, 2020, p. 3).

3 Cisgender betekent ‘niet-transgender’: mensen die geboren zijn als jongen die zich ook jongen voelen, of die geboren zijn als meisje en zich ook meisje voelen. Hun genderidentiteit komt dan dus overeen met het geslacht dat hen toegewezen is bij de geboorte.

Bron: Rathenau Instituut

Figuur 3.4 Online haat

Bedreiging en intimidatie

Doordat groepen zich **online** gemakkelijk kunnen mobiliseren, kan bepaald gedrag al snel escaleren. Groepen mensen kunnen zich daardoor op één individu gaan richten, die het groepsgedrag als intimiderend kan ervaren (Blackwell et al., 2017). De intimidatie kan dan vormen aannemen variërend van het herhaaldelijk sturen van berichten, bellen met iemands werkgever tot het dreigen bepaalde foto's of informatie openbaar te maken. Alleen al het gevolgd worden door bepaalde accounts op sociale media, kan als intimiderend worden ervaren.

Het dreigen met seksueel geweld is iets waar vooral jonge vrouwen online mee te maken krijgen (Plan International, 2020). Uit onze interviews blijkt dat slachtoffers van collectieve online haat dit vaak in veel verschillende vormen meemaken. Bedreiging met bepaalde zware misdrijven is in Nederland strafbaar volgens artikel 285 van het Wetboek van Strafrecht als een delict tegen de vrijheid van het slachtoffer. Het maakt daarbij niet uit of het slachtoffer zich ook daadwerkelijk bedreigd heeft gevoeld. Het dreigen met seksueel geweld is iets waar vooral jonge vrouwen online mee te maken krijgen (Plan International, 2020). Uit onze interviews blijkt dat slachtoffers van collectieve online haat dit vaak in veel verschillende vormen meemaken.

Omvang

Over de totale omvang van online bedreiging en intimidatie in Nederland zijn weinig cijfers gevonden. Wel is bekend wie hier het meeste last van ondervindt. Van de jongeren van 12 tot 18 jaar heeft 5% last van niet-seksueel getinte vormen van bedreiging of intimidatie. Meisjes in deze leeftijdsklasse worden er vaker mee geconfronteerd dan jongens (7% tegen 4%) (CBS, 2018).

Circa 1,4% van internetgebruikers was in 2018 slachtoffer van één of meer incidenten, zoals stalking, bedreiging met geweld en laster (CBS, 2018). Het meest voorkomende niet-seksuele incident is laster (0,9 procent). Hierbij gaat het onder meer om roddels, foto's of filmpjes verspreiden en pesten. Stalking en bedreiging met geweld komen met 0,4 en 0,3 procent minder voor (CBS, 2018).

De aangiftepercentages zijn opvallend laag: bij laster 11%, bij stalking 11% en bij bedreiging 6%. De belangrijkste reden om geen melding of aangifte te doen van stalking, laster en bedreiging is dat vermoed wordt dat het niet helpt, waarbij vaak wordt aangegeven dat het geen zaak voor de politie is (CBS, 2019c). Uit onderzoek van WODC blijkt ook dat slechts een minderheid van de slachtoffers naar de politie stapt bij een online delict: bij slachtoffers van online bedreiging is dit 20,2% (Sipma & Leijsen, 2019).

Uit onderzoek van Amnesty International (zie infographic 3.4) blijkt dat in 2017 ongeveer 23% van de vrouwelijke deelnemers aan een survey in acht landen ten

minste één keer online beschimping of bedreiging had meegemaakt, variërend van 16% in Italië tot 33% in de VS (Amnesty, 2017). Nederland zat niet in deze studie.

Tot slot is het opvallend dat onder journalisten het aantal online en offline bedreigingen toeneemt (SVDJ, 2021). In Nederland heeft het nieuwe meldpunt PersVeilig in 2020 in totaal 121 incidenten geregistreerd, waarvan 36 via sociale media en 9 per e-mail. Op 19 april 2021 waren er bij PersVeilig al 95 melding binnengekomen, waarvan 25 via sociale media en 6 per e-mail.[4]

3.5 Online pesterij en geweld

De categorie ‘online pesterij en geweld’ bevat de fenomenen **cyberpesten**, **trolling en griefing, shame-sexting, sextortion en wraakporno, grooming**, en **extreme pranks**. De gemene deler is dat het hier gaat om het doelbewust kwetsen en schaden van individuen, waarbij ***sadisme*** of doelbewust willen kwetsen de belangrijkste drijfveer is van daders. Online pesterij en geweld omvat gedrag dat tot doel heeft om anderen online lastig te vallen en (seksueel) te intimideren zonder dat eigenrichting of xenofobie hierbij een rol spelen. Dit gedrag komt grotendeels overeen met de Engelse term *online harassment* (treiteren) waarvoor ook middelen uit andere categorieën (zoals sock puppetting of hacking) kunnen worden ingezet. Net als bij online haat, is online pesterij en geweld veel voorkomend, maar lijkt het aantal meldingen bij instanties hiervan een beperkt beeld te geven.

Cyberpesten

Wanneer pesten online gebeurt, wordt dat vaak cyberpesten genoemd. Cyberpesten is herhaaldelijk en doelbewust schadelijk gedrag online door een groep of individu tegen een slachtoffer dat zich moeilijk kan verdedigen (Juvonen & Gross, 2008). Dat maakt cyberpesten een parapluterm voor veel vormen van online schadelijk gedrag. Stalking, doxing en sock puppeting worden soms ook als cyberpesten gezien. Deze overlap is logisch, omdat pesten nu eenmaal in verschillende vormen plaats kan vinden. Wanneer iemand stelselmatig en doelbewust wordt buitengesloten uit online groepen op sociale media, heet dat ook cyberpesten.

Veel onderzoek naar cyberpesten richt zich op kinderen en jongeren. Van alle vormen van online schadelijk en immoreel gedrag, is cyberpesten binnen de wetenschap het gedrag dat al het langst onderzocht wordt. Bij volwassen wordt herhaaldelijk doelbewust schadelijk gedrag online vaak geen pesten genoemd, maar gezien als bijvoorbeeld lastigvallen of haat. Sociaalpsychologische factoren, zoals eenzaamheid en onzekerheid onder jongeren, spelen bij zowel online als

4 Bron: e-mail van vertegenwoordiger van Persveilig op 19 april 2021

offline pesten een rol. Wat pesten **online** anders maakt dan offline pesten, is dat het voor slachtoffers moeilijker is om hieraan te ontsnappen. Doordat het pesten niet meer gebonden is aan een fysieke plek, kunnen slachtoffers zich ook thuis onveilig voelen.

Omstanders van cyberpesten grijpen vaak niet in, omdat berichten online moeilijker te interpreteren zijn. Is hier echt sprake van cyberpesten of vindt het slachtoffer het zelf ook grappig? Ook de asynchroniteit van het internet speelt een rol in het niet ingrijpen, omdat niet altijd duidelijk is of iets al is ‘opgelost’. Anders dan offline hoeft cyberbullying namelijk niet plaats te vinden op het moment dat ook anderen op het internet zitten, maar kunnen berichten al langer online staan voordat omstanders ermee geconfronteerd worden (Cleemput et al., 2014).

De **schade** van cyberpesten voor slachtoffers kan enorm zijn. Onderzoek uit 2017 laat bijvoorbeeld zien dat jongeren die slachtoffer zijn van cyberpesten, significant meer suïcidale gedachten hebben en ook meer zelfmoordpogingen doen dan jongeren die niet worden gepest (Nikolaou, 2017). Cyberpesten heeft ook een negatief effect op de mentale gezondheid van jongeren en kan leiden tot depressies. Cyberpesten is net als offline pesten niet strafbaar (Shaikh et al., 2020).

Omvang
Cyberbullying of cyberpesten wordt door wetenschappers al onderzocht sinds de millenniumwisseling. Veel studies tonen aan dat cyberpesten een veelvoorkomend probleem is. Aangezien cyberpesten niet altijd op dezelfde manier wordt gemeten, lopen de percentages hiervan sterk uiteen. Zo varieerde het percentage van de jongeren dat aangaf ooit slachtoffer te zijn geweest van cyberpesten in 2012 van 4 tot 57 procent (Dehue et al., 2012). Het percentage mensen dat cyberpesten een keer heeft meegemaakt ligt tussen de 20% en 40%; 3% tot 10% maakte dit meerdere malen mee in het voorafgaande jaar (Völlink et al., 2016).

Uit de Veiligheidsmonitor van het CBS blijkt dat 4,2% van de Nederlanders slachtoffer was van een of meerdere vormen van cyberpesten in 2019: een toename ten opzichte van 2017 en 2012 (beide jaren 3,1%) (CBS, 2019c). Uit Statline-cijfers van het CBS blijkt dat maar 22,2% melding en aangifte deed van cyberpesten (CBS, 2021).

Uit de Monitor Sociale Veiligheid die scholen ieder jaar houden, leek pesten (zowel online als offline) af te nemen in het basisonderwijs en voortgezet onderwijs tussen 2014 en 2018 (zie Figuur 3.5). Op de basisschool zou bijna 8% worden gepest via e-mail, chatgesprekken of sms. Pesten via internet komt in het voortgezet onderwijs meer voor dan in het basisonderwijs (NJi, 2019).

Bron: Rathenau Instituut

Figuur 3.5 Online pesterij en geweld

Trolling en griefing

In de meest enge definitie van **trolling** gaat het om het opzettelijk dwars zitten van mensen in online gemeenschappen met gedrag dat niet als acceptabel wordt gezien, zoals mensen uitschelden, ruzie zoeken of zich negatief uitlaten over anderen (Cheng et al., 2017). Het begrip trolling wordt sinds het begin van de 21ste eeuw gebruikt om vormen van asociaal gedrag in online gemeenschappen te beschrijven, zoals op Wikipedia of op online fora. In de media wordt trolling vaak beschreven als gedrag dat wordt geuit door mensen die sadistische en asociale persoonlijkheidskenmerken hebben. Uit recent onderzoek blijkt echter dat iedereen een troll kan zijn en dat onder andere iemands stemming en de context van discussies online grote invloed heeft op ‘trolgedrag’ van mensen.

Trolling gebeurt vaak zonder na te denken over de gevolgen voor slachtoffers, meestal onder het mom van humor. Slachtoffers die ingaan tegen mensen die aan trolling doen, kunnen juist nog meer trolling verwachten; trolling is voor veel mensen ook een soort spel is om de ander zoveel mogelijk te traineren (Phillips, 2015). Al in 2002 bleek uit onderzoek dat feministische en andere non-mainstream online gemeenschappen sneller slachtoffer worden van trolling, waardoor bepaalde groepen in de samenleving meer **schade** zullen hebben van trolling dan anderen (Herring et al., 2002).

Daarnaast is er ook een bredere interpretatie van het begrip trolling, waarbij ook bepaalde vormen van informatiemanipulatie onder deze noemer worden geschaard. Een voorbeeld is het gebruik van nepaccounts om desinformatie te verspreiden en het publieke debat te beïnvloeden. Hierbij wordt de definitie van trolling opgerekt, omdat het hier niet alleen gaat om het opzettelijk dwarszitten van (specifieke) mensen. Bij bijvoorbeeld het gebruik van nepaccounts is het gedrag meer gericht op het manipuleren van informatie en het publieke debat.

Trolling is niet strafbaar, tenzij het gepaard gaat met strafbare feiten als bedreiging, smaad of laster.

Griefing is een vorm van trolling die vooral in de gaming community tot uiting komt, en waarbij andere spelers opzettelijk lastiggevallen worden. Griefing bevat elementen van cyberpesten maar is vaak niet specifiek op één persoon gericht. Ook is vaak de ‘avatar’ van spelers in games het doelwit en niet de persoon achter de avatar. Anders dan bij cyberpesten kunnen spelers gemakkelijker ‘ontkomen’ aan griefing, door bijvoorbeeld te stoppen met gamen of de griefer te blokkeren (Coyne et al., 2009).

Omvang

Trollen op het internet worden al sinds eind jaren negentig onderzocht. Er bestaan ook empirische studies naar de aard en omvang van sommige vormen van trolling

in Nederland. Zo blijkt er in ons land geen gecoördineerd trollennetwerk te bestaan, maar er zijn wel voorbeelden van trollennetwerkjes.

Uit onderzoek van *NRC* bleek in 2017 dat diverse internettrollen op het gebied van desinformatie (zie ook hierboven) actief waren voor de politieke partij Denk. Gezamenlijk verspreidden zij 1.636 berichten en deelden ze 2.171 likes uit ten behoeve van de Denk-campagne (Kouwenhoven & Logtenberg, 2017). Bij GroenLinks in Zeist werd in februari 2021 bekend dat een partijlid trollen had ingezet om sociale media te beïnvloeden. Hij zou verschillende nepprofielen hebben aangemaakt waarmee hij zich in discussies mengde en zijn mening door een niet bestaand iemand liet verkondigen (Knieriem, 2021). Uit weer een ander onderzoek in 2017 bleek dat de trollen voornamelijk negatieve tweets over linkse partijleiders schreven en juist positief over rechts (Borra et al., 2017). Bij big-data-onderzoek van Pointer naar trollen in verkiezingstijd zijn 476 Nederlandstalige trollen gevonden in de periode rondom de laatste Amerikaanse verkiezingen (Pointer, 2021a). Of er sprake is van een toename en hoe dit cijfer zich verhoudt ten opzichte van andere landen, werd niet door Pointer vermeld.

De Universiteit van Amsterdam bestudeerde ruim vijf miljoen tweets waarin tussen 1 januari 2020 en 31 december 2020 Nederlandse politieke leiders werden genoemd (Sabel & Verhagen, 2021). Daarbij keken de onderzoekers ook naar ‘trolachtig gedrag’ en de ‘klassieke pestkoppen’, de intensieve internetgebruikers die soms geautomatiseerd discussies proberen te kapen en politici bestoken met tweets. Een kleine 0,4% van alle accounts bleek dit soort gedrag te vertonen (Sabel & Verhagen, 2021). Deze circa 1.000 accounts zijn wel verantwoordelijk voor 15% van de vijf miljoen onderzochte politieke tweets (zie Figuur 3.5). Van deze groep beweegt de overgrote meerderheid zich aan de rechterkant van het politieke spectrum (Sabel & Verhagen, 2021).

Een voorbeeld van een Nederlandse troll-account is @PeterBrekelmans, die op Twitter sinds halverwege 2020 al bijna 125.000 tweets verstuurde. Bij voorkeur reageert hij op tweets van linkse politici of waarin linkse politici worden genoemd. Twee andere troll-acounts zijn @loweshenny (tussen oktober 2020 en maart 2021 ruim 26.500 tweets en retweets) en @ChrisHagenviet (16.500 tweets sinds 2013) (Sabel & Verhagen, 2021). Verder bleek ook in 2018 dat de zanger Dotan een leger van circa 140 trollen (verzonnen profielen op Facebook en Twitter) had ingezet om zijn imago en inkomen te verbeteren (Miserus & Van der Noordaa, 2018).

Shame-sexting, sextortion en wraakporno

Shame-sexting is volgens kennisinstituut Movisie ‘het ongevraagd doorsturen van seksueel getinte beelden met als doel om de afgebeelde persoon aan de schandpaal te nagelen’ (Movisie, 2019). Sexting is op zichzelf niet schadelijk of strafbaar en past volgens Bureau Jeugd en Media bij gezond experimenteergedrag

van jongeren (Kleijer, 2015). Volgens Rutgers, het Nederlandse kenniscentrum seksualiteit, wordt sexting pas schadelijk als anderen *ongevraagd* met die seksueel getinte beelden aan de haal gaan (Rutgers, 2018). Dan noemen we het shame-sexting.

In combinatie met andere fenomenen als hacking en phishing en door versterking van **online** mechanismen (viraliteit, schaalbaarheid) kan shame-sexting leiden tot **schade** die veel groter is dan offline mogelijk zou zijn. Dat bleek recentelijk bijvoorbeeld in het geval van het dertienjarige meisje Desteny, dat van een flat sprong nadat online seksueel getinte filmpjes van haar circuleerden. De verspreiding van seksueel beeldmateriaal kan bovendien leiden tot sextortion of wraakporno.

Sextortion is een vorm van afpersing waarbij een dader dreigt om zonder toestemming seksueel beeldmateriaal te openbaren van een slachtoffer, om deze te dwingen om meer van dit soort foto's te sturen, te betalen of om (seksueel getinte) opdrachten uit te voeren (Patchin & Hinduja, 2020; Wolak et al., 2018). Sextortion kan als chantage strafbaar zijn onder artikel 318 van het Wetboek van Strafrecht (afpersing en afdreiging).

Wraakporno is het zonder toestemming bezitten, openbaar maken en verspreiden van (gestolen) seksueel beeldmateriaal bijvoorbeeld door hackers, (ex)partners, kindermisbruikers, verkrachters en mensenhandelaren (Rijksoverheid, z.d.). Anders dan bij sextortion gaat het hier niet om afpersing, maar over doelbewust schade toebrengen aan slachtoffers door het naar buiten brengen van de beelden. Sinds 1 januari 2020 is wraakporno strafbaar onder artikel 139h van het Wetboek van Strafrecht. Zowel het bezitten als het verspreiden van seksueel beeldmateriaal zonder toestemming is hierdoor strafbaar gesteld. Minister Grapperhaus heeft in 2020 de opdracht gegeven aan het WODC om te laten onderzoeken of het creëren en verspreiden van 'deepnudes', waarbij met behulp van kunstmatige intelligentie naaktbeelden van iemand gemaakt kunnen worden, ook strafbaar zou moeten worden.

Sommige meisjes met verstoord eetgedrag vallen ten prooi aan zogenaamde pro-ana-coaches die uit zijn op seksueel getinte beelden. Volgens het Centrum tegen Kinderhandel en Mensenhandel (CKM) gaat het met name om mannen tussen de 20 en 30 jaar die zich voorstellen als coach om gewicht te verliezen, maar in de praktijk al snel vragen om seksueel getinte foto's of filmpjes (Simons et al., 2020). Het opzettelijk schaden van iemands gezondheid kan strafbaar zijn als mishandeling onder artikel 300 van het Wetboek van Strafrecht, waardoor online pro-ana-coaches mogelijk strafbaar handelen.

Omvang

Experts maken zich zorgen over grensoverschrijdend gedrag online en het lekken van privé-beelden onder minderjarigen. Toch zijn er weinig concrete cijfers en veel blijft onder de radar (Wagemakers & Toksöz, 2021). Alleen wat wordt gemeld, wordt bijgehouden. Bij de Amsterdamse politie ging het om 224 gevallen van sextortion tussen 2019 en eind 2020 en bij Helpwanted.nl om 2.683 meldingen van sextortion (Wagemakers & Toksöz, 2021).

Uit een recent onderzoek van het CBS bleek dat in 2020 bijna 30% van de 16 tot 18-jarige vrouwen en 23% van de 18 tot 24-jarige vrouwen aangeeft, in de afgelopen 12 maanden online te zijn geconfronteerd met ongewenst seksueel gedrag (CBS, 2020a). Daarnaast gaat het respectievelijk om 9% en 8% van hun mannelijke leeftijdgenoten (Wagemakers & Toksöz, 2021). Onder het begrip seksuele intimidatie of ongewenst gedrag vallen in het onderzoek van CBS zaken zoals: seksueel getinte opmerkingen, ongewenste aanrakingen of gedwongen worden om seksuele dingen te doen. Ruim een derde (36%) heeft deze ervaring van seksuele intimidatie voor zich gehouden (CBS, 2020a).

Helpwanted.nl, de hulplijn voor online seksueel misbruik, zag tijdens de eerste lockdown een verdubbeling van het aantal meldingen. De hulplijn liet onderzoeken of jongeren inderdaad meer negatieve ervaringen op het internet hadden dan voor de lockdown. Ze lieten een peiling doen door Safer Internet Centre Nederland onder 1.164 jongeren van 12 tot en met 25 jaar. Hieruit bleek eerder een daling dan een toename. Vóór corona had een derde van de jongeren weleens een negatieve online seksuele ervaring, terwijl dit tijdens corona ruim een kwart was (28%) (ECP, 2021). Van de pubers gaf 64% vóór corona aan dat ze online sexting als vervelend ervaren; tijdens corona zakte dit percentage naar 39%. Het percentage jongeren tussen 12 en 18-jaar dat sexting als iets positiefs ziet, is tijdens corona gestegen van 19% naar 28% (ECP, 2021).

Waar via sexting beelden worden verspreid, liggen shamesexting en wraakporno op de loer. Ongeveer 14% van de jongens en 9% van de meisjes heeft een online naaktfoto of seksfilmpje van een bekende gezien. Dit kan ook 23% en 24% zijn, afhankelijk van het opleidingsniveau: hoe lager het opleidingsniveau, hoe hoger het percentage (Rutgers & Soa Aids Nederland, 2019). Uit internationaal onderzoek blijkt dat de laatste zes maanden voor hun ondervraging 3,1% van de 4.453 kinderen tussen 11 en 18 jaar zelf (bijna) naaktfoto’s van zichzelf had verstuurd (Lewis, 2018).

Uit cijfers van het Ministerie van Justitie en Veiligheid (2021) blijkt dat de politie tussen 2015 en 2019 bijna 16.500 incidenten van seksueel geweld tegen kinderen heeft behandeld. In 2017 ging bijna 10% van de geregistreerde incidenten over

ongewenste sexting. In 2019 nam dit toe tot bijna 14% (Ministerie van Justitie en Veiligheid, 2021).

Wat zorgelijk lijkt bij online seksuele incidenten en stalking, is dat een hoog percentage, 40% tot 50% van de daders, totale vreemden zijn van het slachtoffer (CBS, 2018). Bureau Halt signaleert dat er steeds meer vraag is naar hun voorlichting en preventieprogramma's op het gebied van online veiligheid en sexting. Een oorzaak hiervan is dat kinderen steeds jonger een smartphone hebben, en dat voorlichting over de risico's van het gebruik van een mobiele telefoon volgens één van de geïnterviewde experts dus te laat komt: vaak pas in groep 8 of de brugklas.

Grooming

Grooming is een proces waarbij een volwassene een relatie van seksueel kindermisbruik ontwikkelt door middel van technologie, zoals in chatrooms en op sociale media (Lorenzo-Dus, 2017). In dit verband wordt ook gesproken van 'digitale kinderlokkerij' (EOKM, 2020). Alleen de interactie tussen volwassene en kind kan al voor seksueel genot van de volwassene zorgen – wat het een vorm van seksueel misbruik maakt. **Online** grooming verschilt van offline grooming, omdat volwassenen veel gemakkelijker (anoniem) contact kunnen leggen met kinderen. Ook blijkt uit onderzoek dat jongeren online sneller risicovol gedrag vertonen (Whittle et al., 2013).

Grooming is **schadelijk** omdat het kan leiden tot seksueel misbruik van minderjarigen. Ook wanneer het misbruik zich alleen online afspeelt, zorgt het voor een onveilige situatie voor minderjarigen, ongeacht of ze tot handelingen gedwongen werden. Sinds 2010 is grooming strafbaar in Nederland onder artikel 248e van het Wetboek van Strafrecht. Voor de strafbaarheid is het noodzakelijk dat de dader een 'voorstel tot ontmoeting' heeft gedaan richting het slachtoffer.

Omvang

Over de totale omvang van grooming in Nederland zijn geen cijfers gevonden, omdat hiervan net als bij veel andere fenomenen vaak geen melding wordt gedaan bij de instanties. Het EOKM en Helpwanted duiden het fenomeen 'grooming' als 'kinderen online benaderen voor seks'. Van het totale aantal keren dat contact is gezocht met Helpwanted (6.318 keer in 2020), ging het in circa 9% van de gevallen om een hulpvraag met betrekking tot grooming of het online benaderd worden voor seks (EOKM, 2020, p. 13). In 2019 was het totale aantal meldingen van kinderpornografie vijf keer zo hoog als in 2015. Het steeg van 5.534 (in 2015) naar 25.628 (in 2019) (Ministerie van Justitie en Veiligheid, 2021).

Extreme pranks

Extreme pranks zijn een vorm van vernedering die plaatsvinden tussen dader, slachtoffer en omstanders (Hobbs & Grafe, 2015). Het doel is vaak om emoties uit te lokken. Vaak is bij extreme pranks sprake van een ongelijke gezagsverhouding, zoals bij ouders die hun kinderen voor de gek houden. **Online** nemen extreme pranks vaak de vorm aan van gefilmde ‘offline’ pranks, waarbij de reactie van slachtoffers breed wordt uitgemeten. Een voorbeeld van een extreme online prank die in 2002 begon is de Scary Maze Game. In het spel moesten deelnemers een doolhof oplossen waarvoor veel concentratie vereist was. Middenin het spel verschenen uit het niets beelden uit horrorfilms, gepaard met een harde gillende stem. Filmpjes van kinderen die huilend en schreeuwend reageren op dit soort video’s zijn populair op sites als YouTube.

In januari 2019 kondigde YouTube aan om schadelijke pranks en pranks die kunnen leiden tot serieus letsel voortaan te verbieden op hun platform (YouTube, 2019). Daaronder vallen ook pranks die mensen laten geloven dat ze in groot gevaar zijn. In de praktijk blijkt het lastig om te beoordelen wanneer pranks **schadelijk** zijn. Onderzoekers zijn het niet altijd eens over wanneer pranks een onschuldige grap zijn, of schadelijk en sadistisch van aard. Vaak plaatsen daders en omstanders van extreme pranks hun eigen gevoelens boven die van het slachtoffer. Uit een onderzoek uit 2017 blijkt dat zowel daders als omstanders vaak genieten van extreme pranks, ook als zij weten dat het slachtoffer geschaad wordt (Burris & Leitch, 2018).

Het schrikeffect van slachtoffers is een essentieel onderdeel van extreme pranks. Dat betekent dat slachtoffers vaak geen toestemming geven om deel te nemen aan een extreme prank. Ook is het online niet te verifiëren of slachtoffers weten dat ze online te zien zijn als onderdeel van een extreme prank. De fysieke en emotionele schade van de extreme prank zelf wordt online versterkt, omdat de vernedering van het slachtoffer online doorgaat. Dat maakt online extreme pranks mogelijk nog schadelijker dan offline extreme pranks, die niet met een groot publiek gedeeld worden. Extreme online pranks kunnen strafbaar zijn als ze gepaard gaan met fysiek geweld.

Omvang

Over extreme online pranks vonden wij weinig wetenschappelijke publicaties. Voor Nederland is het daarom heel lastig om vat te krijgen op de omvang van het fenomeen. Bovendien zijn veel van de voorbeelden van gewelddadige pranks die in de nieuwsmedia genoemd worden, niet Nederlands. De omvang van extreme pranks in Nederland en hoeveel slachtoffers ze maken, lijkt dus nog niet goed in beeld te zijn gebracht. Er is ook geen nationaal meldpunt of stichting voor, zoals bij racisme of shaming.

In Nederland zijn wel minderjarigen opgepakt die zich schuldig maakten aan de zogenaamde *happy-slapping*-pranks (Nu.nl, 2020). Daarbij wordt een willekeurige persoon fysiek aangevallen, en komen er beelden van het incident online te staan. De gevallen van *happy slapping* in Nederland vonden plaats tussen 21 augustus en 7 oktober 2020, in de omgeving van het Osdorpplein en de Tussen Meer in stadsdeel Nieuw-West van Amsterdam.

In de VS zijn meer voorbeelden van gewelddadige en schadelijke pranks bekend. Daar haalden Heather en Michael Marting schadelijke pranks uit met hun kinderen en werden hun filmpjes hiervan honderdduizenden keren bekeken. Het gezin verdiende geld aan de video's en raakte mede hierdoor uiteindelijk de voogdij kwijt over hun kinderen. In 2017 zijn de accounts van de familie Marting verwijderd van YouTube (RTL Nieuws, 2017).

Internationaal bekende, omstreden (Engelstalige) prank-vloggers zijn Roman Atwood (met 16,5 miljoen YouTube-abonnees), Ken Duchamp (met 514.000 YouTube-abonnees) en Sam Pepper (2 miljoen YouTube abonnees). Zij verdien(d)en tonnen per jaar door pranks uit te halen met hun familie of andere slachtoffers. De Zweedse PewDiePie, de YouTuber die op dit moment het allermeeste abonnees ter wereld heeft, zit op 109 miljoen volgers.

3.6 Cyberbedrog

Cyberbedrog is het inzetten van technologische middelen om te bedriegen voor eigen gewin. Onder cyberbedrog vallen hier hacking, phishing, catfishing en cryptofraude. Het gaat om bedrog met ***hebzucht*** (financieel of persoonlijk gewin) als de belangrijkste drijfveer van daders. Bij deze vorm van bedrog wordt online vaak een andere identiteit aangenomen via hacking, phishing of catfishing. Veel fenomenen die binnen deze categorie vallen zijn (deels) strafbaar en worden soms onder de noemer 'cybercriminaliteit' geschaard. Alle fenomenen in deze categorie zouden niet los van het internet kunnen bestaan; vandaar de toevoeging van het woord *cyber* aan deze categorie. We bespreken dus ook niet systematisch het **online** karakter van de fenomenen in deze categorie.

Cyberbedrog wordt niet enkel ingezet voor financieel gewin, maar kan ook worden ingezet om informatie te ontfutselen. Door hacking kunnen bijvoorbeeld naaktfoto's worden buitgemaakt, waardoor iemand vervolgens kan worden afgeperst. Opvallend is dat phishing relatief veel meer voorkomt dan hacking en dat de schade hiervan aanzienlijk kan zijn. Tijdens de coronacrisis is ook het aantal malafide webwinkels en slachtoffers hiervan aanzienlijk gegroeid.

Onwetendheid is bij cyberbedrog een groot probleem. Veel slachtoffers van cyberbedrog doen geen aangifte bij de politie. In Nederland wordt van ‘traditionele’ vormen van misdaad veel vaker aangifte gedaan dan van bijvoorbeeld hacking en phishing (van de Weijer et al., 2019). Cyberbedrog heeft vaak betrekking op het bedriegen van een individu (behalve bij hacking, waarbij kwetsbaarheden in computersystemen worden misbruikt in plaats van individuen). Dit is een verschil met informatiemanipulatie, dat vaak betrekking heeft op het bedriegen van grotere groepen mensen in de samenleving. Onwetendheid is bij cyberbedrog een groot probleem. Veel slachtoffers van cyberbedrog doen geen aangifte bij de politie.

Hacking

Hacking omvat allerlei activiteiten rondom het ongeautoriseerd toegang (proberen te) verkrijgen tot computersystemen (Furnell, 2009, p. 173). Niet alle vormen van hacking zijn schadelijk. Ethische hackers of *white hat hackers* brengen geen schade toe aan systemen. Hun motivatie is juist het veiliger maken van computersystemen. Zogenaamde *black hat hackers* zijn wel specifiek uit op het toebrengen van **schade** aan systemen of het ontfutselen van vertrouwelijke informatie (Aiken et al., 2016). Zowel uit wetenschappelijk onderzoek als uit onze interviews blijkt dat jonge hackers het hacken vaak zien als een spel: een uitdaging om technische barrières te overbruggen en fouten in systemen op te sporen. Vaak zijn zij zich niet voldoende bewust van de schade die hacking kan veroorzaken en de gevolgen voor henzelf wanneer ze worden ontdekt.

Sommige hackers hebben als motivatie het openbaar maken van geheime gegevens, zoals hackers die WikiLeaks van vertrouwelijke informatie voorzagen uit de e-mails van Hillary Clinton in 2016. De term ‘hacktivist’ duidt op mensen die vanuit activistische overtuigingen hacken. Een voorbeeld hiervan is Aaron Swartz, die alle academische artikelen van JSTOR openbaar maakte, omdat hij vond dat academische kennis voor iedereen gratis beschikbaar zou moeten zijn (Naughton, 2015). In afwachting van zijn veroordeling pleegde hij zelfmoord.

In de praktijk wordt hacking vaak ook ingezet als middel voor andere vormen van schadelijk gedrag. Zo kunnen met hacking persoonlijke gegevens buitgemaakt worden die gebruikt kunnen worden voor bedreiging. Of jongeren kunnen worden gepest door iemands Instagram te hacken en zo in diens naam berichten te plaatsen. Net als bij traditionele vormen van inbraak, verschilt de motivatie van hacking. Deze is niet altijd financieel.

Hacking is strafbaar onder artikel 138ab lid 1 van het Wetboek van Strafrecht als computervredebreuk. Hiervoor moet iemand opzettelijk en zonder toestemming een computersysteem binnendringen. Als bij het hacken ook gegevens worden gestolen, is artikel 138ab lid 2 ook van toepassing, waarin het stelen of kopiëren van gegevens uit computersystemen strafbaar wordt gesteld.

Omvang
In Nederland lijkt de omvang van hacking licht toe te nemen. Bij hacken was het slachtofferpercentage in 2017 4,9% en in 2019 5,5% (CBS, 2019a). De leeftijdsgroep die het meest is getroffen zijn de 25 tot 44-jarigen, met 6,4% in 2019. Ouderen boven de 65 zijn het minst slachtoffer met 4,1% in 2019 (CBS, 2019a).
Slechts een klein deel van de slachtoffers van hacking doet hiervan aangifte. Van alle gevallen van identiteitsfraude, koop- en verkoopfraude, hacken en cyberpesten samen is in 2019 ongeveer 1 op de 8 (13%) gemeld bij de politie (CBS, 2019a).

Phishing

Phishing is een manier om via het internet allerlei soorten informatie te stelen over personen en organisaties, bijvoorbeeld via valse e-mails of websites (Vayansky & Kumar, 2018). Anders dan bij hacking wordt bij phishing gebruik gemaakt van *social engineering*: het gebruik van misleidende tactieken om gegevens te achterhalen. Door bijvoorbeeld e-mails helemaal af te stemmen op de persoonlijke situatie van de ontvanger, kan de indruk worden gewekt dat het om een betrouwbare e-mail gaat. Voor phishing hoeft dus niet ingebroken te worden in computersystemen, maar worden juist menselijke kwetsbaarheden misbruikt.

Slachtoffers van phishing lopen niet alleen vaak financiële **schade** op, maar kunnen ook schaamtegevoelens ontwikkelen of minder vertrouwen in mensen krijgen. Een bekend voorbeeld van hedendaagse phishing is WhatsApp-fraude, waarbij criminelen zich voordoen als iemands kind met acute geldproblemen. Door in te spelen op de psychologie van slachtoffers en door de urgentie te onderstrepen, maken mensen soms duizenden euro's over naar criminelen die zich voordoen als naaste familie. Slachtoffers stappen niet altijd naar de politie omdat ze zich schamen voor wat er is gebeurd.

In Nederland is phishing strafbaar onder Artikel 326 (oplichting) en/of Artikel 225 (valsheid in geschrifte) van het Wetboek van Strafrecht.

Omvang
Volgens de politie zit de groei van cyberbedrog de afgelopen jaren met name bij phishing (zie Figuur 3.6). Dit wordt bevestigd door cijfers uit eerdere jaren (Lastdrager, 2018). Tussen 2012 en 2014 ging het om 0,4% van de Nederlandse bevolking boven de 14 jaar die identiteitsfraude had ervaren (Lastdrager, 2018, p. 2). Volgens een andere meting had in 2015 ongeveer 4,5% van de Nederlanders dit had ervaren (Paulissen & Van Wilsem, 2015).

Het aantal aangiften van phishing bedroeg 14% van het totale aantal gerapporteerde slachtoffers in 2014 (Lastdrager, 2018, p. 2)**.** Er is een relatief lage drempel voor daders van phishing en zij zijn lastig te traceren. Eind 2020 werd in

Nederland pas de eerste bouwer van dit soort software aangehouden: een 19-jarige oplichter die met phishing ongeveer een ton verdiende (Heck, 2020).

Bron: Rathenau Instituut

Figuur 3.6 Cyberbedrog

Catfishing

Catfishing is het opzettelijk misleiden van anderen door delen van je eigen identiteit te verbergen of te veranderen, meestal in de context van online dating en soms zonder de intentie om iemand ooit in het echt te ontmoeten (Mosley et al., 2020). Door zich anders voor te doen, proberen catfishers hun kansen op de datingmarkt te vergroten. We scharen catfishing onder ‘cyberbedrog’ omdat het hier gaat om het bedriegen van individuen voor eigen gewin, terwijl het bij informatiemanipulatie (waaronder bij sock puppeting) gaat om het bedriegen van grotere groepen mensen uit de samenleving. Afhankelijk van iemands intenties zou catfishing ook onder ‘online pesterij en geweld’ kunnen vallen.

Uit onderzoek blijkt dat het vooral mannen zijn die catfishen. Wanneer catfishers in het echt afspreken, kan dit tot nare en gevaarlijke situaties leiden voor slachtoffers, die de catfishers vaak helemaal niet herkennen. Ook het expres gebruiken van verouderde foto’s of het liegen over iemands leeftijd kunnen als catfishing worden beschouwd. Doordat het **online** gemakkelijk is om een andere identiteit aan te nemen en anoniem te opereren, is catfishing een typisch internetfenomeen. Catfishing kan ook als tool ingezet worden voor andere vormen van online schadelijk gedrag, zoals cyberpesten.

De mentale **schade** voor slachtoffers van catfishing is groot, omdat hun vertrouwen ernstig wordt geschaad en catfishing ook tot onveilige situaties kan leiden. Uit een onderzoek naar kwetsbare lgbtq+-mannen in Amerika bleek dat catfishing kan bijdragen aan een onveilige online omgeving voor groepen die toch al kwetsbaar zijn. Een respondent van dit onderzoek gaf het voorbeeld van bekenden van het slachtoffer, die zich online anders voordoen en daardoor aan gevoelige persoonlijke informatie konden komen (Lauckner et al., 2019). De onderzoekers raden hulpverleners aan extra alert te zijn op catfishing bij seksuele minderheden, vanwege hun kwetsbare positie en hun veelvuldige gebruik van online dating apps.

Catfishers gebruiken vaak foto’s van anderen die hiervoor geen toestemming hebben gegeven. Dat is ook schadelijk voor de eigenaar van die foto’s, omdat zijn of haar identiteit hiermee wordt gestolen en misbruikt. Daardoor kunnen bij catfishing twee soorten slachtoffers bestaan: degenen van wie de identiteit is gestolen door catfishers, en degenen die zijn bedrogen door catfishers.

In Nederland kan identiteitsfraude strafbaar zijn, afhankelijk van het type delict, onder bijvoorbeeld artikel 310 van het Wetboek van Strafrecht (diefstal van identiteit). Maar slachtoffers van catfishing zijn niet altijd slachtoffer van identiteitsfraude en ook niet altijd van oplichting (als daders geen financieel motief hebben). Dat maakt catfishing strafrechtelijk moeilijk te bestrijden.

Omvang

De politie heeft geen cijfers over cyberbedrog door middel van catfishing in Nederland, zo schrijft het *Algemeen Dagblad* in februari 2021 (Quekel, 2021). Wel

lijkt deze vorm van online bedrog toe te nemen, stelt de krant op basis van gesprekken met Stop Pesten Nu, Helpwanted.nl en het Centraal Meldpunt Identiteitsfraude (CMI).

Cryptofraude

Cryptofraude is een vorm van bedrog waarbij mensen overgehaald worden om cryptovaluta te kopen, waardoor de prijs van een bepaalde cryptomunt stijgt. Vervolgens gaan de daders snel hun valuta verkopen, waardoor slachtoffers financiële **schade** lijden achterblijven met verliezen. De stijgende populariteit van cryptovaluta als nieuwe vorm van beleggen zorgt ook voor meer nieuwe vormen van cryptofraude. Al in 2018 waarschuwde de AFM in het rapport *Crypto's: Aanbevelingen voor een regelgevend kader* voor de omvangrijke risico's op fraude die cryptovaluta met zich meebrengen (AFM, 2018). De AFM noemt onder andere als risico's het zeer technische karakter van cryptovaluta, koersmanipulaties, en hacks van online opslagdiensten van cryptovaluta.

Op de aandelenmarkt is het verboden om de koers te manipuleren met zogenaamde *pump-en-dump*-praktijken, waarbij grote groepen mensen de waarde van cryptovaluta kunstmatig omhoog brengen om daarna alles met winst te verkopen. De NOS wijdde hier op 6 mei 2021 een artikel aan, waarin het signaleert dat de populariteit van cryptofraude lijkt te stijgen (NOS, 2021b). Door de decentrale werking van cryptovaluta en de anonimiteit die het internet biedt, is het voor daders gemakkelijk om aan handhaving te ontkomen. Uitnodigingen om mee te doen aan *pump-en-dump*-praktijken worden verspreid op socialemediaplatformen als Twitter, Discord en Telegram. Daders richten zich vooral op kleinere en onbekende cryptovaluta, omdat deze gemakkelijker te manipuleren zijn dan bijvoorbeeld Bitcoin.

Wanneer cryptofraude zich uit in verduistering en oplichting, is het strafbaar onder artikel 321 en 326 van het Wetboek van Strafrecht. Tot op heden wordt cryptofraude weinig bestraft, onder andere door het gebrek aan regulering op cryptovaluta. Recent onderzoek van de Universiteit van Rome laat wel zien dat cryptobeurzen zelf ook meer verantwoordelijkheid zouden kunnen nemen om cryptofraude tegen te gaan, bijvoorbeeld door strengere controle en monitoring (La Morgia et al., 2021). De Europese Commissie werkt aan voorstellen om de cryptomarkt aan banden te leggen als onderdeel van het Digital Finance Package.

Omvang

In 2020 is de wereldwijde handel in cryptovaluta door de coronacrisis nog verder toegenomen. Het manipuleren van de aandelenmarkt is van alle tijden, maar staat de laatste tijd door cryptovaluta wel meer in de schijnwerpers. Zo ontstond op Reddit in januari 2021 een beweging van activisten die het opnamen tegen grote investeringsmaatschappijen en samen de koers van GameStop met 1.900% deden

stijgen (La Morgia et al., 2021). Alhoewel het hier niet om cryptovaluta ging, zijn de mechanismen erachter vergelijkbaar.

De Correspondent berichtte in 2020 op basis van een analyse van juridische en mediaberichten dat wereldwijd sinds 2011 voor 15 miljard euro aan cryptovaluta is gestolen (Tokmetzis, 2020). Het is grotendeels onduidelijk hoeveel Nederlanders slachtoffer zijn van cryptofraude. In specifieke fraudezaken komt soms aan het licht welke schade Nederlanders hebben geleden, zoals wanneer in 2019 volgens RTL Nieuws blijkt dat honderdduizenden euro's van Nederlanders zijn buitgemaakt door cryptofraude (RTL, 2019).

3.7 Online zelfbeschadiging

Onder online zelfbeschadiging valt gedrag waarbij iemand zichzelf schade toebrengt zonder dat daar een dader bij betrokken hoeft te zijn. Het gaat hier om **extreme challenges**, **cyberverslavingen** en **verstoord eetgedrag**. In al deze gevallen dragen online mechanismen aan het gedrag bij, en brengen ze schade toe aan degene die het gedrag vertoont. Zo zijn sociale media bijvoorbeeld een bron van de verslaving en kan gevaarlijke pro-anorexia-content online leiden tot verstoord eetgedrag. We spreken hier nadrukkelijk niet van drijfveren, omdat het gaat om schade die mensen zichzelf toebrengen en er geen klassieke slachtoffer-dader relatie bestaat bij deze fenomenen. Het is daarbij belangrijk om te vermelden dat zowel verstoord eetgedrag als cyberverslavingen psychische aandoeningen zijn waarvoor slachtoffers mentale hulp (zouden) moeten krijgen. Alle fenomenen in deze categorie zijn inherent verbonden met het internet. Daarom bespreken we niet per fenomeen wat het online bijzonder maakt.

Volgens experts speelt de uitzichtloze situatie van de lockdown waarbij jongeren veel meer dan gewoonlijk thuis zitten, een belangrijke rol bij de toename van het aantal jongeren met psychische klachten begin 2021 (zie Figuur 3.7). Ook het aantal uren dat jongeren online zijn, is toegenomen door de coronacrisis. Er is een correlatie aangetoond tussen enerzijds blootstelling aan online content met riskant gedrag en anderzijds offline riskant gedrag door internetgebruikers. Deze relatie werd gevonden voor drugsgebruik, excessief alcoholgebruik, eetstoornissen, zelfbeschadiging en geweld en extreme challenges (Branley & Covey, 2017).

Bron: Rathenau Instituut

Figuur 3.7 Online zelfbeschadiging

Extreme challenges

Bij online challenges worden mensen aangemoedigd om bepaalde opdrachten zelf uit te voeren en ze vervolgens online te delen. Een bekend voorbeeld is de Ice Bucket Challenge die in de zomer van 2014 viral ging en waarbij mensen uitgenodigd werden om een emmer koud water over zichzelf heen te gooien, een

donatie te doen aan de ALS-stichting en anderen aan te moedigen om ook mee te doen. Online challenges verschillen enorm in het risico voor deelnemers. De Blue Whale Challenge spoorde deelnemers aan tot automutilatie en uiteindelijk zelfs tot zelfmoord (Khasawneh et al., 2020). Media-aandacht voor soortgelijke challenges komt vaak op gang nadat er slachtoffers vallen. We scharen extreme challenges onder zelfbeschadiging, omdat ze mensen vaak aanzetten tot gevaarlijke opdrachten die tot zelfbeschadiging kunnen leiden.

Anders dan pranks worden challenges uitgevoerd en gedeeld door de mensen zelf die op de challenge ingaan, en speelt vernedering geen centrale rol. Vooral op TikTok zijn challenges een populaire manier van vermaak en bieden ze een manier om met leeftijdsgenoten een gedeelde ervaring te ondergaan. Ook marketingbureaus springen slim in op de trend van online challenges, door campagnes te lanceren rondom online challenges en zo aandacht te genereren voor hun producten.

Niet alle online challenges zijn dus **schadelijk**, maar de gevolgen van risicovolle challenges kunnen wel groot zijn. Op internet zijn het vooral opvoedkundige websites met informatie voor ouders en traditionele media die waarschuwen voor de gevaren van online challenges. Artikelen met titels als “21 Dangerous TikTok Trends Every Parent Should Be Aware Of” waarschuwen ouders om alert te zijn bij hun kinderen op bepaalde challenges die gevaarlijk kunnen zijn (Morris, 2021). Voorbeelden hiervan zijn de Cinnamon Challenge, waarbij mensen worden uitgedaagd grote hoeveelheden kaneel door te slikken of de Choking Challenge, waarbij jongeren worden uitgedaagd elkaar te wurgen tot ze bewusteloos raken.

Extreme challenges kunnen strafbaar zijn, afhankelijk van de challenge. In 2018 was de “Keke challenge” bijvoorbeeld een tijdje populair, waarbij mensen werden uitgedaagd om uit een rijdende auto te stappen, een dansje te doen en weer in te stappen. Een woordvoerder van het Openbaar Ministerie bestempelde de challenge als strafbaar onder artikel 5 van de wegenverkeerswet doordat mensen zichzelf of anderen in het verkeer in gevaar brengen (NOS, 2018). Het verbieden van challenges en het aanpakken van de “aanstichters” is moeilijk, doordat het vaak lastig vast te stellen is wie met een bepaalde challenge of hashtag begon, en alle deelnemers onderdeel zijn van het aanmoedigen tot deelname.

Omvang

We hebben geen Nederlandse cijfers gevonden over de totale aantallen slachtoffers van zelfbeschadiging bij het uitvoeren van extreme online challenges. In Nederland zijn wel een aantal incidenten bekend van kinderen die zijn overleden aan een online challenge. Het gaat om Clay Haimé en Tim Reynders, beiden overleden aan een zogenoemde *chocking game* (wurgspel) in 2017 (RTL, 2018; Stichting Internet Challenges, 2021). De Choking challenge kreeg wereldwijde aandacht toen in Italië een tienjarig meisje overleed aan deze challenge, waarna de

Italiaanse overheid tijdelijk de toegang tot TikTok blokkeerde voor accounts van ongeverifieerde jonge gebruikers (France-Presse, 2021).De totale omvang van schadelijke online challenges is moeilijk in kaart te brengen, omdat schade vaak niet gedocumenteerd is en alleen media-aandacht oplevert bij zwaar letsel, terwijl schade zich ook kleiner kan manifesteren.

Cyberverslavingen

Cyberverslaving is excessieve en ongecontroleerde online activiteit met langdurig gebruik van internet in het algemeen en sociale media, gaming en pornosites in het bijzonder (Liu et al., 2020; Müller et al., 2015). Ander compulsief en dwangmatig gedrag dat hieruit kan voortvloeien is bijvoorbeeld excessief shoppen, gamen of obsessief zoeken naar gezondheidsinformatie (cyberchondrie) (Aiken, 2016).

Socialemediaplatformen hebben vaak verslavende kenmerken in zich, zoals ‘endless scrolling’ content op TikTok die nooit stopt en waarbij gebruikers urenlang naar beneden kunnen scrollen zonder dat ooit een einde wordt bereikt (Montag et al., 2021). Ook mechanismen zoals likes en gepersonaliseerde content kunnen ervoor zorgen dat gebruikers langer op sociale media blijven dan ze eigenlijk zelf zouden willen. Bovendien zorgen ze op hun beurt voor steeds extremere content. Daarmee werken platformen verslavingsgedrag in de hand.

Cyberverslavingen zijn **schadelijk** omdat ze in wetenschappelijk onderzoek gelinkt worden aan onder andere chronisch slaaptekort, angst en emotionele problemen (Alimoradi et al., 2019; Cerniglia et al., 2017). Uiteraard is cyberverslaving, zoals alle mentale problemen, niet strafbaar. Wel zijn er wereldwijd initiatieven om verslavende kenmerken van sociale media met wetgeving aan banden te leggen. Een (nog niet goedgekeurd) wetvoorstel uit 2019 in de Amerikaanse staat Missouri verbiedt bedrijven bijvoorbeeld om de menselijke psychologie te misbruiken en daarbij de keuzevrijheid van mensen in te perken.

Omvang

Internetverslaving en *internet gaming disorders* komen steeds meer voor (Chia & Zhang, 2020). Naar schatting ligt het percentage van internetverslaving in Europa en de VS tussen 1,5% en 8,2 % (Weinstein & Lejoyeux, 2010).

De omvang van *internet gaming disorder* schommelt tussen 1,6% en 5,1%, gebaseerd op een internationale studie met 12.938 adolescente proefpersonen (Müller et al., 2015). Circa 3% van de online gamers kan worden gezien als ‘gameverslaafd’: ze kunnen er moeilijk mee stoppen, spelen meer dan ze van plan waren, slapen door het gamen te weinig en hun huiswerk lijdt eronder (van Rooij et al., 2012). Van het aantal hulpzoekers voor gameverslaving is 82% jonger dan 25 jaar. Het aantal verslaafden of misbruikers van internetgamen in Nederland wordt geschat op 16.000. 537 daarvan zijn in behandeling (Jellinek, 2021).

Meer tijd achter de schermen verhoogt het risico op obesitas onder jongeren en hangt eveneens samen met internetverslaving. Verder is er ook Pediatric venous thromboembolism (VTE), oftewel gamertrombose, een aandoening die dodelijk kan zijn bij excessief gamen. Die is de afgelopen twee decennia toegenomen onder adolescenten (Kohorst et al., 2018). Voor Nederland zijn geen cijfers gevonden over VTE.

Verstoord eetgedrag

Eetstoornissen zijn psychische stoornissen die worden gekenmerkt door verstoord eetgedrag en/of inadequaat compensatiegedrag (braken, laxeren). Mensen met een eetstoornis hebben een verstoord lichaamsbeeld, zijn erg bezig met hun gewicht of lichaamsvorm en zijn erg bang om aan te komen (NJi, 2019).

Online kunnen mensen met een eetstoornis elkaar ontmoeten in zogenaamde pro-ana-communities. ‘Pro-ana’ en ‘pro-mia’ zijn respectievelijk de afkorting van *professional anorexia* en *professional boulimia*. Dit zijn de namen van online groeperingen die (vooral) bestaan uit jongeren met problemen die zich actief manifesteren op internetfora, chats en websites waar informatie wordt verstrekt en interactie plaatsvindt die vooral gericht is op de promotie, ondersteuning en het volhouden van eetstoornisgebonden gedrag (van Furth et al., 2011).

In 2020 rapporteerde het online magazine *Wired* dat TikTok vol zou staan met pro-ana-content, waar jonge meisjes door aanbevelingsalgoritmes al snel in kunnen belanden (Gerrard, 2020). *De Volkskrant* schrijft ook over challenges als “Kun je het kabeltje van je EarPods twee keer om je taille wikkelen en er dan nog een knoop in leggen?” en eetstoornismemes die op het eerste gezicht onschuldig lijken (Bouyeure, 2020). Op deze manier doen TikTok-gebruikers pro-ana-content eigenzinnig en daarom aantrekkelijk lijken.

Verstoord eetgedrag kan online nog meer **schade** aanrichten als online content en communities het beeld versterken dat mensen over hun lichaam en eetgedrag hebben. Het baart vooral zorgen dat pro-ana-content anno 2021 wordt verpakt als grappig en herkenbaar, en ook niet altijd bewust kwaadwillend is, waardoor het moeilijker te classificeren is als schadelijke content. Mensen die pro-ana-content promoten zijn niet altijd strafbaar bezig. In 2019 gaf minister Hugo de Jonge aan niks te zien in een verbod op pro-ana content op internet, omdat dat volgens hem juist de zoektocht naar hulp zou kunnen belemmeren (Ministerie van Justitie en Veiligheid, 2019). Wel kunnen gedragingen gelinkt aan pro-ana content strafbaar zijn: het opzettelijk benadelen van iemands gezondheid kan strafbaar zijn als mishandeling onder artikel 300 van het Wetboek van Strafrecht. Als het daarbij om minderjarigen gaat, kan zelfs sprake zijn van kindermishandeling, zo stelt het Centrum tegen Kinderhandel en Mensenhandel samen met eetstoorniskliniek Ursula in een recent onderzoek naar pro-ana coaches (Simons et al., 2020).

Omvang
Thinspiration (content die inspiratie biedt voor mensen die willen afvallen, zoals dieet- of sportinstructies), TikTok-challenges en eetstoornismemes worden steeds populairder en kinderen worden steeds slimmer om algoritmen te omzeilen met varianten op hashtags (Bouyeure, 2020). Ook de opkomst en het gebruik van allerlei gezondheidsapps, stappentellers en andere trackers op mobiele telefoons en horloges werken eetstoornissen in de hand.

Eind 2020 behandelden kinderartsen in Amsterdam een derde meer patiënten met eetstoornissen. Zij vermoeden dat er een verband is met de coronacrisis (Kootstra, 2020). Landelijke cijfers zijn er niet (Kootstra, 2020). Sinds de tweede lockdown wacht een recordaantal kinderen, sommigen nog geen 10 jaar oud, op een behandelplek in een eetstoorniskliniek, constateerde ook *NRC* (Van der Poel & Luyendijk, 2021). De wachttijd varieert van zes weken tot soms een half jaar in het midden van het land. Het aantal inbewaringstellingen van minderjarige anorexiapatiënten neemt ook toe. In de eerste negen maanden van 2020 waren er 24 jongeren met een inbewaringstelling (Van der Poel & Luyendijk, 2021).

Algemeen onderzoek (niet specifiek over de online omgeving) laat zien dat ongeveer 0,3% van de 13 tot 18-jarigen lijdt aan anorexia nervosa (Verhulst et al, 1997). Eenzelfde percentage lijdt volgens dit onderzoek aan boulimia nervosa (NJi, 2019). Anorexia komt veel vaker voor bij vrouwen dan bij mannen (95% is vrouw), ontstaat meestal in de puberteit en bij jongvolwassenen. De website van hulpverleners voor eetstoornissen, Proud2Bme, krijgt circa 11.000 bezoekers per dag (Simons et al., 2020). Het aantal online berichten waarin specifiek wordt gevraagd naar een pro-ana-coach steeg met 400% na aandacht hiervoor van de traditionele media (Simons et al., 2020).

3.8 Conclusie

De taxonomie van schadelijk en immoreel gedrag onderscheidt zes categorieën met daaronder 22 fenomenen. De beschrijving hiervan levert een breed scala aan nieuwe en oudere vormen van gedrag op, waarmee alle gebruikers van het internet te maken kunnen krijgen – variërend van desinformatie, tot online discriminatie, tot shaming.

Het Rathenau Instituut doet hiermee een eerste poging om een overzicht te creëren van de aard en omvang van schadelijk en immoreel gedrag online in Nederland. Uit de veelzijdigheid en beschikbare gegevens over omvang wordt duidelijk dat alle Nederlanders het risico lopen om als slachtoffer, dader of omstander betrokken te raken bij dit gedrag. Iedereen kan te maken krijgen met één of meerdere vormen van de fenomenen die in dit hoofdstuk uiteen zijn gezet. Toch lopen bij bepaalde

fenomenen sommige groepen meer risico dan anderen, afhankelijk van hun leeftijd, geslacht, etniciteit, seksuele voorkeur, geloofsovertuiging of opleidingsniveau. Het is op basis van de beschikbare gegevens echter lastig om hierover algemene uitspraken te doen.

Door de diversiteit van de behandelde fenomenen, het gebrek aan nauwkeurige definities en systematische metingen hiervan, is het in de context van dit onderzoek onmogelijk om vast te stellen welk fenomeen nu het snelst groeit of het meest zorgelijk is. Het antwoord op die vraag hangt ook af van de gekozen criteria. Gaat het bijvoorbeeld om het aantal slachtoffers, de ernst, omvang of het risico van de schade in de toekomst? Alle fenomenen zijn op hun eigen wijze zorgelijk – voor de maatschappij, voor individuen of groepen individuen. We onthouden ons hier dus van prioritering van de fenomenen uit de taxonomie.

Casus: desinformatie

De hieronder beschreven casus is fictief en bedoeld om een beeld te schetsen van mogelijke risico's die kunnen ontstaan door het fenomeen desinformatie. De casus is wel deels gebaseerd op een combinatie van voorvallen in Nederland en het buitenland. In de reflectieparagraaf wordt ingegaan op de betrokken mechanismen en actoren in deze casus. De mechanismen worden uitgebreid besproken in hoofdstuk 4. De actoren komen in hoofdstuk 5 aan bod en hoofdstuk 6 bevat suggesties om situaties als deze te voorkomen en adresseren.

Casus
Mirjam is een tiener en is geïnspireerd geraakt door een wereldwijde beweging van jonge mensen die besloten hebben zich volop in te zetten voor het stoppen van klimaatverandering. Ze is bijzonder digitaal vaardig. Al snel raakt ze in contact met gelijkgestemden van over de hele wereld die haar denkbeelden bevestigen. Eerst via publieke Facebook groepen en Twitter, maar al snel vindt ze ook haar weg in besloten chatgroepen, op platformen als Signal en Telegram. Via deze kanalen worden campagnes op sociale media voorbereid, om bijvoorbeeld hashtags tot trending topic te promoten of het op te nemen voor medestanders op sociale media op het moment dat die worden bekritiseerd. Ook worden via deze kanalen fysieke demonstraties gepland. De grenzen van het toelaatbare worden daarbij soms opgezocht of overtreden, waarbij inspiratie wordt geput uit voorbeelden uit het buitenland. Denk hierbij aan het onaangekondigd bezetten van verkeersaders, intimidatie of het saboteren van bedrijven. Ondanks dat Mirjam de beheerders van de groepen nog nooit in levenden lijve heeft ontmoet en sommigen onder een schuilnaam opereren, vertrouwt ze hen volledig. Ze staan immers allemaal wereldwijd voor hetzelfde doel.

Op een gegeven moment wordt in een internationale Telegramgroep het bericht verspreid dat de Nederlandse overheid rapportages over de stikstofuitstoot door de veehouderij moedwillig naar beneden bijstelt. Volgens de berekeningen van buitenlandse onderzoekers die in de chats worden gedeeld, is het allemaal veel erger. Er wordt ook gesuggereerd dat hier een samenzwering achter kan zitten. Nederlandse mainstream media krijgen ook lucht van deze berichten, maar trekken de alternatieve berekeningen in twijfel. Haar internationale contacten moedigen Mirjam echter aan om daar niet in te trappen. De media spelen vast onder één hoedje met de overheid. Mirjam vindt dat het tijd wordt om deze vermeende misstand aan de kaak te stellen en om over te gaan tot actie. Ze organiseert via de chat een onaangekondigde betoging in Den Haag. De gemeente is hierop voorbereid, want die is getipt door de binnenlandse inlichtingendienst die ook

meekijkt in de chatkanalen. Desondanks loopt de betoging uit op een gewelddadige confrontatie met veehouders die er juist van overtuigd zijn dat de overheid de rapportages overdrijft. Zij waren blijkbaar ook opgetrommeld via besloten kanalen waarin alleen zij actief zijn, maar die niet vooraf zijn opgemerkt door veiligheidsdiensten. Wat Mirjam niet weet, is dat de internationale coördinatoren van de chatkanalen geen klimaatactivisten zijn, maar in werkelijkheid werken bij een buitenlandse inlichtingendienst. Voor de kanalen waarin de veehouders berichten uitwisselen zou zomaar hetzelfde kunnen gelden.

Enige tijd na de gewelddadige confrontatie wordt het bedrog door de buitenlandse inlichtingendienst ontmaskerd. De onrust wordt daarmee niet weggenomen. De onthulling leidt ertoe dat mensen nog sceptischer zijn geworden over berichtgeving. Ze weten nu helemaal niet meer wie ze wel of niet kunnen vertrouwen. Het vertrouwen in elkaar is fundamenteel geschaad.

Reflectie

In deze casus zien we hoe desinformatie leidt tot maatschappelijke schade. Het klimaatbeleid is al bron van maatschappelijke onrust en het publieke debat toont vaak kenmerken van polarisatie. Dat maakt dat dit onderwerp zich sterk leent voor desinformatiecampagnes. Het is bekend dat met name statelijke actoren erop uit kunnen zijn om bij hun geopolitieke tegenstanders onrust op te stoken en polarisatie te versterken. Zo is uit onderzoek naar de activiteiten van de zogenaamde “Russische trollenfabriek” Internet Research Agency (IRA) in St. Petersburg gebleken, dat ‘trollen’ gedurende de Amerikaanse presidentsverkiezingen in 2016 demonstraties over bijvoorbeeld racisme organiseerden in de VS. Er is daarbij ook aangetoond dat zij zowel een protest van aanhangers als critici van de islam op dezelfde plek op hetzelfde moment hebben gecoördineerd (Bertrand, 2017). De motivatie achter dit soort inmengingen is waarschijnlijk het bereiken van strategische voordelen op het wereldtoneel. De aandacht van de getroffen democratische samenlevingen wordt afgeleid naar interne conflicten, waardoor zij geen oog hebben voor wat er geopolitiek speelt, of daar niet op reageren.

In de casus zien we verschillende mechanismen en fenomenen. De klimaatactivisten in deze casus dragen zonder hun medeweten bij aan een **desinformatie**campagnes van een statelijke actor. Het gebruik van het internet is voor hen een **dagelijkse gewoonte**, maar ondanks hun digitale vaardigheden zijn ze toch vatbaar voor misleiding. Ze hebben het gevoel het morele gelijk aan hun kant te hebben en vertonen de karakteristieken van **digitaal vigilantisme**. Dankzij **syndicatie** en **hyper-connectiviteit** vinden ze eenvoudig gelijkgestemden. Ze snappen bovendien hoe algoritmen werken en weten dus ook hoe **viraliteit** kan worden bereikt en hoe via bespeling van algoritmen een publiek discours kan

worden beïnvloed. Het opereren in **anonimiteit** is onder de jonge activisten volstrekt normaal, omdat ze beseffen dat een reputatie als activist mogelijk schadelijk kan zijn voor hun latere carrière. Wat verder bijdraagt aan het gevoel van **schijnbare wetteloosheid** onder de jonge activisten is dat acties die onbestraft blijven in andere landen, als voorbeeld worden gesteld.

We zien in deze casus ook diverse actoren. De klimaatactivisten worden niet enkel beïnvloed door de buitenlandse inlichtingendienst. De mainstream media proberen helderheid te scheppen, maar hebben slechts beperkt effect op degenen die hun informatie putten uit besloten kanalen. In zekere zin kan de verslaggeving juist worden opgevat als een bevestiging van een mogelijke samenzwering van de overheid. Verder spelen de platformen die besloten kanalen aanbieden een faciliterende rol. Chat apps als Signal staan bekend om de privacybescherming van gebruikers, bijvoorbeeld door anonieme accounts en sterke versleuteling aan te bieden. Andere platformen modereren wel, maar voor hen is het een uitdaging om de kanalen van oprechte activisten te onderscheiden van kwaadaardige gebruikers.

We zien verder dat het bestaan van desinformatie aanleiding kan geven tot meer wantrouwen in de samenleving. De binnenlandse inlichtingendiensten infiltreren in de chatkanalen. Deze vorm van surveillance kan op zichzelf bepaalde mensen al verontrusten. Tenslotte is een nog onbenoemde, maar niet minder belangrijke actor, het algemene publiek dat van de gewelddadige confrontatie op de hoogte wordt gesteld. Naar alle waarschijnlijkheid daalt voor hen het algemene gevoel van veiligheid.

4 Mechanismen van immoreel en schadelijk gedrag

Het internet als omgeving bevat mechanismen die van invloed zijn op menselijk gedrag. Mensen kunnen door die online mechanismen bijvoorbeeld anders met normen en regels omgaan. Uiteraard zijn er naast de mechanismen van het internet nog veel meer factoren van invloed op menselijk gedrag, zoals sociale, psychologische, culturele of economische oorzaken. Al deze factoren spelen een rol bij de totstandkoming van schadelijk en immoreel gedrag online. Dit onderzoek richt zich op de mechanismen die kenmerkend zijn voor het internet.

4.1 Opmerkingen vooraf

Voordat we de mechanismen induiken, zijn enkele opmerkingen op hun plaats. In hoofdstuk 3 over de taxonomie is een zeer breed pallet aan schadelijke en immorele fenomenen beschreven. Zoals ook blijkt uit de casussen, spelen bij de meeste fenomenen meerdere online mechanismen een rol, en sommige mechanismen worden weer beïnvloed door andere factoren. De mechanismen zijn dus niet alles bepalend.

Voor diverse fenomenen geldt bovendien dat de oorzaken en onderliggende mechanismen niet volledig bekend zijn. Zelfs van haast alledaagse fenomenen zoals pesten kan niet met zekerheid worden gezegd hoe ze precies ontstaan. Het gaat de doelstelling van deze studie dan ook te boven om fenomenen volledig te doorgronden. We pretenderen hier ook niet een uitputtende lijst met mechanismen te beschrijven, maar richten ons op het verband tussen mechanismen die kenmerkend zijn voor internet en het schadelijke en immorele gedrag.

Binnen het internet zijn er verschillende soorten omgevingen waarin een gebruiker zich kan begeven: er zijn verschillende soorten socialemediaplatformen, websites en fora. Dit betekent dat de beschikbaarheid en invloed van mechanismen per online omgeving sterk kunnen verschillen. Een platform als Instagram bevat door haar ontwerp bijvoorbeeld heel andere mechanismen dan een shock site. Het is praktisch onhaalbaar om hier de volledige diversiteit aan omgevingen apart te beschrijven. Het doel van dit hoofdstuk is namelijk om een breder begrip van immoreel en schadelijk online gedrag mogelijk te maken, door algemene mechanismen te beschrijven die echt kenmerkend zijn.

Het is tevens belangrijk om te beseffen dat veel van de hier beschreven mechanismen ook positieve effecten hebben. Ze kunnen dus niet eenzijdig als schadelijk worden weggezet. Dat het internet voor iedereen toegankelijk is en informatie snel en ver kan worden verspreid, is bijvoorbeeld ook juist een groot voordeel van het internet. Dit betekent dat het simpelweg bestrijden van een mechanisme kan leiden tot nieuwe, mogelijk grote negatieve consequenties. Online anonimiteit, bijvoorbeeld, kan als mechanisme aan negatief gedrag bijdragen, maar het biedt tegelijkertijd ook bescherming tegen schade.

Tenslotte zijn de mechanismen geen natuurverschijnselen, maar vaak het gevolg van ontwerpkeuzes die gemaakt zijn door partijen die hun belangen of een bepaald doel nastreven. Zowel gebruikers als aanbieders van internetdiensten en producten willen bijvoorbeeld vaak dat informatie zo ongehinderd en efficiënt mogelijk kan worden uitgewisseld. De mogelijk nadelige effecten hiervan worden door hen vaak gezien als een ongelukkige externaliteit.

Mechanismen

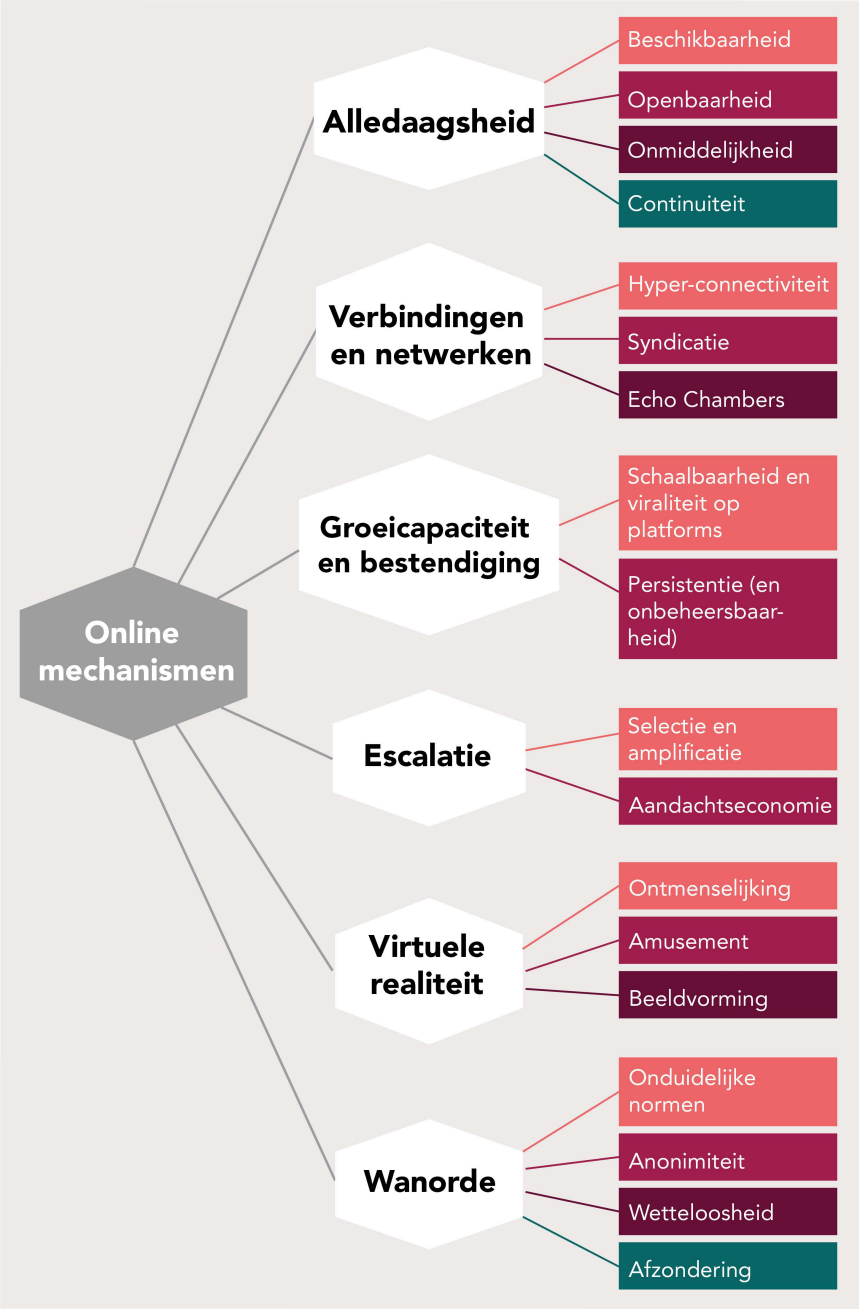
De hieronder beschreven mechanismen verklaren deels hoe het gebruik van het internet kan bijdragen aan immoreel of schadelijk gedrag. De mechanismen staan ook wel bekend als factoren die bijdragen aan concepten als ‘morele mist’ (Cocking & van den Hoven, 2018), ‘digitale drift’ (Goldsmith & Brewer, 2015) of het ‘online ontremmingseffect’ (Suler, 2004). Deze concepten worden in de literatuur vaak aangehaald en beschrijven hoe mensen online minder dan offline de gevolgen kunnen overzien van hun daden, lastiger een ethisch oordeel kunnen vellen, of minder geneigd zijn om ethisch te handelen.

De mechanismen in dit hoofdstuk bestaan deels uit deze concepten, aangevuld met mechanismen die bij dit onderzoek genoemd zijn door experts in interviews en die naar voren kwamen uit de literatuurstudie.

Om de leesbaarheid te vergroten zijn aan elkaar verwante mechanismen samengebracht onder zes beschrijvende kenmerken van het internet:

1. Alledaagsheid
2. Verbindingen en netwerken
3. Groeicapaciteit en bestendiging
4. Escalatie
5. Virtuele realiteit
6. Wanorde

Een overzicht van alle mechanismen en hun indeling is te vinden in Figuur 4.



Bron: Rathenau Instituut

Figuur 4 Overzicht van online mechanismen

4.2 Alledaagsheid

Het gebruik van het internet is een alledaags verschijnsel. Veel Nederlanders besteden het grootste deel van hun dag achter een beeldscherm (SCP, 2016). Tijdens de coronalockdown steeg beeldschermtijd van kinderen zelfs van ongeveer 3 uur naar ruim 7 uur per dag (Van Baars, 2020). Internetgebruik is voor veel mensen dus een routine. Hierin schuilt al het risico op onethisch gedrag. Het is namelijk bekend dat wanneer mensen handelen vanuit gewoonte, zij minder aandacht besteden aan ethische kwesties (Vince, 2018). Een foto is snel geplaatst en een bericht stuur je zomaar door, zonder te reflecteren op de mogelijke consequenties.

Door het alledaagse van internetgebruik zijn de consequenties ervan haast onontkoombaar geworden. Zelfs wanneer iemand geen (frequent) internetgebruiker is, is de kans groot dat informatie over die persoon online geplaatst wordt en vindbaar is. Daarmee is het dus denkbaar dat deze persoon alsnog schade kan ondervinden.

De volgende kenmerken dragen bij aan de alledaagsheid van het internet.

Beschikbaarheid
Het internet is voor vrijwel alle mensen in Nederland beschikbaar. Volgens het CBS gebruikt 95% van de 12 tot 55-jarigen het internet dagelijks, meestal via de smartphone (CBS, 2020b). Een logisch gevolg van deze beschikbaarheid van internettechnologie is dat mensen het internet bij van alles betrekken, dus ook bij immoreel of schadelijk gedrag zoals discriminatie, racisme of andere vormen van haat (Guan & Subrahmanyam, 2009). De toegankelijkheid van het internet maakt het bovendien erg eenvoudig om normen te overschrijden (Haspels-Goudriaan, 2020).

Voor veel mensen en organisaties is de beschikbaarheid van het internet van cruciaal belang geworden. Zo slaan we belangrijke documenten op in de cloud of zijn bedrijfsprocessen afhankelijk geworden van internetverbindingen. Het gebruik van het internet is daarmee niet altijd meer een vrijwillige keuze, maar iets waar we niet meer omheen kunnen.

Openbaarheid
Het internet maakt het mogelijk om veel van wat vroeger privé was openbaar te maken. 97% van de huishoudens is aangesloten op het web (Digitale Overheid, 2020) en maar liefst 84% van de Nederlanders is online met een smartphone (CBS, 2019b). Het delen van informatie via sociale media zoals WhatsApp, Facebook, YouTube en Instagram is voor de meerderheid gemeengoed (Oosterveer, 2021).

Ondanks dat het binnen deze platformen mogelijk is om informatie te delen met slechts een besloten groep, is het ook erg eenvoudig om die informatie vervolgens verder te openbaren. Ook kan informatie onbedoeld of onbewust openbaar worden gemaakt. Een foto die bedoeld was voor vrienden kan daardoor zomaar door kwaadwillenden worden misbruikt in een andere context. Een ruzie of geschil komt ook sneller in de openbaarheid. Deze openbaarheid leidt ook tot meer zichtbaarheid.

Mensen reageren ongeremd op bijvoorbeeld openbare sociale media, omdat ze zich in hun privésfeer wanen. Techniekfilosoof Jan Bats laat in zijn proefschrift door middel van drie experimenten zien dat mensen zich 'thuisvoelen' op sociale media door sterke personalisatie van die online omgevingen. Mensen wanen zich dan in een privésfeer, terwijl de berichten die ze delen vaak wel openbaar beschikbaar zijn. Dat zorgt ervoor dat mensen ongeremder reageren en anderen harder veroordelen (Bats, 2019). Het laatste wordt ook wel het online ontremmingseffect (*disinhibition effect*) genoemd (Aiken, 2016; Suler, 2004).

Het openbare karakter van het internet is een gevolg van ontwerpkeuzes die sinds de begintijd van het wereldwijde web ongewijzigd zijn gebleven. Het internet is in feite een netwerk van computers, waarop elke computer onder normale omstandigheden zichtbaar is voor alle anderen. Het voordeel is dat informatie snel met alle deelnemers kan worden gedeeld. Het nadeel is dat bijvoorbeeld een slecht beveiligde computer in Nederland door elke willekeurige internetgebruiker waar ook ter wereld kan worden gehackt. Binnen internetdiensten als sociale media is het vaak wel mogelijk om de openbaarheid van deelname te beperken. Zo kun je je LinkedInprofiel onvindbaar maken voor onbekenden. Er zijn overigens ook alternatieve ontwerpen mogelijk, waarbij deelnemers vanaf de basis meer controle hebben over hun zichtbaarheid (zie hoofdstuk 5 en 6).

Onmiddellijkheid

Interacties via het internet hebben doorgaans een direct effect. Wanneer een bericht op een openbaar platform zoals Twitter geplaatst wordt, is het onmiddellijk zichtbaar. Ook reacties kunnen direct worden geplaatst. Op het moment dat mensen snel handelen is dit vaak gebaseerd op instinctieve of emotionele gedachten en wordt vooringenomenheid sneller zichtbaar. Dit hoge tempo kan volgens Kahneman (2011) weloverwogen reacties met oog voor ethische kwesties in de weg staan.

Door experts wordt ook aangegeven dat de onmiddellijkheid van het web impulsiviteit in de hand werkt; een karaktereigenschap die sterk speelt met name bij jongeren met ADHD bijvoorbeeld (Aiken, 2016, p. 72; Kaakinen et al., 2020). Ook wordt gesuggereerd dat er sprake is van een soort ‘tempocratie’; wat wil zeggen dat

degene die het snelst of vaakst informatie creëert, de teneur van het gesprek bepaalt. Dit haastige gedrag kan verklaren dat mensen normen en regels uit het oog verliezen.

Bepaalde platformen, zoals 4chan of Snapchat, waar content maar kortstondig zichtbaar is en daarna helemaal verdwijnt (in tegenstelling tot veel andere platformen en websites), versterken dit negatieve effect van onmiddellijkheid. Het niet-permanente karakter van berichten lijkt gebruikers te stimuleren om de inhoud hiervan zo opruiend mogelijk te maken, juist doordat zij weten dat deze berichten binnen enkele minuten of uren zijn verdwenen (Ludemann, 2018, p. 93).

Alle internetgebruikers hebben uiteraard ook veel baat bij de onmiddellijkheid van het internet. Het grote voordeel van chat, bijvoorbeeld, is dat het een stuk sneller gaat dan brieven posten. Aanbieders van internetdiensten en producten zijn er dan ook op gebrand om snelheid voorop te stellen. Zo wordt 5G sneller dan 4G, games moeten zo snel mogelijk laden, en dankzij snelle verbindingen kunnen we direct reageren op live video streams.

Continuïteit
Het internet gaat 24 uur per dag en zeven dagen per week door. Het feit dat het internet geen pauze kent, betekent ook dat het risico op schade voortdurend aanwezig is. Dit kan ook verklaren waarom mensen voortdurend het internet opgaan, om te controleren of hen online iets is overkomen. Meer dan de helft van de smartphone gebruikers zet dit apparaat vaker dan 25 keer per dag aan, en een kwart van de gebruikers meer dan vijftig keer (Stil, 2020). Voor sommigen voelt de smartphone haast als een onmisbare ledemaat en zij vermoeden dat ze lijden aan een verslaving (zie ook paragraaf 3.7 over cyberverslaving) (RTL nieuws, 2020). Een vervelende consequentie van de continuïteit van het web is dat zodra iemand schade lijdt, dit ook voortdurend door kan gaan. Er is geen moment om tot rust te komen.

4.3 Verbindingen en netwerken

Het internet is technisch gezien in de eerste plaats een netwerk dat apparaten en daarmee gebruikers met elkaar in contact brengt. De volgende aspecten van deze netwerken kunnen schade en immoreel gedrag in de hand werken.

Hyperconnectiviteit
Volgens de populaire ‘six degrees of separation’-theorie zijn twee willekeurige personen op de wereld gemiddeld via zes tussenliggende personen met elkaar verbonden. Facebook stelde in 2016 vast dat voor verbinding tussen twee van de

1,6 miljard gebruikers op hun platform gemiddeld slechts 3,5 tussenpersonen nodig zijn (Edunov et al., 2016). Online lijken mensen dus nauwer met elkaar in contact te staan. Vaak kunnen mensen elkaar ook rechtstreeks bereiken, door simpelweg een persoon te zoeken in een zoekmachine. Dit betekent dat gebruikers van sociale netwerken ook heel eenvoudig het doelwit kunnen worden van kwaadwillenden of onbedoeld betrokken kunnen worden bij immorele of schadelijke activiteiten. Bij fenomenen als grooming of sextortion speelt dat kinderen of andere slachtoffers eenvoudig kunnen worden benaderd.

Aanbieders van internetdiensten hebben er sterk baat bij om de connectiviteit zo groot mogelijk te maken. Meestal geldt: hoe mee gebruikers, des te hoger de inkomsten.

Syndicatie

Door de opkomst van sociale media en andere online platformen is het gemakkelijker geworden om gelijkgestemden te vinden en groepen te vormen. In de literatuur wordt deze werking van het internet aangeduid als homofilie en het ‘online syndication effect’ ofwel ‘syndicatie’ (Aiken, 2016, p. 332), en ook wel als de coördinerende werking van het web. Dit mechanisme wordt versterkt door de aanzuigende werking van content waar al veel aandacht voor is. Dit lijkt op het zogenaamde Mattheüseffect (Merton, 1968) dat beschrijft hoe een succesvol of beroemd individu (bijvoorbeeld een wetenschapper) vaak meer geloofwaardigheid geniet dan een relatief onbekende, zelfs wanneer de kwaliteit van het werk van deze twee mensen vergelijkbaar is. Veel platformen tonen indicatoren als het aantal likes, volgers of views, die syndicatie in de hand werken. Dit kan negatieve gevolgen hebben. Zo blijkt dat shamers op Twitter veel sneller volgers aantrekken, dan *non-shamers* (Basak et al., 2019). In een grote groep vertroebelt tevens het gevoel voor verantwoordelijkheid van individuen (De Vries, 2021).

Syndicatie kan ook allerlei andere gevolgen hebben. Allereerst kan een omgeving van gelijkgestemden leiden tot normalisatie (‘iedereen doet het’) van bepaald gedrag (LaFrance, 2020). Door groepen te vormen kan ook eerder een kritische massa worden bereikt die nodig is om over te gaan tot schadelijke gedrag (Munn, 2021). Daarnaast kan ook sprake zijn van het *bystander effect*: namelijk dat niemand ingrijpt op het moment dat een norm, wet of regel wordt overtreden. In de online omgeving is het lastiger dan offline om te bepalen wanneer je als internetgebruiker nou wel of geen *bystander* (omstander) bent.

Bij digitaal vigilantisme is de wens om bij een groep te horen één van de onderliggende redenen om hieraan deel te nemen. Vigilantisten zouden ook gedreven kunnen worden door de status die kan worden ontleend aan de deelname (Afuah, 2013). Ten slotte betekent syndicatie ook dat kwetsbare groepen

eenvoudiger te vinden zijn voor kwaadwillende actoren. Wanneer mensen bijvoorbeeld zelf informatie hebben gepubliceerd waaruit blijkt dat zij worstelen met hun zelfbeeld, kunnen zij op die manier gevonden worden door pro-ana coaches .

Echo chambers
Verder kan door syndicatie ook het psychologische effect van de *confirmation bias* worden versterkt. Dit betekent dat mensen die enkel in aanraking komen met informatie die zij voor waar aannemen, telkens worden bevestigd in hun eigen wereldbeeld (Sternisko et al., 2020). In groepen met gelijkgestemden kan dat soort informatie meer worden gedeeld, ook wanneer het desinformatie betreft (Marwick, 2018). Daarnaast komen mensen hierdoor online terecht in zogenaamde *rabbit holes* (O'Callaghan et al., 2015) of *echo chambers* (Auxier & Vitak, 2019). Doordat zij van algoritmen steeds meer aangeboden krijgen van hetzelfde soort content dat hen het langste boeit, krijgen zij geen content met tegengeluiden meer voorgeschoteld (Sternisko et al., 2020). Dit kan leiden tot vernauwing en verstarring van het denken en kan radicalisering in de hand werken (Faddoul et al., 2020; NRC, 2021). Over het algemeen zijn met name aanbieders van internetdiensten met een advertentieverdienmodel geneigd om het syndicatie-effect te gebruiken, omdat zij aan adverteerders een internetomgeving willen bieden waar gebruikers zo lang mogelijk en zo vaak mogelijk aanwezig (willen) zijn (zie ook 5.2.2) (Kist, 2020).

4.4 Groeicapaciteit en bestendigheid

Eén van de kenmerkende eigenschappen van digitale informatie is de eenvoud waarmee die kan worden vermenigvuldigd. Terwijl het reproduceren van immoreel of schadelijk gedrag in de offline wereld nog beperkt wordt door de moeite die een actor daarvoor moet doen, kunnen digitale acties machinaal – haast moeiteloos – worden herhaald. Om iemand te pesten zal je in de offline wereld diegene bijvoorbeeld fysiek op moeten zoeken. Online hoeft geen fysieke afstand te worden overbrugd. Door het gebrek aan dit soort beperkingen kan immoreel of schadelijk gedrag dus snel groeien.

De voordelen van snelle groei zijn asymmetrisch verdeeld tussen hen die schade aanrichten en hen die schade proberen te voorkomen of te bestrijden. Een plagerige *meme* kan bijvoorbeeld heel eenvoudig via chatgroepen worden verspreid, maar het slachtoffer kan niet of nauwelijks achterhalen waar dat gebeurt en zou voor de bestrijding iedere deelnemer van die chatgroepen moeten benaderen met het verzoek om de meme te verwijderen en niet verder te verspreiden.

De onderstaande mechanismen dragen bij aan de groeicapaciteit en bestendigende werking van het internet.

Schaalbaarheid en viraliteit op platformen
Schaalbaarheid is een van de ontwerpprincipes van het internet. Dit betekent dat er geen intrinsieke beperkingen zijn die verhinderen dat nieuwe computers op het internet kunnen worden aangesloten. Het internet is ontworpen voor groei. Alle aangesloten computers kunnen in principe ook vrij deelnemen aan het uitwisselen van informatie.

Websites proberen vaak wel informatie af te schermen, bijvoorbeeld door middel van een inlogsysteem, maar voor gebruikers is het dan vaak alsnog heel eenvoudig om die informatie over te hevelen en zo verder te verspreiden. Bovendien kan door het hierboven beschreven effect van syndicatie snel een geïnteresseerd publiek worden gevonden. Hoe klein en eigenaardig een bepaalde voorkeur ook kan zijn, online kan er al snel een publiek worden gevonden met een omvang dat het rendabel maakt om te bedienen.

Verder zijn veel platformen ontworpen om snelle verspreiding van informatie te bevorderen. Een hashtag kan ‘trending’ zijn op Instagram, een video ‘viraal’ gaan op TikTok of een website kan veelvuldig bovenaan de resultaten van een zoekmachine terechtkomen. Dit zijn allemaal voorbeelden van hoe algoritmen bijdragen aan het versneld verspreiden van informatie.

Persistentie (en onbeheersbaarheid)
Zodra informatie wordt verspreid op het internet kan iedere ontvanger die informatie opslaan en zelfstandig verder verspreiden. In combinatie met het openbare en onmiddellijke karakter van het internet, kan het voor de oorspronkelijke afzender vrijwel direct na publicatie onmogelijk zijn om informatie nog te verwijderen. Dus zelfs als iemand na het verspreiden van een misleidende *meme* over vaccinaties via WhatsApp dit ongedaan maakt door het bericht te verwijderen, bestaat de kans dat een van de ontvangers deze afbeelding al heeft opgeslagen of zelfs alweer heeft doorgestuurd. Dit onomkeerbare karakter van de publicatie van informatie op het internet kan flink bijdragen aan de omvang van de schade. Een verdachtmaking of shaming op het internet kan iemand bijvoorbeeld voor het leven tekenen, omdat het praktisch niet kan worden uitgewist (Duin, 2020).

Alternatieve netwerkstructuren, waarin de originele auteurs volledige controle houden over hun data, zijn denkbaar, maar vergen een compleet herontwerp van het internet. Tot die tijd is de beste manier van controle houden simpelweg het niet delen van informatie op het internet. Want zelfs bij platformen waarop content

slechts kortstondig zichtbaar blijft, is het mogelijk om screenshots te maken die permanent online kunnen blijven circuleren (Ludemann, 2018).

4.5 Escalatie

De verspreiding van informatie op het internet kan een sterk escalerend effect hebben. Iets overweldigends of indrukwekkends kan snel veel aandacht trekken. Dat heeft te maken met de volgende mechanismen op het internet.

Selectie en amplificatie
Om te navigeren op het internet gebruiken mensen allerlei hulpmiddelen, zoals zoekmachines en aanbevelingsalgoritmes. Deze hulpmiddelen zijn in feite selectiemiddelen. Het gebruik ervan kan immoreel en schadelijk gedrag in de hand werken, bijvoorbeeld doordat het aantal gebruikers dat met uitingen van immoreel of schadelijk gedrag in contact komt, kan toenemen (amplificatie).

Sommigen beweren dat de selectiviteit van deze hulpmiddelen de pluriformiteit van de media bedreigt, en dat mensen erdoor raken opgesloten in een of ‘filterbubbel’ (Pariser, 2012). Onderzoek door het Commissariaat voor de Media naar dit fenomeen in Nederland toont dit echter niet aan (Commissariaat voor de media, 2019). Wel erkent het onderzoek de opiniemacht die platformen hebben. Aanbevelingsalgoritmen bevatten bijvoorbeeld redactionele keuzes en zijn zo van invloed op het politieke discours. Platformen kunnen informatie meer of minder aanprijzen of zelfs volledig verwijderen.

Diverse experts die zijn geconsulteerd voor dit onderzoek uiten hun zorgen over een mogelijk polariserend effect, dat het gevolg kan zijn van selectie en amplificatie. Zij vinden dat de makers van deze hulpmiddelen meer verantwoordelijkheid moeten nemen en de negatieve gevolgen moeten beperken.

Tegelijkertijd uiten zij zorgen over de tendens om niet de hulpmiddelen aan te passen, maar de informatie en gebruikers van platformen te verwijderen. Het gevolg kan zijn dat deze gebruikers hun toevlucht zoeken op andere platformen waar meer gelijkgestemden en minder tegengeluiden te vinden zijn, of waar andere gebruiksvoorwaarden gelden. Parler biedt bijvoorbeeld vergelijkbare functionaliteiten als Twitter, maar wordt steeds vaker aangeduid als ‘uiterst rechts populair alternatief’ vanwege de minder stringente gebruiksvoorwaarden (Algemeen Dagblad, 2021).

Er tekenen zich ook de contouren af van een kat-en-muis-spel tussen de makers van content en moderatoren. De makers van video’s die mogelijk in strijd zijn met

voorwaarden van een platform als YouTube plaatsen daar bijvoorbeeld alleen een introductievideo. De kijkers worden vervolgens via een link naar een platform als Bitchute geleid, waar andere voorwaarden gelden.

Aandachtseconomie

Gebruikers van online platformen kunnen geld verdienen door content via de platformen te ontsluiten. Zo kunnen makers van video's geld verdienen per duizend views (CPM, Cost Per Mille views/impressies), door het aanbieden van steeds extremere content (zie ook paragraaf 5.2). Ook kunnen zij producten of diensten aanprijzen (*branded content*) of kijkers vragen om geld te doneren of een abonnement of lidmaatschap af te sluiten. Daarnaast zijn er indirecte manieren om geld te verdienen. 'Like farming' is een fenomeen waarbij grote Facebookgroepen worden gecreëerd, waarna de gebruikers worden doorverwezen naar websites die advertenties tonen of producten of diensten verkopen. De beheerders van dit soort groepen zijn niet noodzakelijkerwijs begaan met het onderwerp. Zo bleken Macedonische jongeren opvallend veel desinformatieberichten te maken gericht op de Amerikaanse presidentsverkiezingen in 2016. Dit deden zij niet vanwege hun voorkeur voor een bepaalde kandidaat, maar vanwege de advertentieopbrengsten (Rathenau Instituut, 2020b). Adverteerders weten vaak niet waar hun berichten terechtkomen, waardoor ze onbedoeld immoreel of schadelijk gedrag kunnen financieren (Stop Hate for For Profit, 2021).

Aangezien gebruikers van het internet in de praktijk maar een beperkt aantal kanalen gebruiken om producten en diensten te verkennen, concurreren aanbieders in die kanalen om de aandacht. Deze economische dynamiek wordt ook wel beschreven als de aandachtseconomie-theorie (Davenport & Beck, 2001). De gedachte hierbij is dat het financiële succes van aanbieders sterk wordt bepaald door de aandacht die ze weten te trekken en vast te houden, vaak door slim in te spelen op selectie- en amplificatiemechanismen. De theorie is ook verwant aan theorieën als het surveillancekapitalisme (Zuboff, 2019), waarin wordt gesteld dat platformen er een groot belang bij hebben om zoveel mogelijk te weten te komen over gebruikers, om zodoende de selectie en amplificatie te kunnen optimaliseren en de aandacht van gebruikers zo goed mogelijk te kunnen vasthouden.

Het gevolg van deze dynamiek kan zijn dat immorele en schadelijke content en gedragingen zo worden gepresenteerd, dat zij veel aandacht trekken (Brady et al., 2017). Dit kan een verklaring zijn voor de extreme vormen die allerlei immorele en schadelijke fenomenen kunnen aannemen (Bishop, 2019). Zo is een extreme prank-video waarin een kind te zien is dat een computerspel speelt en ontzettend schrikt van horrorbeelden die daarin verstopt zitten, tientallen miljoenen keren bekeken (Hobbs & Grafe, 2015).

De advertentieverkoop op basis van gebruikersprofielen houdt vaak rekening met de *total watch time* (hoe lang een bezoeker van een website naar bijvoorbeeld een filmpje kijkt). Vaak wordt een verband gelegd tussen deze *total watch time* en mechanismen als filterbubbels, echo chambers en radicalisering (Auxier & Vitak, 2019; Faddoul et al., 2020; O'Callaghan et al., 2015). Hoe dit werkt, beschrijft ook een uitgebreide podcastreeks van *The New York Times* (Roose, 2020) onder de titel 'Rabbit Hole'. Deze gaat over een Amerikaanse jongen die radicaliseerde en in een 'fabeltjesfuik' (Lubach, 2020) belandde via het YouTube-algoritme dat hem bepaalde filmpjes aanbeveelde.

Voor een deel hebben de beheerders van internetplatformen er zelf belang bij om immoreel en schadelijk gedrag te weren en een veilige en prettige online omgeving te creëren, bijvoorbeeld om adverteerders of gebruikers tevreden te stellen. De belangen van adverteerders en gebruikers kunnen echter ook met elkaar conflicteren (Gabszewicz et al., 2001). Wanneer het businessmodel van een platform of beheerder afhankelijk is van advertentie-inkomsten, zullen de belangen van adverteerders zwaar wegen. Dit heeft gevolgen voor de content (Sanders, 2021, p. 61). Zo kan een internetplatform ertoe geneigd zijn om meer tijd en aandacht van gebruikers te verlangen dan in het belang is van de gebruiker, en doet de kwaliteit van de content er minder toe dan bij een abonnementmodel waarbij de gebruiker betaalt voor hoogwaardige content.

4.6 Virtuele realiteit

Het internet is niet tastbaar en kan daardoor als onecht en kunstmatig worden ervaren. Tegenwoordig is het internet echter een belangrijk deel van het dagelijks leven en hebben acties online verstrekkende gevolgen in de fysieke wereld (Rathenau Instituut, 2020a). Verwarring over wat echt of niet echt is, kan leiden tot schade, bijvoorbeeld wanneer de normen die gelden voor een fantasiespel worden toegepast op de realiteit. De onderstaande mechanismen spelen in deze problematiek een rol.

Ontmenselijking
Is een bedreiging gericht aan @minpres op Twitter werkelijk een bedreiging gericht op de minister-president van Nederland of enkel gericht op een virtueel Twitteraccount? Dit is een relevante vraag, omdat moraliteit afhankelijk is van hetgeen in het geding is. Het beschimpen van een niet-levend object, zoals een online account, is iets anders dan een levend mens kwetsen. Het verwarrende is echter dat het internet tegelijkertijd heel kunstmatig en heel persoonlijk kan zijn. Als medium brengt het internet mensen in contact, maar ontdoet het hen ook grotendeels van menselijke kenmerken (De Vries, 2021). De dader ziet het

slachtoffer van online pesten bijvoorbeeld meestal niet, en hoeft het slachtoffer niet eens te kennen. Wanneer dit wel het geval zou zijn, zou de dader zich wellicht anders, sociaal wenselijker, gedragen.

Amusement
Het amusementsgehalte van het internet speelt bij diverse fenomenen ook een rol. Ruim 7 miljoen Nederlanders spelen immers gemiddeld een uur per dag spelletjes op een computer, tablet, smartphone of game console (Multiscope, 2020). Vaak zijn games online. Het internet heeft dus ook een belangrijke entertainmentfunctie. Het maakt voor de interpretatie van gedrag alleen wel veel verschil of er sprake is van een entertainmentcontext of niet. Een uitspraak als “Ik ruk je kop eraf” heeft in een vechtspelletje een heel andere betekenis dan op een socialemediaplatform.

Het lastige is dat games en gameplatformen vaak dezelfde kenmerken vertonen als sociale media. Op gameplatform Steam heb je ook profielen, vrienden en chats, net als op Facebook. Sommige socialemediaplatformen zijn van origine specifiek gericht op het bedienen van gamers, maar worden nu ook voor allerlei andere doeleinden gebruikt, zoals Discord of Twitch. Het is dus begrijpelijk dat fenomenen uit games, zoals het opzettelijk overtreden van normen en regels (trolling en griefing), maar ook onschuldig bedoelde bedreigingen of gewelddadige uitingen, zichtbaar worden buiten de context van games. Immoreel of schadelijk gedrag kan dus ook voortkomen uit verwarring over de ernst van de context.

Socialemediaplatformen zijn vaak ook ontworpen voor een hoog amusementsgehalte. Een platform als TikTok staat bol van de humoristische en amusementsvideo’s, maar een gebruiker kan ook video’s aantreffen met misleidende of haatdragende informatie (Weimann & Masri, 2020). Deze content kan zonder serieuze of kwade intenties zijn gemaakt. Zo blijkt cyberpesten vaak ook een vorm van vermaak voor de daders (Raskauskas & Stoltz, 2007) en amusement blijkt tevens een belangrijke motivatie te zijn voor vandalen die Wikipedia bekladden (Shachaf & Hara, 2010).

Beeldvorming
Informatie die wordt verspreid via het internet, is van grote invloed op het mentale beeld dat mensen hebben over de werkelijkheid. Experts die voor deze studie werden geconsulteerd, wijzen erop dat het voor steeds meer beroepen en andere posities in de samenleving praktisch een noodzaak is om op enigerlei wijze aanwezig te zijn op het web, een online imago of reputatie op te bouwen. Veel mensen hebben dus een LinkedIn- of Facebookprofiel waar ze gunstig op worden geportretteerd, om zo mee te kunnen doen aan de maatschappij.

Deze tendens kent een aantal risico's. Het belang van positieve beeldvorming op het internet kan een verklaring zijn voor de soms heftige reacties op kritiek. Een aantasting van iemands reputatie kan grote gevolgen hebben. Een negatieve reactie in de privésfeer kan zomaar gevolgen hebben voor iemands professionele praktijk.
Jongeren maken niet langer onderscheid tussen wie ze online en offline zijn en spannen zich in om online een sociale wenselijke versie van zichzelf te presenteren en zichzelf op te poetsen (Cocking & van den Hoven, 2018, p. 30) (Cocking en Van den Hoven, 2018, p.30). Op sociale media is bijna niemand saai of ongelukkig. Vooral rooskleurige verhalen en de mooiste vakanties komen voorbij.

De zogenaamde 'app-generatie' is meer met zichzelf bezig dan alle voorgaande generaties. Sociale media versterken dit doordat ze zijn ontworpen rondom persoonlijke online profielen (Gardner & Davis, 2013, pp. 69–71). Er is een positief verband gevonden tussen narcisme en de kans dat meer berichten worden geplaatst op sociale media (Gardner & Davis, 2013, p. 76). Waar men in offline gesprekken voor 30% tot 40% over zichzelf praat, gaat dit bij online socialemediaberichten om ongeveer 80% (Gardner & Davis, 2013, p. 76).

4.7 Wanorde

Het internet wordt ook wel cyberspace genoemd, alsof het om een omgeving gaat die los staat van grenzen en de soevereiniteit van landen. In werkelijkheid zijn bij het gebruik van het internet al snel partijen uit allerlei landen betrokken. Een video uit China kan op een server in Duitsland staan en bijvoorbeeld getoond worden via software onder een licentie van een partij uit de Verenigde Staten. Deze complexiteit kan leiden tot de praktische onmogelijkheid om orde te handhaven. De onderstaande kenmerken en mechanismen spelen daarin een rol.

Onduidelijke normen
Voor velen is het nog onduidelijk hoe beschaafd gedrag op het internet er eigenlijk uit hoort te zien. Hoe ga je bijvoorbeeld welgemanierd met elkaar om op het internet? Sinds het begin van interacties op internet zoeken mensen naar de etiquette voor bijvoorbeeld e-mails, chats, games, forums en andere online omgevingen (Shea, 1994). Telkens wanneer zich een nieuwe interactiefunctie aandient, moet dit socialisatieproces opnieuw worden doorlopen. Mensen gedragen zich in de *virtual reality* omgeving van VRChat bijvoorbeeld weer heel anders dan in de audio chatdienst Clubhouse. Mensen komen op het internet dus vaak in een omgeving terecht waar onduidelijkheid kan bestaan over de geldende normen.

De onduidelijkheid over de normen betekent ook dat veel mensen niet weten wanneer en hoe ze anderen aan moeten spreken op overschrijdingen (Movisie, z.d.). Uit sommige experimenten blijkt dat tegenspraak effectief kan zijn, ook als dit wordt geautomatiseerd (Machkovech, 2016), maar dan moet dat natuurlijk wel plaatsvinden. Zonder correctie is het – zeker voor jongeren – lastig om een volwaardig besef van normen te ontwikkelen.
Sommige platformen laten het vormen van de etiquette helemaal over aan de gebruikers, maar andere dwingen dit technisch af. Zo wordt voor de aanstaande Horizon VR dienst van Facebook bijvoorbeeld actief nagedacht over hoe dicht avatars bij elkaar in de buurt mogen komen, om intimidatie te voorkomen (Rabkin, 2021). In veel online omgevingen zijn sociale normen niet alleen onduidelijk, er is ook een gebrek aan begeleiding in het aanleren of toezicht op normen. Dit alles kan een oorzaak zijn van immoreel of schadelijk gedrag.

Anonimiteit
Het is op het internet vaak onduidelijk met wie men communiceert en of de andere partij een mens of robot is (Christopherson, 2007). Het verhullen van je identiteit of het aannemen van de identiteit van een ander is erg eenvoudig. Vaak volstaat het aanmaken van een gratis e-mailadres, waarna accounts kunnen worden aangemaakt bij allerlei andere dienstverleners. Deze anonimiteit wordt door kwaadwillenden vaak gebruikt om ongestraft normen te kunnen overtreden. Regelmatig melden internetplatformen bijvoorbeeld dat zij tien- of soms wel honderdduizenden nepaccounts tegelijkertijd verwijderen (Van Bemmel, 2020). Zelfs met geavanceerde middelen kan het achterhalen van de identiteit van een internetgebruiker praktisch onmogelijk zijn. Dit betekent dat niet alleen daders, maar ook slachtoffers anoniem kunnen zijn. Het probleem is dat de anonimiteit van slachtoffers immoreel gedrag door daders meer in de hand werkt dan situaties waarin slachtoffers niet anoniem zijn (Yam & Reynolds, 2016). Anderzijds leidt anonimiteit ertoe dat daders het risico op schade voor zichzelf laag inschatten en eerder geneigd zijn om onethische keuzes te maken (Vince, 2018).

(Schijnbare) wetteloosheid
Door te internetten doorkruisen gebruikers met het grootste gemak een veelvoud aan jurisdicties. Onrechtmatig gedrag blijft vaak onbestraft, omdat de daders lastig kunnen worden achterhaald – niet alleen vanwege de anonimiteit, maar ook omdat opsporing en berechting van daders complexe samenwerking tussen internationale opsporings- en handhavingsinstanties vergen. Zelfs wanneer een dader kan worden geïdentificeerd, kan deze alsnog blijken te verblijven in een land waarop geen invloed kan worden uitgeoefend Nederland. Het internationale karakter van het internet maakt handhaving kortom erg complex. Het uitblijven van consequenties vertaalt zich voor slachtoffers van immoreel of schadelijk gedrag naar machteloosheid (zie Casus online shaming).

De schijnbare wetteloosheid van het internet wordt ook door commerciële partijen gefaciliteerd. Sommige partijen adverteren zelfs met hun bereidheid om niet mee te werken aan handhaving. Ook zijn er internetdiensten die technisch zo worden vormgegeven, dat er geen enkele partij verantwoordelijk kan worden gehouden. Dan kan het gaan om diensten of producten die zijn gebaseerd op blockchain of *distributed ledger* technologie (DLT).[5]

Afzondering
Internetten is vaak een solistische activiteit. Veel mensen hebben een eigen telefoon of computer, en maken daarop zelfstandig in afzondering van anderen gebruik van het internet. Dit betekent dat er ook minder mogelijkheden zijn voor (ouderlijke) begeleiding, toezicht of correcties (Peterson & Densley, 2017).

Dit kan problematisch gedrag veroorzaken, omdat moraliteit voor een belangrijk deel een sociale dimensie heeft (Ellemers et al., 2019). Dit kan ook de reden zijn waarom sommige mensen zich online en offline zo verschillend gedragen.

5 DLT is een containerbegrip voor systemen, zoals blockchain, waarin meerdere partijen opereren in een digitale omgeving waarin een centrale autoriteit of operator ontbreekt. Blockchain gebruikt een datastructuur die bestaat uit een ketting van hash-linked blokjes van data.

Casus: verstoord eetgedrag

De hieronder beschreven casus is fictief en bedoeld om een beeld te schetsen van mogelijke risico's die ontstaan door online challenges die aanzetten tot verstoord eetgedrag. Hoewel fictief is de casus deels gebaseerd op een combinatie van voorvallen in Nederland en het buitenland.

Casus
Tijdens de lockdown brengt Sam dagenlang in haar eentje door op haar tienerkamer met als enige afleiding haar laptop en telefoon. Het begint allemaal met een *challenge* van Kim, iemand die ze nog kent van de camping waar ze was vorige zomer. Kim laat op TikTok meerdere keren per dag zien hoeveel calorieën ze die dag eet. Ze plaatst regelmatig filmpjes van haar maaltijden: soms gaat het om een portie ijsblokjes. Ze moedigt anderen aan om deel te nemen aan deze extreme challenge en haar te overtreffen met nog lagere cijfers van hun dagelijkse calorie-inname. Sam raakt verslaafd aan het volgen van de eindeloze stroom updates van Kim. Ze verveelt zich en het lijkt haar wel een grappige uitdaging om zelf ook haar calorieën bij te houden, dus ze besluit om mee te doen. Ze krijgt online meteen al veel positieve reacties op haar berichten en dat geeft haar voldoening. Met sommige van haar nieuwe volgers heeft ze leuk contact, waardoor ze zich minder eenzaam voelt. Zij begrijpen haar tenminste.

Sam leest in een krantenartikel over pro-ana-communities en gaat online zoeken naar de namen en websites die in dit artikel staan. Pro-ana websites zijn niet verboden, dus er is genoeg te vinden. Sam durft offline op school of met haar ouders niet te praten over haar nieuwe obsessie. Online kan ze iemand anders zijn en hier wel open over zijn. Dat lucht op. Sam maakt een nieuw e-mailadres aan onder de alias Rox, en maakt hiermee op TikTok en Instragram een sock puppet of nepaccount aan onder dezelfde naam met een valse geboortedatum en profielfoto van een vreemde die totaal niet op haar lijkt. Binnen de zoekmachines en socialemediaplatformen gebuikt ze verkeerd gespelde zoektermen en hashtags zoals anoreixa, anorexiia en annorexia om blokkades van deze online platformen te omzeilen. Binnen een paar klikken vindt ze al een pro-anagroep met nog extremere foto's dan ze ooit had gezien. Sam sluit zich ook aan bij de WhatsAppgroep van deze community waarin de hele dag appjes over afvallen over en weer gaan.

Ze zit in een groep waarvan sommige accounts soms worden verwijderd. Deze duiken dan meteen elders weer op en hebben binnen een mum van tijd weer meer dan vijfhonderd volgers, omdat iedereen via WhatsApp in contact blijft. Er is geen moderator in deze groep, niemand die geen eetstoornis heeft en geen volwassene

die een ander geluid kan laten horen. Sommige meisjes plaatsen foto's van hun opnames in het ziekenhuis en vertellen hier heel openhartig over. Sam kijkt tegen deze meisje op en wil al hun YouTube video's zien. De *content recommendation* systemen hebben snel door welke soort video's Sam graag tot het einde bekijkt en welke beelden ze likes geeft. Deze systemen bieden haar hier steeds meer van aan. De *recovery-accounts* van mensen in haar netwerk die vertellen hoe ze anorexia overwonnen, verdwijnen steeds meer uit beeld.

Adverteerders die producten promoten om af te vallen zijn verboden op TikTok, maar Sam heeft geen banners nodig om kwakzalvers die laxeermiddelen verkopen te kunnen vinden. Voor haar verjaardag krijgt ze een duur horloge met een stappenteller en op haar telefoon koopt ze afval-apps met calorieëntellers. Iedere dag deelt ze de metingen van deze statistieken op sociale media. Ze wordt steeds meer aangemoedigd door een groeiend aantal volgers en ze verdient er zelfs wat aan. Dan heeft Sam haar streefgewicht gehaald en gaat ze er voorbij. Met zoveel volgers kan ze nu niet zomaar stoppen. In haar groepje roept niemand haar tot de orde.

Sam wordt ziek, ook mentaal. Haar ouders willen dat ze hulp krijgt, maar ze komt op een wachtlijst. Pas weken later, als ze op een schooldag meerdere malen flauwvalt, wordt ze opgenomen. Ze mag haar telefoon gewoon meenemen naar de kliniek. Niemand vraagt haar daar welke apps ze heeft, wie ze volgt en of ze contact heeft met mensen die ze offline niet kent. Wat ze overdag hoort van haar hulpverleners vergeet ze allemaal weer zodra ze door haar vertrouwde WhatsAppgroep, Instagram en TikTok scrollt. Eigenlijk komen er nooit meer beelden langs van mensen die niet hyperslank zijn. Het lukt haar niet om uit de online pro-anagroep te stappen. Dit zijn nog de enige leeftijdsgenoten met wie ze contact heeft. Van die paar vriendinnen van school – de enigen met wie ze offline contact had – hoort ze nog zelden iets.

Reflectie
We zien in dit scenario meerdere mechanismen tegelijk aan het werk. In dit voorbeeld gaat het om de fysieke en emotionele **afzondering** van een potentieel slachtoffer in combinatie met **anonieme openbaarheid** online en de **continue** aanwezigheid van technologie in de privésfeer. Ook speelt **syndicatie** in combinatie met **echo chambers** een belangrijke rol: een kleine groep potentiële slachtoffers kan elkaar online veel sneller vinden. Zij horen hierdoor geen geluiden meer die hen op andere ideeën kunnen brengen. De echo chambers, **selectiviteit** en syndicatie worden in de hand gewerkt door *content recommendation*-algoritmen die zijn ingesteld om bijvoorbeeld aantallen minuten kijktijd te verhogen. In de **aandachtseconomie** zijn dit maatstaven voor adverteerders bij **traditionele media** en online platformen. Individuele gebruikers worden eveneens beloond voor

content die veel volgers of abonnees genereert. Zij streven daarom naar **amplificatie** en **viraliteit**, waarvoor sensationele content en amusement meestal het beste werkt. Soms spoort dit minderjarigen aan tot schadelijk gedrag zoals extreme dieetchallenges.

Bij deze casus zijn ook diverse actoren betrokken. Sommigen spelen een zeer actieve rol, zoals minderjarige gebruikers en volgers die elkaar aanzetten tot verstoord eetgedrag. Zij doen dit in een omgeving zonder begeleiding en toezicht van ouders, andere volwassenen en zonder hulpverleners of moderatoren. De ouders, school, vrienden en hulpverleners in de offline omgeving spelen in dit voorbeeld ook een passieve rol waar het gaat om de bevordering van offline contact, hulp en begeleiding bij de omgang met online mechanismen. Tot slot spelen ook kwakzalvers en aanbieders van afvalproducten en diensten, de platformen en traditionele media een belangrijke rol. Hun focus op conversie, clicks, aantallen minuten kijktijd en de verzameling van individuele dataprofielen werkt in dit geval schadelijk gedrag in de hand.

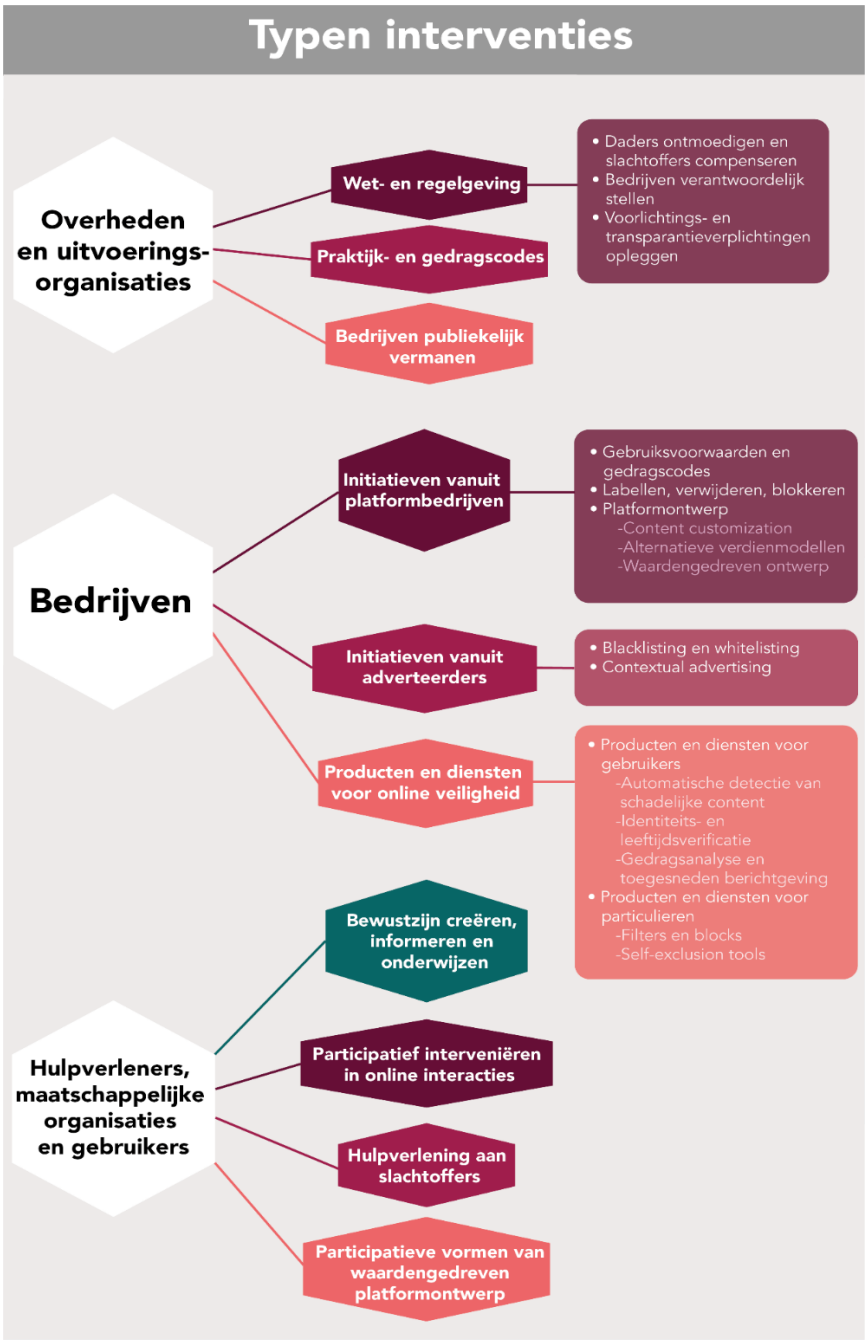
5 De huidige aanpak van schadelijk gedrag online

In de afgelopen jaren zijn verscheidene initiatieven van de grond gekomen om schadelijk en immoreel gedrag online te beperken of zelfs te voorkomen. In dit hoofdstuk inventariseren we de bestaande aanpak van schadelijk gedrag vanuit *actorperspectief*, en op het niveau van *types van interventies*. Dit overzicht van bestaande initiatieven geeft inzicht in de interventies die al werken en veelbelovend zijn. Maar in de aanpak vallen nog gaten; er is dus ruimte voor extra interventies. Dit dient als inspiratie voor de strategische agenda die in hoofdstuk 6 gepresenteerd wordt.

In dit hoofdstuk onderscheiden we drie groepen actoren: overheden en uitvoeringsorganisaties; bedrijven (waaronder platformbedrijven, maar ook aanbieders van andere producten en diensten); en tot slot, een (brede) categorie van hulpverleners, maatschappelijke organisaties en internetgebruikers. Voor elk van deze groepen benoemen we de belangrijkste typen interventies, inclusief voorbeelden (zie Figuur 5). Vervolgens bespreken we per type interventie de belangrijkste lessen die te trekken vallen uit de bestaande initiatieven.

5.1 Overheden en uitvoeringsorganisaties

We beginnen het overzicht met initiatieven die overheden en hun uitvoeringsorganisaties tot dusver genomen hebben om schadelijk en immoreel gedrag online tegen te gaan of te voorkomen. Eerst bespreken we interventies op het gebied van wet- en regelgeving. Daarna benoemen we nog twee andere sturingsopties: het maken van meer vrijblijvende afspraken met marktpartijen in de vorm van praktijk- en gedragscodes, en het publiekelijk vermanen van bedrijven, om ze te stimuleren tegen problematisch gedrag op te treden.



Bron: Rathenau Instituut

Figuur 5 Taxonomie van interventies

Wet- en regelgeving

Waar het regulering betreft, was er in Europa tot nu toe vooral sprake van nationale initiatieven. Verschillende landen hebben in de afgelopen jaren wetgeving aangenomen of voorbereid die problematisch gedrag online en daaruit voortvloeiende schade moet helpen terugdringen. Maar inmiddels is er ook

relevante Europese regelgeving in de maak. Zo moeten de *Digital Services Act (DSA)* en *Digital Markets Act (DMA)* die in 2020 door de Europese Commissie gepresenteerd zijn, onder meer bijdragen aan de bestrijding van illegale content en desinformatie en de marktmacht van grote platformen aan banden leggen.

In een discussiestuk over de aanpak van *online harm* en manipulatie stelt het Britse Behavioural Insights Team, een groep gedragswetenschappers die het Britse kabinet adviseert, dat de gereedschapskist van de wetgever traditioneel drie opties biedt (Costa & Halpern, 2019). Ten eerste kunnen overheden interventies doen om problematisch gedrag te ontmoedigen. Ten tweede kunnen overheden maatregelen nemen om bedrijven, als eigenaars of beheerders van de online omgevingen waar het gedrag plaatsvindt, ertoe te bewegen zich aan minimale standaarden te houden. Daartoe kunnen overheden deze bedrijven bepaalde verantwoordelijkheden en verplichtingen opleggen. En ten derde kunnen overheden bedrijven stimuleren om consumenten of gebruikers beter voor te lichten, door bedrijven te dwingen transparanter te zijn over wat ze doen. Hieronder bespreken we deze opties.

Daders ontmoedigen en slachtoffers compenseren
Sommige van de fenomenen die in dit rapport besproken worden, zijn onder de huidige wetgeving al strafbaar. Het gaat dan vaak om wetten die ontworpen zijn met ‘offline’ varianten van het gedrag in gedachten, die eveneens gelden voor online gedragingen. Veel Europese landen hebben bijvoorbeeld strafwetgeving die betrekking heeft op *hate speech* en haatdelicten, die ook inzetbaar is voor online varianten (zie bv. Policy Department for Citizens’ Rights and Constitutional Affairs, 2020).

Daarnaast kunnen overheden ervoor kiezen om online gedragingen strafbaar te stellen. In veel landen gebeurt dit voor specifieke fenomenen, zoals cyberterrorisme of (door het internet gefaciliteerd) seksueel misbruik van kinderen. Zo koos Nederland er recentelijk voor om het zonder toestemming vervaardigen of verspreiden van seksuele afbeeldingen van anderen, wat gebeurt bij wraakporno of sextortion, strafbaar te stellen (Ministerie van Justitie en Veiligheid, 2020). Demissionair minister Grapperhaus wil dat dit ook gaat gelden voor doxing, of het op sociale media delen van privégegevens van mensen (Bakker, 2021). Ook wordt in Europees verband gewerkt aan een initiatief om haatdelicten beter strafrechtelijk aan te kunnen pakken (Policy Department for Citizens’ Rights and Constitutional Affairs, 2020).

Daarnaast biedt ook het privaatrecht al mogelijkheden om tegen vormen van online schadelijk gedrag te procederen. Civiele procedures zijn er niet zozeer op gericht daders te bestraffen, als wel een partij verantwoordelijk te stellen voor de door het

slachtoffer geleden schade. Ze kunnen bijdragen aan erkenning van of compensatie voor de opgelopen schade (bijvoorbeeld middels een schadevergoeding) of bijdragen aan eerherstel (denk aan excuses of een rectificatie). Bureau Clara Wichmann, een stichting die zich inzet voor de rechtspositie van vrouwen, zocht bijvoorbeeld uit welke mogelijkheden dit recht biedt in gevallen van online *hate speech* (Bureau Clara Wichmann, 2020).

Een punt van zorg, zowel bij de toepassing van bestaande wetten als het ontwerp van nieuwe, is dat het vaak lastig is om ze te handhaven. Ten eerste, omdat rechtsstaten online vooralsnog minimaal aanwezig zijn (Adviesraad Internationale Betrekkingen (AIV), 2020; zie ook Bantema et al., 2018) en handhavers vaak de kennis en (technische) middelen niet hebben om er efficiënt te opereren (zie bv. Politie et al., 2020). Ten tweede, omdat de juridische instrumenten en procedures die in de offline wereld voorhanden zijn, niet altijd geschikt zijn voor het handhaven van dit soort wetten en regels. Politie en Openbaar Ministerie hebben last van dit probleem, maar ook slachtoffers, die allerlei drempels ervaren als ze bijvoorbeeld onrechtmatige content waarvan ze schade ondervinden, willen laten verwijderen (Adviesraad Internationale Betrekkingen (AIV), 2020; IVIR, 2020). En ten derde is handhaving lastig, omdat wetten en regels betrekking hebben op zeer diverse diensten (IVIR 2020) van bedrijven die vaak internationaal actief zijn – terwijl er geen bevoegde, grensoverschrijdende rechtsmacht is (Adviesraad Internationale Betrekkingen (AIV), 2020; Aiken, 2016). Juristen zien dan ook veel heil in het verleggen van de aandacht van nationale, naar internationale regelgeving (Policy Department for Citizens' Rights and Constitutional Affairs, 2020).

Bedrijven verantwoordelijk stellen

In het verleden waren het internet en internetgebonden activiteiten grotendeels ongereguleerd. Daar lijkt nu verandering in te komen. Overheden lijken zich er steeds vaker rekenschap van te geven dat de kenmerken van de online omgeving en de actoren die er actief zijn, een rol spelen bij het ontstaan of katalyseren van schadelijk gedrag.

De Europese *Richtlijn inzake elektronische handel* (2000) die nu nog geldt maar op termijn vervangen zal worden door de *Digital Services Act (DSA)*, vrijwaart socialemediaplatformen en internet-access- en webhostingproviders in principe van aansprakelijkheid voor de content die hun gebruikers uploaden, op voorwaarde dat ze daarbij louter als 'doorgeefluik' fungeren (Europese Raad, 2000). Alleen als ze in kennis gesteld zijn van de aanwezigheid van onrechtmatige content, moeten ze die verwijderen (het zogenoemde *notice-and-take-down*-systeem). In de afgelopen jaren hebben verschillende Europese landen aanvullende regelgeving geïntroduceerd, hetzij om af te dwingen dat bedrijven beter gehoor geven aan deze verplichting, hetzij om ze anderszins verantwoordelijkheid te laten nemen voor het

schadelijke gedrag dat via hun kanalen plaatsvindt. In veel gevallen is de achterliggende redenering dat zelfregulering tot dusver onvoldoende heeft gewerkt (zie bv. UK Government, 2019).

Hieronder geven we een vergelijkend overzicht van de belangrijkste nationale wetgevende initiatieven in Europa: de Duitse *Network Enforcement Act* (2017), de Franse *loi Avia* (2020) en de Britse *Online Safety Bill* (wetsvoorstel).[6] Hoewel ze op onderdelen verschillen (bijvoorbeeld in het type content waarop ze betrekking hebben), hebben deze wetten met elkaar gemeen dat ze bedrijven allerlei verantwoordelijkheden opleggen, bijvoorbeeld op het gebied van contentverwijdering, klachtenprocedures of rapportage. Onderaan in het overzicht nemen we ook de voor dit onderzoek meest relevante punten op uit het voorstel voor de *Digital Services Act* dat afgelopen december door de Europese Commissie gepresenteerd werd. Deze voorgestelde *act* zal op termijn gelden voor alle lidstaten van de Europese Unie. In de tussentijd kunnen lidstaten, waaronder Nederland, en het Europees Parlement, nog invloed uitoefenen op de inhoud van het voorstel.

Uit reacties op deze initiatieven blijkt dat de keuze voor het reguleren van content lastig is. Er moeten immers afwegingen gemaakt worden tussen verschillende vrijheden en grondrechten, zoals de vrijheid van meningsuiting, het recht op toegang tot informatie, de persvrijheid of de vrijheid van ondernemerschap, versus het recht op persoonlijke integriteit en veiligheid en allerlei democratische en rechtsstatelijke beginselen. Al te grote fricties tussen deze rechten en vrijheden kunnen ertoe leiden dat initiatieven falen. Dat gebeurde bijvoorbeeld in Frankrijk, waar de *loi Avia* na bezwaar van juristen en maatschappelijke organisaties door het grondwettelijk hof op onderdelen werd afgekeurd, en daarom na goedkeuring door het parlement alsnog sterk afgeslankt is (Vie publique, 2020). De argumentatie daarbij was dat de wet een al te grote inbreuk maakte op de vrijheid van meningsuiting.

Op de volgende pagina's:
Tabel 2 Vergelijkend overzicht van wetgevende initiatieven in Europa

6 Voluit: *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz - NetzDG)*, respectievelijk *LOI n° 2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur internet*. Het wetsvoorstel voor de Britse Online Safety Bill is in mei 2021 openbaar gemaakt; de uitgangspunten voor de wet werden eerder uiteengezet in twee versie van de *Online Harms White Paper* (UK Government 2019 en UK Government 2020). Daarnaast is ook een Ierse initiatief het nog noemen waard: de kaderwet *General Scheme for an Online Safety and Media Regulation Bill*, die nu in de pre-legislatieve fase zit (Government of Ireland 2021). De *Online Safety Bill* wordt naar verwachting de meest ambitieuze van alle nationale initiatieven, zowel in termen van typen content als bedrijven die daaronder vallen (*in scope*).

Wet	Bedrijven *in scope*	Type content	Verplichtingen (o.a.)	Sancties
Network Enforcement Act (D)	Sociaalnetwerk-sites	Onrechtmatige content	- Content verwijderen of toegang blokkeren (binnen specifieke termijn), na melding - Toegankelijke en efficiënte procedures instellen voor melding en terugkoppeling - Rapportage-plicht voor bedrijven vanaf een specifiek aantal meldingen	- Geldboetes (hoogte afhankelijk van type inbreuk en grootte/bereik van bedrijf of site)
Loi Avia (F, oorspronkelijke versie[7])	Webhosting-providers	Sommige onrechtmatige content: haatcontent, terroristisch en kinderporno-grafisch materiaal	- Content verwijderen (binnen specifieke termijn), ook zonder melding - Toegankelijke en efficiënte procedures instellen voor melding en terugkoppeling	- Geldboetes (+ celstraffen voor daders)
Online Safety Bill (UK, wetsvoorstel)	Alle hosters van *user-generated* content; bedrijven die publieke *en* private online interactie faciliteren; zoek-machines	Onrechtmatige content; content die schadelijk is voor kinderen; rechtmatige content die schadelijk is voor volwassenen[8]	Wettelijke zorgplicht, verankerd in door toezichthouder afdwingbare praktijkcode. Die behelst o.a.: - maatregelen nemen om schade bij gebruikers te voorkomen, o.a. door bepaalde content te verwijderen (binnen specifieke termijn); - toegankelijke en efficiënte procedures instellen voor melding en terugkoppeling - heldere bezwaar-procedures instellen - rapportageplicht - beschermings-plicht t.a.v. journalistieke content voor bedrijven in category 1 (*high-risk*, *high reach*)	- Geldboetes (hoogte afhankelijk van type inbreuk en grootte/bereik van bedrijf of site) - Maatregelen gericht op ontwrichting bedrijfsvoering (bv. onderbreken of blokkeren services) of strafrechtelijke vervolging van senior managers

7 De tabel geeft de inhoud van de wet weer zoals die op 13 mei 2020 door het parlement goed gekeurd is. Enige tijd later is hij sterk afgezwakt, na een oordeel van het grondwettelijk hof dat hij op bepaalde punten onconstitutioneel was (zie verderop in hoofdtekst).

8 Vaak genoemde voorbeelden uit de categorie ‘rechtmatig maar schadelijk voor volwassenen’ zijn *content promoting self-harm*, *hate content*, *online abuse* (in zoverre niet strafbaar) of*‘content encouraging or promoting eating disorders* (UK Government 2020).

Wet	Bedrijven *in scope*	Type content	Verplichtingen (o.a.)	Sancties
Digital Services Act (EU, in voorbereiding)	Webhosting-providers, incl. online platformen; aanbieders van infrastructurele tussendiensten (bv. internet-access-providers, cloudhosters)	Uitsluitend onrechtmatige content	Afhankelijk van het type dienst: - wettelijke notice-and-action-verplichting; toegang blokkeren bij recidiverende gebruikers - toegankelijke en efficiënte procedures instellen voor melding en terugkoppeling; incl. voorrang geven aan 'trusted flaggers' (deskundig op het gebied van verspreiding illegale content) - heldere bezwaar-procedures instellen - rapportageplicht	Geldboetes en meer 'structurele' maatregelen (zoals de verplichting om onderdelen van het bedrijf af te stoten)

Als oplossing voor dit probleem stellen sommige experts voor om in plaats van content te reguleren, platformmechanismen te reguleren: de algoritmische principes voor rangschikking van berichten die ervoor zorgen dat bepaalde content meer wordt aanbevolen dan andere (Pomerantsev, 2019; Rathenau Instituut, 2021a). Zo wordt het recht van mensen op vrije meningsuiting niet te zeer beperkt, maar kan het bereik van schadelijke uitingen (ook wel *freedom of reach* genoemd (Diresta, 2018)) toch enigszins ingedamd worden. Daarmee wordt getornd aan een belangrijk mechanisme achter schadelijk en immoreel gedrag.

Een ander punt van kritiek is dat de genoemde wetten en richtlijnen veel beslissingsmacht leggen bij bedrijven. Het risico is dat platformen, uit vrees voor boetes, intensief content gaan verwijderen en zodoende aan een vorm van censuur gaan doen (Index on Censorship, 2019). In het geval van wetten die ook betrekking hebben op 'schadelijke' content (in tegenstelling tot alleen onrechtmatige) kunnen begripsvaagheid en cultuurverschillen dit risico nog versterken (Adviesraad Internationale Betrekkingen (AIV), 2020; Pomerantsev, 2019). In het Verenigd Koninkrijk zien we nu dat het wetsvoorstel voor de *Online Safety Bill*, in reactie op dergelijke bezwaren, als onderdeel van de zorgplicht voor bedrijven ook verplichtingen heeft opgenomen ter vrijwaring van de vrijheid van meningsuiting en ter bescherming van content die van belang is voor het democratische proces. Dat stelt critici echter nog niet gerust (Hern, 2021). Voor deze problemen bestaan dan ook geen pasklare oplossingen, maar toezicht op het verwijderingsbeleid is in elk geval van groot belang. Het Rathenau Instituut stelt dan ook voor dat het Nederlandse parlement in de aanloop naar de *Digitale Services Act*, aandringt op

een sterke toezichtsstructuur en onafhankelijk publiek toezicht op contentmoderatie (Rathenau Instituut, 2021b).

In reactie op de kritiek dat onduidelijk zou zijn wat ‘schadelijke content’ precies betekent, heeft de Britse overheid ertoe besloten dit begrip in haar wetvoorstel beter te definiëren (UK Government, 2020b). Maar ook dat is geen perfecte oplossing. Want in een democratische rechtsstaat, zo valt te betogen, zouden definities als deze voorwerp moeten zijn van een inclusieve maatschappelijke dialoog (vgl. bv. Helberger et al., 2018). Bovendien zou een dergelijke werkwijze kunnen leiden tot misbruik van bevoegdheden – zeker als ook minder vrije staten ze gaan hanteren. Een alternatief is om wet- en regelgeving zo te ontwerpen dat controle over content juist meer in handen komt te liggen van gebruikers. Bijvoorbeeld door hen zelf de criteria te laten bepalen voor de selectie en rangschikking van de berichten of advertenties, of door hen te laten kiezen hoe hun data daarbij ingezet worden (Costa & Halpern, 2019). Op deze opties gaan we later verder in (zie paragraaf 5.2).

Voorlichtings- en transparantieverplichtingen opleggen
Sommige van de genoemde wetten en richtlijnen leggen bedrijven behalve verantwoordelijkheden ten aanzien van content, ook voorlichtings- en transparantieverplichtingen op. Zo wil de Britse *Online Safety Bill* openheid afdwingen over de prevalentie van online schadelijk gedrag, door de branche rapportages met cijfermateriaal publiekelijk beschikbaar te laten stellen. Dit moet de overheid en haar uitvoeringsorganisaties helpen om beter vat te krijgen op de omvang en aard van online schadelijk gedrag. Daarnaast zullen bedrijven informatie moeten geven over wat ze doen om het gedrag tegen te gaan of gebruikers ertegen te beschermen (UK Government, 2020b).

Een andere vorm van transparantie is die waar de *Digital Services Act* op aanstuurt. Met de *act* wil Europa met name grote platformbedrijven (bedrijven met een zogenaamde ‘poortwachtersfunctie’) ertoe verplichten om inzicht te bieden in hoe hun aanbevelingssystemen werken, en meer in het algemeen, in de rol die data en kunstmatige intelligentie spelen in de diensten die deze platformbedrijven aanbieden. De gedachte daarachter is dat bedrijven zelf, maar ook onderzoekers, al in een vroeger stadium zullen kunnen evalueren wat voor maatschappelijke impact het gebruik van specifieke systemen heeft (Tokmetzis & Bol, 2020), zodat ze daarop kunnen anticiperen. Grote platformenbedrijven zullen voor dit doel waarschijnlijk ook risicoanalyses moeten gaan uitvoeren.[9]

9 Ook op nationaal niveau wordt een begin gemaakt met dergelijke transparantieverplichtingen – zij het meestal in het kader van regulering gericht op specifieke fenomenen. Een Nederlands voorbeeld is een wetgevend initiatief rond micro-targeting in politieke advertentiecampagnes (zie Rathenau Instituut 2020).

Critici van regulering die veel beslissingsmacht legt bij bedrijven, bijvoorbeeld maatschappelijke organisaties die opkomen voor burgerlijke vrijheden, zijn over het algemeen veel meer te spreken over transparantieverplichtingen (zie bv. Pomerantsev 2019). Het Rathenau Instituut benadrukt wel dat transparantie veel facetten heeft, en dat bij regulering duidelijk moet zijn waar platformen zich precies voor moeten verantwoorden en wat eventuele toezichthouders dan gaan toetsen (Rathenau Instituut, 2021a).

Praktijk- en gedragscodes

In het reeds genoemde discussiestuk over de aanpak van *online harm* merkt het Britse Behavioural Insights Team op dat overheden, naast de drie besproken reguleringsopties, steeds vaker ook andere mogelijkheden onderzoeken om online schadelijk gedrag tegen te gaan (Costa & Halpern, 2019). Daarbij gaat het doorgaans om vrijblijvende afspraken die tot stand komen in samenwerking met bedrijven in de sector zelf. Eén type afspraken die overheden in de afgelopen jaren geïnitieerd hebben, zijn praktijk- en gedragscodes. Ze bestaan in allerlei varianten: nationaal en internationaal, met medewerking van verschillende soorten bedrijven, en toegespitst op verschillende fenomenen en typen content. We bespreken enkele voorbeelden.

Een bekende Europese gedragscode is die tegen online *hate speech*, die in 2016 gelanceerd werd.[10] De Europese Commissie sprak toen met Facebook, Microsoft, Twitter en YouTube af, dat deze bedrijven minstens de helft van alle meldingen van content die aanzet tot haat of geweld op basis van etniciteit, huidskleur, religie, nationaliteit of etnische afkomst, binnen 24 uur zouden opvolgen en eventuele onrechtmatige content verwijderen. De code vormde een aanvulling op een reeds bestaande Europese richtlijn voor racisme en vreemdelingenhaat.[11] Twee jaar later werd ook een praktijkcode ondertekend op het gebied van desinformatie (Stolton, 2020). Hier sloten niet alleen grote techbedrijven zich bij aan, maar ook brancheorganisaties van de reclamesector en adverteerders. Zij spraken samen af om meer transparantie te bieden over politieke advertenties, nepaccounts en het sluiten daarvan, de inzet van *fact checkers*, en het beter zichtbaar maken van getoetste informatie (AIV 2020).

De meeste nationale praktijk- en gedragscodes hebben specifiek betrekking op onrechtmatige content. Zo kent het Verenigd Koninkrijk, in afwachting van de invoering van de *Online Safety Bill*, een praktijkcode op het gebied van terrorisme en seksueel misbruik van kinderen (UK Government, 2020a). In Nederland heeft de internetbranche op verzoek van de overheid een gedragscode opgesteld die

10 Voluit: *EU Code of conduct on countering illegal hate speech online*.
11 Het *Kaderbesluit 2008/913/JBZ van de Raad van 28 november 2008 betreffende de bestrijding van bepaalde vormen en uitingen van racisme en vreemdelingenhaat door middel van het strafrecht*.

webhostingbedrijven ertoe aan moet zetten om na meldingen van kinderporno of ander verboden materiaal, dit materiaal (sneller) offline te halen (noticeandtakedowncode.nl/, 2018). Bij de meeste gedragscodes blijven de aan het problematische gedrag onderliggende mechanismen, zoals de werking van aanbevelingsalgoritmes, vooralsnog onbesproken.

Overheidsinitiatieven die een sterk beroep doen op de bereidheid van bedrijven tot zelfregulering, worden wisselend ontvangen. Veel commentatoren vinden dat bedrijven verantwoordelijkheid moeten nemen voor het gedrag dat ze faciliteren, maar trekken de effectiviteit en betrouwbaarheid van vrijwillige praktijk- en gedragscodes in twijfel. De Europese gedragscode tegen online haat, bijvoorbeeld, wordt ervaren als weinig effectief, onder meer door zijn vrijblijvende karakter (Rathenau Instituut, 2020b; Stolton, 2020). Bovendien bestaat ook bij gedragscodes de zorg dat platformbedrijven te veel beslissingsmacht naar zich toe trekken, ten koste van de vrijheden en grondrechten van gebruikers (Adviesraad Internationale Betrekkingen (AIV), 2020). Co-regulering, waarbij onafhankelijke toezichthouders de inspanningen van bedrijven beoordelen en in het slechtste geval dwangmaatregelen op kunnen leggen, worden daarom gezien als veelbelovender (Rathenau Instituut, 2020b).

De Adviesraad Internationale Vraagstukken (AIV) constateert dat landen bij het maken van reguleringskeuzes veelal ‘hun eigen weg’ kiezen, afhankelijk van heersende opvattingen op het gebied van nationale veiligheid, of commerciële en individuele vrijheden (Adviesraad Internationale Betrekkingen (AIV), 2020, p. 6). Het Nederlandse beleid is ‘van oudsher gericht op minimale regulering en het vrijlaten van de internetmarkt die grotendeels in private handen is’; traditioneel stuurt de overheid dan ook aan op ‘zelfregulering door de techsector’ (Adviesraad Internationale Betrekkingen (AIV), 2020). De raad vindt echter dat de tijd nu gekomen is voor een herijking van dit beleid; het Rathenau Instituut heeft dit advies onderschreven (Rathenau Instituut, 2021a).

Bedrijven publiekelijk vermanen

Overheden en uitvoeringsorganisaties kunnen bedrijven ook op meer informele wijze ertoe aanzetten om bij te dragen aan het terugdringen van schadelijk en immoreel gedrag online. Dat kan op ‘positieve’ wijze, door ze te stimuleren om verantwoordelijk te handelen. Zo kunnen overheden bijvoorbeeld investeren in de ontwikkeling van technische middelen om schadelijk gedrag terug te dringen. Hier gaan we in de volgende paragraaf verder op in. Maar het kan ook op ‘negatieve’ wijze, door juist een gebrek aan actie aan de kaak te stellen.

Een beproefde strategie is het publiekelijk vermanen van bedrijven die onvoldoende doen om problematisch gedrag te ontmoedigen of relevante content te bestrijden.

Na de zelfmoord van een tiener die in verband werd gebracht met de beschikbaarheid van online beelden van zelfbeschadiging, bijvoorbeeld, vermaande de Britse staatssecretaris voor Volksgezondheid en Zorg, Facebook en Instagram dat zij onvoldoende deden om kinderen van dit soort materiaal af te schermen. De gedachte achter zo'n stap kan zijn dat bedrijven, uit vrees voor reputatieschade, meer gaan doen om het genoemde probleem aan te pakken. Idealiter stellen ze daarmee ook een voorbeeld aan andere bedrijven. In het Britse geval reageerde Instagram meteen schuldbewust en kondigde het platformbedrijf aan meer maatregelen te zullen nemen (Costa & Halpern, 2019).[12]

In Nederland zijn dergelijke voorbeelden schaars. Kamerleden stellen af en toe kritische vragen aan een minister, bijvoorbeeld over de rol van sociale media bij het verspreiden van desinformatie of over het gebrek aan transparantie vanwege bedrijven over de achterliggende mechanismen (Facebook, 2021)(Tweede Kamer, 2020). Bewindspersonen dreigen soms met publiekelijk vermanen. Zo dreigde demissionair minister Grapperhaus van Justitie en Veiligheid vorig jaar een lijst vrij te zullen geven van webhostingbedrijven die te weinig ondernemen tegen de verspreiding van kinderpornografisch materiaal, als zij zich niet aan de eerder genoemde gedragscode zouden houden (Houtekamer & Wassens, 2021).

5.2 Bedrijven

Ook buiten overheidssturing om spannen bedrijven zich al in om schadelijk en immoreel gedrag online te beperken of te voorkomen. Grote sociaalnetwerk-bedrijven lopen daarbij het meest in de kijker, omdat zij gezien hun invloed te maken hebben met maatschappelijke en politieke druk om in te grijpen. Maar ook andere, kleinere partijen leveren al inspanningen. We bespreken hier initiatieven vanuit drie hoeken: platformen die online interacties faciliteren, adverteerders en hun tussenpersonen, en producenten of aanbieders van innovatieve producten of diensten die bijdragen aan de reductie van immoreel en schadelijk gedrag online. Bij die laatste categorie maken we nog het onderscheid tussen producten en diensten gericht op bedrijven (bijvoorbeeld webhosters) en op particulieren.

Initiatieven vanuit platformbedrijven

Er zijn grofweg drie typen initiatieven die platformbedrijven nemen om problematisch gedrag online te beperken, dan wel te voorkomen. Ten eerste proberen ze voorkeursgedrag te stimuleren door 'spelregels' te formuleren voor deelname aan online interacties, bijvoorbeeld in de vorm van gebruiksvoorwaarden

[12] *Targeted exhortation* hanteert dus eigenlijk hetzelfde principe als het fenomeen van *shaming*, maar zet het in voor een ander doel. Daarbij maakt het gebruik van de gevoeligheid van bedrijven voor de kracht van het mechanisme van online beeldvorming.

of gedragscodes. Ten tweede nemen ze maatregelen om immoreel en schadelijk gedrag te onderdrukken, bijvoorbeeld door het verwijderen of blokkeren van bepaalde content of door plegers te sanctioneren. En ten derde maken ze aanpassingen in hun platformontwerp, zodat gebruikers beter beschermd worden tegen schade door online gedrag.

Gebruiksvoorwaarden en gedragscodes

Gebruiksvoorwaarden (*terms of use*) bevatten regels die gebruikers geacht worden na te leven als ze zich op een platform willen begeven. Doorgaans betreffen ze het type content dat geplaatst mag worden en/of de omgang met andere gebruikers. De meeste platformen verbieden in elk geval het plaatsen van onrechtmatige content – ook de kanalen die juist liefhebbers van ‘extreme’ uitingen aantrekken. Welk materiaal of gedrag daarnaast ook nog in de voorwaarden opgenomen worden, en of het gedrag ontraden dan wel verboden wordt, verschilt echter per platform (zie Tabel 3). Ook zijn er grote verschillen in de manier waarop potentieel schadelijke content of handelingen omschreven worden (bijvoorbeeld in de mate van concreetheid of detail). Ook de taal of toon van de verschillende gedragscodes verschilt sterk, afhankelijk van het beoogde publiek.

In reactie op maatschappelijke of politieke druk hebben grote sociaalnetwerksites in de afgelopen jaren hun gebruiksvoorwaarden aangescherpt; recentelijk bijvoorbeeld in het kader van controverses rond desinformatie (Tumber & Waisbord, 2021). Doorgaans stellen ze daarbij zelf de kaders voor wat gebruikers al dan niet mogen doen; gebruikers stemmen daar *de facto* mee in. Dat doen ze hetzij expliciet (door bij hun eerste gebruik van de dienst de gebruiksvoorwaarden en bijbehorende definities te accorderen) hetzij meer impliciet (bijvoorbeeld na een wijziging van de voorwaarden). Gebruiksvoorwaarden zijn leidend voor het beleid van platformen ten aanzien van blokkering of verwijdering van content.

Gebruiksvoorwaarden en gedragscodes worden doorgaans geëvalueerd als weinig effectieve middelen om online gedrag in goede banen te leiden. Een bekend probleem is dat gebruikers ze meestal niet lezen en zonder meer accepteren. En daar is reden toe, want meestal zijn dit soort teksten weinig gebruiksvriendelijk (Costa & Halpern, 2019; UK Government, 2019). Bovendien zijn er terugkerende zwakheden. Veel platformen hebben bijvoorbeeld nog geen robuust beleid rond zelfbeschadiging en zelfdoding – terwijl internetplatformen juist de plek zijn waar mensen in nood gaan zoeken naar informatie of hulp (Newton, 2021b).

Op de volgende pagina’s:
Tabel 3 Gebruiksvoorwaarden van platformen

Voorbeeld	Karakterisering	Verbiedt…	Beperkt…	Ontraadt…
Facebook (Facebook, 2021)	Sociaal netwerk Breed publiek Meer dan 2,8 miljard maandelijks actieve gebruikers.	- Illegale content/gedrag - Potentieel schadelijke content (bv. hate speech, gewelddadige content, seksuele content) en het faciliteren, organiseren of promoten van schadelijk gedrag		
Twitter (Twitter, 2021)	Microbloggingsite Breed publiek Meer dan 330 miljoen maandelijks actieve gebruikers	- Illegale content/gedrag - Potentieel schadelijke content en het promoten van schadelijk gedrag (bv. geweld en pesterijen; zelfmoord en zelfbeschadiging) en ‘gevoelige’ content (bv. gruwelijk of seksueel getint materiaal)		
4Chan (4Chan, 2021)	Discussieforum (vooral reacties op beeldmateriaal) Jonge gebruikers, doorgaans anoniem Meer dan 20 miljoen maandelijks actieve gebruikers	- Illegale content/gedrag - Het vragen naar of verspreiden van persoonlijke informatie (doxing) of oproepen tot aanvallen (raids) - Klagen over 4Chan	- Bepaalde potentieel schadelijke content is alleen toegelaten op specifieke kanalen (bv. troll posts, racistische uitlatingen, pornografische content)	- Spammen of het doen van onbegrijpelijke uitingen - Andere gebruikers aanvallen (ook verbaal)

Parler (Parler, 2021)	Microbloggingsite Appelleert aan liefhebbers van vrijheid van meningsuiting; staat bekend om zijn leden met rechtse en extreemrechtse denkbeelden. Bereik varieert sterk (recentelijk enkele miljoenen actieve gebruikers per maand)	- Illegale content en ermee dreigen van illegale daden te begaan	- Bepaalde content moet gelabeld worden: potentieel schadelijk (bv. gewelddadig) of 'gevoelig' materiaal (bv. met naaktheid)	- Spammen
Reddit (Reddit, 2021a)	Sociaal netwerk en discussieplatform met 430 miljoen maandelijkse actieve gebruikers. Bestaat uit meer dan 100.000 actieve 'subreddits' gericht op specifieke interesses, van politiek tot koffieliefhebbers tot hardlopers. Subreddits kunnen eigen aanvullende regels stellen en hebben eigen moderatoren	- Illegale content/gedrag - Schadelijke content (pesten, geweld, haatzaaien en discriminatie) - Informatiemanipulatie (waaronder beïnvloeden van het verkiezingsproces) - Doxing, wraakporno - Sockpuppeting (maar anoniem zijn mag)	- Seksueel expliciete content moet gelabeld worden en verschijnt niet in de algemene tijdlijn van Reddit waarin populaire berichten te zien zijn	Extreme subreddits die schadelijk kunnen zijn worden door Reddit soms 'in quarantaine' geplaatst. Ze zijn dan moeilijk vindbaar en niet te zien voor mensen zonder account. Voorbeelden zijn complotten over 9/11 en pro-ana subreddits.
TikTok (Tik Tok, 2020)	Sociaal netwerk gericht op het delen van korte video's met 1,7 miljoen gebruikers in Nederland. Vooral populair onder kinderen en jongeren	- Illegale content - Gewelddadig extremisme, haatzaaien, zelfmoord, zelfbeschadiging en gevaarlijk gedrag - Lastigvallen en pesten - Naaktheid in elke vorm - Grooming, kindermishandeling - Misinformatie en sock puppeting		

Labelen, verwijderen, blokkeren

Een meer repressieve vorm van interventie is contentmoderatie: het labelen, en soms verwijderen, van onrechtmatig of anderszins schadelijk materiaal. Sommige platformen, zoals Reddit, zetten daartoe gebruikers in die op vrijwillige basis toezien op de naleving van gebruiksvoorwaarden (Reddit, 2021b). Andere, zoals Facebook of Twitter, huren professionele moderatoren in of gebruiken gespecialiseerde technologie om content op geautomatiseerde wijze te monitoren (Facebook, 2019). Met name grote platformen doen daarbij een beroep op *fact checkers* ter bestrijding van desinformatie. Recente gebeurtenissen zoals de coronapandemie hebben een impuls gegeven aan deze praktijk. Facebook en YouTube, bijvoorbeeld, hebben de afgelopen maanden miljoenen onbetrouwbare berichten gelabeld of verwijderd (Griffin, 2021; Wagner, 2020). Tegenwoordig worden er ook technische middelen ingezet om het *fact checken* deels te automatiseren (Rathenau Instituut, 2020a).

Contentmoderatie gecombineerd met *fact checking* heeft al enige bemoedigende resultaten opgeleverd. Uit eigen onderzoek van Facebook blijkt bijvoorbeeld dat bezoekers bij het zien van waarschuwingslabels bij onbetrouwbare berichtgeving over corona, in 95% van de gevallen niet doorklikten naar de originele content (Zuckerberg, 2020). Het probleem is echter dat moderatie lastig op te schalen is als daarvoor mensen ingezet worden. Algoritmische detectie en andere technische hulpmiddelen zijn dan weer minder betrouwbaar, en kunnen vooroordelen bij gebruikers reproduceren of zelfs versterken. Sowieso zijn de criteria die bij moderatie gehanteerd worden, sterk contextspecifiek. Wat een Amerikaans bedrijf schadelijk vindt is dat niet per se elders in de wereld, en vice versa. Er is dus eigenlijk kennis nodig van lokale cultuur om goed aan contentmoderatie te kunnen doen. Onderzoekers wijzen er ook nog op dat ook traditionele media – geschreven pers en tv – een rol spelen bij het stimuleren van problematisch gedrag door er ruchtbaarheid aan te geven (Kaiser et al., 2020). Moderatie moet dus altijd gecombineerd worden met andere maatregelen.

In extreme gevallen kunnen platformen gebruikers ook sanctioneren. Zo kunnen ze gebruikers schorsen door hun account op te heffen (*deplatforming*) of door de toegang vanaf een bepaald IP-adres te blokkeren (*blacklisting*). Google weert daarnaast ernstige overtreders van zijn gebruiksvoorwaarden uit zijn advertentienetwerk, bijvoorbeeld websites waar complottheorieën rondgaan. (Kist, 2020). Om plegers te identificeren, werken bedrijven soms samen met ontwikkelaars van specialistische technologie. Crisp, bijvoorbeeld, is een Amerikaans bedrijf dat met behulp van kunstmatige intelligentie tracht te achterhalen wat de relaties zijn tussen verschillende platformgebruikers. Op basis daarvan schat het in welke contacten schadelijk zouden kunnen zijn (Crisp

Thinking, 2021). Platformen zetten dit soort software in bij de strijd tegen online kindermisbruik (UK Government, 2019). Hier gaan we in paragraaf 5.2 nader op in.

Een zorg die zowel speelt bij het verwijderen van content als bij het blokkeren van gebruikers, is dat platformen bij dit soort ingrepen veel beslissingsmacht naar zich toe trekken. De gebruiksvoorwaarden die als uitgangspunt dienen bij dergelijke besluiten, hebben maar zelden een basis in (inter)nationale wetgeving. Bovendien hanteren platformen onduidelijke definities en hebben gebruikers niet altijd de mogelijkheid om bezwaar te maken tegen verwijderingsbesluiten (Adviesraad Internationale Betrekkingen (AIV), 2020). Ook in dit opzicht raken platformen dus aan de rechten en vrijheden van burgers. Critici vinden dat ze in de praktijk te vaak op de stoel gaan zitten van een rechter (bv. Bureau Clara Wichmann, 2020).

Een manier waarop platformen deze risico's enigszins kunnen beperken, is door samen te werken met maatschappelijke organisaties bij het bepalen van beleid, of door daartoe adviesraden of toezichthouders op te zetten. Een bekend voorbeeld is het door Facebook ingestelde en gefinancierde Oversight Board. De toezichtsraad moet waken over de rechten van gebruikers, en moet er in het bijzonder op toezien dat Facebook en Instagram hun vrijheid van meningsuiting vrijwaren. De raad houdt zich enerzijds bezig met het beoordelen van bezwaarschriften tegen beslissingen over content, bijvoorbeeld verwijderingsbesluiten. Als ze daarbij tegen een beslissing van Facebook of Instagram ingaat, is haar oordeel bindend. Anderzijds heeft de Board een adviserende rol en draagt de raad aanbevelingen aan voor (aanpassingen aan) het contentbeleid van beide platformen.

Commentatoren reageren gemengd op het functioneren van dergelijke toezichtsraden. Enerzijds is het een goede zaak dat platformen, die soms tegengestelde belangen moeten dienen (bijvoorbeeld die van adverteerders en gebruikers, of van een autoritaire overheid en haar burgers), bepaalde beslissingen uit handen geven. Anderzijds stuiten concrete besluiten ook op kritiek. Een voorbeeld is de beslissing van Facebooks Oversight Board om de verwijdering van (toenmalig) president Trump van het platform, te bekrachtigen (Paul, 2021). Voor onderzoekers op het gebied van media en governance tonen dit soort gevallen vooral aan dat de bedrijven achter sociale media vooralsnog onvoldoende sterk gereguleerd worden (bv. MacCarthy, 2021). Anderen benadrukken dat advies- en toezichtsraden alleen goed kunnen functioneren als ze transparant zijn over wat er met hun input gebeurt. In de praktijk blijkt dat nog niet altijd het geval (bv. Helberger et al., 2018; Sánchez Montañés, 2021).

Tot slot roepen praktijken van *deplatforming* en *blacklisting* bij de experts die we consulteerden ook de zorg op dat problematisch gedrag zich verplaatst van de grotere platformen met strikte gedragscodes naar kleinere, alternatieve platformen

waar gebruikers meer vrijheden ervaren, maar zich ook onttrekken aan maatschappelijke controle. Gebruikers krijgen er nog minder tegengeluiden te horen voor de boodschappen die ze opzoeken of verspreiden. Dit kan mechanismen als syndicatie versterken, en zodoende, weer bijdragen aan het problematische gedrag. Daarnaast houdt het risico's in op maatschappelijke fragmentatie: extreme standpunten en gedragingen leven voort, maar buiten het zicht van (veel) anderen.

Platformontwerp

Een heel andere strategie waar platformen voor kunnen kiezen, is het maken van ontwerpkeuzes die immoreel of schadelijk gedrag of daaruit voortvloeiend slachtofferschap kunnen reduceren of helpen voorkomen. We noemen drie soorten initiatieven. Ten eerste het faciliteren van *content customization*, waarbij gebruikers zelf beslissingen kunnen nemen over wat voor materiaal ze te zien krijgen, hoe, en wanneer. Ten tweede, alternatieve (onder meer advertentievrije) verdienmodellen. En ten derde, diverse vormen van waardengedreven platformontwerp. Dit houdt in dat (nieuwe) platformen zo vorm krijgen dat mechanismen die schadelijk en immoreel gedrag bevorderen, juist worden afgeremd.

Content customization

Opties voor *content customization* houden bijvoorbeeld in dat gebruikers kunnen controleren hoe hun data verzameld of gedeeld worden, dat ze meer zeggenschap hebben over de criteria voor het selecteren, rangschikken of presenteren van berichten op hun tijdlijnen, of dat ze de hoeveelheid of het type reclame dat ze te zien krijgen, zelf kunnen bepalen (Costa & Halpern, 2019). De veronderstelling is dat dit soort keuzes gebruikers kunnen helpen zichzelf te wapenen tegen immoreel of schadelijk gedrag, of dat ze achterliggende mechanismen, zoals selectiviteit en amplificatie, kunnen helpen doorbreken.

Grote platformbedrijven experimenteren vooralsnog maar weinig met dit soort oplossingen. Dat is ook niet vreemd, want zij opereren binnen een speelveld waarin zowel hun gebruikers als zijzelf financieel of anderszins baat kunnen hebben bij het genereren van (veel) aandacht (zie hoofdstuk 4). Toch komen de initiatieven nu stilaan van de grond – deels wellicht onder druk van de publieke opinie. Zo wil Facebook zijn gebruikers meer inspraak geven in de wijze waarop berichten in de News Feed gerangschikt worden (Benton, 2021; Newton, 2021a). Twitter denkt zelfs aan een soort app store voor algoritmen, waarin gebruikers voor meerdere sociale netwerken kunnen aangeven welk gewicht ze willen geven aan specifieke ordeningsprincipes (Kastrenakes, 2021). In alle gevallen is het echter nog de vraag, of gebruikers genoeg kennis zullen hebben om dit soort keuzes op geïnformeerde wijze te maken (Newton, 2021a).

Commentatoren voorzien dat dit soort ingrepen hoe dan ook een uitzondering zullen blijven, als bedrijven niet ook onderworpen worden aan strengere regulering. Overheden zouden ook kunnen eisen dat platformen meer mogelijkheden voor *content customization* invoeren, of dat ze reeds bestaande functies zichtbaarder maken (Costa & Halpern, 2019). Daarnaast zouden ze actiever bij kunnen dragen aan een gunstig ontwikkelingsklimaat voor nieuwe intermediairs, zoals ontwikkelaars van innovatieve diensten (ibid.). Dat kunnen bedrijven zijn, maar ook organisaties zonder winstoogmerk. Okuna, bijvoorbeeld, is een in Nederland opgestart alternatief sociaal netwerk dat de bewegingen van gebruikers niet monitort, maar het geheel aan hen overlaat om te bepalen wat ze te zien krijgen, en deels ook hoe (Okuna, 2021). Dergelijke producten maken het mogelijk om zonder de tussenkomst van regulering, gebruikers toch meer controle te geven op hun online sociale ervaring.

Alternatieve verdienmodellen

Okuna is een gecrowdfund initiatief, en hanteert daarnaast op onderdelen een abonnementsmodel: gebruikers betalen voor extra functies (Okuna, 2021). Zo genereert het platform de inkomsten die nodig zijn om te kunnen opereren zonder afhankelijk te zijn van reclame-inkomsten. Ook dat kan bevorderlijk zijn voor de sociale wenselijkheid van de interacties die er plaatsvinden. De manier waarop de online advertentiemarkt nu ingericht is, past immers in de logica van de aandachtseconomie, die een voedingsbodem vormt voor schadelijk gedrag. Binnen een advertentievrij model is bereik juist een minder belangrijke factor. Bovendien appelleert een dergelijk model aan gebruikers die afkomen op kwaliteitscontent. Mensen die voor kwaliteitscontent betalen, blijken zich over het algemeen ook netter en beleefder te gedragen.

Ook grote technologiebedrijven beginnen langzamerhand heil te zien in abonnementsmodellen. Een voorbeeld daarvan is het initiatief van Facebook om *fan subscriptions* aan te bieden (Ha, 2020) of het plan van Apple voor een betaalde podcastservice (Kafka, 2021). Een verbod op microtargeting (een vorm van adverteren op basis van specifieke doelgroepen, waarbij data van individuele gebruikers ingezet worden) zou een stimulans kunnen bieden aan dit soort initiatieven. Er bestaat op dit moment in Europa een politieke discussie over een dergelijk verbod (zie bv. Vinocur, 2021). Alphabet (bedrijf achter Google) lijkt hier al op voor te sorteren. In maart 2021 kondigde het bedrijf aan dat het voortaan niet meer aan microtargeting wil gaan doen, maar dat het wil inzetten op het bereiken van cohorten (groepen ingedeeld op basis van hun klikgedrag) in plaats van individuen (Morozov, 2021). Of dit positief zal uitpakken voor privacy en het ontstaan van *rabbit holes* en *echo chambers* wordt echter betwijfeld (Newton, 2021c).

Het nadeel van advertentievrije verdienmodellen is echter dat ze, door de financiële barrières die ze opwerpen, ook bepaalde groepen uitsluiten (Chen & Thorson, 2021; Van den Berg, 2021). Daarmee dreigen ze online omgevingen waar minder schadelijk gedrag voorkomt, tot iets 'exclusiefs' te maken. Slechts een minderheid van huishoudens heeft geld voor meer dan één abonnement (Reuters, 2020) en dan krijgt Netflix vaak voorrang boven een krant.

Waardengedreven ontwerp

De keuze voor waardengedreven ontwerp (*value sensitive design*) houdt in dat publieke waarden leidend zijn bij de ontwikkeling van online omgevingen. Vaak zijn bestaande grond- of mensenrechten, zoals het recht op privacy of veiligheid, daarbij het uitgangspunt. Bedrijven kunnen dit soort waarden bijvoorbeeld implementeren door bij de bouw van een platform of systeem een zogenaamd *human rights impact assessment* uit te voeren (Adviesraad Internationale Betrekkingen (AIV), 2020). Dat houdt in dat ze al in de ontwikkelingsfase de potentieel nadelige gevolgen van het project identificeren en vervolgens adresseren bij de uitwerking ervan (Danish Institute for Human Rights, The, 2020).

Zo zouden makers bij het ontwerp van een nieuwe microbloggingdienst zich kunnen afvragen wat het voor de sociale veiligheid van gebruikers betekent, als iemand eenvoudig een uitspraak over een andere gebruiker aan diens online profiel kan koppelen (Twitters @mention). Enerzijds geeft deze functie aanleiding tot levendige gedachtenuitwisselingen tussen gebruikers; anderzijds faciliteert het ook praktijken als shaming, of de escalatie van online haat. Zijn er ontwerpkeuzes te maken die hogere barrières opwerpen voor het ontstaan van dit soort fenomenen?

Ook de aan online schade onderliggende mechanismen kunnen inspiratie bieden voor waardengedreven platformontwerp. Zo zijn er al websites en apps die gebruikers stimuleren om berichten te lezen waar ze zelf niet zo gauw naar zouden zoeken, om te voorkomen dat ze alleen in aanraking komen met gelijkgestemden (Costa & Halpern, 2019). Daarbij zetten ze analyses van gebruikspatronen dus heel anders in dan mainstreamplatformen dat doen. Daarnaast zijn er initiatieven om online sociale netwerken in te bedden in bestaande, locatiegebonden gemeenschappen. Gebiedonline bijvoorbeeld, is een website die abonnees verbindt met mensen en bedrijven in hun buurt (Gebiedonline, 2021). De gedachte hierachter is dat dit helpt om normen en waarden uit de fysieke wereld te bestendigen in online interacties. Ze vormen daarmee een tegengewicht voor de ontmenselijking en de schijnbare 'regelloosheid' van de online omgeving.

Veel platformen die waardengedreven zijn ontworpen, zijn kleinschalig. Enerzijds is dat een sterkte: de schaal van grote platformen – die leidt tot hyperconnectiviteit – draagt immers bij aan veel mechanismen achter schadelijk en immoreel gedrag.

Maar anderzijds is de kleinschaligheid een zwakte, want zo vormen ze geen volwaardig alternatief voor de dominante netwerken waar gebruikers vaak al vertoeven, en waar ze veel van hun bekenden kunnen treffen. Ook daarom, vinden veel commentatoren, is sterkere regulering van platformen noodzakelijk. Eén stap die in dit verband kan helpen is het verleggen van data-eigenaarschap naar gebruikers (bv. Döpfner, 2021) en het afdwingen van interoperabiliteit, datastandaarden en dataportabiliteit (Costa & Halpern, 2019). Als mensen hun data mee kunnen nemen en in contact kunnen blijven met gebruikers op platformen waar ze zelf geen gebruik van maken, wordt het immers makkelijker om over te stappen op een andere dienst, en zo de groei van alternatieve platformen te stimuleren. Het Rathenau Instituut adviseerde in 2021 het Nederlandse parlement om in de beleidsdiscussie omtrent de Digital Services Act ook oog te hebben voor aanvullende maatregelen voor het reguleren van poortwachters, omdat interoperabiliteit niet alle netwerkeffecten wegneemt die ervoor zorgen dat nieuwe spelers snel naar een dominante positie kunnen bewegen (Rathenau Instituut, 2021b).

Initiatieven vanuit adverteerders (tussenpartijen)

Behalve online platformen kunnen ook bedrijven die hun waren of diensten op deze platformen adverteren, een bijdrage leveren aan het beperken van schadelijk of immoreel gedrag online. We bespreken hier twee typen interventies. Geen van deze opties heeft het terugdringen van schadelijk en immoreel gedrag als *hoofddoel*; in alle gevallen is het alleen een bijeffect van de gekozen advertentiestrategie. Maar gezien de rol die reclame speelt in het bestendigen van onderliggende mechanismen (zie hoofdstuk 4), zijn ze toch het noemen waard.

Blacklisting en whitelisting

Online gedrag kan schadelijke gevolgen hebben voor de gebruikers van online omgevingen, maar bij uitbreiding kan het ook vervelend zijn voor bedrijven die er adverteren (en die zo bijdragen aan het verdienmodel van platformen). Bedrijven die zich bekommeren om hun publieke imago (oftewel hun *brand safety*), willen namelijk niet geassocieerd worden met problematische content, bijvoorbeeld omdat hun advertenties ernaast komen te staan. Om ervoor te zorgen dat dit niet gebeurt, kunnen ze een tussenpartij inschakelen die werkt met *whitelists* en *blacklists*: lijsten van ‘veilige’ en ‘schadelijke’ content. Om tot die lijsten te komen, maken bedrijven bijvoorbeeld gebruik van transcripten van de audio in online video’s. Deze transcripten worden gescand op problematische termen (bijvoorbeeld scheldwoorden, of woorden die in verband gebracht kunnen worden met zedendelicten). Als ze weinig voorkomen, komen de filmpjes, het kanaal waar ze aangeboden worden, of hun makers op een *whitelist* te staan.

Whitelisting en *blacklisting* kunnen indirect de productie en het delen van niet-schadelijke content stimuleren, en zodoende bijdragen aan het voorkomen van online schadelijk gedrag. Makers van filmpjes of andere content (*content creators*) die op een *blacklist* terechtkomen, lopen namelijk het risico dat ze advertentie-inkomsten mislopen. Dit kan een reden zijn om meer content te produceren die *wel* aan de eisen voldoet. *Whitelisting* vormt juist een 'positieve' prikkel om dergelijke 'veilige' content te produceren of te uploaden.

Een kanttekening bij deze methode is dat ze adverteren duur maakt. Kleine en middelgrote bedrijven hebben vaak niet de middelen om er gebruik van te maken, en kiezen daarom voor kwantiteit in plaats van kwaliteit (dus dat hun advertenties een groot bereik hebben, in plaats van dat ze op 'goede' plekken terechtkomen). Bovendien zijn de algoritmen waar tussenpartijen gebruik van maken om transcripten te scannen op problematische termen, niet feilloos. Soms halen ze er termen uit die niet problematisch zijn, of zien ze juist problematische termen over het hoofd. Daar komt ook nog bij dat ongeveer de helft van de content die ten gevolge van blacklisten verwijderd wordt, vervolgens weer elders opduikt. In die zin vormen *whitelisting* en *blacklisting* dus maar een tijdelijke oplossing.

Contextual advertising

Veel platformen maken voor het plaatsen van advertenties gebruik van een geautomatiseerd systeem dat werkt op basis van gebruikersprofielen (*programmatic advertising*). De bewegingen van bezoekers worden bijgehouden (*tracking*) met behulp van *cookies*. Aan de hand van de gegevens die dit oplevert, koppelt het systeem hen aan een specifiek gebruikersprofiel. De partij die een advertentie wil plaatsen, doet een bieding op gebruikers met een bepaald profiel. Het systeem zorgt er vervolgens voor dat de gebruikers die passen bij dit profiel, de advertentie te zien krijgen. Men spreekt bij deze werkwijze ook wel van 'gepersonaliseerde advertenties', omdat het de data van individuele gebruikers zijn die bepalen waar advertenties komen te staan.

Een alternatief voor deze methode is *contextual advertising (*of *contextual targeting*). Hierbij wordt geen gebruik gemaakt van *cookies*, maar kiest de adverteerder ervoor zijn reclameboodschap te combineren met een bepaald type content (zoals tot nu toe gebruikelijk was in bijvoorbeeld papieren kranten). De veronderstelling daarbij is dat deze content een publiek aantrekt dat ook affiniteit heeft met het geadverteerde product. Een voorloper op het gebied van *contextual advertising* was de NPO, die in 2018 samen met STER begon te experimenteren met deze werkwijze (Ster, 2020). De reden voor het initiatief was dat de overgrote meerderheid van de bezoekers van de NPO-website, zelf aangaf geen cookies te willen delen. Ook de twee grootste krantenuitgevers van Nederland (Mediahuis en

DPG Media) zijn inmiddels van plan om te gaan adverteren op basis van inhoud (NLProfiel, 2020).

Platformen kunnen van *contextual advertising* gebruik maken om hun bezoekers meer privacy te geven, maar ook om advertentie-inkomsten terug te winnen die anders naar andere platformen gaan. Traditionele media die geverifieerde content bieden, zijn immers met nieuwere platformen in een concurrentiestrijd verwikkeld om adverteerders. Door een andere reclamestrategie te hanteren, mikken ze ook op andere klandizie: bedrijven die zelf willen kunnen bepalen naast wat voor content hun advertenties staan. Als platformen deze strategie inzetten, kan dit een gunstig effect hebben op de kwaliteit van de aangeboden content. Bovendien kan deze vorm van adverteren ervoor zorgen dat reclame-inkomsten minder naar frauduleuze websites gaan, en vaker terechtkomen bij de makers van content.

Een nadeel van ‘cookieloos’ werken is dat er geen cijfermateriaal beschikbaar is over wat bezoekers precies op een website doen. Adverteerders vinden dat soms lastig, omdat ze graag inzicht willen krijgen in het bereik van hun advertenties. Bovendien zou *contextual advertising* vooral goed werken voor traditionele content van bijvoorbeeld kranten en omroepen. Hier is de aard van de content immers bekend – anders dan bij bijvoorbeeld sociale netwerken.

Producten en diensten voor online veiligheid

Een derde groep bedrijven die nu al een bijdrage leveren aan de bestrijding van schadelijk en immoreel gedrag online, zijn producenten en aanbieders van zogenaamde *online safety tech*: producten of diensten die gebruikers beschermen tegen (potentieel) problematische content, contact of gedrag (Department for Digital, Culture, Media and Sport, 2020). We maken een onderscheid tussen producten en diensten ontwikkeld voor bedrijven, en voor particulieren.

Producten en diensten voor bedrijven

Automatische detectie van schadelijke content

In paragraaf 5.2 noemden we al de mogelijkheid van automatische detectie van problematische berichten of beelden. Platformen dragen zelf bij aan de ontwikkeling van dergelijke technologie (Rathenau Instituut, 2020a; Sánchez Montañés, 2021), maar er zijn ook bedrijven die zich erin specialiseren. Ze bieden hun producten aan platformbeheerders aan, maar ook aan webhosters en adverteerders (Costa & Halpern, 2019; Department for Digital, Culture, Media and Sport, 2020).

Een techniek om op geautomatiseerde wijze problematische content op te sporen, is *hash-based* detectie. Ze maakt gebruik van zogenaamde ‘hashcodes’: datablokken in bijvoorbeeld beeldmateriaal, die beschouwd kunnen worden als een

soort unieke digitale ‘vingerafdrukken’. In Nederland wordt de techniek al ingezet voor de opsporing van kinderpornografisch materiaal. Zo stelt het Meldpunt Kinderporno een hashcheckserver beschikbaar waar bedrijven, bijvoorbeeld webhosters, zich gratis op kunnen aansluiten. Het Ministerie van Justitie en Veiligheid stimuleert hen middels de eerder genoemde gedragscode (zie paragraaf 5.1) om er gebruik van te maken, en bij een ‘hit’ problematische content te verwijderen (Tweede Kamer, 2018).

Identiteits- en leeftijdsverificatie
Een andere relevante specialisatie is de ontwikkeling van tools voor leeftijdsverificatie of identiteitsverificatie. Vooral hulpmiddelen die gebruik maken van *attribute-based identitiy management* zijn beloftevol. Die laten gebruikers toe aan verschillende identificatievereisten te voldoen, zonder daarbij onnodig veel privacygevoelige gegevens (zogenaamde ‘attributen’) prijs te geven. Een Nederlands voorbeeld is het identiteitsplatform IRMA. Gebruikers kunnen er een soort ‘digitaal paspoort’ mee aanmaken, waarmee ze in kunnen loggen in afgesloten online omgevingen.

Leeftijdsverificatietechnologie wordt daarnaast ook gebruikt voor het matchen van content met specifieke groepen van gebruikers. Het internationaal actieve technologiebedrijf SuperAwesome, bijvoorbeeld, helpt contenteigenaren, platformbeheerders en adverteerders ervoor te zorgen dat jeugdige gebruikers niet in aanraking komen met ‘ongepaste’ inhoud, zoals voor volwassenen bedoelde reclameboodschappen (Superawesome, 2021).

Gedragsanalyse en toegesneden berichtgeving
Bedrijven en organisaties ontwikkelen ook methoden voor de analyse van gedragingen van gebruikers. Die kunnen ingezet worden om kwetsbare groepen, bijvoorbeeld internetgebruikers met neiging tot gokverslaving of zelfbeschadiging, te identificeren, en vervolgens hun zoekresultaten aan te passen of banners te genereren die verwijzen naar specialistische hulp (bv. Costa & Halpern, 2019). De Redirect Method, ontwikkeld door onder anderen *tech start-up* Moonshot en Google-*incubator* Jigsaw, is hier een voorbeeld van. Het is een open-source-methode om aan de hand van een analyse van online zoektermen, personen te identificeren die op zoek zijn naar schadelijk materiaal. Zij krijgen dan gerichte advertenties toegestuurd met constructieve alternatieve berichten (Moonshot, 2021). Het Poolse Samurai Labs bouwde laatst zelfs een *reasoning machine*: een bot die kan interveniëren in online conversaties en voorkomen dat ze ontaarden in online haat en cyberpesten (Konopka, 2021).

Een andere manier om gezonde interactie te stimuleren, is door gebruikers af te remmen in hun neiging tot impulsief online gedrag, door ze op gezette tijden uit te

nodigen tot reflectie op hun eigen handelen. Zo wordt er software ontwikkeld die potentieel schadelijke uitingen automatisch detecteert, en nog voordat ze geplaatst worden, de publicatie ervan vertraagt. Het is een subtiele manier om gebruikers in elk geval de kans te geven zich tijdig op hun intenties te bezinnen (Costa & Halpern, 2019). Andere systemen zetten ook *prompts* of *reminders* in die de gebruikers expliciet tot reflectie aanmanen. Uit onderzoek blijkt dat dit een potentieel nuttige aanpak is voor schadelijk gedrag. Zo kunnen verzoeken aan gebruikers om de kwaliteit van berichten te evalueren, helpen voorkomen dat ze die achteloos doorsturen (Pennycook et al., 2021). Zodoende kunnen ze bijdragen aan het afremmen van online mechanismen zoals viraliteit.

Commentatoren benadrukken echter dat online *safety tech* pas een vlucht kan nemen, als er een goed ontwikkelingsklimaat is voor nieuwe intermediairs, zoals commerciële softwarebedrijven (Costa & Halpern, 2019). Overheidssturing is in dit verband cruciaal. Onderzoek wijst uit dat grote platformen vooralsnog weinig gebruik maken van technologieën voor leeftijdsverificatie en -validatie (Aiken, 2016) – hoewel de technologie stilaan wel voorhanden komt. Door platform- of hostingbedrijven te stimuleren of te dwingen er gebruik van te maken, of door zelf in de ontwikkeling ervan te investeren, kunnen overheden een impuls geven aan onderzoek en ontwikkeling (Helberger et al., 2018). Het Britse kabinet wil hierin een voortrekkersrol spelen, door parallel aan het werk aan de *Online Safety Bill*, de sector van de *online safety tech* te stimuleren en te ondersteunen (Department for Digital, Culture, Media and Sport, 2020; UK Government, 2020b).

Tegelijkertijd houdt het inschakelen van private ondernemingen voor het bestrijden van online gedrag of daaruit voortvloeiende schade ook gevaren in. Dat geldt met name bij de inzet van kunstmatige intelligentie, bijvoorbeeld om kwetsbare groepen te identificeren (zoals Instagram doet om de leeftijd van gebruikers vast te stellen (Instagram, 2021)) of hun zoekresultaten aan te passen. Onderzoekers wijzen erop dat dit risico's inhoudt, bijvoorbeeld op het gebied van privacy (Costa & Halpern, 2019). Bovendien zouden dergelijke praktijken ook in strijd kunnen zijn met de AVG, waarin een uitlegbaarheidsprincipe verankerd is. Gebruikers moeten dus in staat gesteld worden om te beoordelen hoe het algoritme 'redeneert'. Met name bij zelflerende algoritmen is het zelfs voor informatici echter niet altijd mogelijk om dit te achterhalen.

Producten en diensten voor particulieren

Een andere categorie van *online safety tech* zijn producten en diensten waarmee gebruikers zichzelf of anderen kunnen behoeden voor schadelijk of immoreel gedrag of de gevolgen daarvan. We onderscheiden hier twee typen: filters en *blocks*, en *self-exclusion tools*.

Filters en blocks
In het eerste geval valt bijvoorbeeld te denken aan filters die ouders op hun computer installeren om te voorkomen dat kinderen bepaalde content te zien krijgen. Bekende voorbeelden zijn het Nederlandse *Kliksafe* of het Amerikaanse *Net Nanny* (Kliksafe, 2021; Net Nanny, 2021). Vooralsnog blijken dit soort producten echter niet altijd effectief: er is zowel sprake van onder- als overblokkering (Oosterwijk & Fischer, 2017).

Self-exclusion tools
Experts zien ontwikkelingsmogelijkheden voor allerlei middelen voor zelfuitsluiting (*self-exclusion tools*) om de negatieve impact van online mechanismen zoals beschikbaarheid en continuïteit tegen te gaan. Voorbeelden zijn programma's die gebruikers in kunnen zetten om zichzelf te beschermen tegen verslaving op het internet (bijvoorbeeld *Pluckeye*, *Cold Turkey* of *LeechBlock*), of software die het gebruik van specifieke websites of typen netwerken blokkeren (altijd, of op bepaalde momenten van de dag). Ook bedrijven die bepaalde risicovolle handelingen faciliteren, bieden soms dit soort producten aan. Veel Britse banken, bijvoorbeeld, geven hun klanten de optie om transacties op hun rekening te blokkeren als het betaalverzoek afkomstig is van een website voor online gokken. Ze kunnen de blokkering dan pas ongedaan maken als er een bepaalde wachttijd verstreken is (Costa & Halpern, 2019).

Hoewel dit soort opties het risico op problematisch gedrag niet elimineert, zet het gebruikers wel aan tot reflectie op hun handelen. Belangrijk daarbij is dat de filters inzetbaar zijn op verschillende platformen tegelijk (Costa & Halpern, 2019). Bovendien moeten gebruikers er zelf voor kunnen kiezen. Een overmaat aan controle en toezicht kan immers de autonomie en persoonlijke levenssfeer van mensen schaden.

5.3 Hulpverleners, maatschappelijke organisaties, gebruikers

De strijd tegen schadelijk en immoreel gedrag online wordt behalve door overheden en bedrijven, ook gevoerd door hulpverleners, maatschappelijke organisaties en (collectieven van) burgers. Ter afsluiting van dit overzicht van bestaande interventies stippen we vier strategieën aan die daarbij gehanteerd worden: bewustzijn creëren, informeren en onderwijzen; participatief interveniëren in online interacties; hulpverlening aan slachtoffers; en participatieve vormen van waardengedreven platformdesign.

Bewustzijn creëren, informeren en onderwijzen

In de afgelopen jaren hebben diverse maatschappelijke organisaties en groepen van burgers campagne gevoerd voor het verbeteren van de kwaliteit van online interacties. Een aantal bekende Marokkaanse Nederlanders namen in februari 2021 bijvoorbeeld het initiatief voor een campagne tegen online shaming. Politici, acteurs en auteurs startten een petitie en spraken zich op sociale media op georganiseerde wijze onder de hashtag #StopShaming uit tegen de online cultuur van intimidatie (Redactie NOS, 2021). Burgerbeweging DeGoedeZaak lanceerde daarnaast een oproep tegen online haat, die vergezeld ging van een *toolkit* met inzichten en adviezen waarmee geïnteresseerden zich kunnen beschermen of zelf actie ondernemen (DeGoedeZaak, z.d.).

Commentatoren benadrukken het belang van dergelijke initiatieven, waarbij burgers zich publiekelijk over problematisch gedrag uitspreken. Alle gebruikers van online diensten hebben immers een aandeel in de interacties die op het internet plaatsvinden; het verloop ervan is dus ook hun gezamenlijke verantwoordelijkheid (Rasch, 2021).

In de categorie ‘bewustzijn creëren’ valt ook te denken aan initiatieven waarmee maatschappelijke organisaties platformbedrijven onder druk zetten om actie te ondernemen tegen online schadelijk gedrag. Ranking Digital Rights bijvoorbeeld, is een collectief van onderzoekers en activisten die zich inzetten voor online burgerrechten. Ze maakten onder meer een ranglijst van platformen – van Twitter tot Amazon – aan de hand van relevante indicatoren. Daarbij kwam onder meer naar boven dat er onder bedrijven nog maar weinig bereidheid is om openheid te geven over hoe ze gebruikersdata verzamelen en online interacties modereren, en over de algoritmen die ze daarbij hanteren (Brouillette, 2020). Middels onderzoek en publicaties brengt het collectief dit soort problemen onder de aandacht van de bedrijven zelf, maar ook van overheden (beleidsmakers) en investeerders.

Gespecialiseerde mediawijsheidsorganisaties leveren een meer structurele bijdrage aan het creëren van bewustzijn van de risico’s van online interacties. Netwerk Mediawijsheid bestaat inmiddels uit meer dan 1000 organisaties die in Nederland actief zijn (Netwerk Mediawijsheid, 2021). Een groot deel daarvan ontwikkelt projecten op het gebied van cyberveiligheid en cyberweerbaarheid, die doorgaans gericht zijn op kinderen of hun opvoeders (Bureau Jeugd en Media, 2021). Daarnaast zijn dergelijke organisaties met name actief op het gebied van nepnieuws en desinformatie, en omgangsvormen op sociale media. Bij projecten gericht op kinderen worden soms ook spelelementen ingezet. Een voorbeeld is de game *Slecht nieuws* (over nepnieuws), waarin de gebruiker gevraagd wordt in de huid te kruipen van een ‘boosdoener’, om zodoende te leren hoe valse berichtgeving tot stand komt en zich verspreidt (Slecht Nieuws, 2021). Een ander

interessant project, op het grensgebied van voorlichting, kunst en activisme, is *TheirTube*: een filterbubbelsimulator die gebruikers laat ervaren welke rol data en algoritmen spelen in de gebruikservaring op de bekende videostreamingdienst YouTube (Their Tube, 2021).

De behoefte aan inspanningen en investeringen op het gebied van mediawijsheid en digitale vaardigheden is groot – niet alleen onder kinderen (die zich op steeds vroegere leeftijd op het internet begeven, maar ook onder (kwetsbare) volwassenen (Aiken et al., 2016; Rathenau Instituut, 2020a). Bovendien gebeurt er op dit gebied nog te weinig onderzoek. Van veel programma's rond schadelijk gedrag is bijvoorbeeld niet duidelijk of ze wel goed werken (zie bv. Oosterwijk & Fischer, 2017). Maar er is ook nog weinig bekend over wat jongeren online doen en welke consequenties dit heeft voor hun ontwikkeling (Aiken, 2016).

Los daarvan benadrukken experts dat projecten op het gebied van mediawijsheid alleen kunnen slagen, als ze opgezet zijn met voldoende voeling met de (online) leefwereld van het beoogde publiek. Organisaties met relevante expertise zijn daarom vaak geschikter voor de uitvoering van dit soort projecten dan overheden of hun uitvoeringsorganisaties. Wel kunnen overheden deze organisaties stimuleren of (financieel) ondersteunen. Deskundigen adviseren daarnaast een constructieve – in plaats van repressieve – aanpak, omdat met name jongeren zich maar zelden bewust zijn van de gevolgen van hun online gedrag. Het gesprek over online normen en waarden zou daarbij volgens onze respondenten centraal moeten staan.

Tot slot wordt er veel belang gehecht aan het brede sociale netwerk van jongeren; behalve docenten kunnen ook anderen betrokken worden bij het creëren van bewustzijn, informeren en onderwijzen in de strijd tegen immoreel en schadelijk gedrag online (Oosterwijk & Fischer, 2017). Combinaties van strategieën worden eveneens gezien als bevorderlijk voor het welslagen van projecten. Een voorbeeld van zo'n integrale aanpak is een reeks initiatieven vanuit de gemeente Amsterdam ter bestrijding van problemen rond seksuele intimidatie en geweld. Daartoe wordt geld uitgetrokken voor onderzoek, en tegelijkertijd campagne gevoerd tegen shaming (#jijstaatnietalleen) en voorlichting gegeven op scholen (Wagemakers & Toksöz, 2021). Ook onderzoekt de gemeente of het mogelijk is om aan daders een 'online straatverbod' op te leggen (Katawazi & Wagemakers, 2021; Wagemakers & Toksöz, 2021).

Participatief interveniëren in online interacties

Behalve organisaties doen ook burgers al inspanningen om de kwaliteit van online interacties te verbeteren. In het Engels heet dit *technological placekeeping*: de praktijk van het actief 'onderhouden' van digitale ruimtes. Mensen kunnen zichzelf ermee beschermen tegen schadelijke fenomenen, maar ze kunnen er ook de

gezondheid van digitale conversaties mee bevorderen, en zodoende het vertoeven in online omgevingen voor iedereen aangenamer te maken (Wong, 2021).

Een recent Nederlands initiatief op dit gebied is de hashtag-campagne #DatMeenJeNiet van Movisie, een kennisinstituut dat zich buigt over sociale vraagstukken (Movisie, z.d.). Deelnemende jongeren beloven om zich in gevallen van online discriminatie niet als een *bystander* te gedragen (zie hoofdstuk 4), maar als een *upstander*: iemand die anderen aanspreekt op hun problematische gedrag. Een ander voorbeeld op het gebied van desinformatie en misinformatie is *Make Media Great Again*. Dit project stelt tools beschikbaar waarmee vrijwilligers annotaties kunnen maken in online artikelen en audiovisuele producties (Make Media Great Again, 2021). De annotaties dienen als suggesties aan de redacteur, die daarmee de kwaliteit van zijn berichtgeving kan verbeteren. NU.nl was de eerste mediapartner voor het initiatief. Ook de *virtual neighbourhood watch*, een vorm van online buurtpreventie waarbij technisch onderlegde internetgebruikers samenwerken met rechtshandhaving om kwetsbaarheden voor cybercriminaliteit in software te identificeren en beperken (zie bv. Oosterwijk & Fischer, 2017), zou gezien kunnen worden als een vorm van technologisch *placekeeping*.

Net als de eerdergenoemde campagnes vanuit burgers, stimuleren ook dit soort initiatieven de gebruikers van online omgevingen tot het nemen van eigen verantwoordelijkheid voor de kwaliteit van online interacties. Commentatoren benadrukken echter dat die verantwoordelijkheid altijd een gedeelde is: ook overheden en bedrijven moeten hun steentje bijdragen (Helberger et al., 2018). Met name bedrijven zouden burgers die zich actief inzetten om veilige digitale ruimtes te creëren, veel beter kunnen ondersteunen of belonen voor hun inspanningen (Wong, 2021).

Hulpverlening aan slachtoffers

Er is op dit moment nog weinig gespecialiseerde hulp voor slachtoffers van schadelijk of immoreel gedrag online. Alle kinderen en hun begeleiders kunnen terecht bij het portal Meldknop.nl, dat in 2012 gelanceerd werd op initiatief van het Meldpunt Kinderporno en Digibewust (een programma van het toenmalige Ministerie van Economische Zaken).[13] Ze kunnen er terecht als ze iets vervelends meemaken op het internet, zoals geweld, pesten, oplichting of seksuele intimidatie. Op de site is informatie en advies te vinden, in de vorm van uitleg en tips over 21 fenomenen in de categorieën pesten, seks, oplichting en lastig vallen. Daarnaast kunnen ze er direct per mail, chat of telefoon (en/of een app die ze kunnen downloaden) de hulp inroepen van deskundigen bij aangesloten organisaties, zoals

13 Sinds de lancering is het portal overgenomen door Veiliginternetten.nl, een gezamenlijk initiatief van het ministerie van Economische Zaken en Klimaat, het ministerie van Justitie en Veiligheid / het Nationaal Cyber Security Centrum, ECP | Platform voor de InformatieSamenleving, en het bedrijfsleven.

Helpwanted.nl, Vraaghetdepolitie.nl, het Meldpunt Internet Discriminatie (MiND) of Pestweb. Meldknop.nl krijgt jaarlijks ongeveer 50.000 bezoekers op de homepage en dit bleef de afgelopen jaren redelijk constant. De website heeft geen overzicht van de aantallen meldingen bij de organisaties waar de portal naar verwijst (Meldknop.nl, 2021).[14]

Slachtoffers van gedragingen die een 'offline' tegenhanger hebben, kunnen daarnaast soms terecht bij organisaties die zich toeleggen op het meeromvattende fenomeen. Iemand die last heeft van cyberpesten, kan bijvoorbeeld voor informatie terecht bij de Stichting Stop Pesten Nu. En iemand die worstelt met cyberverslaving, kan aankloppen bij een GGD of een andere organisatie die zich toelegt op verslavingsproblematiek.

Toch vinden de experts die we voor dit onderzoek spraken, dat slachtoffers van online gedrag anders geholpen moeten worden. Ze benadrukken bijvoorbeeld dat hulpverleners of anderen die steun bieden, ook online aanwezig moeten zijn. Het internet is namelijk vaak de plek waar het slachtofferschap in stand gehouden wordt. Mensen die gevoelig zijn voor een eetstoornis, bijvoorbeeld, kunnen er op zoek gaan naar hulp of inspiratie, maar ze kunnen er ook in aanraking komen met mensen die hun gevoeligheden juist uitbuiten (zie Casus verstoord eetgedrag). Om dit soort mechanismen te doorbreken, moeten slachtoffers ondersteund worden in zowel de on- als de offline wereld (zie hoofdstuk 6).

Daarnaast heeft hulp aan slachtoffers alleen kans van slagen, als de juiste mensen erbij betrokken worden. Net als in het geval van initiatieven op het gebied van mediawijsheid, wordt voeling met de leefwereld van slachtoffers daarbij cruciaal geacht. Experts pleiten er bijvoorbeeld voor dat organisaties intensiever gebruik maken van ervaringsdeskundigen: mensen die zelf ooit slachtoffer zijn geweest van schadelijk gedrag online, die vanuit begrip voor anderen hun beschermingsstechnieken of herstelstrategieën kunnen delen. Eveneens belangrijk is de opbouw van online netwerken en *safe havens* ter ondersteuning van slachtoffers, bijvoorbeeld bij online haat of pesterijen.

Ook hier weer, oordelen onze respondenten, kan de overheid beter bestaande initiatieven ondersteunen en helpen op te schalen, dan al te veel taken naar zich toe te trekken. Toch wijzen ze ook op het belang van goed (flankerend) beleid. Slachtofferschap online komt immers vaak voort uit 'offline' kwetsbaarheden, zoals een zwakke socio-economische positie.

14 Bron: e-mail woordvoerder Meldknop.nl bij ECP, op 23 juni 2021

Participatieve vormen van waardengedreven platformontwerp
Naast het actief 'onderhouden' van digitale ruimtes, kunnen burgers er direct of indirect ook zelf vorm aan geven. Dat kunnen ze doen door financieel of anderszins bij te dragen aan waardengedreven platformontwerp (zie paragraaf 5.2). We noemden al Okuna: een 'alternatief' sociaal netwerk dat geen gebruik maakt van advertenties, niet aan *tracking* doet, en de persoonlijke informatie van gebruikers niet te gelde maakt (Okuna, 2021). Behalve dat het platform gebruikers meer vrijheid geeft om zelf te bepalen wat ze zien (in plaats van spectaculaire, 'virale' content centraal te stellen) – betrekt het hen ook bij het ontwerpen van gedragsregels. Zo draagt het netwerk bij aan de betrokkenheid van gebruikers. Gebruikers worden ondersteund en gestimuleerd om (het sturen op) wenselijk gedrag te bevorderen. Een ander voorbeeld is het eerdergenoemde Nederlandse Gebiedonline: een coöperatieve online omgeving die verbinding tussen mensen probeert te creëren rond specifieke (offline) buurten of thema's, en gericht is op het bestendigen of versterken van sociale cohesie, waarbij on- en offline in elkaars verlengde liggen (Gebiedonline, 2021).

Onderzoekers op het gebied van platformdesign en publieke waarden vinden het belangrijk dat gebruikers hoe dan ook een grotere rol gaan spelen in het ontwerp van de online omgevingen waar ze hun tijd doorbrengen – ook als dat platformen zijn van grote technologiereuzen. Als duidelijk is dat een bedrijf beslissingen neemt die niet in het belang van de gebruikers zijn, of de gezondheid van online interacties dreigen te schaden, dan kunnen zij collectief druk uitoefenen op het achterliggende bedrijf. Bijvoorbeeld via sociale media of door problemen aan te kaarten bij een toezichthouder (Dijck et al., 2018; zie ook Helberger et al., 2018). Overheden kunnen gebruikers daarbij op weg helpen.

5.4 Conclusie

Het overzicht van bestaande initiatieven maakt duidelijk dat al verscheidene stappen genomen worden om schadelijk en immoreel gedrag online tegen te gaan of te voorkomen. Deze initiatieven bieden inspiratie voor de strategische agenda waar dit rapport naartoe werkt (hoofdstuk 6). Tegelijkertijd vertoont de huidige aanpak ook een aantal opvallende lacunes. De belangrijkste observatie die we in dit verband kunnen maken, is dat veel van de huidige initiatieven vrij *reactief* zijn. Ze zijn met name gericht op de bestrijding van symptomen van schadelijk en immoreel gedrag, en nauwelijks op de onderliggende mechanismen. Daarbij zien we wel verschillen tussen de diverse actoren. Met name overheden en platformbedrijven zijn vooralsnog weinig proactief bezig. Bij platformbedrijven is dat niet zo verwonderlijk. Sleutelen aan mechanismen betekent immers dat de keuze gemaakt moet worden voor een alternatieve vorm van platformontwerp. Maar dit

brengt onzekerheden mee ten aanzien van verdienmodellen – en bedrijven bewegen zich nu eenmaal binnen een bestaand speelveld. In de praktijk zijn het dan ook vooral andere, kleinschaliger partijen die met alternatieve vormen van ontwerp experimenteren.

Voor overheden geldt dat zij pas in actie komen als gedrag uit de hand loopt – en dus ingetoomd moet worden. Meer inzicht in de mechanismen onder fenomenen van online schadelijk gedrag, zoals we in dit rapport trachten te bieden, kan overheden en andere partijen helpen om pro-actiever op te treden. In hoofdstuk 6 doen we nog meer suggesties voor het versterken van de kennispositie van de overheid.

Het belang van online mechanismen wordt door de hulpverleners die wij spraken erkend, maar het blijkt nog lastig om hun aanpak hierop af te stemmen. De bestaande hulporganisaties hebben met name aandacht voor fenomenen waarvan ook ‘offline’ varianten bestaan, en die dus al een plek hebben in het zorglandschap. De professionals die er werken, blijken te worstelen met de mechanismen die onder de online variant van het gedrag liggen, zoals beschikbaarheid en continuïteit, syndicatie of schaalbaarheid.

Maatschappelijke organisaties en (collectieven van) bezorgde burgers zijn wat betreft hun aandacht voor specifieke fenomenen enigszins complementair aan overheden en hun uitvoeringsorganisaties. Overheden hebben met name aandacht voor informatiemanipulatie en online haat, en in wat mindere mate ook voor fenomenen op het gebied van zelfbeschadiging. Maatschappelijke organisaties ageren juist tegen allerlei vormen van pesterij en geweld of digitaal vigilantisme. In hun aanpak zijn ze ook meer gericht op de onderliggende mechanismen, zoals ontmenselijking of onduidelijke normen. Dat geldt ook voor sommige ondernemers op het gebied van *online safety tech*.

Een andere trend die uit ons overzicht naar voren komt, is dat hulpverleners en maatschappelijke organisaties in de praktijk een belangrijke signalerende functie vervullen. Vaak krijgen zij, vanuit hun kennis van de online wereld of hun expertise op een specifiek fenomeen, problemen die uit online gedrag voortvloeien eerder in het vizier dan overheden of bedrijven. Voor overheden en hun uitvoeringsorganisaties is dit reden om hulpverleners en maatschappelijke organisaties te koesteren en hun inspanningen te stimuleren.

In het volgende hoofdstuk formuleren we een strategische agenda voor de overheid in samenwerking met actoren in markt en maatschappij op basis van geleerde lessen en de gesignaleerde lacunes.

6 Strategische agenda

Dit onderzoek brengt voor het eerst schadelijk en immoreel gedrag online in Nederland in beeld in al zijn facetten. Het Rathenau Instituut ontwikkelde een taxonomie met zes categorieën van schadelijk en immoreel gedrag online met daaronder 22 verschillende fenomenen waar alle internetgebruikers in Nederland vroeg of laat mee te maken kunnen krijgen (hoofdstuk 3). Deze taxonomie is een momentopname; nieuwe fenomenen zullen blijven opduiken.

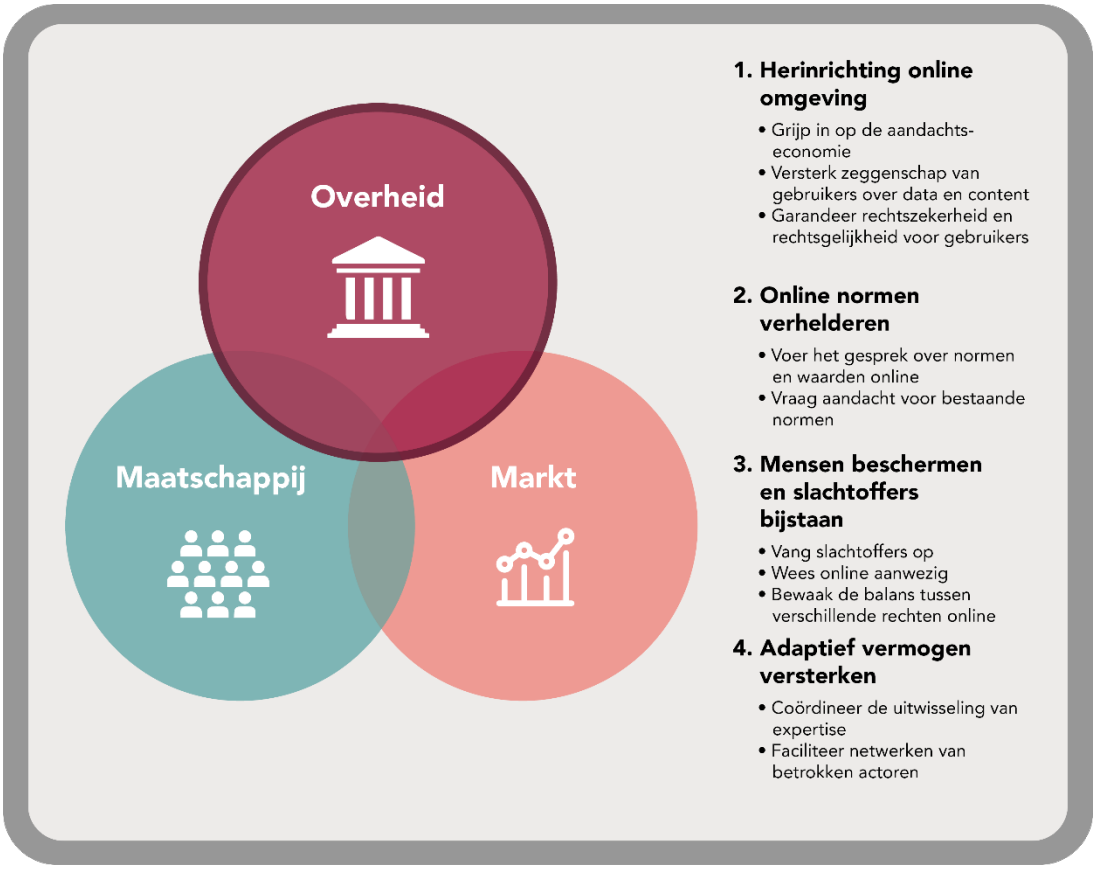
De schade die het gedrag veroorzaakt, kan ernstig zijn voor individuen, groepen en de samenleving als geheel. Deze kan variëren van een tienermeisje dat zichzelf uithongert omdat ze in een extreme challenge terecht komt met leeftijdsgenoten online (zie Casus verstoord eetgedrag), tot vrouwelijke journalisten en wetenschappers die zichzelf niet meer durven uitspreken (zie ‘haatzaaien’ in 3.4), of maatschappelijke ontwrichting door de verspreiding van complottheorieën en desinformatie (zie Casus desinformatie). Uit het beeld dat experts en literatuur schetsen over de omvang van dit soort fenomenen, blijkt dat Nederlanders online een niet te onderschatten risico lopen om slachtoffer te worden van schadelijk gedrag, of er zelf in af te glijden. Het garanderen van een veilige online omgeving is dus een grote maatschappelijke opgave waarin de overheid samen met andere actoren een belangrijke voortrekkersrol kan vervullen.

Naast inzicht in de aard en omvang van schadelijk en immoreel gedrag online, brengt dit onderzoek ook systematisch de kenmerken en mechanismen van het internet in kaart die een rol spelen bij het initiëren, faciliteren en versterken van schadelijk en immoreel gedrag online (hoofdstuk 4). Zo bespraken we anonimiteit, viraliteit, de aandachtseconomie en nog vijftien andere mechanismen die hierbij een rol spelen. Toch blijkt er nog nauwelijks actie ondernomen te worden om deze mechanismen ten goede te keren (of hun nadelige effecten te beperken), en volstaan bestaande interventies niet om online veiligheid te garanderen (hoofdstuk 5).

Het internet leek altijd een domein van zelfregulering en zelfredzaamheid van de samenleving, waar de overheid geen rol had en gebruikers zichzelf wel zouden redden. Uit ons onderzoek blijkt dat burgers op het internet onvoldoende beschermd zijn en dat daardoor grondrechten in het geding komen. Bedrijven, maatschappelijke organisaties en burgers hebben gecoördineerde collectieve actie nodig om schadelijk en immoreel gedrag online tegen te gaan. Ingrijpen in de mechanismen die op internet bijdragen aan het versterken van schadelijk en immoreel gedrag is nodig om deze schade te beperken. De overheid heeft de

verantwoordelijkheid om ook online de grondrechten van burgers te beschermen. Deze agenda biedt daar op strategisch niveau handvatten voor.

Op basis van interviews en gesprekken met deskundigen uit beleid, wetenschap en praktijk, vele wetenschappelijke, journalistieke bronnen en beleidsstukken, en de achtergrond en analyse van het Rathenau Instituut, identificeren we vier strategische thema's waarop de Rijksoverheid een sturende, coördinerende en faciliterende rol kan vervullen. Zo kan de Rijksoverheid in samenwerking met actoren uit de markt en maatschappij, schadelijk en immoreel gedrag online aanpakken en een veilige online omgeving bevorderen.



Bron: Rathenau Instituut

Figuur 6 Strategische agenda voor de aanpak van schadelijk en immoreel gedrag online

Het eerste thema – *Herinrichting online omgeving* – bevat handvatten voor de Rijksoverheid om de mechanismen die kenmerkend zijn voor het internet en die schadelijk en immoreel gedrag online mede veroorzaken, ten goede te keren. Het tweede thema – *Online normen verhelderen* – gaat in op de rol van de Rijksoverheid, markt en maatschappij bij het vernieuwen van de maatschappelijke

afspraken over normen en waarden online. De handelingsopties binnen dit thema beogen een bredere bewustwording en groter begrip van schadelijk en immoreel gedrag in de samenleving te bewerkstelligen. Het derde thema – *Mensen beschermen en slachtoffers bijstaan* – bevat suggesties voor de Rijksoverheid, handhaving en uitvoeringsorganisaties om beter te reageren op de fenomenen van schadelijk en immoreel gedrag online en de schade die de fenomenen veroorzaken. Het vierde thema – *Adaptief vermogen versterken* – bevat suggesties voor de Rijksoverheid om grip te krijgen en te houden op schadelijk en immoreel gedrag online dat continu in beweging is. Deze suggesties zijn gericht op het toekomstbestendig maken van deze vier strategische thema's op de actieagenda.

Onder elk van de vier thema's zijn een aantal opgaven geformuleerd. We geven daarbij telkens een aantal handelingsopties mee ter overweging, die uit het onderzoek naar voren kwamen.

6.1 Thema 1: herinrichting van de online omgeving

Er is een overheid nodig die beleid maakt voor de herinrichting van onze internetomgeving, waarin de mechanismen die schadelijk en immoreel gedrag mede veroorzaken worden aangepakt en waarmee schade voorkomen kan worden. Actoren in de markt en maatschappij blijken onvoldoende in staat om online ontsporingen tegen te gaan. Nationale en internationale overheden lijken de onderliggende mechanismen van online ontsporing op dit moment nog onvoldoende te adresseren. Het Ministerie van Justitie en Veiligheid zou dit inzicht kunnen aangrijpen om aan te dringen op actie op het niveau van mechanismen in de Digitaliseringstrategie van het aanstaande kabinet. Ook kan de Nederlandse Rijksoverheid de Europese wetgevingstrajecten van de Digital Services Act Package, de Artificial Intelligence Framework, de Data Act en de Data Governance Act aangrijpen als kans om deze mechanismen bij te sturen.

Het Rathenau Instituut stelde in haar Manifest (2020c) dat de overheid effectiever tegenwicht moet bieden aan de macht van de grote technologiebedrijven, die nu dominant zijn in de inrichting van de online omgeving. Tegelijkertijd zouden bedrijven meer verantwoordelijkheid moeten nemen om de rechten van hun gebruikers te beschermen. Op basis van de gesprekken met experts en deskundigen en de bestudering van vele bronnen, formuleren we voor de inrichting van de online omgeving de volgende drie opgaven voor de overheid:
1) grijp in op de aandachtseconomie;
2) versterk de zeggenschap van burgers over hun data en content; en
3) garandeer rechtszekerheid en rechtsgelijkheid voor gebruikers.

Grijp in op de aandachtseconomie

De aandachtseconomie die op internet is ontstaan, speelt een belangrijke rol bij het aanjagen van schadelijk en immoreel gedrag online. Gebruikers en online platformen zijn in competitie om aandacht van andere internetgebruikers, en content die verontwaardiging of schok oproept, is daar goed toe in staat. Die schadelijke en immorele content heeft weer een nauwe relatie met schadelijk en immoreel gedrag online; schadelijk gedrag leidt tot schadelijke content (denk aan haatzaaiende berichten of bedreiging en intimidatie door internetgebruikers aan andere gebruikers), en schadelijke content kan schadelijk gedrag uitlokken. Denk aan extreme challenges die jongeren aanzetten zichzelf iets aan te doen. Uit de literatuur en gesprekken met experts komt een aantal opties naar voren die de Rijksoverheid in dit kader kan overwegen:

Meer grip krijgen op het bereik van content

Het is belangrijk om onderscheid te maken tussen de regulering van de inhoud van content enerzijds en de verspreiding hiervan anderzijds (Rathenau Instituut, 2021a). Bij de verspreiding van content spelen de verdienmodellen en algoritmen van platformen een cruciale rol. De overheid zou kunnen proberen om meer grip te krijgen op het bereik van content. Dit kan door van techbedrijven meer transparantie te eisen over de werking van hun algoritmen die content selecteren voor gebruikers, zoals binnen de context van de DSA al besproken wordt. De algoritmen zijn bij socialemediaplatformen ontworpen om de betalende klant (vaak de adverteerder) goed te kunnen bedienen. Wanneer er meer inzicht bestaat in de werking van deze algoritmen door transparantievereisten op te leggen aan platformen, kan een onafhankelijke toezichthouder de werking van algoritmen monitoren en controleren. De Rijksoverheid kan de beleidsdiscussie over de DSA en het AI Framework aangrijpen om meer transparantie en toezicht op algoritmen te bewerkstelligen.

Marktmacht van poortwachters aanpakken

De overheid kan de marktmacht van poortwachters aanpakken om meer ruimte te geven aan nieuwe spelers die bijvoorbeeld een meer ethisch platformdesign aanbieden. Dit kan worden gedaan door mededingingsregels aan te scherpen. Politieke discussies hierover vinden al op Europees niveau plaats via het wetsvoorstel van de Digital Market Act (DMA). Een andere optie om de marktmacht aan te pakken, is via de verdienmodellen van de poortwachters. Zo is het mogelijk om de exploitatie van persoonlijke data en microtargeting te beperken of te verbieden op Europees niveau. Dit kan de keuze voor advertentievrije verdienmodellen stimuleren. Tot slot kan de overheid verkennen in hoeverre andere vormen van eigendom dan de beursnotering (zoals coöperaties van gebruikers) ervoor kunnen zorgen dat online platformen hun ‘nutsfunctie’ beter kunnen vervullen.

Digitale autonomie van publieke diensten vergroten
De Rijksoverheid kan ervoor zorgen dat de technische infrastructuur van publieke diensten zoals de media, onderwijs, en gezondheidszorg meer in publieke handen komt. Wat wij offline als de publieke ruimte beschouwen (denk aan: marktpleinen, scholen of wegen) is online in de private handen van grote Amerikaanse en Chinese technologiebedrijven. Nederland kan hierdoor niet zelf de regels bepalen. Dit kan worden tegengaan door te werken aan nationale of Europese digitale autonomie. Dit kan bijvoorbeeld gerealiseerd worden door strengere eisen in de inkoopvoorwaarden aan leveranciers van digitale producten en diensten te stellen, en door meer eigen internetbedrijvigheid in Nederland en Europa te creëren. Welke opties voor het vergroten van digitale autonomie realistisch zijn voor een klein land als Nederland zal verder onderzocht moeten worden, maar striktere voorwaarden voor digitale infrastructuur en onafhankelijk toezicht kunnen daar vast onderdeel van zijn. Binnen Europa wordt ‘strategische autonomie’ ook steeds belangrijker gevonden (Vanheste, 2021). De nationale overheid kan voor haar investering in binnenlandse technologie in Brussel aanspraak maken op onder meer het EU-fonds voor Recovery and Resilience, dat miljarden vrij maakt voor de digitale transitie.

Waardengedreven technisch ontwerp stimuleren
De overheid kan waardengedreven ontwerp van de digitale infrastructuur meer stimuleren. Mechanismen zoals schaalbaarheid, hyperconnectiviteit en viraliteit kunnen zo worden aangepakt. Voorbeelden van waardengedreven ontwerp die naar voren kwamen uit de literatuur en gesprekken met experts, zijn *content customization* en *value sensitive design*. Deze kunnen bijdragen aan online oplossingen voor de bescherming van slachtoffers. Nu zijn het vooral kleine bedrijven en platformen die dit aanbieden, maar die niet kunnen doorgroeien door de marktmacht van de grote ecosystemen. Overheden zouden hun groei kunnen stimuleren via subsidies voor nationale of Europese techbedrijven. Gebruikers zouden met meer waardengedreven technologie ook meer mogelijkheden kunnen krijgen om hun eigen online ruimte in te richten met een eigen selectie van aanbod van content. Online schade kan echter niet voorkomen worden met enkel technische middelen.

Andere manieren van adverteren bevorderen
De overheid kan online ontsporing helpen tegengaan door te stimuleren dat adverteerders bij het meten van hun bereik de focus verleggen van kwantiteit naar kwaliteit. Zo kunnen mechanismen als schaalbaarheid, viraliteit en de aandachtseconomie deels aangepakt worden. Adverteerders zouden de inhoud van content – in plaats van de omvang van bereik en data – meer centraal kunnen stellen. Voorbeelden hiervan die door experts in dit onderzoek zijn aangedragen, zijn *whitelisting* of *contextual advertising*. De overheid zou haar invloed kunnen

aanwenden om deze (nu nog) relatief dure vormen van adverteren meer toegankelijk en betaalbaar te maken voor kleine adverteerders. Toen grote adverteerders zich vorig jaar met een boycot publiekelijk verzetten tegen haatzaaien en racisme naar aanleiding van het overlijden van George Floyd, leek dit niet genoeg effect te hebben. Grote adverteerders vertegenwoordigen maar een klein deel van de omzet van sociale mediaplatformen.

Versterk zeggenschap van gebruikers over data en content

De overheid heeft verschillende mogelijkheden om de zeggenschap van burgers over hun eigen data en content te versterken. Gebruikers zitten nu vaak ingesloten (*locked-in*) in een bepaald platform, omdat ze daarbinnen hun eigen online netwerken met vrienden en familie hebben opgebouwd. Zij kunnen dit netwerk niet zomaar overhevelen of meenemen naar een andere online omgeving waar zij een nieuw account aanmaken. Dit beperkt hun autonomie en dus hun mogelijkheden om voor een online omgeving te kiezen die aansluit bij hun waarden en behoeften, waar online mechanismen zijn aangepast om sociaal wenselijk gedrag in de hand te werken. Ook hebben gebruikers relatief weinig invloed op het gebruik van hun persoonlijke data en de contentaanbevelingen die zij en andere personen in hun netwerk van platformen krijgen. De overheid kan de volgende opties verkennen om de autonomie en controle van gebruikers te versterken.

Dataportabiliteit en interoperabiliteit

De overheid kan gebruikers meer zeggenschap geven over data, en dataportabiliteit en interoperabiliteit stimuleren. Dataportabiliteit en interoperabiliteit maken deel uit van de beleidsdiscussie over het Europese wetsvoorstel van de Digital Markets Act. Zo kan de marktmacht van grote platformen worden ingeperkt, zodat meer ruimte ontstaat voor nieuwe aanbieders die gebruikers meer inzage geven in en zeggenschap geven over de exploitatie van hun persoonlijke data(profielen). Bij dataportabiliteit en interoperabiliteit is het echter altijd van belang om de bescherming van privacy te blijven garanderen.

Subsidies en aanbestedingen

De overheid zou via subsidies en aanbestedingen de ontwikkeling van ethische tools kunnen stimuleren, die het ontstaan of escaleren van problematisch gedrag kunnen voorkomen. Voorbeelden zijn tools voor leeftijdsindicatie, (upload)content-filters of detectiesoftware. Ook kan de overheid verdienmodellen met abonnementen of lidmaatschappen stimuleren. Gebruikers gedragen zich doorgaans netter in een omgeving met redactionele content die niet gratis is. Gebruikers zijn hier ook minder anoniem, omdat zij via betalingen te traceren zijn. Als gebruikers de belangrijkste inkomstenbron zijn van online platformen, komen

hun belangen automatisch centraler te staan dan die van adverteerders. Zo zal ook de veiligheid en het welzijn van de gebruiker meer prioriteit krijgen.

Garandeer rechtszekerheid en rechtsgelijkheid voor gebruikers

Om de beleving van wanorde in de online omgeving, gecreëerd door onduidelijke normen, anonimiteit en schijnbare wetteloosheid, aan te pakken, zou de rechtszekerheid en rechtsgelijkheid voor gebruikers online moeten verbeteren. Dit onderzoek heeft verschillende opties opgeleverd voor de Rijksoverheid om de rechtszekerheid en rechtsgelijkheid voor internetgebruikers te helpen versterken.

Voor- en nadelen van online identificatie verkennen

De overheid kan de voor- en nadelen van een vorm van online identificatie verkennen. Een online surfbewijs of toegangsbewijs zou preventief kunnen werken door online anonimiteit van daders weg te nemen. Online identificatie maakt hun opsporing, bestraffing of vergelding beter mogelijk en kan daarmee een afschrikkende werking hebben, maar vergroot tegelijkertijd ook risico's voor groepen die juist bescherming nodig hebben, zoals slachtoffers, journalisten of klokkenluiders. Een dergelijke maatregel verdient zorgvuldige afweging van de voor- en nadelen (zie ook thema 2).

Internationale afspraken online jurisdictie

De overheid kan aansturen op internationale afspraken over de handhaving van wet- en regelgeving op het internet. Op dit moment vallen nationale jurisdicties en het globale internet moeilijk te rijmen. Het is bijvoorbeeld moeilijk om een website met desinformatie of andere schadelijke content te verwijderen, als de organisatie hierachter niet op het Nederlandse grondgebied ligt. Internationale afspraken lijken noodzakelijk om dergelijke grensoverschrijdende problemen aan te pakken.

6.2 Thema 2: online normen verhelderen

Overheid, markt en maatschappij hebben alle een rol te vervullen en verantwoordelijkheid te nemen bij het verhelderen en bewaken van online normen. Een grote uitdaging bij het voorkomen en bestrijden van schadelijk en immoreel gedrag, is het gebrek aan duidelijke normen in de online omgeving. De mechanismen die we in hoofdstuk 4 beschrijven, creëren samen een vorm van morele mist. Die mist leidt ertoe, dat we gedrag dat we in de offline wereld onacceptabel vinden, online veel lastiger kunnen herkennen. Bovendien ontstaat in digitale omgevingen ook nieuw gedrag (denk aan grooming in virtual-reality-omgevingen). In dit soort gevallen moet het sociale proces waarin normen gevormd

worden, nog op gang komen, of moeten bestaande grenzen opnieuw onderhandeld worden voor de online omgeving.

Moraliteit is het resultaat van een sociaal contract. Noch de overheid, noch commerciële partijen zoals platformbedrijven hebben de legitimiteit om alleen te bepalen hoe burgers zich ten opzichte van elkaar zouden moeten gedragen. Tegelijkertijd hebben al deze actoren onafhankelijk van elkaar een verantwoordelijkheid om zich in het sociale verkeer sociaal wenselijk te gedragen, en niet de grenzen van de wet op te zoeken. Er is behoefte aan een grootschalig maatschappelijk debat om normen online te expliciteren en te verhelderen. Op basis van de gesprekken met experts en deskundigen en bestudering van de literatuur, formuleren we voor het verhelderen van online normen twee opgaven voor overheid, markt en maatschappij:
1) voer het gesprek over online normen en waarden; en
2) forceer aandacht voor bestaande normen.

Voer het gesprek over online normen en waarden

Het grootschalige debat waaraan behoefte is, ontstaat niet spontaan. De samenleving heeft ondersteuning en stimulans nodig om het gesprek over online normen te voeren. De overheid kan hier een faciliterende en stimulerende rol vervullen, maar ook bedrijven en maatschappelijke partijen kunnen ervoor zorgen dat allen die belang hebben bij de leefbaarheid van online omgevingen, bijdragen aan het gesprek. Het onderzoek leverde de volgende opties op, waarbij verschillende maatschappelijke actoren aan zet zijn.

Deliberatieve processen

Overheden hebben in het gesprek over online normen met name een faciliterende rol. Een optie is om zelf de dialoog te initiëren en te organiseren. Dit kan nuttig zijn bij de voorbereiding van nieuw beleid, bijvoorbeeld over anonimiteit op online platformen. Anonimiteit is een belangrijk mechanisme achter schadelijk en immoreel gedrag, omdat dit ontremming bij gebruikers in de hand werkt. Het wegnemen of inperken van anonimiteit zou dus preventief kunnen werken. Maar tegelijkertijd biedt de anonimiteit van de online omgeving ook bescherming aan slachtoffers, journalisten of klokkenluiders. Bovendien zijn er veel verschillende manieren om de anonimiteit van gebruikers in te perken (met behulp van persoonsgegevens of alleen identiteitskenmerken – door ze online te publiceren of alleen te gebruiken voor verificatie door platformen). De samenleving en politiek zullen zich dus eerst moeten buigen over de vraag: bij welke interacties vinden we openheid over iemands identiteit noodzakelijk? En onder welke voorwaarden vinden we (een mate van) anonimiteit acceptabel, of zelfs noodzakelijk?

Bottom-up gedragsregels
Platformbedrijven werken vaak met gedragsregels, bijvoorbeeld in de vorm van gebruiksvoorwaarden. Maar als ze gedragsregels zelf ontwerpen, trekken deze bedrijven veel macht naar zich toe bij het bepalen van sociale normen. Overheden zouden platformbedrijven kunnen stimuleren of verplichten om hun gebruikers veel actiever te betrekken bij het opstellen van gedragsregels. Zo gaan gebruikers ook meer ‘eigenaarschap’ ervaren van de geldende regels en zullen ze sterker geneigd zijn zich ernaar te gedragen. Om het gesprek over normen te stimuleren, kunnen platformbedrijven ook mogelijkheden creëren voor online deliberatie, bijvoorbeeld door daar goed zichtbare en makkelijk toegankelijke online ruimtes voor in te richten. Ook voor de bedrijven zelf kan het aantrekkelijk zijn, omdat ze hun diensten dan beter op gearticuleerde maatschappelijke wensen kunnen afstemmen.

In gesprek met jongeren
Het huidige mediawijsheidsonderwijs is er hoofdzakelijk op gericht kinderen en jongeren de vaardigheden bij te brengen die ze nodig hebben om zichzelf te beschermen tegen online schadelijk gedrag. Maar ze kunnen ook actiever betrokken worden in het gesprek over online moraliteit. Jonge mensen die nu opgroeien, maken een deel van hun socialisatie door in online omgevingen, waar autoriteitsfiguren minder sterk aanwezig zijn dan in de offline wereld (zie bv. Aiken, 2016; Cocking & van den Hoven, 2018). Begeleiders moeten dus het gesprek met hen aangaan over (on)wenselijk gedrag in dit soort omgevingen.

In dat gesprek is het aan te raden ook *victim blaming* – de schuld van het slachtofferschap wordt bij de gedupeerde gelegd in plaats van bij de dader – onderdeel van gesprek te maken. Slachtoffers van online schadelijk gedrag (en met name catfishing, grooming en shame-sexting) zijn vaak het voorwerp van victim blaming. Victim blaming is een duidelijk symptoom van de normatieve wanorde die in online omgevingen kan ontstaan. Internetgebruikers zouden namelijk niet bij voorbaat hun gedrag moeten aanpassen uit vrees dat ze door anderen veroordeeld worden, terwijl daders vaak vrijuit gaan. Bij slachtoffers leidt de ervaring van victim blaming vaak tot diepe schaamte. Begeleiders kunnen zich richten op het wegnemen van die schaamte, door met slachtoffers te praten over hoe normen in hun geval overschreden zijn.

Actoren die zich goed in de belevingswereld van jongeren kunnen inleveren, blijken het meest geschikt te zijn om dit gesprek te initiëren. Ook hier doet de overheid er dus het beste aan om de expertise te stimuleren die al bij uitvoerings- en maatschappelijke organisaties aanwezig is. Ze kan hen ook meer slagkracht geven door ruchtbaarheid te geven aan hun initiatieven of ze financieel te ondersteunen.

Vraag aandacht voor bestaande normen

Als een samenleving zich eenmaal achter een set normen heeft geschaard, moet die vervolgens ook aandacht krijgen – zowel online als offline. Dat gaat niet vanzelf. De overheid en haar uitvoeringsorganisaties hebben een rol bij het forceren van die aandacht, maar marktpartijen en maatschappelijke actoren evenzeer. Op dit vlak zien we nog veel potentie: veel méér partijen dan nu het geval is, zouden hun positie hiertoe kunnen gebruiken.

Slecht gedrag veroordelen, goed gedrag stimuleren

Een uitvoeringsorganisatie als de politie begeeft zich nu online, bijvoorbeeld om op te treden in omgevingen waar sprake is van strafbare feiten (zie thema 3). Maar de politie heeft naast het handhaven van wetten en regels ook nog een andere taak: problemen signaleren en voorkomen. In de online omgeving kan de politie dit doen door onwenselijk gedrag publiekelijk te veroordelen. Binnen specifieke gemeenschappen, zoals de gaming community (waar sommige korpsen nu al actief zijn), kan de politie in situaties die dreigen te escaleren, ook gewenst gedrag stimuleren. Bijvoorbeeld door zelf het goede voorbeeld te geven. Om dit soort interventies op enige schaal te kunnen uitvoeren, is er behoefte aan meer mankracht.

Zichtbaarheid van gedragsregels verbeteren

Gebruiksvoorwaarden van platformen zijn vaak lastig te vinden en doorgaans ook moeilijk te begrijpen. De overheid kan bedrijven ertoe dwingen de gedragsregels op allerlei manieren toegankelijker te maken. Het is daarbij aan te raden verschillende groepen te betrekken bij het opstellen van gedragsregels, zodat deze voor alle doelgroepen, ook kwetsbare groepen, goed te begrijpen zijn. Daarnaast kan de overheid op meer informele wijze druk uitoefenen, bijvoorbeeld door een gebrek aan inspanning aan de kaak te stellen. In hoofdstuk 5 concludeerden we dat dit in Nederland nog maar zelden gebeurt. Bewindspersonen gaan er nauwelijks toe over om bedrijven publiekelijk te vermanen. Maatschappelijke organisaties doen dit vaker, en de overheid zou op hun inspanningen kunnen bouwen. Ook adverteerders kunnen bijdragen, bijvoorbeeld door makers van content te wijzen op bestaande regels, zoals de Nederlandse Reclame Code.

Instituties mobiliseren

Een breed scala aan instituties is nu al betrokken bij het stellen of ontwikkelen van maatschappelijke normen – ook als ze zich vanuit hun specifieke missie niet zozeer identificeren met een online problematiek. Politieke partijen, maar ook kerken, jeugdbewegingen of zelfs sportverenigingen zouden hun blik kunnen verruimen naar aspecten van online moraliteit, en een rol kunnen vervullen in het agenderen daarvan. Een goed voorbeeld is het recente initiatief van een aantal Britse voetbalclubs en de internationale bonden UEFA en FIFA, om drie dagen lang

sociale media te boycotten. Zo vragen zij aandacht voor online racistisch verbaal geweld. Zij spreken al tot een specifieke demografie en kunnen hun strategie dus goed afstemmen op wat voor die groep werkt. De overheid kan de dialoog met dergelijke instituties aangaan om ze te mobiliseren aandacht te genereren voor online normen.

Digitale ruimtes onderhouden
Internetgebruikers doen nu al aan allerlei vormen van technological placekeeping, of spreekwoordelijk ‘onderhoud’ van online omgevingen. Toch spreken maar weinig mensen zich publiekelijk uit over de wenselijkheid van andermans gedrag. Platformen kunnen dit stimuleren door bestaande inspanningen te belonen. Zo zouden ze berichten van groepen die zich actief inzetten voor de strijd tegen desinformatie, een prominentere plek kunnen geven. Daarmee dienen ze ook hun eigenbelang. Platformen hebben ook zelf baat bij een gezonde online sfeer. Overheden of toezichthouders kunnen bedrijven zo nodig tot dit soort initiatieven verplichten.

6.3 Thema 3: mensen beschermen en slachtoffers bijstaan

Het is een primaire taak van de overheid om de rechten van burgers te beschermen. Ook online zal de overheid hier invulling aan moeten geven. Op het moment dat mensen slachtoffer worden, verdienen zij bijstand. Specifieke groepen die online kwetsbaar zijn, zoals minderheden en jongeren, hebben het meest baat bij bescherming van de overheid in hun online activiteiten.

Uit de fenomenen en casuïstiek die in dit rapport worden beschreven, blijkt dat mensen in de online omgeving vaak onbeschermd zijn. Slachtoffers ervaren een gebrek aan aanwezigheid van zowel overheid als hulpverleners en hebben het gevoel dat ze niet op dezelfde manier beschermd worden als in het fysieke domein. Hoewel de online omgeving niet feitelijk wetteloos is, wordt hij wel vaak zo ervaren. Het bestraffen van mensen die zich onrechtmatig of strafbaar gedragen, is belangrijk voor genoegdoening, maar gebeurt online nog weinig (Van De Weijer et al., 2020). De schaal van de fenomenen laat dat niet toe. Ook bij fenomenen die strafbaar (kunnen) zijn, is een juridische aanpak gericht op straffen dus niet voldoende om mensen te beschermen.

Het gebrek aan bescherming online heeft verschillende redenen, bijvoorbeeld dat mensen gebruik maken van diensten die worden geleverd vanuit een ander land, waar kwaadwillenden vrij spel hebben, of omdat online platformen vaak slecht bereikbaar zijn voor hulp. Ook verschilt het per fenomeen sterk, of er door private

partijen of de overheid al bijstand voor slachtoffers georganiseerd is. Rondom een thema als online haat zien we bijvoorbeeld een toename aan betrokkenheid van maatschappelijke organisaties, internetdienstverleners en de overheid. Voor fenomenen als extreme challenges, desinformatie en shaming is dat weer veel minder het geval.

Ook hier hebben internetdienstverleners de maatschappelijke verantwoordelijkheid om schade helpen te voorkomen en bij te dragen aan herstel. In het ontwerp van hun diensten moeten zij daar rekening mee houden en wanneer schade optreedt, mag van hen een actieve houding worden verwacht (zie thema 1). Naast dat het gesprek over online normen gevoerd moet worden (zie thema 2), heeft de overheid dus een duidelijke verantwoordelijkheid om mensen te beschermen en slachtoffers bij te staan. Vanuit haar positie kan de overheid ook andere actoren aansporen om mensen beter te beschermen en slachtoffers bij te staan.

Op basis van het onderzoek formuleren we drie opgaven voor de overheid om mensen online beter te beschermen en slachtoffers bij te staan:
1) vang slachtoffers op;
2) wees online aanwezig; en
3) bewaak de balans tussen verschillende rechten online.

Vang slachtsoffers op

De overheid kan investeren in de opvang van slachtoffers van schadelijk en immoreel gedrag online. Slachtoffers kunnen op dit moment hun ervaringen vaak moeilijk kwijt, zeker als de strafbaarheid niet evident is. Naast de vervolging van eventuele daders is het voor slachtoffers ook belangrijk dat ze serieus worden genomen en dat naar hen wordt geluisterd. Uit dit onderzoek blijkt dat private meldpunten, zoals de Stichting Online Shaming of de Stichting Stop Kindermisbruik, nauw contact hebben met de platformen. De politie kan hiervan leren. Om slachtoffers op te vangen kan de overheid ook de volgende twee handelingsopties overwegen.

Richt een meldpunt online misstanden op

De overheid kan zelf een nationaal meldpunt oprichten waar slachtoffers van online immoreel en schadelijk gedrag terechtkunnen. Het Meldpunt Internetdiscriminatie (MiND) kwam in dit onderzoek naar voren als een bruikbaar model voor de hulp aan slachtoffers en beperking van schadelijk en immoreel gedrag online. Bij het MiND zijn officieren van justitie betrokken, die een melding kunnen bekijken en die goed kunnen inschatten of het gaat om onrechtmatige gedrag. Vaak wordt vervolgens eerst aan het betreffende internetplatform gevraagd om de gemelde content te verwijderen. Als dat niet lukt, kan de rechter worden gevraagd om dit te verplichten. Meldknop.nl (zie hoofdstuk 5) functioneert op dit moment al als een

portal voor kinderen en hun begeleiders om in contact te komen met hulpverlenende instanties gespecialiseerd in verschillende fenomenen uit de taxonomie. De modellen van MiND en Meldknop.nl zouden verder kunnen worden uitgebouwd en opgeschaald, zodat het ook voor andere fenomenen uit dit onderzoek kan worden ingezet, en zodat het meldpunt breder ingezet kan worden. Een centraal meldpunt kan ook bruikbaar zijn voor de registratie van schadelijk en immoreel gedrag online, waardoor de samenleving beter zicht krijgt op de aard en omvang hiervan in Nederland. Een meldpunt is verder gebaat bij effectieve samenwerking met online platformen, vanwege hun mogelijkheid tot het beperken van de impact van schadelijk en immoreel gedrag online. Bij het verwijderen van content is dat ook in het belang van de ‘auteur’ van die content. Procedures zouden daarom gepaard moeten gaan met bezwaar- en verhaal-mogelijkheden.

Luister naar slachtoffers en registreer alle meldingen
De Casus online shaming illustreert dat het voor het rechtvaardigheidsgevoel van slachtoffers essentieel is dat naar hen geluisterd wordt. We merken op dat slachtoffers van online immoreel en schadelijk gedrag vaak geen melding of aangifte doen. Dat draagt bij aan een gevoel van wetteloosheid op internet. Actiever luisteren naar slachtoffers en meldingen registreren, kunnen bijdragen aan het beter bijstaan van slachtoffers.

Wees online aanwezig

Als op straat een onveilige situatie ontstaat, weten mensen over het algemeen wat ze aan de overheid hebben. Politie en andere hulpverleners zullen na melding van burgers de situatie beoordelen en actie ondernemen. Wanneer mensen op internet schade ondervinden, is het voor burgers niet altijd duidelijk wat ze aan de overheid hebben. Dit draagt bij aan het gevoel van online wetteloosheid en straffeloosheid dat slachtoffers ervaren. De online aanwezigheid van de overheid kan dit wellicht verhelpen. De literatuuranalyse en gesprekken met experts leveren de volgende handelingsopties op om de online aanwezigheid van de overheid te versterken.

Online jongerenwerkers en wijkagenten inzetten
Online jongerenwerkers en wijkagenten hebben zicht op wat jongeren online bezighoudt en kunnen daarmee schadelijk en immoreel gedrag zowel offline als online tegengaan. Tijdens de coronacrisis zijn door lokale wijkagenten positieve ervaringen opgedaan met online gamen met jongeren. Zo konden zij vroegtijdig escalaties signaleren en voorkomen. De overheid kan maatschappelijke (hulpverlenings)organisaties stimuleren om online actiever te worden en ook andere doelgroepen dan jongeren aan te spreken. Zij zouden binnen socialemediaplatformen aanwezig kunnen zijn, om door slachtoffers ingeschakeld te kunnen worden. Daarbij is het uiteraard van belang dat het handelen van

hulpverleners en politie op transparante wijze gebeurt en bevoegdheden daarvoor helder zijn.

Verstorend optreden door de politie
Verstorend optreden door de politie kan helpen om de impact van schadelijk en immoreel gedrag zoveel mogelijk te beperken. Bij mogelijk strafbare feiten zoals cybercriminaliteit is de politie daar reeds toe bevoegd. Criminelen kunnen worden tegengewerkt of platformen gesloten. Dergelijke interventies zouden ook effectief kunnen zijn bij de bestrijding van andere fenomenen uit de taxonomie, zoals bijvoorbeeld bij sock puppeting of online discriminatie. De overheid moet zich wel bewust zijn van de spanning tussen verschillende rechten online.

Zichtbaarheid van de politie in de online omgeving
Meer zichtbaarheid van de politie online zou kunnen bijdragen aan een verlaging van online ontsporing. Uit gesprekken met experts blijkt dat alleen al de zichtbaarheid van online advertenties van de politie, cybercriminele jongeren kan laten reflecteren op hun gedrag. In de fysieke wereld is de politie zichtbaar op straat, waardoor de fysieke omgeving geen wetteloze ruimte lijkt. Door ook online ruimte in te nemen, bijvoorbeeld met advertenties of via afspraken met platformen, worden gebruikers er online aan herinnerd dat het internet geen wetteloze omgeving is.

Toegankelijkheid van de hulpverlening in de online omgeving
Slachtoffers van online schadelijk gedrag moeten makkelijk de weg kunnen vinden naar hulp – ook online. De website Meldknop.nl, waar ook de overheid bij betrokken is, fungeert als een laagdrempelige ‘toegangspoort’ naar organisaties die deze hulp kunnen bieden. Het initiatief is echter exclusief gericht op kinderen, en biedt geen expertise bij fenomenen op het gebied van informatiemanipulatie en zelfbeschadiging. Een dergelijk portal, mits goed gepromoot, zou ook nuttig kunnen zijn voor volwassen slachtoffers, als zo’n website de expertise over een breed scala aan fenomenen bijeenbrengt.

Bewaak de balans tussen verschillende rechten online

De afgelopen jaren is de spanning tussen verschillende rechten online steeds sterker onderdeel geworden van het publieke debat. De grote online platformen doen meer dan ooit aan contentmoderatie en overheden wereldwijd worstelen met de macht die platformen hebben over uitingen op internet. Tegelijkertijd uiten mensenrechtenorganisaties hun zorgen over de groeiende invloed van (autoritaire) regimes op het vrije internet. De overheid moet de balans bewaken tussen verschillende rechten online om haar burgers te beschermen tegen online schadelijk en immoreel gedrag, en tegelijkertijd hun vrijheden te waarborgen. Ook

platformen vragen om regulering en stellingname vanuit de overheid. In de praktijk betekent dit vooral, dat de overheid zich niet afzijdig moet houden in de maatschappelijke discussie rondom online mensenrechten en contentmoderatie.

Contentmoderatie is bedoeld om gebruikers te beschermen tegen schadelijke en immorele content online, maar kan ook verregaande inbreuk maken op grondrechten. Gebruikers hebben vaak geen inspraak op de zogenaamde ‘community standards’ van online platformen en kunnen door grote marktmacht ook niet zomaar overstappen naar alternatieven. Contentmoderatie verschilt per platform en is niet altijd in lijn met geldende culturele en maatschappelijke normen, omdat veel platformbedrijven zich niet in Nederland bevinden. De overheid kan door regulering en dialoog zorgen dat platformen de bescherming van rechten serieus nemen in hun keuzes rondom contentmoderatie. Op basis van het onderzoek doen we een aantal suggesties voor de Rijksoverheid om de balans tussen mensenrechten online te bewaken.

Zorg voor democratische controle op contentmoderatie
Op dit moment bepalen grote online platformen de facto wat wel en niet als schadelijk en immoreel gezien wordt op het internet. Amerikaanse normen over schadelijkheid en immoraliteit bepalen op deze manier grotendeels wat Nederlandse burgers in online omgevingen te zien krijgen. De ongemakkelijkheid met deze rol groeit ook bij platformen zelf. Het Rathenau Instituut schreef in mei 2021 al aan de Tweede Kamer dat onafhankelijk publiek toezicht een manier kan zijn om contentmoderatie niet eenzijdig bij bedrijven te beleggen. De overheid kan dit punt inbrengen in de Europese beleidsdiscussie rondom de Digital Services Act (DSA).

Focus niet alleen op illegale content
Focus niet alleen op illegale content, omdat dat burgers onvoldoende beschermt tegen alle vormen van schadelijk en immoreel gedrag die in dit onderzoek naar voren zijn gekomen. Nieuwe Europese wetgeving zoals de Digital Services Act lijkt alleen te gaan over het verwijderen van illegale content, en niet over schadelijke content. Uit ons onderzoek blijkt dat veel gedragingen online wel een illegale component kunnen hebben, maar dat dit voor zowel slachtoffers als daders niet altijd duidelijk is. Wanneer wordt online shaming laster? En wanneer is kwakzalverij verboden? Die onduidelijkheid over de grenzen van online gedrag maakt dat slachtoffers zich niet altijd gesteund voelen door bestaande juridische kaders. Een pasklare oplossing is er niet, omdat we als samenleving pas recent geconfronteerd worden met de vraag hoe we ieders rechten online op een goede manier kunnen beschermen. De vrijheid van meningsuiting kan botsen met de bewegingsruimte en veiligheid van mensen online en van minderheden in het bijzonder. Het is belangrijk dat de overheid dit dilemma niet uit de weg gaat en verder onderzoekt wat

mogelijkheden zijn om hiermee om te gaan. Burgers worden namelijk nog steeds onvoldoende beschermd met een nauwe focus op het verwijderen van illegale content op internet.

Verbeter klachtenprocedures bij online platformen om rechtszekerheid en rechtsgelijkheid te garanderen
Uit ons onderzoek blijkt dat nog lang niet alle platformen transparante en consistente klachtenprocedures aanbieden en dat vooral kleine platformen hier tekort op schieten. De overheid kan de onderhandelingen over de DSA aangrijpen om meer eisen te stellen aan de gebruikersvoorwaarden en klachten- en verhaalprocedures van platformen. Platformen zouden snelle en transparante mechanismen moeten hebben met een mogelijkheid tot collectieve klachtenprocedures voor grotere groepen burgers. Dergelijke maatregelen zijn ook onderdeel van het voorstel voor een Online Safety Bill die besproken wordt binnen het VK, en zouden ter inspiratie kunnen dienen.

Wees terughoudend bij contentmoderatie
Wees terughoudend bij technische oplossingen voor contentmoderatie. Het modereren van de online omgeving zou niet uitsluitend uitbesteed moeten worden aan kunstmatige intelligentie, maar zou tot stand moeten komen in dialoog met de samenleving. Slechts in beperkte gevallen kan algoritmische contentmoderatie op zichzelf staan, zoals bij het verwijderen van beelden van kindermisbruik. Zulke uploadfilters worden door mensenrechtenorganisaties vaak als een te grote inbreuk op de vrijheid van meningsuiting gezien. De overheid zou dus terughoudend moeten zijn in het stimuleren van technische ‘quick-fixes’ voor contentmoderatie, zeker omdat het niets doet aan de onderliggende mechanismen achter schadelijk en immoreel gedrag online, en omdat we als samenleving het gesprek over online normen nog moeten voeren.

6.4 Thema 4: adaptief vermogen van samenleving versterken

Actoren binnen de overheid, markt en maatschappij reageren op schadelijk en immoreel gedrag, maar lopen daarmee het risico voortdurend achter de feiten aan te lopen. Het internet bestaat slechts enkele decennia, en heeft in die korte tijd in hoog tempo allerlei nieuwe vormen van gedrag mogelijk gemaakt. Maar schadelijk en immoreel gedrag online wordt vaak pas zichtbaar op het moment dat die een kritieke massa heeft bereikt. Een meer proactieve, adaptieve en preventieve procesinrichting is gewenst. Zo wordt de aanpak van schadelijk en immoreel online gedrag toekomstbestendig.

Op basis van dit onderzoek en de constatering dat fenomenen en mechanismen achter schadelijk en immoreel gedrag continu in beweging zijn, formuleert het Rathenau Instituut twee opgaven voor de overheid om de aanpak van schadelijk en immoreel gedrag langdurig te organiseren:
1) coördineer de uitwisseling van expertise over schadelijk en immoreel gedrag online; en
2) faciliteer netwerken van betrokken actoren.

Coördineer de uitwisseling van expertise

Dit onderzoek heeft zichtbaar gemaakt dat de kennis over de fenomenen in hoofdstuk 3 nog onvolledig is, en dat het veld van actoren een lappendeken is. De taxonomie van schadelijk en immoreel gedrag online is met dit onderzoek voor het eerst in beeld. Nu dit overzicht er is, kunnen netwerken van deskundigen en hulpverleners gevormd worden en de fenomenen in de taxonomie van schadelijk en immoreel gedrag online blijvend gemonitord, uitgebreid en aangepast worden.

De overheid kan een coördinerende rol vervullen bij het samenbrengen en organiseren van de kennis en kunde. Opties om dat vorm te geven zijn de volgende.

Stel een kenniscoördinator in

De kenniscoördinator ‘schadelijk en immoreel gedrag online’ heeft als taak om continu kennis bijeen te brengen of te vergaren over de aard en omvang van online misstanden. Het doel van de coördinator is om belangrijke ontwikkelingen te signaleren en beleidsmakers op basis daarvan te adviseren. Wanneer meer systematisch gegevens worden verzameld over de verschillende fenomenen, wordt het ook beter mogelijk de urgentie van fenomenen op specifieke momenten vast te stellen en te prioriteren. Het is daarbij wel van belang dat eventuele dataverzameling over het gedrag van burgers op internet aan kaders wordt gebonden, zodat onnodige surveillance of inbreuk op de persoonlijke levenssfeer wordt voorkomen.

Bevorder onderzoeksprogramma's en samenwerking

De overheid kan de ontwikkeling van onderzoeksprogramma’s en samenwerking met onderzoeksinstellingen bevorderen, om de mechanismen en fenomenen beter te doorgronden en slachtoffers en daders beter in beeld te krijgen. Een doelgroepenbenadering kan helpen om vanuit (kwetsbare) groepen te begrijpen wat hen beweegt, kwetsbaar maakt en zou ondersteunen, en zo doelgroepgerichte programma’s te ontwikkelen.

Aan de hand van onze taxonomie van immoreel en schadelijk gedrag online zou de overheid een kennisagenda kunnen opstellen om per fenomeen meer zicht te krijgen op dader- en slachtofferschap en onderliggende sociaalmaatschappelijke factoren. Veel fenomenen uit dit onderzoek hebben disproportioneel effect op bepaalde groepen in de samenleving. Dat geldt bij haatzaaien en bedreiging bijvoorbeeld voor vrouwen en minderheden, maar ook jongeren zijn vaker slachtoffer van fenomenen als cyberpesten en shame-sexting. Daarbij moet gezegd worden dat veel onderzoek zich ook specifiek richt op jongeren, waardoor juist volwassenen uit het zicht kunnen raken. Het is daarom belangrijk dat de overheid meer zicht krijgt op welke groepen in de samenleving disproportioneel vaak slachtoffer zijn van online schadelijk en immoreel gedrag. Daarvoor zou de overheid bij uitstek kunnen samenwerken met platformen en maatschappelijke organisaties om deze gegevens boven tafel te krijgen, en daar beleid en hulpverlening op af te stemmen.

Investeren in capaciteit en kennis van professionals
De overheid kan investeren in de capaciteit en kennis van beleidsmakers, handhavers, hulpverleners en andere professionals. Zij hebben middelen nodig om zich te verdiepen in de fenomenen en mechanismen van schadelijk en immoreel gedrag in de online omgeving en daarop te kunnen reageren. Investeringen in meer mankracht, specialisatie en samenwerking tussen deze professionals kan helpen om beter toegerust te raken op de uitdagingen gekoppeld aan schadelijk en immoreel gedrag online. Door professionals zelf te laten bepalen welke aanpak werkt en hierover met elkaar in gesprek te gaan, benutten we de ervaring en professionaliteit van betrokkenen.

Faciliteer netwerken van betrokken actoren

De taxonomie van fenomenen (hoofdstuk 3) bestrijkt een veelheid aan domeinen die raken aan de verantwoordelijkheid van de overheid, en bestrijkt daarmee het werkveld van vele beleidsmakers en medewerkers verspreid over verschillende ministeries. Zo waren bij dit onderzoek ambtenaren vanuit het Ministerie van Justitie & Veiligheid, Economische Zaken, Binnenlandse Zaken en Onderwijs, Cultuur en Wetenschap betrokken, omdat zij allen op een deelonderwerp werken, zoals desinformatie, mediawijsheid of bescherming van de publieke ruimte.

In Nederland zijn daarnaast talrijke organisaties actief die expertise ontwikkelen over online immoreel en schadelijk gedrag of de daaraan onderliggende mechanismen. In hoofdstuk 5 kwamen vele actoren voorbij, zoals organisaties die internetgebruikers ondersteunen met advies en tools om zich tegen schadelijk gedrag te wapenen. Daarnaast zijn er organisaties die gebruikers mobiliseren om wenselijk gedrag te bevorderen, bijvoorbeeld door het internet zo in te richten dat

het wenselijk gedrag (van bijvoorbeeld *upstanders*) stimuleert en beloont. Ook zijn er collectieven die opkomen voor de belangen van slachtoffers van bepaalde fenomenen, organisaties die bedrijven onder druk zetten om tegen problematisch gedrag te ageren of hun gebruikers ertegen te beschermen, en partijen die bij overheden lobbyen voor betere regulering. Daarnaast spelen online platformen natuurlijk een centrale rol bij verspreiding van schadelijke content, en worden er in de markt producten ontwikkeld om gebruikers en bedrijven tegen schadelijke fenomenen te beschermen.

Het veld aan actoren dat betrokken is bij schadelijk en immoreel online gedrag is nu nog een lappendeken. Alle betrokken partijen zouden beter kunnen samenwerken en zo hun adaptief vermogen kunnen versterken om snel te reageren op nieuwe ontwikkelingen. De overheid kan hier een coördinerende en ondersteunende rol in spelen. Een concrete manier om dit te doen is de volgende.

Investeer in ambtenarennetwerk
De overheid zou kunnen investeren in een netwerk van ambtenaren die zich bezighouden met deelaspecten van schadelijk en immoreel online gedrag. Dit onderzoek heeft al een eerste aanzet van een dergelijk netwerk opgeleverd, door ambtenaren van verschillende ministeries te betrekken. Door deze groep ambtenaren met elkaar te verbinden, kan geleerd worden van ervaringen en beleid steeds integraler worden.

Stimuleer samenwerking tussen platform en maatschappelijk middenveld
De overheid zou samenwerking tussen platformen en maatschappelijk middenveld kunnen stimuleren via subsidies of andere programma's. Hulpverlenende organisaties kunnen actief hulp aanbieden op socialemediaplatformen wanneer de content daar aanleiding toe geeft. In Nederland gebeurt dit bijvoorbeeld al op traditionele en sociale media bij berichten over zelfmoord met een verwijzing naar 113. Door platformen met meer maatschappelijke instanties samen te laten werken, kunnen zij online hulp op een laagdrempelige manier faciliteren. Denk aan verwijzingen naar lgbtq+ organisaties bij content die homofoob van aard is; of samenwerkingen tussen platformen en organisaties die zich inzetten voor mentale gezondheid. Dat kan slachtoffers bijstaan als ze te maken krijgen met online immoreel en schadelijk gedrag.

Coördineer samenwerking toezichthouders
De overheid zou de samenwerking tussen toezichthouders kunnen coördineren om het toezicht op fenomenen en mechanismen van schadelijk en immoreel gedrag online te versterken en verbeteren. Toezicht op schadelijk en immoreel gedrag online bestrijkt het domein van verschillende toezichthouders, zoals de Autoriteit Consument en Markt, de Reclame Code Commissie en de Autoriteit

Persoonsgegevens. Samenwerking en afstemming tussen toezichthouders lijkt cruciaal om toezicht op de verschillende aspecten van schadelijk en immoreel gedrag te coördineren en versterken.

6.5 Conclusie

Dit rapport presenteert een taxonomie van schadelijk en immoreel online gedrag, waarmee 22 fenomenen van dergelijk gedrag in samenhang in beeld komen (hoofdstuk 3). Ook brengt dit rapport systematisch de kenmerken en mechanismen van het internet in kaart die een rol spelen bij het initiëren, faciliteren en versterken van schadelijk en immoreel online gedrag (hoofdstuk 4). Uit het onderzoek naar bestaande interventies blijkt dat nauwelijks actie is ondernomen om deze mechanismen ten goede te keren (hoofdstuk 5). Daarom formuleerde het Rathenau Instituut in dit hoofdstuk vier strategische thema's en opgaven die de overheid, in samenwerking met markt en maatschappij, in staat stellen om meer gericht actie te ondernemen tegen ontsporingen en moreel en wenselijk gedrag online te bevorderen.

Het eerste thema – Herinrichting van de online omgeving – bevat handvatten voor de Rijksoverheid om de mechanismen die kenmerkend zijn voor het internet ten goede te keren. Het tweede thema – Online normen verhelderen – formuleert aanbevelingen om maatschappelijke afspraken over normen en waarden online te vernieuwen. Het derde thema – Mensen beschermen en slachtoffers bijstaan – bevat suggesties voor de Rijksoverheid en haar uitvoeringsinstanties om beter te reageren op schadelijk en immoreel gedrag online en de schade daarvan. Het vierde thema – Adaptief vermogen versterken – bevat suggesties voor de Rijksoverheid om grip te krijgen en te houden op schadelijk en immoreel gedrag online dat continu in beweging is, en is gericht op het toekomstbestendig maken van de strategische agenda.

Waar het internet eerst een vrijplaats was voor gebruikers, is nu toch echt een actievere rol van de overheid nodig. De problemen zijn urgent en de schade reëel, de markt vraagt om regulering en de maatschappij verdient ondersteuning en bescherming. Er ligt een rol voor verschillende ministeries binnen de Rijksoverheid, handhaving en uitvoeringsorganisaties om maatregelen te nemen. Nieuwe kansen dienen zich daarvoor aan, bijvoorbeeld in de digitaliseringsstrategie van het nieuwe kabinet en binnen de lopende onderhandelingen over Europese beleidskaders. Het Rathenau Instituut hoopt met dit onderzoek een bijdrage te hebben geleverd aan het ontwikkelen van een toekomstbestendige aanpak van schadelijk en immoreel online gedrag in Nederland.

Literatuurlijst

4Chan. (2021). *rules*. Https://www.4chan.org/rules

Adviesraad Internationale Betrekkingen (AIV). (2020). *Regulering van online content Naar een herijking van het Nederlandse internetbeleid*. Adviesraad Internationale Betrekkingen (AIV).

AFM. (2018). *Crypto's: Aanbevelingen voor een regelgevend kader*. Autoriteit Financiële Markten.

Afuah, A. (2013). Are network effects really all about size? The role of structure and conduct. *Strategic Management Journal*, *34*(3), 257–273. https://doi.org/10.1002/smj.2013

Aiken, M. (2016). *The Cyber Effect: A Pioneering Cyberpsychologist Explains How Human Behaviour Changes Online*. John Murray Press.

Aiken, M., Mc Mahon, C., Haughton, C., O'Neill, L., & O'Carroll, E. (2016). A consideration of the social impact of cybercrime: examples from hacking, piracy, and child abuse material online. *Contemporary Social Science*, *11*(4), 373–391. https://doi.org/10.1080/21582041.2015.1117648

Algemeen Dagblad. (2021, 20 april). *Omstreden Twitter-alternatief Parler keert terug in appwinkel*. Algemeen Dagblad. https://www.ad.nl/tech/omstreden-twitter-alternatief-parler-keert-terug-in-appwinkel~a169770b/

Alimoradi, Z., Lin, C.-Y., Broström, A., Bülow, P. H., Bajalan, Z., Griffiths, M. D., Ohayon, M. M., & Pakpour, A. H. (2019). Internet addiction and sleep problems: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *47*, 51–61. https://doi.org/10.1016/j.smrv.2019.06.004

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, *6*(14), eaay3539. https://doi.org/10.1126/sciadv.aay3539

Amnesty. (2017, 20 november). *Amnesty reveals alarming impact of online abuse against women*. Amnesty International. https://www.amnesty.org/en/latest/news/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women/

Auxier, B. E., & Vitak, J. (2019). Factors Motivating Customization and Echo Chamber Creation Within Digital News Environments. *Social Media + Society*, *5*(2), 2056305119847506. https://doi.org/10.1177/2056305119847506

Bakker, A. (2021, 9 april). *Grapperhaus wil retweeten van privégegevens van agenten strafbaar stellen*. https://www.limburger.nl/cnt/dmf20210409_97571764

Bantema, W., Twickler, S. M. A., Munneke, S. A. J., Duchateau, M., & Stol, W. Ph. (2018). *Handhaving van de openbare orde door bestuurlijke maatregelen in een digitale wereld*. Sdu. https://www.politieenwetenschap.nl/publicatie/politiewetenschap/2018/burgemeesters-in-cyberspace-313/

Basak, R., Surah, S., Ganguly, N., & Ghosh, S. (2019). Online Public Shaming on Twitter: Detection, Analysis, and Mitigation. *IEEE Transactions on Computational Social Systems*, *6*(2), 208–220. https://doi.org/10.1109/TCSS.2019.2895734

Bats, J. (2019). *The moral matter of an interactive online domain: a philosophical and empirical exploration of how our relation with the online domain mediates online morality*. University of Twente.

Bellingcat. (2021, 7 januari). *The Making of QAnon: A Crowdsourced Conspiracy*.

Bellingcat. https://www.bellingcat.com/news/americas/2021/01/07/the-making-of-qanon-a-crowdsourced-conspiracy/

Benton, J. (2021, 22 april). Facebook is going to ask you more often what you want in your News Feed. *Nieman Lab*. https://www.niemanlab.org/2021/04/facebook-is-going-to-ask-you-more-often-what-you-want-in-your-news-feed/

Bertrand, N. (2017, 1 november). *Russia organized 2 sides of a Texas protest and encouraged 'both sides to battle in the streets'* [Businessinsider]. https://www.businessinsider.nl/russia-trolls-senate-intelligence-committee-hearing-2017-11

Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs Conspiracy: Collective Narratives in the Age of Misinformation. *PLOS ONE*, *10*(2), e0118093. https://doi.org/10.1371/journal.pone.0118093

Bhikhie, A. (2020, 29 juni). *CU en GroenLinks dienen hatecrimewet in om*

discriminatie harder te straffen. NU.nl. https://www.nu.nl/politiek/6061053/cu-en-groenlinks-dienen-hatecrimewet-in-om-discriminatie-harder-te-straffen.html

Billingham, P., & Parr, T. (2019). Online Public Shaming: Virtues and Vices. *Journal of Social Philosophy*, *51*(3), 371–390. https://doi.org/10.1111/josp.12308

Billingham, P., & Parr, T. (2020). Enforcing social norms: The morality of public shaming. *European Journal of Philosophy*, *28*(4), 997–1016. https://doi.org/10.1111/ejop.12543

Bishop, S. (2019). Managing visibility on YouTube through algorithmic gossip. *New Media & Society*, *21*(11–12), 2589–2606. https://doi.org/10.1177/1461444819854731

Blackwell, L., Dimond, J., Schoenebeck, S., & Lampe, C. (2017). Classification and Its Consequences for Online Harassment: Design Insights from HeartMob. In *Proc. ACM Hum.-Comput. Interact.* (Vol. 1, Nummer CSCW, p. Article 24). Association for Computing Machinery.

Bliuc, A.-M., Faulkner, N., Jakubowicz, A., & McGarty, C. (2018). Online networks of racial hate: A systematic review of 10 years of research on cyber-racism. *Computers in Human Behavior*, *87*, 75–86. https://doi.org/10.1016/j.chb.2018.05.026

Bond, S. (2021, 14 mei). *Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows*. NPR.Org. https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes

Borra, E., Niederer, S., Preuß, J., & Weltevrede, E. (2017). *Mapping troll-like practices on twitter*. https://dare.uva.nl/search?identifier=c9824b9e-e0e5-4342-83c7-9f5237727f16

Bouma, R. (2020, 25 september). *Amerikaanse complottheorie QAnon ook in Nederland in opkomst*. NOS. https://nos.nl/l/2349814

Bouyeure, L. (2020, 29 mei). *Op TikTok ligt de pro-anacontent voor het oprapen*. de Volkskrant. https://www.volkskrant.nl/gs-b00bd886

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(28),

7313–7318. https://doi.org/10.1073/pnas.1618923114

Branley, D. B., & Covey, J. (2017). Is exposure to online content depicting risky behavior related to viewers' own risky behavior offline? *Computers in Human Behavior*, *75*, 283–287. https://doi.org/10.1016/j.chb.2017.05.023

Brouillette, A. (2020). *Key findings from the 2020 RDR Corporate Accountability Index*. Ranking Digital Rights. https://rankingdigitalrights.org/index2020/key-findings

Brumfiel, G. (2021, 12 mei). *For Some Anti-Vaccine Advocates, Misinformation Is Part Of A Business*. NPR.Org. https://www.npr.org/sections/health-shots/2021/05/12/993615185/for-some-anti-vaccine-advocates-misinformation-is-part-of-a-business

Bureau Clara Wichmann. (2020). *Onderzoeksrapport Online Gendered Hate Speech: Civiel procederen tegen online hate speech*. Bureau Clara Wichmann. https://clara-wichmann.nl/content/uploads/2021/02/Onderzoeksrapport-Online-Gendered-Hate-Speech.pdf

Bureau Jeugd en Media. (2021). *Wij zijn we*. https://www.bureaujeugdenmedia.nl/. https://www.bureaujeugdenmedia.nl/project/deinternethelden

Burris, C. T., & Leitch, R. (2018). Harmful fun: Pranks and sadistic motivation. *Motivation and Emotion*, *42*(1), 90–102. https://doi.org/10.1007/s11031-017-9651-5

CBS. (2017, 17 oktober). *More than 12M 'Me Too' Facebook posts, comments, reactions in 24 hours*. CBS. https://www.cbsnews.com/news/metoo-more-than-12-million-facebook-posts-comments-reactions-24-hours/

CBS. (2018). *Digitale Veiligheid & Criminaliteit 2018* [Webpagina]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/publicatie/2019/29/digitale-veiligheid-criminaliteit-2018

CBS. (2019a). *1,2 miljoen slachtoffers van digitale criminaliteit* [Webpagina]. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/nieuws/2019/29/1-2-miljoen-slachtoffers-van-digitale-criminaliteit

CBS. (2019b). *Internet*. Nederland langs de Europese meetlat. https://longreads.cbs.nl/europese-meetlat-2019/internet/

CBS. (2019c). *Veiligheidsmonitor 2019* [Webpagina]. Centraal Bureau voor de Statistiek. https://doi.org/10/veiligheidsmonitor-2019

CBS. (2020a). *Online seksuele intimidatie - Prevalentiemonitor Huiselijk Geweld en Seksueel Geweld 2020 | CBS* [Webpagina]. CBS. https://longreads.cbs.nl/phgsg-2020/online-seksuele-intimidatie

CBS. (2020b, 1 april). *453 duizend Nederlanders hadden in 2019 thuis geen internet*. CBS. https://www.cbs.nl/nl-nl/nieuws/2020/14/453-duizend-nederlanders-hadden-in-2019-thuis-geen-internet

CBS. (2021). *CBS Statline*. https://opendata.cbs.nl/#/CBS/nl/dataset/83095NED/table

Cerniglia, L., Zoratto, F., Cimino, S., Laviola, G., Ammaniti, M., & Adriani, W. (2017). Internet Addiction in adolescence: Neurobiological, psychosocial and clinical issues. *Neuroscience & Biobehavioral Reviews*, *76*, 174–184. https://doi.org/10.1016/j.neubiorev.2016.12.024

Champlin, E. (1998). Nero Reconsidered. *New England Review*, *19*(2), 97–108.

Chen, W., & Thorson, E. (2021). Perceived individual and societal values of news and paying for subscriptions. *Journalism*, *22*(6), 1296–1316. https://doi.org/10.1177/1464884919847792

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1217–1230. https://doi.org/10.1145/2998181.2998213

Chia, D. X. Y., & Zhang, M. W. B. (2020). A Scoping Review of Cognitive Bias in Internet Addiction and Internet Gaming Disorders. *International Journal of Environmental Research and Public Health*, *17*(1), 373. https://doi.org/10.3390/ijerph17010373

Christopherson, K. M. (2007). The positive and negative implications of anonymity in Internet social interactions: “On the Internet, Nobody Knows You’re a Dog”. *Computers in Human Behavior*, *23*(6), 3038–3056. https://doi.org/10.1016/j.chb.2006.09.001

Cocking, D., & van den Hoven, J. (2018). *Evil Online*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119471219

Commissariaat voor de media. (2019). *Filterbubbels in Nederland*. Commissariaat voor de media.

Common, M., & Kleis Nielsen, R. (2021, 19 februari). *How to respond to disinformation while protecting free speech*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/risj-review/how-respond-disinformation-while-protecting-free-speech

COMPACT Education Group. (2020). *Guide to Conspiracy Theorists*. https://conspiracytheories.eu/_wpx/wp-content/uploads/2020/03/COMPACT_Guide-2.pdf

Costa, E., & Halpern, D. (2019). *The behavioural science of online harm and manipulation, and what to do about it: An exploratory paper to spark ideas and debate*. Behavioural Insights Team. https://www.bi.team/publications/the-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it/

Council of Europe. (2021). *Hate Speech*. Council of Europe. https://www.coe.int/en/web/freedom-expression/hate-speech

Coyne, I., Chesney, T., Logan, B., & Madden, N. (2009). Griefing in a Virtual Community: An Exploratory Survey of Second Life Residents. *Zeitschrift Für Psychologie / Journal of Psychology*, *217*(4), 214–221. https://doi.org/10.1027/0044-3409.217.4.214

Crisp Thinking. (2021). *We understand and identify emerging threats from online groups*. Crisp Thinking. https://www.crispthinking.com/our-approach/
Danish Institute for Human Rights, The. (2020). *Introduction to human rights impact assessment | The Danish Institute for Human Rights*. Danish Institute for Human Rights, The. https://www.humanrights.dk/tools/human-rights-impact-assessment-guidance-toolbox/introduction-human-rights-impact-assessment

Davenport, T. H., & Beck, J. C. (2001). *The attention economy: understanding the new currency of business*. Harvard Business School Press.

De Vries, A. (2018, 23 april). Opsporen? Doe het zelf! *Social Media DNA*. https://socialmediadna.nl/opsporen-doe-het-zelf/

De Vries, N. (2021, 22 februari). *Digital corpses: waarom mensen online naar lijken kijken* [Trouw]. https://www.trouw.nl/religie-filosofie/digital-corpses-waarom-mensen-online-naar-lijken-kijken~b13b49a3/

DeGoedeZaak. (z.d.). *Stop Shaming!* DeGoedeZaak. https://campagnes.degoedezaak.org/campaigns/stopshaming

Dehue, F., Bolman, C., Vollink, T., & Pouwelse, M. (2012). Cyberbullying and traditional bullying in relation to adolescents' perception of parenting. *Journal of CyberTherapy and Rehabilitation*, *5*(1), 25–34.

Delgado-López, P. D., & Corrales-García, E. M. (2018). Influence of Internet and Social Media in the Promotion of Alternative Oncology, Cancer Quackery, and the Predatory Publishing Phenomenon. *Cureus*, 1–11. https://doi.org/10.7759/cureus.2617

Denef, S., De Vries, A., Hadjimatheou, K., & Roosendaal, A. (2017). *DIY policing*. Fraunhofer IAO.

Department for Digital, Culture, Media and Sport. (2020). *Safer technology, safer users: The UK as a world-leader in Safety Tech; A Sectoral Analysis of UK Online Safety Technology*. UK Government. https://www.gov.uk/government/publications/safer-technology-safer-users-the-uk-as-a-world-leader-in-safety-tech

Digan, K. (2021, 23 maart). *President KNAW: Universiteiten, bescherm je medewerkers*. ScienceGuide. https://www.scienceguide.nl/2021/03/president-knaw-universiteiten-bescherm-je-medewerkers/

Digitale Overheid. (2020, 14 april). *97 procent Nederlanders heeft thuis internet*. Rijksoverheid.nl. https://www.digitaleoverheid.nl/nieuws/97-procent-nederlanders-had-in-2019-thuis-internet/

Dijck, J. van, Poell, T., & Waal, M. de. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press.

Diresta, R. (2018, 30 augustus). Free Speech Is Not the Same As Free Reach. *Wired*. https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/

Döpfner, M. (2021, 27 januari). *It's time for Europe to take private data from the hands of powerful tech monopolies and give it back to the people*. Business Insider. https://www.businessinsider.com/big-tech-private-data-facebook-google-apple-europe-eu-2021-1

Duin, R. J. (2020, 9 maart). *Eenmaal online gaat de foto van een verdachte nooit meer weg*. Het Parool. https://www.parool.nl/nieuws/eenmaal-online-gaat-de-foto-van-een-verdachte-nooit-meer-weg~ba172078/

ECP. (2021, 9 februari). Internet blijkt lichtpuntje in coronajaar: geen toename van negatieve online ervaringen onder jongeren. *ECP | Platform voor de InformatieSamenleving*. https://ecp.nl/actueel/internet-blijkt-voor-jongeren-lichtpuntje-in-coronajaar/

ECRI. (2019). *ECRI report on the Netherlands* (p. 65). Council of Europe.

Edunov, S., Bhagat, S., Burke, M., Diuk, C., & Onur Filiz, I. (2016, 4 februari). Three and a half degrees of separation. *Facebook Research*. https://research.fb.com/blog/2016/02/three-and-a-half-degrees-of-separation/

Eindhovens Dagblad. (2019, 17 december). *Boete voor arts uit Eindhoven vanwege reclame*. ed.nl. https://www.ed.nl/eindhoven/boete-voor-arts-uit-eindhoven-vanwege-reclame~ab38a423/

Ellemers, N., Van Der Toorn, J., & Paunov, Y. (2019). The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017. *Personality and Social Psychology Review*, *23*(4), 332–366. https://doi.org/10.1177/1088868318811759

EOKM. (2020). *Factsheet Grooming*. Expertisebureau Online Kindermisbruik. https://www.eokm.nl/wp-content/uploads/2020/08/EOKM-Factsheet-Grooming_update_aug_2020v2.pdf

Espinoza, J. (2020, 11 november). *Former White Congressional Candidate Tweets 'I'm a Black Gay Guy,' Explanation Leads to Bizarre Series of Events*. Complex. https://www.complex.com/life/2020/11/white-gop-candidate-dean-browning-tweets-im-a-gay-black-guy-bizarre-explanation

Europese Raad. (2000, 8 juni). *Richtlijn inzake elektronische handel*. Europese Unie. https://eur-lex.europa.eu/legal-content/NL/TXT/?uri=celex%3A32000L0031

Eurostat. (2021). *Individuals - internet activities*. https://ec.europa.eu/eurostat/databrowser/product/page/ISOC_CI_AC_I

Facebook. (2019). *Rules Enforcement*. https://transparency.twitter.com/en/reports/rules-enforcement.html#2019-jan-jun

Facebook. (2021). *Community Standards*. Facebook. Https://www.facebook.com/communitystandards/.

Faddoul, M., Chaslot, G., & Farid, H. (2020). A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos. *arXiv:2003.03318 [cs]*. http://arxiv.org/abs/2003.03318

Fox, J., Cruz, C., & Lee, J. Y. (2015). Perpetuating online sexism offline: Anonymity, interactivity, and the effects of sexist hashtags on social media. *Computers in Human Behavior*, *52*, 436–442. https://doi.org/10.1016/j.chb.2015.06.024

France-Presse, A. (2021, 23 januari). *Italy blocks TikTok for certain users after death of girl allegedly playing 'choking' game*. The Guardian. http://www.theguardian.com/world/2021/jan/23/italy-blocks-tiktok-for-certain-users-after-death-of-girl-allegedly-playing-choking-game

Freckelton QC, I. (2020). COVID-19: Fear, quackery, false representations and the law. *International Journal of Law and Psychiatry*, *72*, 101611. https://doi.org/10.1016/j.ijlp.2020.101611

Furnell, S. (2009). Hackers, viruses and malicious software. In *Handbook of internet crime* (pp. 173–193). Willan.

Gabszewicz, J. J., Laussel, D., & Sonnac, N. (2001). Press advertising and the ascent of the 'Pensée Unique'. *European Economic Review*, *45*(4–6), 641–651. https://doi.org/10.1016/S0014-2921(01)00139-8

Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering Online Hate Speech* (p. 71). UNESCO. http://en.unesco.kz/countering-online-hate-speech

Gardner, H., & Davis, K. (2013). *The App Generation: How Today's Youth Navigate Identity, Intimacy, and Imagination in a Digital World*. Yale University Press. https://www.jstor.org/stable/j.ctt5vm7dh

Gebiedonline. (2021). *Ons platform*. Gebiedonline. Https://gebiedonline.nl/ons-platform

Geerts, G., & Den Boon, C. A. (1999). *Van Dale Groot woordenboek van de Nederlandse taal* (13de dr.). Van Dale Uitgevers.

Gelfert, A. (2018). Fake News: A Definition. *Informal Logic*, *38*(1), 84–117. https://doi.org/10.22329/il.v38i1.5068

Gerrard, Y. (2020, 3 september). TikTok Has a Pro-Anorexia Problem. *Wired*. https://www.wired.com/story/opinion-tiktok-has-a-pro-anorexia-problem/

Ghaffary, S. (2021, 13 mei). *How angry Apple employees' petition led to a controversial new hire's departure*. Vox. https://www.vox.com/recode/2021/5/13/22435266/apple-employees-petition-controversial-antonio-garcia-martinez-new-hire-departure

Goldsmith, A., & Brewer, R. (2015). Digital drift and the criminal interaction order. *Theoretical Criminology*, *19*(1), 112–130. https://doi.org/10.1177/1362480614538645

Grant, H. (2020a, 10 december). *Pornhub to ban unverified uploads after child abuse content claims*. The Guardian. http://www.theguardian.com/global-development/2020/dec/10/pornhub-to-ban-unverified-uploads-after-child-abuse-content-claims

Grant, H. (2020b, 15 december). *How extreme porn has become a gateway drug into child abuse*. The Guardian. http://www.theguardian.com/global-development/2020/dec/15/how-extreme-porn-has-become-a-gateway-drug-into-child-abuse

Griffin, A. (2021, 27 januari). *YouTube reveals full scale of coronavirus misinformation on its platform*. The Independent. https://www.independent.co.uk/life-style/gadgets-and-tech/youtube-covid-19-coronavirus-misinformation-b1793492.html

Guan, S. A., & Subrahmanyam, K. (2009). Youth Internet use: risks and opportunities. *Curr Opin Psychiatry*, *22*(4), 351–356. https://doi.org/10.1097/YCO.0b013e32832bd7e0

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour*, *4*(5), 472–480. https://doi.org/10.1038/s41562-020-0833-x

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), 1–8. https://doi.org/10.1126/sciadv.aau4586

Ha, A. (2020, 29 juni). Facebook expands its fan subscription program.

TechCrunch. https://social.techcrunch.com/2020/06/29/facebook-expands-fan-subscriptions/

Hadjimatheou, K. (2019). Citizen-led digital policing and democratic norms: The case of self-styled paedophile hunters. *Criminology & Criminal Justice*, 1748895819880956. https://doi.org/10.1177/1748895819880956

Harambam, J. (2017). De politisering van de Waarheid. *Sociologie*, *13*(1), 73–92. https://doi.org/10.5117/SOC2017.1.HARA

Haspels-Goudriaan, J. (2020, 16 maart). *Geen overval, winkelinbraak of diefstal: Jongeren beginnen criminele loopbaan vaker in cyberspace*. Algemeen Dagblad. https://www.ad.nl/den-haag/geen-overval-winkelinbraak-of-diefstal-jongeren-beginnen-criminele-loopbaan-vaker-in-cyberspace~a61e4bd5/

Heck, W. (2020, 5 november). *Maker aangehouden van'phishing panels'voor grootschalige oplichting*. NRC. https://www.nrc.nl/nieuws/2020/11/05/maker-van-phishing-panels-aanghouden-voor-grootschalige-oplichting-a4018890

Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, *34*(1), 1–14. https://doi.org/10.1080/01972243.2017.1391913

Hern, A. (2021, 12 mei). *Online safety bill 'a recipe for censorship', say campaigners*. The Guardian. https://www.theguardian.com/media/2021/may/12/uk-to-require-social-media-to-protect-democratically-important-content

Herring, S., Job-Sluder, K., Scheckler, R., & Barab, S. (2002). Searching for Safety Online: Managing 'Trolling' in a Feminist Forum. *The Information Society*, *18*(5), 371–384. https://doi.org/10.1080/01972240290108186

Herweijer, K., & Ververs, C. (2020, 12 november). *Fenomeen pedojagen groeit: 'We komen even wat vragen stellen vriend'*. Telegraaf. https://www.telegraaf.nl/nieuws/1340573470/fenomeen-pedojagen-groeit-we-komen-even-wat-vragen-stellen-vriend

Hinson, L., Mueller, J., O'Brien-Milne, L., & Wandera, N. (2018). Technology-facilitated gender-based violence: What is it, and how do we measure it? *International Center for Research on Women*, 8.

Hobbs, R., & Grafe, S. (2015). YouTube pranking across cultures. *First Monday*, *20*(7). https://doi.org/10.5210/fm.v20i7.5981

Houtekamer, C., & Wassens, R. (2021, 2 april). *Het afvoerputje van het internet zit in een Noord-Hollands dorp*. NRC. https://www.nrc.nl/nieuws/2021/04/02/het-afvoerputje-van-het-internet-zit-in-een-noord-hollands-dorp-a4038329

Husting, G., & Orr, M. (2007). Dangerous Machinery: 'Conspiracy Theorist' as a Transpersonal Strategy of Exclusion. *Symbolic Interaction*, *30*(2), 127–150. https://doi.org/10.1525/si.2007.30.2.127

Index on Censorship. (2019, 5 april). The UK government's online harms white paper shows disregard for freedom of expression. *Index on Censorship*. https://www.indexoncensorship.org/2019/04/uk-government-online-harms-white-paper-shows-disregard-freedom-expression/

Instagram. (z.d.). *@pedohunterznl*. Instagram. https://www.instagram.com/hunterzprotectnl/

Instagram. (2021). *Continuing to Make Instagram Safer for the Youngest Members of Our Community*. Instagram. Https://about.instagram.com/blog/announcements/continuing-to-make-instagram-safer-for-the-youngest-members-of-our-community

Ipsos. (2020). *Trust misplaced?* https://www.ipsos.com/en/trust-misplaced

IVIR. (2020). *WODC-onderzoek: Voorziening voor verzoeken tot snelle verwijdering van onrechtmatige online content*. Instituut voor Informatierecht. https://www.ivir.nl/publicaties/download/WODC_voorziening_onrechtmatige_content.pdf

Jellinek. (2021, 21 april). *Hoeveel mensen in Nederland zijn verslaafd en hoeveel zijn er in behandeling?* Jellinek. https://www.jellinek.nl/vraag-antwoord/hoeveel-mensen-zijn-verslaafd-en-hoeveel-zijn-er-in-behandeling/

Juvonen, J., & Gross, E. V. (2008). Extending the School Grounds?—Bullying Experiences in Cyberspace. *Journal of School Health*, *78*(9), 496–505. https://doi.org/10.1111/j.1746-1561.2008.00335.x

Kaakinen, M., Sirola, A., Savolainen, I., & Oksanen, A. (2020). Impulsivity, internalizing symptoms, and online group behavior as determinants of online hate. *PLOS ONE*, 1-17. https://doi.org/10.1371/journal.pone.0231052

Kafka, P. (2021, 20 april). *Apple will let podcasters sell subscriptions and keep a cut*

for itself. Vox. https://www.vox.com/recode/2021/4/20/22394032/apple-podcast-subscription-plans

Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin Books.

Kaiser, J., Schmidt, C., Benkler, Y., Tilton, C., Etling, B., Roberts, H., Clark, J., & Faris, R. (2020, 21 oktober). *Mail-In Voter Fraud: Anatomy of a Disinformation Campaign | Berkman Klein Center*. https://cyber.harvard.edu/publication/2020/Mail-in-Voter-Fraud-Disinformation-2020

Kastrenakes, J. (2021, 9 februari). *Twitter's Jack Dorsey wants to build an app store for social media algorithms*. The Verge. https://www.theverge.com/2021/2/9/22275441/jack-dorsey-decentralized-app-store-algorithms

Katawazi, G., & Wagemakers, T. (2021, 11 mei). *Met 'online straatverbod' hoopt burgemeester Halsema online shamers harder aan te pakken*. AT5. https://www.at5.nl/artikelen/208616/met-online-straatverbod-hoopt-burgemeester-halsema-online-shamers-harder-aan-te-pakken

Khasawneh, A., Madathil, K. C., Dixon, E., Wiśniewski, P., Zinzow, H., & Roth, R. (2020). Examining the Self-Harm and Suicide Contagion Effects of the Blue Whale Challenge on YouTube and Twitter: Qualitative Study. *JMIR Mental Health*, *7*(6), e15973. https://doi.org/10.2196/15973

Kist, R. (2020, 25 september). *Adverteerders financieren valse informatie over corona*. NRC. https://www.nrc.nl/nieuws/2020/07/06/adverteerders-financieren-valse-info-corona-a4005111

Kist, R., & Van den Bos, M. (2021, 8 maart). *Hoe Nederlandse complotdenkers en virussceptici sociale media telkens te slim af zijn*. NRC. https://www.nrc.nl/nieuws/2021/03/08/onderzoek-valse-informatie-van-sociale-media-weren-lukt-lang-niet-altijd-a4034651

Kleijer, J. (2015, 16 april). *Shame sexting is een groepsding*. Bureau Jeugd & Media. https://www.bureaujeugdenmedia.nl/shame-sexting-is-een-groepsding/

Kliksafe. (2021). *homepage*. Kliksafe. https://www.kliksafe.nl/

Knieriem, P. (2021, 2 februari). *Politicus in Zeist zou trollen gebruiken om sociale media te beïnvloeden: 'Die Noortje bestaat helemaal niet'*. RTV Utrecht. https://www.rtvutrecht.nl/nieuws/2133571/?fb=true

Kohorst, M. A., Warad, D. M., Nageswara Rao, A. A., & Rodriguez, V. (2018). Obesity, sedentary lifestyle, and video games: The new thrombophilia cocktail in adolescents. *Pediatric Blood & Cancer*, *65*(7), e27041. https://doi.org/10.1002/pbc.27041

Konopka, B. (2021, 15 april). *Gdynia firm's 'Cyber Guardian' leading the way in combatting online violence*. https://www.thefirstnews.com/article/gdynia-firms-cyber-guardian-leading-the-way-in-combatting-hate-speech-online-21268

Kootstra, J. (2020, 10 december). *Sterke stijging anorexiapatiënten die helemaal stoppen met eten en drinken*. https://nos.nl/l/2360082

Kouwenhoven, A., & Logtenberg, H. (2017, 10 februari). *Hoe Denk met 'trollen' politieke tegenstanders monddood probeert te maken*. NRC. https://www.nrc.nl/nieuws/2017/02/10/de-trollen-van-denk-6641045-a1545547

Kraak, H. (2020, 22 oktober). *Hoe een 27-jarige rapper op pedofielen jaagt vanuit zijn huiskamer in Deventer*. de Volkskrant. https://www.volkskrant.nl/columns-opinie/hoe-een-27-jarige-rapper-op-pedofielen-jaagt-vanuit-zijn-huiskamer-in-deventer~bac60f27/

La Morgia, M., Mei, A., Sassi, F., & Stefa, J. (2021). The Doge of Wall Street: Analysis and Detection of Pump and Dump Cryptocurrency Manipulations. *ArXiv:2105.00733 [Cs]*. http://arxiv.org/abs/2105.00733

Laato, S., Islam, A. K. M. N., Islam, M. N., & Whelan, E. (2020). What drives unverified information sharing and cyberchondria during the COVID-19 pandemic? *European Journal of Information Systems*, *29*(3), 288–305. https://doi.org/10.1080/0960085X.2020.1770632

LaFrance, S. by A. (2020, 15 december). Facebook Is a Doomsday Machine. *The Atlantic*. https://www.theatlantic.com/technology/archive/2020/12/facebook-doomsday-machine/617384/

Lastdrager, E. (2018). *From Fishing to Phishing*. Gildeprint Drukkerijen. https://www.fraudehelpdesk.nl/vergroot-uw-kennis/from-fishing-to-phishing/

Lauckner, C., Truszczynski, N., Lambert, D., Kottamasu, V., Meherally, S., Schipani-McLaughlin, A. M., Taylor, E., & Hansen, N. (2019). "Catfishing," cyberbullying, and coercion: An exploration of the risks associated with dating app

use among rural sexual minority males. *Journal of Gay & Lesbian Mental Health*, *23*(3), 289–306. https://doi.org/10.1080/19359705.2019.1587729

Levey, T. G. (2018). *Sexual harassment online: shaming and silencing women in the digital age*. Lynne Rienner Publishers, Inc.

Lewis, R. (2018). Literature review on children and young people demonstrating technology-assisted harmful sexual behavior. *Aggression and Violent Behavior*, *40*, 1–11. https://doi.org/10.1016/j.avb.2018.02.011

Linnemann, E., & Melchior, M. (2017, 3 maart). *Zo gaan vrouwelijke opiniemakers om met online haat en discriminatie*. De Volkskrant. https://www.volkskrant.nl/wetenschap/zo-gaan-vrouwelijke-opiniemakers-om-met-online-haat-en-intimidatie~b1764a77/

Liu, W., Mirza, F., Narayanan, A., & Souligna, S. (2020). Is it possible to cure Internet addiction with the Internet? *AI & SOCIETY*, *35*(1), 245–255. https://doi.org/10.1007/s00146-018-0858-0

Lorenzo-Dus, N. (2017). “cause ur special”: Understanding trust and complimenting behaviour in online grooming discourse. *Journal of Pragmatics*, *112*, 67–82.

Lubach, A. (2020). *De online fabeltjesfuik | Zondag met Lubach (S12)*. https://www.youtube.com/watch?v=FLoR2Spftwg

Ludemann, D. (2018). /pol/emics: Ambiguity, scales, and digital discourse on 4chan. *Discourse, Context & Media*, *24*, 92–98. https://doi.org/10.1016/j.dcm.2018.01.010

MacAllister, J. M. (2016). The Doxing Dilemma: Seeking a Remedy for the Malicious Publication of Personal Information. *Fordham Law Review*, *85*, 2451–2483.

MacCarthy, M. (2021, 11 mei). The Facebook Oversight Board’s failed decision distracts from lasting social media regulation. *Brookings*. https://www.brookings.edu/blog/techtank/2021/05/11/the-facebook-oversight-boards-failed-decision-distracts-from-lasting-social-media-regulation/

Machkovech, S. (2016, 15 november). *Twitter bots can reduce racist slurs—if people think the bots are white*. Arstechnica. https://arstechnica.com/science/2016/11/twitter-bots-can-reduce-racist-slurs-if-people-think-the-bots-are-white/

Make Media Great Again. (2021). *Doe mee aan onze collaboratieve media beweging!* Make Media Great Again. https://www.mmga.io/

Marwick, A. (2018). WHY DO PEOPLE SHARE FAKE NEWS? A SOCIOTECHNICAL MODEL OF MEDIA EFFECTS. *GEORGETOWN LAW TECHNOLOGY REVIEW*, *2*(2), 474–512.

Mendes, K., Ringrose, J., & Keller, J. (2018). #MeToo and the promise and pitfalls of challenging rape culture through digital feminist activism. *European Journal of Women's Studies*, *25*(2), 236–246. https://doi.org/10.1177/1350506818765318

Merton, R. K. (1968). The Matthew Effect in Science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

MiND. (2020). *Discriminatiecijfers in 2019* (p. 88). Art.1. https://www.mindnederland.nl/wp-content/uploads/2020/04/Discriminatiecijfers-in-2019-1.pdf

Ministerie van Justitie en Veiligheid. (2019). *Kamerbrief 2018D24515*. https://zoek.officielebekendmakingen.nl/kst-31015-177.html

Ministerie van Justitie en Veiligheid. (2020, 3 december). *Strafvorderingsrichtlijn misbruik seksueel beeldmateriaal - Nieuwsbericht - Openbaar Ministerie* [Nieuwsbericht]. Openbaar Ministerie. https://www.om.nl/actueel/nieuws/2020/12/03/strafvorderingsrichtlijn-misbruik-seksueel-beeldmateriaal

Ministerie van Justitie en Veiligheid. (2021). *Dadermonitor seksueel geweld tegen kinderen 2015-2019 - Rapport - Nationaal Rapporteur* [Rapport]. https://www.nationaalrapporteur.nl/publicaties/rapporten/2021/06/08/dadermonitor-seksueel-geweld-tegen-kinderen-2015-2019

Mink, I., & Van Bon, S. (2017). Discriminatiecijfers 2016. *Artikel 1*. https://www.art1.nl/publicaties/landelijke-meldingen-discriminatie-2016/

Miserus, M., & Van der Noordaa, R. (2018). *Het trollenleger van Dotan*. De Volkskrant. https://www.volkskrant.nl/kijkverder/2018/dotan/#/

Möller, J., Helberger, N., & Makhortykh, M. (2019). *Filter bubbles in the*

Netherlands? https://dare.uva.nl/search?identifier=2d8db249-cb3a-4eae-b514-56897c08a2d6

Montag, C., Yang, H., & Elhai, J. D. (2021). On the Psychology of TikTok Use: A First Glimpse From Empirical Findings. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.641673

Moonshot. (2021). *Redirect Method*. Moonshot. Https://moonshotcve.com/redirect-method/

Morozov, E. (2021, 7 mei). *Spoiler: Zo innovatief zijn Apple en Google helemaal niet*. De Correspondent. https://decorrespondent.nl/12349/spoiler-zo-innovatief-zijn-apple-en-google-helemaal-niet/18470148096642-19468a5f

Morris, S. (2021, 3 juni). *21 Dangerous TikTok Trends Every Parent Should Be Aware of*. Newsweek. https://www.newsweek.com/21-dangerous-tiktok-trends-that-have-gone-viral-1573734

Mortimer, K. (2017). Understanding Conspiracy Online: Social Media and the Spread of Suspicious Thinking. *Dalhousie Journal of Interdisciplinary Management*, *13*(1). https://doi.org/10.5931/djim.v13i1.6928

Mosley, M. A., Lancaster, M., Parker, M. L., & Campbell, K. (2020). Adult attachment and online dating deception: a theory modernized. *Sexual and Relationship Therapy*, *35*(2), 227–243. https://doi.org/10.1080/14681994.2020.1714577

Motivaction. (2021, 2 september). *Ondanks discussie over nepnieuws: groeiende meerderheid vertrouwt journalistiek wél*. https://www.motivaction.nl/kennisplatform/nieuws-en-persberichten/ondanks-discussie-over-nepnieuws-groeiende-meerderheid-vertrouwt-journalistiek-wel

Movisie. (z.d.). *Campagnepagina #DatMeenJeNiet*. Movisie. Geraadpleegd 14 juni 2021, van https://www.movisie.nl/campagnepagina-datmeenjeniet

Movisie. (2019). *Shame sexting bij tienermeiden met een Marokkaans-islamitische achtergrond*. Movisie. https://www.movisie.nl/artikel/shame-sexting-tienermeiden-marokkaans-islamitische-achtergrond

Müller, K. W., Janikian, M., Dreier, M., Wölfling, K., Beutel, M. E., Tzavara, C., Richardson, C., & Tsitsika, A. (2015). Regular gaming behavior and internet gaming disorder in European adolescents: results from a cross-national representative

survey of prevalence, predictors, and psychopathological correlates. *European Child & Adolescent Psychiatry*, *24*(5), 565–574. https://doi.org/10.1007/s00787-014-0611-2

Multiscope. (2020, 13 februari). *Nederlanders gamen dagelijks half miljard minuten*.

Multiscope. http://www.multiscope.nl/persberichten/nederlanders-gamen-dagelijks-half-miljard-minuten.html

Munn, L. (2021). More than a Mob: Parler as Preparatory Media for the Capitol Storming. *First Monday*, *26*(3). https://doi.org/10.5210/fm.v26i3.11574

Naughton, J. (2015, 7 februari). *Aaron Swartz stood up for freedom and fairness – and was hounded to his death*. The Guardian. http://www.theguardian.com/commentisfree/2015/feb/07/aaron-swartz-suicide-internets-own-boy

NCTV, M. van J. en. (2021, 14 april). *Fenomeenanalyse 'De verschillende gezichten van de coronaprotesten' - Publicatie - Nationaal Coördinator Terrorismebestrijding en Veiligheid* [Publicatie]. NCTV. https://www.nctv.nl/documenten/publicaties/2021/04/14/fenomeenanalyse-de-verschillende-gezichten-van-de-coronaprotesten

Nelson, J. L., & Taneja, H. (2018). The small, disloyal fake news audience: The role of audience availability in fake news consumption. *New Media & Society*, *20*(10), 3720–3737. https://doi.org/10.1177/1461444818758715

Net Nanny. (2021). *homepage*. Net Nanny. https://www.netnanny.com/.

Netwerk Mediawijsheid. (2021). *Netwerk Mediawijsheid*. Netwerk Mediawijsheid. https://www.netwerkmediawijsheid.nl

Newton, C. (2021a, 1 april). *Nick Clegg tries to reset the conversation*. Platformer. https://www.platformer.news/p/nick-clegg-tries-to-reset-the-conversation

Newton, C. (2021b, 9 april). *The case of the missing platform policies*. https://www.platformer.news/p/the-case-of-the-missing-platform

Newton, C. (2021c, 16 juni). *Why Google's FLoC flopped*. https://www.platformer.news/p/why-googles-floc-flopped

Nikolaou, D. (2017). Does cyberbullying impact youth suicidal behaviors? *Journal of Health Economics*, *56*, 30–46. https://doi.org/10.1016/j.jhealeco.2017.09.009

NJi. (2019). *Eetstoornissen - Cijfers | NJi*. Nederlands Jeugdinstituut. https://www.nji.nl/nl/Databank/Cijfers-over-Jeugd-en-Opvoeding/Cijfers-per-onderwerp/Eetstoornissen

NLProfiel. (2020, 9 september). *NLProfiel – samenwerkende Nederlandse uitgevers op gebied van targeting in digitale media*. https://nlprofiel.nl/

NOS. (2018, 29 juli). *Challenge met rijdende auto: 'Volslagen idioot en strafbaar'*. https://nos.nl/l/2243726

NOS. (2021a, 11 januari). *Zwarte lijst met namen van artsen verboden door rechter*. https://nos.nl/l/2363920

NOS. (2021b, 6 augustus). *Hoe de cryptohandel gemanipuleerd wordt: 'Het werkt het best bij onervaren mensen'*. https://nos.nl/l/2379563

noticeandtakedowncode.nl/. (2018). *Gedragscode Notice-and-Take-Down*.

noticeandtakedowncode.nl/. https://noticeandtakedowncode.nl/wp-content/uploads/2018/12/ECP_01054-Gedragscode-notice-and-takedown-pdf-2.pdf

NRC. (2021, 14 januari). *Waarom radicaliseren mensen?* NRC. https://www.nrc.nl/nieuws/2021/01/14/gevaar-vernauwt-het-denken-a4027567

Nu.nl. (2020, 18 november). *Vijf minderjarigen aangehouden voor 'happy slapping' in Amsterdam*. NU. https://www.nu.nl/amsterdam/6091203/vijf-minderjarigen-aangehouden-voor-happy-slapping-in-amsterdam.html

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2015). Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems. *Social Science Computer Review*, *33*(4), 459–478. https://doi.org/10.1177/0894439314555329

Okuna. (2021). *Home*. Okuna. https://about.okuna.io/en/home.

Oleshchuk, P. (2020). The Instruments of Modern Media Lobbying. *Future Human Image*, *14*, 48–55. https://doi.org/10.29202/fhi/14/6

Oosterveer, D. (2021, 23 januari). *Social media in Nederland 2021: TikTok-gebruik door jongeren stijgt explosief en passeert Facebook*. Marketingfacts. https://www.marketingfacts.nl/berichten/social-media-in-nederland-2021

Oosterwijk, K., & Fischer, T. F. C. (2017). *Interventies jeugdige daders cybercrime*. WODC. https://veiligheidscoalitie.nl/action/?action=download&id=2386

Ortiz, S. M. (2019). “You Can Say I Got Desensitized to It”: How Men of Color Cope with Everyday Racism in Online Gaming. *Sociological Perspectives*, *62*(4), 572–588. https://doi.org/10.1177/0731121419837588

Pariser, E. (2012). *The filter bubble: what the Internet is hiding from you*. Penguin Books.

Parler. (2021). *Community guidelines*. https://legal.parler.com/documents/guidelines.pdf

Paul, K. (2021, 5 mei). *Facebook ruling on Trump renews criticism of oversight board*. The Guardian. http://www.theguardian.com/technology/2021/may/05/facebook-oversight-board-donald-trump

Paulissen, & Van Wilsem. (2015). *Politie en Wetenschap*. https://www.politieenwetenschap.nl/publicatie/politiewetenschap/2015/dat-heeft-iemand-anders-gedaan-259/

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, *592*(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Peterson, J., & Densley, J. (2017). Cyber violence: What do we know and where do we go from here? *Aggression and Violent Behavior*, *34*, 193–200. https://doi.org/10.1016/j.avb.2017.01.012

Petities.nl. (2021, 31 maart). *Stop chemtrails nu en weermanipulatie nu*. Petities.nl. https://petities.nl/petitions/stop-chemtrails-nu-en-weermanipulatie-nu?locale=nl

Phillips, W. (2015). *This is why we can’t have nice things: mapping the relationship between online trolling and mainstream culture*. The MIT Press.

Plan International. (2020, 5 oktober). *Wereldwijd 58 procent van de meisjes*

slachtoffer van online intimidatie. Plan International. https://www.planinternational.nl/actueel/wereldwijd-58-procent-van-de-meisjes-slachtoffer-van-online-intimidatie

Pointer. (2021a, 21 maart). *De invloed van sociale media op de verkiezingscampagne*. KRO-NCRV. https://pointer.kro-ncrv.nl/de-invloed-van-sociale-media-op-de-verkiezingscampagne

Pointer. (2021b, 20 mei). *Nederlands trollenleger verspreidt en coördineert desinformatie over vaccin*. KRO-NCRV. https://pointer.kro-ncrv.nl/nederlands-trollenleger-verspreidt-en-coordineert-desinformatie-over-vaccin

Policy Department for Citizens' Rights and Constitutional Affairs. (2020). *Hate speech and hate crime in the EU and the evaluation of online content regulation approaches*. Europees Parlement.

Politie, Openbaar ministerie, Regioburgemeesters, & J&V. (2020). *Position paper: de politie van morgen en overmorgen - Publicatie - Openbaar Ministerie* [Publicatie]. https://www.om.nl/documenten/publicaties/om-onderdelen/pag-om/map/position-paper-de-politie-van-morgen-en-overmorgen

Pomerantsev, P. (2019). *A Cycle of Censorship: The UK White Paper on Online: Harms and the Dangers of Regulating Disinformation†* (Working papers of the Transatlantic Working Group on Content Moderation Online and Freedom of Expression). Instituut voor Informatierecht. https://www.ivir.nl/publicaties/download/Cycle_Censorship_Pomerantsev_Oct_2019.pdf

Powell, A. (2015). Seeking rape justice: Formal and informal responses to sexual violence through technosocial counter-publics. *Theoretical Criminology*, *19*(4), 571–588. https://doi.org/10.1177/1362480615576271

Prij, J., & Janssens, M. (2020, 12 juni). *Nepnieuws: graag een beetje bezonnenheid!* Christen Democratische Verkenningen. https://www.tijdschriftcdv.nl/inhoud/tijdschrift_artikel

Quekel, S. (2021, 15 februari). *Nieuwe vorm van identiteitsfraude rukt op: 'Foto's verschijnen op pornosite'*. Algemeen Dagblad. https://www.ad.nl/tech/nieuwe-vorm-van-identiteitsfraude-rukt-op-foto-s-verschijnen-op-pornosite~aafb7609/

Rabkin, M. (2021, 18 maart). *Facebook shows its upcoming social VR avatars for*

Horizon at SXSW [CNET]. https://www.cnet.com/news/facebook-shows-its-upcoming-social-vr-avatars-for-horizon-at-sxsw/

Rasch, M. (2021, 18 april). *Zelfs na je dood laat Big Tech je niet met rust*. Follow the Money - Platform voor onderzoeksjournalistiek. https://www.ftm.nl/artikelen/dood-big-tech-deepfake

Raskauskas, J., & Stoltz, A. D. (2007). Involvement in traditional and electronic bullying among adolescents. *Developmental Psychology*, *43*(3), 564–575. https://doi.org/10.1037/0012-1649.43.3.564

Rathenau Instituut. (2018a). *Digitalisering van het nieuws: online nieuwsgedrag en personalisatie in Nederland.* https://www.rathenau.nl/nl/digitale-samenleving/digitalisering-van-het-nieuws

Rathenau Instituut. (2018b). *Vertrouwen in de wetenschap | Rathenau Instituut*. https://www.rathenau.nl/nl/wetenschap-cijfers/impact/vertrouwen-de-wetenschap/vertrouwen-de-wetenschap

Rathenau Instituut. (2020a). *Cyberweerbaar met nieuwe technologie*. Rathenau Instituut (auteurs: Boheemen, P. van, G. Munnichs, L. Kool, G. Diercks, J. Hamer & A. Vos).

Rathenau Instituut. (2020b). *Digitale dreigingen voor de democratie: Over nieuwe technologie en desinformatie*. Rathenau Instituut (auteurs: Boheemen, P. van, G. Munnichs & E. Dujso).

Rathenau Instituut. (2020c). *Rathenau manifest: Stel nu 10 ontwerpeisen aan de digitale samenleving van morgen*.

Rathenau Instituut. (2021a). *Reactie Rathenau Instituut op Consultatie AIV advies Regulering van online content*. https://www.rathenau.nl/sites/default/files/2021-03/Reactie_Rathenau_Instituut_Consultatie_AIV_advies_Regulering_online_content.pdf

Rathenau Instituut. (2021b). *De toekomst van online platformen: Twee Europese wetsvoorstellen onder de loep*. https://www.rathenau.nl/sites/default/files/2021-05/Rathenau_Instituut_Bericht_aan_parlement_De_toekomst_van_online_platformen.pdf

Redactie NOS. (2021, 25 februari). *Strijden tegen online shaming en expose*

accounts. NPO Radio 1. https://www.nporadio1.nl/binnenland/29898-strijden-tegen-online-shaming-en-expose-accounts

Redactie ScienceGuide. (2021, 28 mei). *Nog voor zomerreces wetsvoorstel tegen "walgelijke" doxing Vizier op Links*. ScienceGuide. https://www.scienceguide.nl/2021/05/nog-voor-zomerreces-wetsvoorstel-tegen-walgelijke-doxing-vizier-op-links/

Reddit. (2021a). *Reddit Content Policy*. Reddit. https://www.redditinc.com/policies/content-policy

Reddit. (2021b). *Ways to become a moderator*. Https://mods.reddithelp.com/hc/en-us/articles/360001745332-Ways-to-become-a-moderator

Reuters. (2020). *Overview and Key Findings of the 2020 Digital News Report*.

Reuters Institute Digital News Report. https://www.digitalnewsreport.org/survey/2020/overview-key-findings-2020/

Rijksoverheid. (z.d.). *Wraakporno*. Rijksoverheid.nl. Geraadpleegd 18 juni 2021, van https://www.rijksoverheid.nl/onderwerpen/seksuele-misdrijven/wraakporno

ROB. (2019). *Adviesrapport Zoeken naar waarheid - Publicatie - Raad voor het Openbaar Bestuur* [Publicatie]. https://www.raadopenbaarbestuur.nl/documenten/publicaties/2019/05/09/zoeken-naar-waarheid

Rogers, R., & Niederer, S. (2019). *The Politics of Social Media Manipulation*. Den Haag: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Roose, K. (2020). Rabbit Hole. *The New York Times*. https://www.nytimes.com/column/rabbit-hole

RTL. (2018, 22 mei). *Clay (15) overleden door 'onnozele challenge': 'Zijn verlies is afgrijselijk'*. RTL Nieuws. https://www.rtlnieuws.nl/nieuws/nederland/artikel/4200111/clay-15-overleden-door-onnozele-challenge-zijn-verlies

RTL. (2019, 28 februari). *Bitcoin-oplichters verdienen tonnen aan Nederlandse slachtoffers*. RTL Nieuws. https://www.rtlnieuws.nl/tech/artikel/4625991/bitcoin-scam-oplichting-broker-trading-investering-4-procent-rendement

RTL. (2021, 19 maart). *RTL Nieuws sluit reacties over Sylvana Simons na baggertsunami*. Mediacourant.nl. https://www.mediacourant.nl/2021/03/rtl-nieuws-sluit-reacties-over-sylvana-simons-na-baggertsunami/

RTL Nieuws. (2017, 4 mei). *Stel raakt voogdij over kinderen kwijt na schokkende prank-video's op YouTube*. RTL Nieuws. https://www.rtlnieuws.nl/editienl/artikel/100621/stel-raakt-voogdij-over-kinderen-kwijt-na-schokkende-prank-videos-op

RTL nieuws. (2020, 24 augustus). *Verslaafd aan je smartphone: 'Het is een ledemaat geworden'*. RTL. https://www.rtlnieuws.nl/editienl/artikel/5179060/verslaafd-aan-smartphone-nomofobie-ledemaat-telefoon

RTL Nieuws. (2021, 4 mei). *Tot 12 maanden cel voor 'pedojagers' na fatale mishandeling 73-jarige man*. RTL Nieuws. https://www.rtlnieuws.nl/nieuws/nederland/artikel/5228920/pedojagen-arnhem-mishandeling-dood-man-73-uitspraak

Rutgers. (2018, 4 september). *Sexting: Praat met jongeren over de gevaren en risico's*. Sexting: Praat met jongeren over de gevaren en risico's. https://www.rutgers.nl/nieuws-opinie/nieuwsarchief/sexting-praat-met-jongeren-over-de-gevaren-en-risico%E2%80%99s

Rutgers, & Soa Aids Nederland. (2019). *Seks onder je 25e VSO 2019*. Seks onder je 25e. https://seksonderje25e.nl/vso

Sabel, P., & Verhagen, L. (2021, 5 maart). *Politici zijn het doelwit van tientallen haattweets per dag – wie zit erachter?* De Volkskrant. https://www.volkskrant.nl/nieuws-achtergrond/politici-zijn-het-doelwit-van-tientallen-haattweets-per-dag-wie-zit-erachter~b6e2744e/

Sánchez Montañés, M. (2021, 14 april). *Big tech cannot crack down on online hate alone*. World Economic Forum. https://www.weforum.org/agenda/2021/04/big-tech-cannot-crack-down-on-online-hate-alone/

Sanders, M. (2021). *Owner Identity and Interdependent Markets: an examination of ownership filters of institutional complexity, coalitional change and value creation in disrupted two sided market categories*. https://repub.eur.nl/pub/135457

Saris, K., & Van de Ven, C. (2021, 3 maart). *De online haat dreigt vrouwen uit de*

politieke arena te verdrijven. De Groene Amsterdammer. https://www.groene.nl/artikel/misogynie-als-politiek-wapen

Schildkamp, V., & Rodenburg, F. (2021, 21 februari). *Begraafplaats Bodegraven doelwit van complotdenkers over pedo-netwerk, ook RIVM deed al aangifte*. Algemeen Dagblad.

SCP. (2016). *Resultaten*. https://www.mediatijd.nl/tijdsbesteding/resultaten

Shachaf, P., & Hara, N. (2010). Beyond vandalism: Wikipedia trolls. *Journal of Information Science*, *36*(3), 357–370. https://doi.org/10.1177/0165551510365390
Shaikh, F. B., Rehman, M., & Amin, A. (2020). Cyberbullying: A Systematic Literature Review to Identify the Factors Impelling University Students Towards Cyberbullying. *IEEE Access*, *8*, 148031–148051. https://doi.org/10.1109/ACCESS.2020.3015669

Shanahan, J. (2021, 5 maart). Support for QAnon is hard to measure — and polls may overestimate it. *Nieman Lab*. https://www.niemanlab.org/2021/03/support-for-qanon-is-hard-to-measure-and-polls-may-overestimate-it/

Shea, V. (1994). *Netiquette* (Ed. 1.0). Albion Books.

Simons, E. I., Nooteboom, F., & Van Furth, E. F. (2020). *De Wereld van Pro-ana Coaches*. https://www.hetckm.nl/nieuws-en-publicaties/pro-ana-coaches-maken-bewust-misbruik-van-meisjes-met-eetstoornis/1

Sipma, T., & Leijsen, E. M. C. van. (2019). Slachtofferschap van online criminaliteit. *Den Haag*. https://repository.wodc.nl/handle/20.500.12832/236

Slecht Nieuws. (2021). *intro*. Slecht Nieuws. https://www.slechtnieuws.nl/#intro
SOS. (2021). *Rechter verbiedt zwarte lijst artsen – Stop Online Shaming* [Stichting Online Shaming]. https://www.stoponlineshaming.org/rechter-verbiedt-zwarte-lijst-artsen/

Ster. (2020, 8 december). *Online adverteren 2,5 jaar na de verscherping van de privacywet: wat zijn de lessen? - Ster reclame* [Blog]. Ster.nl. https://www.ster.nl/nieuws/online-adverteren-2-5-jaar-na-de-verscherping-van-de-privacywet-wat-zijn-de-lessen/

Sternisko, A., Cichocka, A., & Van Bavel, J. J. (2020). The dark side of social movements: social identity, non-conformity, and the lure of conspiracy theories.

Current Opinion in Psychology, *35*, 1–6.
https://doi.org/10.1016/j.copsyc.2020.02.007

Stichting Internet Challenges, W. (2021). *https://www.internetchallenges.nl/de-stichting*. https://www.internetchallenges.nl/de-stichting

Stil, H. (2020, 19 januari). *Hoe de smartphone ons leven in kroop*. Het Parool. https://www.parool.nl/nieuws/hoe-de-smartphone-ons-leven-in-kroop~b35ae634/

Stolton, S. (2020, 5 juni). EU code of practice on disinformation 'insufficient and unsuitable,' member states say. *Www.Euractiv.Com*. https://www.euractiv.com/section/digital/news/eu-code-of-practice-on-disinformation-insufficient-and-unsuitable-member-states-say/

Stop Hate for For Profit. (2021). *#StopHateForProfit*. Stop Hate For Profit. https://www.stophateforprofit.org/

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *7*(3), 321–326. https://doi.org/10.1089/1094931041291295

Superawesome. (2021). *Homepage*. Superawesome. Https://www.superawesome.com/.

SVDJ. (2021, 18 maart). Nick Waters (Bellingcat): 'Geweld tegen journalisten neemt toe'. *SVDJ*. https://www.svdj.nl/nick-waters-bellingcat-geweld-tegen-journalisten-neemt-toe/

The Guardian. (2020, 16 oktober). *QAnon: a timeline of violence linked to the conspiracy theory*. The Guardian. http://www.theguardian.com/us-news/2020/oct/15/qanon-violence-crimes-timeline

The Independent. (2019, 16 maart). *How nonsensical white genocide conspiracy theory cited by alleged gunman is spreading poison around the world*. The Independent. https://www.independent.co.uk/news/world/australasia/new-zealand-christchurch-mosque-attack-white-genocide-conspiracy-theory-a8824671.html
Their Tube. (2021). *homepage*. Their Tube. Http://www.their.tube/.

Tik Tok. (2020 december). *Community Guidelines*. Tik Tok. https://www.tiktok.com/community-guidelines

Tokmetzis, D. (2020, 7 januari). *De schaduwzijde van cryptovaluta: er is al voor 15*

miljard euro opgelicht en gestolen. De Correspondent. https://decorrespondent.nl/10826/de-schaduwzijde-van-cryptovaluta-er-is-al-voor-15-miljard-euro-opgelicht-en-gestolen/527193722-4276042b

Tokmetzis, D., & Bol, R. (2020, 3 november). *De macht van bedrijven als Google en Apple is gigantisch. Zo trekken Europa en de VS de teugels aan*. De Correspondent. https://decorrespondent.nl/11732/de-macht-van-bedrijven-als-google-en-apple-is-gigantisch-zo-trekken-europa-en-de-vs-de-teugels-aan/571313204-b1c40942

Tumber, H., & Waisbord, S. (2021). *The Routledge Companion to Media Disinformation and Populism*. Routlegde.

Tweede Kamer. (2018). *Kamerstuk II, 31015, nr. 175*. https://zoek.officielebekendmakingen.nl/kst-31015-175.html

Tweede Kamer. (2020). *Kamerstuk II, 30821, nr. 120*. https://zoek.officielebekendmakingen.nl/kst-30821-120.html

Twitter. (2021). *Rules and policies*. Https://help.twitter.com/en/rules-and-policies#twitter-rules

UK Government. (2019). *Online Harms White Paper*. GOV.UK. https://www.gov.uk/government/consultations/online-harms-white-paper

UK Government. (2020a). *Interim code of practice on online child sexual exploitation and abuse*. UK Government. https://www.gov.uk/government/publications/online-harms-interim-codes-of-practice/interim-code-of-practice-on-online-child-sexual-exploitation-and-abuse-accessible-version

UK Government. (2020b). *Online Harms White Paper: Full government response to the consultation*. GOV.UK. https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response
University of Oxford. (2020, 22 mei). *Conspiracy beliefs reduce the following of government coronavirus guidance | University of Oxford*. University of Oxford. https://www.ox.ac.uk/news/2020-05-22-conspiracy-beliefs-reduces-following-government-coronavirus-guidance

Van Baars, L. (2020, 25 september). *De tijd die kinderen achter een scherm doorbrengen, is verdubbeld naar ruim 7 uur per dag*. Trouw. https://myprivacy.dpgmedia.nl/consent?siteKey=w38GrtRHtDg4T8xq&callbackUrl=

https%3a%2f%2fwww.trouw.nl%2fprivacy-wall%2faccept%3fredirectUri%3d%252fbinnenland%252fde-tijd-die-kinderen-achter-een-scherm-doorbrengen-is-verdubbeld-naar-ruim-7-uur-per-dag%257ebe89f15d%252f

Van Bemmel, N. (2020, 12 juni). *Twitter verwijdert tienduizenden accounts wegens Chinese, Turkse en Russische staatspropaganda*. De Volkskrant. https://www.volkskrant.nl/nieuws-achtergrond/twitter-verwijdert-tienduizenden-accounts-wegens-chinese-turkse-en-russische-staatspropaganda~bedd1f53/

Van de Weijer, S. G. A., Leukfeldt, E. R., & Van Der Zee, S. (2020). *Slachtoffer van onlinecriminaliteit, wat nu? Een onderzoek naar aangiftebereidheid onder burgers en ondernemers*. https://www.politieenwetenschap.nl/publicatie/politiewetenschap/2020/slachtoffer-van-onlinecriminaliteit-wat-nu-356/

Van de Weijer, S. G. A., Leukfeldt, R., & Bernasco, W. (2019). Determinants of reporting cybercrime: A comparison between identity theft, consumer fraud, and hacking. *European Journal of Criminology*, *16*(4), 486–508. https://doi.org/10.1177/1477370818773610

Van den Berg, I. (2021, 5 februari). *Journalistiek kost geld. Wie betaalt?* OneWorld. https://www.oneworld.nl/lezen/achtergrond/journalistiek-kost-geld-wie-betaalt/

Van der Poel, R., & Luyendijk, W. (2021, 24 maart). *Vader van Nora vraagt zich af: wat als je begint met luisteren naar een meisje met anorexia?* NRC. https://www.nrc.nl/nieuws/2021/03/24/ik-ging-stapje-voor-stapje-mee-en-werd-zo-medeplichtig-a4037178

Van Furth, E., Hemkes, S., & Dingemans, A. (2011). Het fenomeen Pro-ana. *Psychopraktijk*, *3*(5), 35–37. https://doi.org/10.1007/s13170-011-0075-8

Van Houwelingen, K. (2017, 17 augustus). *'Iedereen zal weten wie deze types zijn'*. De Gelderlander. https://advance-lexis-com.proxy.uba.uva.nl:2443/document/?pdmfid=1516831&crid=6022c8b6-8318-4a6d-888f-410788641f18&pddocfullpath=%2Fshared%2Fdocument%2Fnews%2Furn%3AcontentItem%3A5P8C-0CK1-DYRY-X16T-00000-00&pdcontentcomponentid=149018&pdteaserkey=sr43&pditab=allpods&ecomp=5bq2k&earg=sr43&prid=8786c04d-cd46-428e-a032-84325da6df62

Van Noort, W. (2020, 6 augustus). *Je mening is niet slecht, jij bent slecht, waarom*

online shamen niet werkt. NRC. https://www.nrc.nl/nieuws/2020/08/06/je-mening-is-niet-slecht-jij-bent-slecht-waarom-online-shamen-niet-werkt-a4008036

Van Rooij, A. J., Schoenmakers, T. M., van den Eijnden, R. J. J. M., & van de Mheen, D. (2012). Online video gameverslaving: verkenning van een nieuw fenomeen. *Tijdschrift voor gezondheidswetenschappen*, *90*(7), 420–426. https://doi.org/10.1007/s12508-012-0146-1

Vanheste, T. (2021, 30 september). *Hoe vult Europa het verlangen naar technologische soevereiniteit in?* Rathenau Instituut. https://www.rathenau.nl/nl/vitale-kennisecosystemen/hoe-vult-europa-het-verlangen-naar-technologische-soevereiniteit

Vayansky, I., & Kumar, S. (2018). Phishing – challenges and solutions. *Computer Fraud & Security*, *2018*(1), 15–20. https://doi.org/10.1016/S1361-3723(18)30007-1

Veldhuis, P., & Ingabire, S. (2021, 2 mei). *Het was 'sensatiezucht' en 'dom kuddegedrag', maar de pedojacht had een fatale afloop*. NRC. https://www.nrc.nl/nieuws/2021/05/02/het-was-sensatiezucht-en-dom-kuddegedrag-maar-de-pedojacht-had-een-fatale-afloop-a4042134

Vie publique. (2020, 29 juni). *Loi du 24 juin 2020 visant à lutter contre les contenus haineux sur internet*. Vie publique.fr. https://www.vie-publique.fr/loi/268070-loi-avia-lutte-contre-les-contenus-haineux-sur-internet

Vince, G. (2018, 3 april). *Evolution explains why we act differently online*. BBC. https://www.bbc.com/future/article/20180403-why-do-people-become-trolls-online-and-in-social-media

Vinocur, N. (2021, 2 april). *The movement to end targeted internet ads*. POLITICO. https://www.politico.eu/article/targeted-advertising-tech-privacy/

Visser, M. (2020, 15 augustus). *Een op de tien Nederlanders gelooft dat rond corona vieze spelletjes worden gespeeld*. Trouw. https://www.trouw.nl/binnenland/een-op-de-tien-nederlanders-gelooft-dat-er-rond-corona-vieze-spelletjes-worden-gespeeld~bd98ce41/

Vogels, E. A. (2021, 13 januari). The State of Online Harassment. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/

Völlink, T., Dehue, F., & Guckin, C. M. (2016). *Cyberbullying: From Theory to Intervention*. Routledge.

Von Piekartz, H. (2020, 15 oktober). *NCTV: boosheid en ongenoegen over coronamaatregelen leiden vaker tot radicalisering*. De Volkskrant. https://www.volkskrant.nl/nieuws-achtergrond/nctv-boosheid-en-ongenoegen-over-coronamaatregelen-leiden-vaker-tot-radicalisering~baa3e15a/

Von Piekartz, H., & Bahara, H. (2021, 26 maart). *Doxing: hoe online dreigementen hun weg vinden naar de fysieke wereld*. https://www.volkskrant.nl/nieuws-achtergrond/doxing-hoe-online-dreigementen-hun-weg-vinden-naar-de-fysieke-wereld~be39da12/

VRT. (2020, 18 augustus). *Chatgroep die holebi's viseert, is niet enige die oproept tot haat en geweld: 'Er bestaan er tientallen bij ons'*. vrtnws.be. https://www.vrt.be/vrtnws/nl/2020/08/17/chatgroepen-die-aanzetten-tot-homohaat-en-geweld-tegen-lgbt-s/

Wagemakers, T., & Toksöz, Z. (2021, 13 mei). *Als die ene seksfoto van lang geleden je blijft achtervolgen*. NRC. https://www.nrc.nl/nieuws/2021/05/13/als-die-ene-seksfoto-van-lang-geleden-je-nog-steeds-achtervolgt-a4043259

Wagner, K. (2020, 9 november). *Facebook Labeled 167 Million User Posts for Covid Misinformation*. Bloomberg. https://www.bloomberg.com/tosv2.html?vid=&uuid=a2f841f0-a987-11eb-8bca-4308fa876329&url=L25ld3MvYXJ0aWNsZXMvMjAyMC0xMS0xOS9mYWNlYm9vay1sYWJlbGVkLTE2Ny1taWxsaW9uLXVzZXItcG9zdHMtZm9yLWNvdmlkLW1pc2luZm9ybWF0aW9u

Weimann, G., & Masri, N. (2020). Research Note: Spreading Hate on TikTok. *Studies in Conflict & Terrorism*. https://doi.org/10.1080/1057610X.2020.1780027

Weinstein, A., & Lejoyeux, M. (2010). Internet Addiction or Excessive Internet Use. *The American Journal of Drug and Alcohol Abuse*, *36*(5), 277–283. https://doi.org/10.3109/00952990.2010.491880

Whittle, H., Hamilton-Giachritsis, C., Beech, A., & Collings, G. (2013). A review of online grooming: Characteristics and concerns. *Aggression and Violent Behavior*, *18*(1), 62–70. https://doi.org/10.1016/j.avb.2012.09.003

Wiegman, M. (2016, 23 mei). *Alledaags racisme vervuilt het debat*. Het Parool. https://www.parool.nl/nieuws/alledaags-racisme-vervuilt-het-debat~ba98ca17/

Wingfield, N. (2014, 15 oktober). *Feminist Critics of Video Games Facing Threats in "Gamergate" Campaign*. The New York TImes. https://www.nytimes.com/2014/10/16/technology/gamergate-women-video-game-threats-anita-sarkeesian.html

Wong, J. C. (2021, 16 januari). *Banning Trump won't fix social media: 10 ideas to rebuild our broken internet – by experts*. The Guardian. http://www.theguardian.com/media/2021/jan/16/how-to-fix-social-media-trump-ban-free-speech

Yam, K. C., & Reynolds, S. J. (2016). The Effects of Victim Anonymity on Unethical Behavior. *Journal of Business Ethics*, *136*(1), 13–22. https://doi.org/s10551-014-2367-5

YouTube. (2019, 16 januari). *Announcement: Strengthening enforcement of our Community Guidelines - YouTube Community*. Announcement: Strengthening enforcement of our Community Guidelines. https://support.google.com/youtube/thread/1063296/%F0%9F%9A%A9-announcement-strengthening-enforcement-of-our-community-guidelines?hl=en

Zheng, H., Sin, S.-C. J., Kim, H. K., & Theng, Y.-L. (2020). Cyberchondria: a systematic review. *Internet Research*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/INTR-03-2020-0148

Zuboff, S. (2019). *The age of surveillance capitalism: the fight for a human future at the new frontier of power* (First edition). PublicAffairs.

Zuckerberg, M. (2020, 16 april). Facebook. Https://www.facebook.com/zuck/posts/10111806366438811.

Bijlage 1: begeleidingscommissie

1. prof. dr. D.R. Veenstra (voorzitter) – Hoogleraar sociologie, Rijksuniversiteit Groningen
2. dr. T. Völlink – Assistant professor psychologie, Open Universiteit
3. drs. S. van der Waal – Research directeur, Waag
4. dr. J.B. de Jong (aanvrager) – Senior adviseur strategie, Ministerie van Justitie en Veiligheid
5. drs. T.L. van Mullekom (opdrachtgever) – Projectbegeleider, Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC)

Bijlage 2: verkenningsworkshop

Op 21 januari 2021 organiseerde het onderzoeksteam van het Rathenau Instituut een verkenningsworkshop. Het doel van deze workshop was om in kaart te brengen welke kennis over schadelijk en immoreel gedrag online op dat moment al aanwezig was bij ministeries, handhavings- en hulporganisaties, en welke relevante initiatieven er toen al liepen bij deze instanties. Daarnaast was het doel om kennisleemten en -behoeften te identificeren die richting konden geven aan het onderzoek. Er waren vijftien deelnemers bij de workshop betrokken.

Deelnemer	Organisatie
Mirjam Buisman	Ministerie van Onderwijs, Cultuur en Wetenschap
Hidde Brugmans	Ministerie van Economische Zaken en Klimaat
Franca van der Laan	Politie
Janet Lambeck	Ministerie van Justitie en Veiligheid
Joyce de Leij	Politie
Ymke Lugten	Openbaar Ministerie
Maarten Glorie	Ministerie van Onderwijs, Cultuur en Wetenschap
Puck Gorrissen	Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
Pieter van Koetsveld	Ministerie van Onderwijs, Cultuur en Wetenschap
[naam bekend bij onderzoekers]	Nationaal Coördinator Terrorismebestrijding en Veiligheid
[naam bekend bij onderzoekers]	Nationaal Coördinator Terrorismebestrijding en Veiligheid
Jolise Stol	Slachtofferhulp Nederland
Paul Thewissen	Minsterie van Onderwijs, Cultuur en Wetenschap
Bastiaan Winkel	Ministerie van Justitie en Veiligheid
Marcel Woltjes	Ministerie van Onderwijs, Cultuur en Wetenschap

Bijlage 3: respondenten

In de loop van februari en maart 2021 hield het onderzoeksteam interviews met vijftien experts op het gebied van schadelijk en immoreel gedrag online. De interviews zijn, in combinatie met wetenschappelijke artikelen en 'grijze' literatuur, als bron gebruikt in hoofdstuk 3 (taxonomie), hoofdstuk 4 (mechanismen), hoofdstuk 5 (bestaande initiatieven) en hoofdstuk 6 (strategische actieagenda).

Tussen de respondenten zitten onderzoekers, ondernemers, hulpverleners en ervaringsdeskundigen. Sommige respondenten hebben een brede expertise (bijvoorbeeld op het gebied van de mechanismen achter schadelijk en immoreel gedrag online); andere respondenten zijn juist expert op het gebied van een specifiek fenomeen of de aanpak daarvan. Hieronder geven we een overzicht.

Respondent	Rol en organisatie
Emine Uğur	Ervaringsdeskundige op het gebied van online haat
Jan Bats	Docent Sociologie en Techniekfilosofie, Open Universiteit en Haagse Hogeschool
Nick Beentjes	Managing Director Benelux, Channel Factory
Claudia van Diessen	Beleidsadviseur, Stichting Halt
Eric van Furth	Directeur, GGZ Rivierduinen; hoogleraar Eetstoornissen, Leids Universitair Medisch Centrum; lid stuurgroep K-EET (Ketenaanpak Eetstoornissen)
Scarlet Hemkes	Persvoorlichter en communicatieadviseur, 113 Zelfmoordpreventie; oprichter en voormalig hoofdredacteur Proud2Bme.nl
Nina Hoek van Dijke	Eigenaar, Jong & Je Wil Wat
Jeroen van den Hoven	Hoogleraar Techniek en Filosofie, Technische Universiteit Delft
David Nieborg	Assistant professor Mediastudies, Universiteit van Toronto
Richard Rogers	Hoogleraar Nieuwe Media en Digitale Cultuur, Universiteit van Amsterdam

Emma Simons	Beleidsmedewerker en onderzoeker, Centrum Kinderhandel Mensenhandel
Kees Teszelszky	Conservator Digitale Collecties, Koninklijke Bibliotheek
Daniel Trottier	Universitair Hoofddocent, adf. Media en Communicatie, Erasmus Universiteit Rotterdam
Patti Valkenburg	Universiteitshoogleraar Media, Jeugd en Samenleving, Universiteit van Amsterdam
[naam bekend bij onderzoekers]	Nationaal Coördinator Terrorismebestrijding en Veiligheid

Bijlage 4: interviewleidraad

Voor het onderzoek zijn semi-gestructureerde interviews gevoerd. Bij deze interviews is gebruik gemaakt van de volgende interviewleidraad, bestaande uit twaalf thema's en bijbehorende vragen.

1. Hoe kijkt de respondent aan tegen het probleem van immoreel of schadelijk gedrag online, en met welke voorbeelden is hij of zij bekend?
2. Hoe schat de respondent de omvang van het probleem (in Nederland) in? Welke cijfers kent hij of zij, is er sprake van groei of afname, en welke verschuivingen vinden er plaats?
3. Hoe schat de respondent ten aanzien van een specifiek fenomeen de verschillen tussen online en offline omgevingen in? Welke nieuwe ontwikkelingen zijn er te verwachten, in het licht van technologische ontwikkelingen en trends?
4. Hoe schat de respondent de schadelijke effecten van het fenomeen in?
5. Welke online mechanismen spelen een inspirerende, faciliterende of katalyserende rol bij het fenomeen/gedrag?
6. Welke groepen worden door het fenomeen/gedrag geraakt?
7. Welke kanten van het fenomeen blijven volgens de respondent vooralsnog onderbelicht?
8. Welke best practices kent de respondent ten aanzien van het fenomeen?
9. Welke beleidsaanbevelingen zou de respondent willen doen?
10. Naar welke bronnen of andere experts kan de respondent de onderzoekers verwijzen?
11. Waar zou het onderzoek van het Rathenau Instituut de respondent mee kunnen helpen?
12. Ruimte voor aanvulling door de respondent zelf: welke tips of boodschap wil hij of zij de onderzoekers nog meegeven?

Bijlage 5: werksessie

Op 13 april 2021 organiseerde het onderzoeksteam van het Rathenau Instituut een werksessie over mogelijkheden voor de aanpak van schadelijk en immoreel gedrag online. Tijdens het literatuuronderzoek en de interviews kwamen vijf oplossingsrichtingen in beeld waarvoor veel animo was, maar die nog niet uitgewerkt waren in de vorm van concrete initiatieven. Het doel van de werksessie was om deze oplossingsrichtingen verder te concretiseren in een dialoog tussen medewerkers van verschillende ministeries, handhavings- en hulporganisaties, onderzoekers, vertegenwoordigers van maatschappelijke organisaties en anderen met relevante expertise. In totaal zijn 22 deelnemers bij de werksessie betrokken, verdeeld over vijf oplossingsrichtingen:

1. Online toezicht en hulp
2. Het gesprek over normen online
3. Waardengedreven inrichting van platformen
4. Technologische oplossingen
5. Handhaving van wetten en regels in de online omgeving

Deelnemers	**Organisatie**
1. Online toezicht en hulp	
[naam bekend bij onderzoekers]	Team OSINT van de Landelijke Eenheid Politie
Irene van Aarle	Proud2Bme
Willem Bantema	Thorbecke Academie, NHL Stenden
Mirjam Buisman	Ministerie van Onderwijs, Cultuur en Wetenschap
Jolise Stol	Slachtofferhulp Nederland
2. Het gesprek over normen online	
Nick Felix	KRO-NCRV
Maarten Glorie	Ministerie van Onderwijs, Cultuur en Wetenschap
Puck Gorrissen	Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
Fleur Jongepier	Radboud Universiteit

3. Waardengedreven inrichting van platformen	
[naam bekend bij onderzoekers]	Nationaal Coördinator Terrorismebestrijding en Veiligheid
Blanca Harms	Rijksuniversiteit Groningen
Edo Haveman	Facebook
Pieter van Koetsveld	Ministerie van Onderwijs, Cultuur en Wetenschap
Stefan Oude Wesselink	Opt Out Advertising
4. Technologische oplossingen	
[naam bekend bij onderzoekers]	Nationaal Coördinator Terrorismebestrijding en Veiligheid
Michiel Leenaars	Stichting NLnet
Mieke van Heesewijk	SIDN fonds
Roelof Muis	Politie
5. Handhaving van wetten en regels in de online omgeving	
Nicole Lieve	Politie
Jaqueline de Jong	Ministerie van Justitie en Veiligheid
Rolf van Wegberg	Technische Universiteit Delft
Inge Welbergen	Ministerie van Onderwijs, Cultuur en Wetenschap

Bijlage 6: validatiebijeenkomst

Op 26 mei 2021 organiseerde het onderzoeksteam van het Rathenau Instituut ter validatie van de onderzoeksresultaten een expertmeeting. Daarbij waren onderzoekers betrokken, en medewerkers van uitvoeringsorganisaties en maatschappelijke organisaties.

Voorafgaand aan de bijeenkomst kregen de deelnemers een samenvatting van het onderzoek ongestuurd (circa twintig pagina's). Tijdens de bijeenkomst kregen ze de gelegenheid om op de onderzoeksresultaten te reageren. Het ging vooral om de handelingsopties die voortkomen uit de analyse die het rapport maakt. Het doel was om samen met de aanwezigen te komen tot een prioritering van handelingsopties, en te reflecteren op de rol die verschillende partijen kunnen spelen bij de aanpak van schadelijk en immoreel gedrag online.

Bij de validatiebijeenkomst waren zeven personen betrokken.

Deelnemer	Organisatie
[naam bekend bij onderzoekers]	Nationaal Coördinator Terrorismebestrijding en Veiligheid
Linda Hell	Bond van Adverteerders
Heleen Janssen	Universiteit van Amsterdam
Franca van der Laan	Politie
Willem van Lynden	Eigenaar Mediamaze; bestuurslid Stichting Stop Online Shaming
Jan-Willem van Prooijen	Vrije Universiteit; Nederlands Studiecentrum Criminaliteit en Rechtshandhaving
Arnout de Vries	TNO

Contactgegevens
Anna van Saksenlaan 51
Postbus 95366
2509 CJ Den Haag
070-342 15 42
info@rathenau.nl
www.rathenau.nl

Het Rathenau Instituut stimuleert de publieke en politieke meningsvorming over de maatschappelijke aspecten van wetenschap en technologie. We doen onderzoek en organiseren het debat over wetenschap, innovatie en nieuwe technologieën.

Rathenau Instituut