

## Dutch Summary

### *Samenvatting*

### Over de voordelen en randvoorwaarden van vertrouwen en betrouwbaarheid

Vertrouwen wordt ook wel omschreven als 'de lijm' of 'het sociale smeermiddel' dat mensen met elkaar verbindt (Arrow, 1974). Deze vergelijking geeft het grote belang van vertrouwen in sociale interacties aan. Het gaat daarbij niet enkel om het geven van vertrouwen, maar ook om het positief reageren op dit gegeven vertrouwen. Een positieve reactie op gekregen vertrouwen is immers een teken van 'betrouwbaarheid'. Reciprociteit speelt een belangrijke factor in het ontstaan van vertrouwen en betrouwbaarheid. Reciprociteit houdt in dat men (over het algemeen) terugdoet of -geeft wat men in eerste instantie zelf heeft gekregen; 'wie goed doet, goed ontmoet' (Falk & Fishbacher, 2006; Gouldner, 1960). De positieve verwachting dat de ander zich reciproof zal gedragen, geeft vertrouwen.

Er zijn verschillende wetenschappelijke onderzoeksmethoden om vertrouwen en betrouwbaarheid te meten. Voor het meten van interpersoonlijk vertrouwen, het vertrouwen tussen twee personen, is de ontwikkeling van de *trust game* zeer belangrijk geweest (Berg et al., 1995; voor meta-analyses, zie bijvoorbeeld Johnson & Mislin, 2011; Van Den Akker et al., 2020). De *trust game* als onderzoeksparadigma biedt de mogelijkheid om op experimentele wijze daadwerkelijk gedrag te meten in plaats van enkel intenties tot gedrag.

Hoewel het paradigma op verschillende manieren wordt ingezet, omvat het doorgaans een interactie tussen twee anonieme personen (Persoon A en Persoon B) die de mogelijkheid krijgen om geld te verdelen. Dit ziet er bijvoorbeeld als volgt uit: de eerste persoon, Persoon A, krijgt €20,- en kan kiezen uit twee opties. Optie 1: Persoon A verdeelt zelf de €20,- tussen zichzelf en de tweede persoon (Persoon B). Persoon A bepaalt in dit geval hoe het geld verdeeld wordt, Persoon B heeft daarop geen invloed. Of, Optie 2: Persoon B verdeelt het geld tussen A en B. Deze laatste optie heeft het bijkomende voordeel dat het geld verdrievoudigd wordt, B zal dus geen €20,-, maar €60,- verdelen. Persoon A heeft bij deze optie echter geen invloed op de verdeling en krijg enkel wat B hem/haar zal geven. De keuze van Persoon A voor Optie 1 (om het geld zelf te verdelen) wordt beschouwd als het *niet* geven van vertrouwen ('geen vertrouwen'). Persoon A's keuze voor Optie 2 (om het geld door B te laten verdelen) wordt beschouwd als het *wel* geven van vertrouwen ('vertrouwen'). In het laatste geval verdeelt B de €60,- en wordt het deel dat hiervan aan A wordt gegeven beschouwd als een maat van betrouwbaarheid.

Dankzij de eenduidige opzet zijn studies waarin deze onderzoeksmethode wordt gebruikt goed met elkaar te vergelijken. Meta-analyses laten een consistent patroon zien in de resultaten (Johnson & Mislin, 2011; Van den Akker et al., 2020). Zo is veelal gevonden dat de overgrote meerderheid van de Personen A bereid is om de ander (Persoon B) te vertrouwen, en dat de meeste Personen B zich als reactie daarop ook betrouwbaar gedragen.

Ondanks de voordelen die dit paradigma biedt, moet ook worden erkend dat de *trust game* in zijn oorspronkelijke vorm ook beperkingen kent. Ten eerste is de *trust game* zeer geschikt om te onderzoeken hoe mensen reageren als ze worden vertrouwd, maar niet hoe mensen reageren als zij *niet* worden vertrouwd. De standaard *trust game* biedt geen kans om dit scenario te onderzoeken – simpelweg omdat het spel eindigt wanneer Persoon A besluit om B niet te vertrouwen. Ten tweede, de opzet van het paradigma is zeer bevorderlijk voor reciprook gedrag, waardoor het lastig is om de rol van verschillende motieven – vertrouwen, betrouwbaarheid en reciprociteit – van elkaar te onderscheiden. Zo is het meestal volledig duidelijk en transparant voor deelnemers wat de gevolgen van elke keuze zijn. Deelnemers weten precies wat de opbrengsten zijn als ze vertrouwen geven (€60,-) en kunnen daarmee ook precies vaststellen of iemand betrouwbaar gedrag vertoont (of Persoon B al dan niet een gelijke verdeling maakt). Ook zijn de beslissingen van deelnemers sterk gekoppeld in de tijd: als A besluit B te vertrouwen om het geld te verdelen, bepaalt B direct daarna hoeveel geld hij/zij teruggeeft aan A. In het huidige proefschrift beschrijf ik hoe ik in een reeks experimentele studies onderzocht heb of aanpassingen in de traditionele *trust game* op deze karakteristieken invloed hebben op het vertrouwen en de betrouwbaarheid van mensen.

## Belangrijkste bevindingen

### Hoofdstuk 2

In Hoofdstuk 2 onderzocht ik wat de effecten zijn wanneer iemand ervoor kiest om een ander *niet* te vertrouwen. In Deel I van dit hoofdstuk onderzocht ik aan de hand van twee experimenten de affectieve (Experiment 2.1) en gedragsmatige reacties van Persoon B (Experiment 2.1 en 2.2) op het moment dat B erachter komt dat Persoon A hem/haar niet heeft vertrouwd. Om dit te onderzoeken is er onderscheid gemaakt tussen de effecten van *actieve* en *inactieve* vormen van wantrouwen (gelinkt aan bredere theorieën over de effecten van actieve vs. inactieve beslissingen; zie bijvoorbeeld Kahneman & Miller, 1986; Kahneman & Tversky, 1982). In beide experimenten vergeleek ik of het *niet geven* van vertrouwen (inactieve vorm) dezelfde effecten heeft als het *ontnemen* van vertrouwen (actieve vorm).

De studies lieten zien dat een actieve (vs. inactieve) vorm van wantrouwen niet alleen negatievere emotionele reacties bij de gewantrouwde persoon opriep (Experiment 2.1), maar ook dat zij minder pro-sociaal gedrag vertoonden in een daaropvolgende interactie (Experiment 2.2). Een belangrijke bevinding daarbij was dat actief gewantrouwde deelnemers niet alleen minder pro-sociaal gedrag vertoonden ten opzichte van de initiële wantrouwer (Persoon A), maar ook ten opzichte van een persoon die niet betrokken was bij de eerste actie van wantrouwen. Deze resultaten benadrukken het belang van vertrouwen. Vertrouwen is niet alleen gunstig als het er is; als vertrouwen ontbreekt, heeft het schadelijke gevolgen – vooral wanneer mensen op een actieve manier worden gewantrouwd.

Deel II van Hoofdstuk 2 richtte zich op Persoon A. Ik onderzocht de bereidheid van deelnemers (als Personen A) om Persoon B te vertrouwen wanneer deze beslissing actief vs. inactief werd genomen (Experiment 2.3). De resultaten lieten zien dat dit onderscheid geen effect had op de keuze van Persoon A om de ander al dan niet te vertrouwen. Terwijl het voor gewantrouwen dus van groot belang was of deze beslissing actief of inactief werd genomen, bleek dit niet mee te spelen bij de persoon die deze actieve vs. inactieve beslissing maakte (Persoon A).

### Hoofdstuk 3

In Hoofdstuk 3 heb ik middels een reeks experimenten onderzocht wat er gebeurt als de opbrengsten van vertrouwen onzeker zijn. Waar het in de *trust game* veelal gebruikelijk is bij vertrouwen het bedrag te verdrievoudigen, was deze vermenigvuldigingsfactor in deze experimenten onverwacht hoger (verviervoudigd) of onverwacht lager (verdubbeld). Daarbij was er sprake van gelijke informatie voor beide personen (A en B wisten allebei van de veranderde opbrengsten) ofwel ongelijke informatie (alleen Persoon B wist van de veranderde opbrengst, Persoon A werd niet op de hoogte gebracht). In het laatste geval ging A er dus altijd vanuit dat de opbrengsten, zoals standaard, verdrievoudigd werden.

Met deze opzet onderzocht ik of het hebben van een informatievoordeel (wanneer alleen B op de hoogte werd gesteld), leidt tot het ervaren en gebruiken van een zelfgecreëerde *moral wiggle room* (Dana et al., 2006). Een *moral wiggle room* houdt in dat bepaalde situaties speelruimte bieden voor wat eerlijk gedrag is. Van deze ruimte kunnen mensen gebruik maken. Zo kan men gedrag vertonen dat wellicht eerlijk *lijkt* voor de ander, maar waarmee men in werkelijkheid de eigen uitkomsten bevordert.

Deel I van Hoofdstuk 3 is gericht op het gedrag, de betrouwbaarheid, van Persoon B. In Experiment 3.1 werd, *nadat* het vertrouwen van Persoon A aan B was gegeven, de opbrengst van vertrouwen onverwachts hoger (€80,- in plaats van €60,-) óf onverwachts lager (€40,- in plaats van €60,-). Persoon B kreeg daarbij ofwel te horen dat Persoon A ook werd geïnformeerd over deze veranderde opbrengst, ofwel dat Persoon A niet op de hoogte werd gebracht. In dat laatste geval dacht A dus dat de opbrengst nog steeds €60,- was. Wanneer deelnemers (Personen B) dachten dat A ook op de hoogte gebracht zou worden van de onverwachte opbrengst, gaven zij gemiddeld de helft van de daadwerkelijke opbrengsten aan Persoon A (B maakte een gelijke verdeling). Hierbij zorgde informatie gelijkheid tussen A en B dus voor vergelijkbare patronen zoals doorgaans gevonden in *trust game* studies. Echter, wanneer Persoon B werd verteld dat Persoon A *niet* geïnformeerd zou worden over de onverwachte opbrengst, verdeelden ze de opbrengsten minder gelijk. Bij onverwacht hoge opbrengsten, gaven Personen B gemiddeld aanzienlijk minder dan de helft van de opbrengst aan Persoon A. Zij gaven bijvoorbeeld de helft van wat Persoon A dacht dat de opbrengst was, en hielden daarbij de extra opbrengsten voor zichzelf. Deze situatie leek dus te zorgen voor het ervaren en inzetten van de *moral wiggle room* om extra's naar eigen hand te zetten. Opmerkelijk gezien gaven Personen B gemiddeld meer dan de helft aan A wanneer de opbrengst onverwacht

laag was. Zij leken daarmee te compenseren voor het 'verlies' waarvan A niet op de hoogte was gesteld.

In Experiment 3.2 kregen deelnemers zelf de keuze of ze Persoon A wilden informeren over de onverwachte opbrengsten. De resultaten toonden aan dat Personen B minder bereid waren om informatie te delen wanneer de opbrengst onverwacht hoog was in vergelijking met onverwacht laag. En wanneer de opbrengst onverwacht hoog was, hielden deelnemers die Persoon A *niet* informeerden vaker de extra inkomsten voor zichzelf. Ditmaal vond ik geen compensatie bij een onverwacht lagere opbrengst. Bij een onverwacht lagere opbrengst kozen bijna alle deelnemers ervoor om A hiervan op de hoogte te stellen.

In Experiment 3.3 werden opnieuw alle deelnemers toegewezen aan de rol van Persoon B en kregen ze te horen dat Persoon A hen vertrouwen had gegeven. Waar in Experiment 3.1 en 3.2 de opbrengst pas veranderde nadat A vertrouwen had gegeven, kregen deelnemers ditmaal te horen dat A *vooraf* al geïnformeerd was dat de opbrengst van vertrouwen onzeker was, en dat deze €40,-, €60,- of €80,- zou kunnen worden. Bovendien wist Persoon A bij het nemen van de beslissing ook al of hij/zij uiteindelijk zou horen hoe hoog de opbrengst was geworden. In dit onderzoek gaven deelnemers (Personen B) wanneer de opbrengst laag was (€40,-), Persoon A gemiddeld de helft van deze opbrengst, ongeacht of Persoon A wel of niet wist hoe hoog de opbrengst was geworden. Wanneer de opbrengst hoog was (€80,-) maakte het, net als in Experiment 3.1 en 3.2, voor deelnemers wel uit of Persoon A wel of niet op de hoogte zou worden gesteld. Wanneer de deelnemers dachten dat Persoon A *niet* te horen zou krijgen dat er €80,- te verdelen was, gaven ze Persoon A minder dan de helft van deze opbrengst.

Samen laten deze studies zien dat wanneer het voor Persoon A onduidelijk is hoe hoog de opbrengsten van vertrouwen daadwerkelijk zijn, mensen soms gebruik maken van de *moral wiggle room* die ontstaat. Vooral wanneer de opbrengsten hoog zijn, lijken mensen geneigd om gebruik te maken van deze speelruimte en de extra opbrengsten in eigen zak te steken.

Deel II van Hoofdstuk 3 is gericht op het gedrag van Persoon A en een buitenstaander. In Deel I van Hoofdstuk 3, verdeelde Persoon B de opbrengst van vertrouwen vooral gelijk tussen zichzelf en Persoon A wanneer hij/zij wist dat A ook te weten zou komen hoe hoog de opbrengsten daadwerkelijk waren. Wanneer de opbrengst van gegeven vertrouwen echter hoog was zonder dat dit bekend was bij A, hielden deelnemers meer dan de helft hiervan voor zichzelf. Het zou natuurlijk kunnen zijn dat ze dit deden omdat ze dit, dankzij de ontstane *moral wiggle room*, zelf eerlijk vonden – wat niet weet, wat niet deert. Maar de vraag is of dit gedrag ook in het algemeen, bijvoorbeeld door anderen die zelf niet zo'n keuze hoeven maken, als eerlijk wordt gezien. Om dit te onderzoeken, werd in Experiment 3.4 de oordelen van niet-betrokken buitenstaanders onderzocht. Deelnemers aan dit onderzoek kregen verschillende verdelingen te zien die overeenkwamen met verdelingen die Personen B in de eerdere experimenten hadden gemaakt. Vervolgens gaven zij hun oordeel over deze verdelingen.

Resultaten laten zien dat ongelijke verdelingen over het algemeen negatief beoordeeld worden. Opvallend genoeg waren deze beoordelingen ook niet milder wanneer Persoon A niet geïnformeerd was over de opbrengsten, en dus de verdeling enkel ogenschijnlijk eerlijk en gelijk was. Het gedrag van Persoon B (zoals in Deel I, Experiment 3.1 – 3.3) lijkt daarmee hoofdzakelijk te worden gedreven door het inzetten van *moral wiggle room* in hun eigen belang, en niet door een unanimitéit dat dergelijk gedrag algemeen geaccepteerd is.

In Experiment 3.5 onderzocht ik de bereidheid van Persoon A om een ander te vertrouwen onder onzekere omstandigheden (zoals veroorzaakt door onzekere opbrengsten en een (a)symmetrische kennisdeling hierover). De uitkomsten lieten zien dat de bereidheid van Personen A niet beïnvloed werd door of zij zelf wel of niet geïnformeerd zouden worden over hoe groot de opbrengst van vertrouwen zou zijn. Ook verwachtten zij geen verschil in de verdeling die Persoon B zou maken onder de onzekere omstandigheden. Deze resultaten laten zien dat een groot deel van de mensen, zelfs wanneer ze niet op de hoogte worden gebracht van de opbrengsten, bereid is een ander te vertrouwen. Daarmee stellen ze zichzelf welwillend kwetsbaar ten opzichte van een persoon met meer informatie. In tegenstelling tot eerder onderzoek waarin werd gesteld dat men vaak te cynisch is over de betrouwbaarheid van anderen (Fetchenhauer & Dunning, 2009, 2010), biedt dit een meer optimistisch beeld.

#### Hoofdstuk 4

In Hoofdstuk 4 onderzocht ik de effecten van een tijdsinterval in een *trust game* tussen de beslissing van Persoon A (als maat van vertrouwen) en de reactie van Persoon B (als maat van betrouwbaarheid). Eerder onderzoek laat zien dat beslissingen die mensen maken, vaak beïnvloed worden door beslissingen op een eerder tijdstip (e.g., Arkes & Blumer, 1985; Thaler, 1980). Dit geldt ook voor reciprociteit. Naarmate er meer tijd verstrijkt tussen een eerste actie van een ander persoon en de mogelijkheid om hierop te reageren, neemt reciprociteit af (Frederick et al., 2002). Dat komt bijvoorbeeld doordat men zich minder verplicht voelt iets terug te doen, herinneringen aan de eerste actie minder scherp zijn en de waardering voor deze eerste actie afneemt over tijd. Aangezien reciprociteit een belangrijke rol speelt in het ontstaan van vertrouwen, is te verwachten dat een soortgelijk patroon op zal treden bij de bereidheid om een ander te vertrouwen (gedrag van Persoon A). Ondanks dat voor betrouwbaarheid (gedrag van Persoon B) dezelfde motieven kunnen gelden, zijn er ook argumenten waarom betrouwbaarheid niet sterk beïnvloed zal worden door tijd. Eerder onderzoek wijst namelijk uit dat het gedrag van Personen B in een *trust game* ook wordt gedreven door vriendelijkheid en eerlijkheid (Ashraf et al., 2006; Espín et al., 2016).

Om de effecten van een tijdsinterval te onderzoeken werd opnieuw de standaardopzet van de *trust game* (waarin Persoon A en B *direct* na elkaar hun keuze maken) vergeleken met een aangepaste versie, waarin de keuzes van A en B een maand na elkaar plaatsvonden. De resultaten toonden aan dat een interval van een maand de bereidheid van Persoon A om de ander te vertrouwen verminderde, hetgeen in lijn is met de literatuur over reciprociteit en

beslissingen over tijd. Hetzelfde tijdsinterval van een maand had echter geen invloed op de keuze van Persoon B; de mate van betrouwbaarheid bleef gelijk.

In overeenkomst met eerder onderzoek bevestigden deze bevindingen het idee dat Personen A 'te cynisch' zijn over de betrouwbaarheid van Personen B en daarbij het bedrag dat zij terugverwachten van Persoon B onderschatten (Fetchenhauer & Dunning, 2009, 2010). Bovendien toonden de huidige bevindingen aan dat dit ontorechte cynisme versterkt wordt wanneer er meer tijd zit tussen beide beslissingen.

### Slotnoot

Meta-analyses over *trust games* hebben aangetoond dat mensen over het algemeen bereid zijn om een ander te vertrouwen, en dat dat gegeven vertrouwen vaak op een positieve manier wederkerig wordt beantwoord (Johnson & Mislin, 2011; Van den Akker et al., 2020). De studies in dit proefschrift ondersteunen dit, maar tonen tegelijkertijd ook een genuanceerder beeld. Door verschillende kenmerken van de standaard *trust game* systematisch te variëren, liet ik zien dat vertrouwen en betrouwbaarheid ook sterk afhankelijk zijn van de situationele kenmerken. Door de *trust game* op een subtiele wijze aan te passen, kon ik de gevolgen van (actief en inactief) wantrouwen, de rol van informatieve onzekerheid, en het belang van de timing van beslissingen onderzoeken. Als zodanig draagt het huidige proefschrift bij aan ons begrip van de voordelen en randvoorwaarden van vertrouwen en betrouwbaarheid, en hun relaties tot wederkerigheid.