

**Member-to-member generalisation in trust behaviour:**

**How do prior experiences inform prosocial behaviour towards novel ingroup and  
outgroup members?**

## Abstract

Trust in individuals is strongly guided by group membership; ingroup favouritism in trust is a very robust finding in the literature on pro-social behaviour. We know, however, that group attitudes can change based on discrete encounters with group members (i.e., intergroup contact). This research examines how people use previous experiences with ingroup and outgroup members to inform decisions to trust novel, unknown group members. This process, which we refer to as *member-to-member* generalisation, was examined in two studies using a student sample ( $N = 135$ ) and a larger and more representative online sample ( $N = 226$ ). The moderating effects of group membership (ingroup vs outgroup vs unknown) and interaction valence (positive vs negative) on member-to-member generalisation were explored in ten sequential Trust Games. We examined changes in investment behaviour based on feedback from the previous partner, where feedback was either positive (high reciprocation) or negative (no reciprocation). We observed consistent evidence for member-to-member generalisation. People did not just rely on initial group attitudes to guide their trust behaviour, but adjusted their behaviour towards novel individuals based on previous experiences. Generalisation was stronger for interactions that were negative and seemingly unexpected. When people showed strong distrust of the outgroup (Study 1) or were highly identified with the ingroup (Study 2), they changed their behaviour towards novel partners more after experiencing incongruent interactions. These findings are discussed in relation to intergroup contact theory, outgroup homogeneity, and expectancy violation effects.

*Word count:* 237

**Keywords:** trust, intergroup relations, generalisation, ingroup bias, Trust Game

**Member-to-member generalisation in trust behaviour:****How do prior experiences inform prosocial behaviour towards novel ingroup and outgroup members?**

Imagine that you, a non-Muslim, are travelling by train for a business meeting. A Muslim man wearing traditional clothing takes a seat next to you. You start chatting, and before you know it, you have talked to him for the rest of your journey, having a pleasant conversation about many topics. When you get off the train and get to your business meeting, a Muslim woman wearing a headscarf greets you. From intergroup contact theory, it is assumed that the pleasant interaction with the Muslim man on the train changes your attitudes towards Muslims as a group. However, does your behaviour towards other members of the group also change, such as the Muslim woman at the meeting? Does intergroup contact also generalise to improve pro-social behaviour towards novel group members?

The effect of group membership on pro-social behaviour is well-established; people generally demonstrate ingroup favouritism (Balliet, Wu, & De Dreu, 2014), and display distrust and discriminatory behaviour towards disliked or rival outgroups (Dovidio, Gaertner, Kawakami, & Hodson, 2002; Dovidio, Kawakami, & Gaertner, 2002). Research surrounding intergroup contact theory, however, demonstrates that our feelings towards outgroups become more positive or negative based on interactions with individual group members (Brown & Hewstone, 2005; Graf, Paolini, & Rubin, 2014; Pettigrew, 1998). What is yet unknown, is how experiences with individual group members directly influence behaviour towards novel group members whom one has not interacted with before. In this research, we focus specifically on *member-to-member* generalisation, exploring how people generalise their experiences with individual group members to trust other individuals from the same group. We describe two experiments that examined member-to-member generalisation in the context of trust behaviour. We examine how group membership (ingroup vs outgroup vs control) and

valence during interactions (positive vs negative) influence generalisation of trust towards novel group members.

### **Ingroup favouritism in trust**

Trust is essential in interactions and social relationships and is a strong predictor of cooperation between individuals and between groups (Balliet & Van Lange, 2013). Trust indicates that people have positive expectations about the behaviour and intentions of the other person (Lewicki, Tomlinson, & Gillespie, 2006; Rousseau, Sitkin, Burt, & Camerer, 1998). A common way to study trust between individuals is to examine investment behaviour in the Trust Game (Berg, Dickhaut, & McCabe, 1995). Here, a trustor is given an endowment that he/she can invest in a trustee. If the trustor invests his/her endowment, the amount is multiplied and given to the trustee. The trustee then has the choice to reciprocate trust by returning some of the received amount to the trustor, but he/she does not have to. Both players can end the game with more money than they started out with, but only if they both cooperate.

Contrary to the predictions from Game Theory (in which the “rational” behaviour is not to return, and therefore not to invest any money), people generally display both trust and trustworthiness (i.e., investing and returning money) in this game (Camerer, 2003). Trust occurs in single game interactions with strangers, but is particularly prevalent in repeated interactions when the interaction partner reciprocates trust (King-Casas et al., 2005). Reciprocal altruism, in which people may temporarily incur some cost for a behaviour (such as trust) in order to increase the chance that others may return the favour is considered evolutionary sound (Trivers, 1971). Trust is most commonly withdrawn after non-reciprocation, a strategy called the “tit-for-tat strategy” (Axelrod & Hamilton, 1981).

One of the most powerful rules of behaviour is that people are kinder to people in their own groups than to people in their outgroups. In both laboratory and real-world

situations, people tend to show preferential treatment of ingroup members compared to outgroup members (Brewer, 1999; Tajfel & Turner, 1979). In a large-scale meta-analysis, Balliet and colleagues found a robust effect of group membership on cooperative decisions in different types of mixed-motive games, showing higher trust towards ingroup members than outgroup members (Balliet et al., 2014). Ingroup favouritism in trust has been observed in different group contexts, including groups defined on the basis of race (e.g. Burns, 2006), nationality (e.g. Stoddard & Leibbrandt, 2014), or religion (e.g. Rotella, Richeson, Chiao, & Bean, 2013), as well as in a minimal-group setting (Buchan, Johnson, & Croson, 2006).

*Social identity theory* (SIT; Tajfel & Turner, 1979) offers an explanation for the above effects. SIT posits that the groups to which we belong form an important part of our identity and serve to partially fulfil belonging and self-esteem needs. As group memberships form an important part of our self-concept (Smith & Henry, 1996), we are motivated to belong to positively-valued groups and to see our ingroups in a positive light. That is why we often favour ingroup members over individuals from outgroups. Research in this tradition has established that people are motivated to treat ingroup members positively and consider deviants to be aberrations who need to be removed from the ingroup (e.g. Marques, Yzerbyt, & Leyens, 1988).

In addition to ingroup favouritism, intergroup bias also extends to perceptual and judgment biases. Ingroups are generally perceived to be variable and heterogeneous, whereas outgroups are perceived to be more stereotypic, more homogeneous, and thus more similar to each other. This intergroup difference in perceived variability is commonly referred to as the *outgroup homogeneity effect* (Judd & Park, 1988; Ostrom & Sedikides, 1992; Park & Rothbart, 1982). This effect may occur because perceivers have less motivation to attend to differences between outgroup members than to differences within their group (Haslam, Oakes, Turner, & McGarty, 1995).

### Changing group attitudes and person impressions

While people generally tend to favour their ingroup and perceive outgroups as stereotypic and homogenous, attitudes of outgroups change through experience, as *intergroup contact theory* (Allport, 1954; Brown & Hewstone, 2005; Pettigrew, 1998) shows. The central premise underlying intergroup contact theory is that positive interaction with an outgroup member, as a representative of their group, has the potential to reduce prejudice not only towards that specific outgroup member, but also toward the outgroup as a whole. This type of *member-to-group* generalisation has been robustly supported in the literature (Davies, Tropp, Aron, Pettigrew, & Wright, 2011; Lemmer & Wagner, 2015; Pettigrew & Tropp, 2006). Effects are strongest when the contacted person is regarded as typical of the outgroup and when respective group memberships are salient.

In addition to changing group attitudes from personal contact, information about individual group members can also directly transfer to other individuals in a process referred to as *attitude transfer* (Crawford, Sherman, & Hamilton, 2002; Ranganath & Nosek, 2008). Crawford and colleagues (2002) showed that people generalise information about group members of highly entitative groups (i.e., groups that are perceived to a unified entity) to other group members, because these individuals are perceived as interchangeable parts of the group. Such transfer does not occur for low entitativity groups (Crawford et al., 2002). Furthermore, Ranganath and Nosek (2008) showed that generalisation of traits from one group member to another is a process that occurs automatically on an implicit level, even when it is resisted on an explicit level.

Most relevant to the current research, implicit attitude transfer has been found to occur more strongly for outgroup members than ingroup members, particularly when the information presented about outgroup members is negative (Ratliff & Nosek, 2011). These findings are likely due to a combination of outgroup homogeneity and negativity biases.

When members of the group are perceived as similar to each other, information about one group member is more readily transferred to other individuals. *Negativity biases* indicate that in general, people are more sensitive to negative than positive information (Baumeister et al., 2001; Rozin & Royzman, 2001; Skowronski & Carlston, 1989). In particular, negative information has been shown to be generalised more quickly than positive information (Fazio, Eiser, & Shook, 2004; Shook, Fazio, & Eiser, 2007). Therefore, attitudes formed about individual group members are transferred more readily when individuals are outgroup members, and when the information about the individuals is negative (Ratliff & Nosek, 2011).

### The current research

The above literature demonstrates that 1) our feelings about ingroups and outgroups influence responses to individual group members (Balliet et al., 2014), and 2) experiences with individual group members can change our attitudes towards the group as a whole (Brown & Hewstone, 2005). However, intentions and attitudes only have utility if they predict actual, subsequent behaviour (see Herek, 1986). It has been shown that there is sometimes a disconnect between attitudes and behaviour (Sheeran, 2002). People who want to be nonprejudiced may avoid intergroup contact (Plant & Devine, 2003) and may even can behave in a prejudicial or discriminatory manner (Fehr & Sassenberg, 2009). Unfortunately, less intergroup contact research has directly measured behaviour towards an outgroup member (cf. Finseraas, Hanson, Johnsen, Kotsadam, & Torsvik, 2019), and much of this work relies on self-reports of previous contact experiences and attitudes or mentally simulating positive contact (e.g. Meleady & Seger, 2017).

It is yet unknown how experiences with individual group members directly influence behaviour towards novel group members, a process we refer to as *member-to-member generalisation*. In the current research, the Trust Game paradigm was adapted to measure

how people generalise their experiences with group members in their trust behaviour towards novel individuals from the same group. We manipulated two elements of the Trust Game; group membership and interaction valence.

Firstly, group membership was manipulated through political affiliation. Participants either played the game with partners that shared their political affiliation (the ingroup), that supported an opposing political party (the outgroup), or with partners with unknown political affiliation (the control). Political affiliation has frequently been used in research on social identity and intergroup relations (e.g. Brewer, 1999; Deegan, Hehman, Gaertner, & Dovidio, 2015; Hackel, Looser, & Van Bavel, 2014; Riek, Mania, Gaertner, McDonald, & Lamoreaux, 2010). Based on pilot research with student samples,<sup>1</sup> it was predicted that participants would feel strongly towards political groups and that it would be more acceptable to explicitly express dislike of the political outgroup than it is of other outgroups such as national, ethnic, or religious groups (also see Iyengar & Westwood, 2015). We selected a natural group (instead of minimal groups) as we were interested in how pre-existing group knowledge and biases influence member-to-member generalisation. Secondly, valence of game interactions was manipulated by providing feedback about group members' reciprocity behaviour. Some game partners reciprocated trust (positive interactions), while others violated trust (negative interactions).

## Research questions and hypotheses

---

<sup>1</sup> On a 100-point feeling thermometer scale, a sample of  $N = 97$  students (83% female) showed a mean of 31.00 ( $SD = 24.79$ ) towards UKIP voters, compared to  $M = 55.60$  ( $SD = 25.92$ ) for Conservative voters,  $M = 66.66$  ( $SD = 22.94$ ) for Labour, and  $M = 69.50$  ( $SD = 21.14$ ) for Greens.

The main aim of this research was to establish whether people generalise their experiences with previous group members in their trust behaviour towards other, novel group members. In terms of the Trust Game paradigm, this question was analysed by examining how much people change their trust behaviour towards different group members throughout the game. Do people change their trust and expectations in novel group members after their trust is reciprocated or violated by a different member of the same group?

**Hypothesis 1: People will increase their investments in novel group members following trust reciprocation by other group members and decrease their investments in novel group members after trust violation.**

After establishing whether people generalise their experiences with previous group members to novel partners, we aimed to determine if the process of member-to-member generalisation differs for ingroup and outgroup members. Therefore, we examined how both outgroup and ingroup generalisation compare to a control condition where no group membership information was available. Work on group entitativity (Crawford et al., 2002) suggests that member-to-member generalisation would be greater when a group category is provided compared to the control. This effect is predicted to be particularly pronounced for the outgroup, due to the outgroup homogeneity effect (e.g., Judd & Park, 1988).

**Hypothesis 2: Group membership will moderate member-to-member generalisation of trust behaviour. Generalisation will occur more strongly when group membership is provided, and particularly pronounced for the outgroup.**

Are member-to-member generalisation processes different for negative interactions compared to positive interactions? Based on the observations of negativity biases in perception, memory, and generalisation (e.g., Baumeister et al., 2001; Shook et al., 2007), it was predicted that people should change their investments in novel group members more

strongly after experiencing negative interactions with previous group members, compared to positive interactions.

**Hypothesis 3: Interaction valence is predicted to moderate member-to-member generalisation of trust behaviour. Generalisation will occur more strongly for negative experiences than for positive experiences, due to a negativity bias. Generalization for negative experiences will be stronger for the outgroup, in line with Hypothesis 2.**

Are there between-group differences in trust recovery after a violation? Social identity theory (Tajfel & Turner, 1979) states that people are motivated to maintain a positive image of the ingroup and discard of ingroup deviants. Therefore, interactions with untrustworthy ingroup members should not influence trust in novel ingroup members. However, untrustworthy outgroup members might change the perception of the whole group more permanently.

**Hypothesis 4: Following trust violation, trust in novel ingroup members is predicted to be quicker to recover than trust in novel outgroup members.**

## Study 1

The first study examined member-to-member generalisation of trust behaviour in a laboratory setting, utilising a British student sample. Participants indicated their attitudes towards the major British political parties on a feeling thermometer scale, after which they selected their own political affiliation. Next, the Trust Game was played with ten different partners from participants' own political ingroup, from a political outgroup, or with partners with unknown political affiliation (control). Participants always played the role of trustor in the Trust Game, while the partners in the game were pre-programmed by the experimenter. Group membership of the partners was varied between-subjects, while interaction valence was varied within-subjects.

Based on a pilot study with university students which showed a dislike towards supporters of the UK Independence Party (UKIP), but a positive attitude towards many other political outgroups, UKIP was selected as the target outgroup. This disliked group was selected to create the largest intergroup differentiation.

## Method

### Participants and design

The participant pool consisted of 152 University students<sup>2</sup>. As the outgroup was set to consist of UKIP supporters, any participant that supported UKIP, or showed no preference

---

<sup>2</sup> The sample size was based on power calculations using the coefficient and standard error of previous studies using the same paradigm but a within-subject design of group, using the same MatLab script as Vermue, Seger, and Sanfey (2018). In these previous studies, the effect size of the interaction of interest and a sample size of  $N$

of their own selected party over UKIP, was removed from analysis. To measure the amount of preference, a difference score was calculated between the feeling thermometer score for the selected party and the feeling thermometer score for UKIP. A cut-off score of 25 was used to determine party preference; 17 participants were removed from analysis. The remaining 135 participants (75% female,  $M_{age} = 20.96$ ,  $SD_{age} = 5.05$ ) received course credit or a payment of 3 pounds for their time. In addition, participants had a chance to win their earnings in the game, converted to pounds (one token = 50 pence), based on a dice roll at the end of the experiment. Twenty-six participants received the bonus; the average bonus amount was 5.75 pounds.

This study employed a mixed design with the between-subject factor group (ingroup, outgroup, control) and the within-subject variable trial number (1 to 10) as main predictors of investments in the separate phases of the Trust Game. The in- and outgroup were determined based on the self-identified political affiliation of the participant. The outgroup was set to consist of UKIP supporters, and the control group always consisted of partners with an unknown political affiliation. Identification with the ingroup was measured in addition to the Trust Game data. People higher in such identification generally show more preferential treatment toward ingroup members (Brewer, 1999); we expected this to potentially effect generalisation processes for the ingroup and outgroup. The standardized mean scores of this scale was treated as a continuous predictor of investments.

## Materials and procedure

The experiment was programmed using the Python-based program PsychoPy (Peirce, 2007) and data was collected in a laboratory setting, with participants seated in separate

---

= 48 resulted in an obtained power of 0.76. As this study utilised a between-subjects design of group, a tripled sample size was aimed for ( $3 * 50 = 150$  participants).

cubicles. Informed consent was obtained from all individual participants included in this study. The first part of the experiment consisted of feeling thermometer scales, the selection of a political party, and the ingroup identification scale. The feeling thermometer scales consisted of a continuous scale from 0° (very cold) to 100° (very warm), for each of the five major political parties in the UK.<sup>3</sup> The parties were presented by name and with a logo to help with recognition. After participants had rated all the political parties on the feeling thermometer scale, they were asked to select the political party that they identified most with. This selection determined the presented ingroup for the rest of the experiment. Participants next completed a 7-item ingroup identification scale ( $\alpha = 0.83$ ) adapted from Cinnirella (1997). For the identification questionnaire, the name of the group was set based on the selected party.

**Sequential Trust Game.** After completing the questionnaires, participants received instructions about the Trust Game and a few control questions to ensure full understanding of the game. Participants always played the game in the role of trustor/first mover. After completing four practice rounds of the game, participants played 10 rounds of the Trust Game. For each partner in the game, an image of the party symbol was shown to indicate group membership, and two initials were displayed to indicate individual players. These partners were presented to the participants as responses from people who previously played the game, where the original would be matched to the participants' investment decisions. However, in reality all responses were pre-programmed to show either trustworthy or untrustworthy behaviour at specific points in the game.

Each round was played with a different individual, but all rounds were played with

---

<sup>3</sup> The Conservative Party, The Labour Party, The Liberal Democrats, The Green Party, and The UK Independence Party (UKIP).

only ingroup members, outgroup members, or people of which the political affiliation was unknown (control group). Participants were specifically instructed that they would play only one round with each partner. However, they were not aware of the number of rounds that they were going to play, to avoid any possibility of strategic changes in investments during the last rounds.

Each round of the Trust Game consisted of the following stages: introduction of the partner, expectation rating, investment decision, overview of transaction, and feedback. If the participant decided not to invest in the partner, the overview of transaction stage was skipped. Participants were first introduced to the partner of that round. The party symbol was displayed when the participant interacted with an ingroup or outgroup member. No other pictures of the partners was shown, eliminating any effects of facial attractiveness or trustworthiness (van 't Wout & Sanfey, 2008; Wilson & Eckel, 2006). For control partners, an "affiliation unknown" image was shown on the screen.

Participants next indicated their expectations regarding the reciprocity behaviour of the partner, by selecting how much they thought this partner would return to them, in percentage of the investment, on a scale from zero to 100% of the investment. After the expectation question, participants could decide how much of their endowment of 10 tokens they wanted to invest in the current partner. Any number of tokens between zero and 10 tokens could be invested in the partner. After the participant made their choice, an overview screen was displayed for three seconds indicating the selected investment and the partner's received amount (the investment x3). Last, participants were given feedback about how many tokens the current partner returned to them, and how many tokens they had earned in that round.

The ten rounds of the Trust Game were divided into three separate phases. In the first three rounds of the game, (the trust-building phase) all partners reciprocated high amounts,

between 45% and 60% of the amount that the partner receives (investment x3). In these rounds, participants always ended the round with more tokens than they started with, if they invested. In rounds four and five (the trust-violation phase) the partners did not reciprocate any invested tokens, irrespective of the investment of the participant. Finally, in rounds six to 10 of the Trust Game (the trust-recovery phase), all partners reciprocated similar high amounts as in the first rounds. This behaviour was kept constant for all participants, only the information about the partner (ingroup, outgroup, or control) varied between participants.

After completing all rounds of the Trust Game, participants were shown their average earnings in the game and completed a number of final questionnaires<sup>4</sup> including questions regarding the purpose of the study and any suspicions they might have about the experiment<sup>5</sup>. Finally, participants rolled a dice to determine whether they would receive the bonus payment of their average earnings in the game, and they received a written debriefing before exiting the room. This study was approved by the appropriate institutional research ethics committee before data collection.

---

<sup>4</sup> Outgroup attitudes were included as an exploratory measure. No directional hypotheses were formulated for this variable, as participants experienced both positive and negative interactions with the outgroup. As no significant effects were found, this variable is not reported any further.

<sup>5</sup> 60% of all participants (91/152) reported no suspicions about the study. Most participants thought the purpose of the study was related to trust and political affiliations (97 participants mentioned the word “political”, 60 participants mentioned the word “trust”, and 44 participants mentioned both words). Few participants mentioned prejudice, stereotypes, or ingroup favouritism as an aim of the study (7 participants mentioned the word “group”, 7 participants mentioned “prejudice” or “bias”). No participants suspected generalisation or learning about a group as the aim of the study.

## Data analysis

The data was analysed using the statistics program R version 3.4.2 (R Core Team, 2017) in RStudio (RStudio Team, 2015). The analysis was divided into three main strands. Firstly, initial investments and expectations were compared between the three groups using ANOVAs. Secondly, to analyse how group membership of the partners influenced investments in the Trust Game over time, as well as their relation to ingroup identification, a number of different linear multilevel models were created (also known as mixed-effects models; Baayen, Davidson, & Bates, 2008). The effect of group (ingroup, outgroup, control) and trial number (rounds 1 to 10) on the investments were analysed for each phase separately.

The main effect of trial number during a phase of the game examines the first hypothesis about member-to-member generalisation. Interactions between group and trial number indicate differences in generalisation between groups, thus testing hypothesis 2. The interaction between group and trial number in the trust-recovery phase examines the fourth hypothesis.

A random intercept per participant, random slopes for group and trial number, as well as the interaction between these variables were added to obtain a maximal random structure (Barr, Levy, Scheepers, & Tily, 2013). The following planned contrasts were implemented to compare the separate groups: the contrast 1 compared the control group with the ingroup and outgroup combined, and the contrast 2 compared the ingroup with the outgroup. Unstandardized regression coefficients are reported alongside inferential statistics as simple effect sizes (Baguley, 2009).

Thirdly, the amount of change in investments was compared between the different phases by creating individual coefficients of change for each phase and running linear multilevel models on this data. This analysis was performed to be able to directly compare the level of trust generalisation between positive interactions in the trust-building phase and

negative interactions in the trust-violation phase, thus testing hypothesis 3. Moreover, effects of group over phases (instead of specifically for each phase of the game) could be examined with this analysis, allowing for another way to examine the second hypothesis. Lastly, the relation between investments and expectations were examined with correlations and a multilevel model.

## Results

### Initial trust bias

An ANOVA on investments in the first round of the Trust Game showed a significant effect of group,  $F(2, 132) = 25.58, p < .001, \eta_p^2 = 0.28$ . Post-hoc independent  $t$ -tests confirmed that initial investments differed significantly between all groups (see Table 1 for descriptive statistics and  $t$ -test results). Consistent with an ingroup favouritism effect, participants invested the highest amount in ingroup partners and the lowest amount in outgroup partners. The second ANOVA on first round expectations of return confirmed this pattern, showing a significant effect of group,  $F(2, 132) = 50.37, p < .001, \eta_p^2 = 0.43$  (see Table 1).

Table 1

*Mean and standard deviations for initial investments (0 – 10 tokens) and expectations of return in percentage of the received investment (0 to 100%) in partners from the different groups, and t-test comparisons of initial investments between groups*

<b>Investments</b>				
	Mean (SD)	1	2	3
<b>1. Ingroup</b> <i>(N = 46)</i>	5.89 (2.77)	-		
<b>2. Outgroup</b> <i>(N = 45)</i>	2.56 (1.44)	$t(68) = 7.22,$ $p < .001, d = 1.50$	-	
<b>3. Control</b> <i>(N = 44)</i>	4.45 (2.26)	$t(85) = 2.70,$ $p = .008, d = 0.57$	$t(72) = -4.72,$ $p < .001, d = 1.01$	-
<b>Expectations</b>				
	Mean (SD)	1	2	3
<b>1. Ingroup</b>	53.87 (18.70)	-		
<b>2. Outgroup</b>	19.24 (13.93)	$t(83) = 10.03,$ $p < .001, d = 2.10$	-	
<b>3. Control</b>	45.45 (18.30)	$t(87) = 2.16,$ $p = .034, d = 0.45$	$t(80) = 7.59,$ $p < .001, d = 1.61$	-

[Insert Figure 1 here]

*Figure 1.* Mean investments in different group partners for each round of the Trust Game for Study 1. Error bars indicate 95% confidence intervals. The dotted lines indicate the phases of the game (building, violation, recovery).

## Trust-building phase

The first model examined the effect of group, trial number, and average mean-centred ingroup identification score on investments in the trust-building phase of the game (Round 1 to Round 4, see Figure 1). All main effects and interactions between these three variables were included as fixed effects in the model. In addition, a random per subject intercept was added, and a random slope for trial number. The fixed effects of this model explained 22% of the variance within the data ( $R^2_{fixed} = 0.22$ ). Adding the random effects to the model increased the amount of variance explained to 84% ( $R^2_{total} = 0.84$ ).

Significant main effects of group,  $F(2, 129) = 28.00, p < .001, b_1 = -0.13, t(129) = -0.82, p = .412, b_2 = 1.94, t(129) = 7.45, p < .001$ , and trial number,  $F(2, 129) = 86.76, p < .001, b = 0.49$ , were observed. The positive coefficient for the main effect of trial indicates that people increased their investments during the trust-building phase, confirming hypothesis 1. Post-hoc multiple comparisons based on the model show that investments differed significantly between all groups: ingroup – outgroup,  $t(129) = 5.29, p < .001, d = 0.93$ , ingroup – control,  $t(129) = -2.62, p = .027, d = 0.46$ , outgroup – control,  $t(129) = 2.61, p = .028, d = 0.46$ . The least-square means as predicted from the model for each group are:  $M_{control} = 5.05 (SE = 0.31, 95\% CI [4.42, 5.68]), M_{ingroup} = 6.20 (SE = 0.30, 95\% CI [5.60, 6.79])$ , and  $M_{outgroup} = 3.90 (SE = 0.31, 95\% CI [3.28, 4.51])$ , respectively.

The Group x Trial interaction was significant as well,  $F(2, 129) = 12.86, p < .001, b_1 = 0.05, t(129) = 1.31, p = .192, b_2 = -0.32, t(129) = -4.92, p < .001$ . Post-hoc comparisons of the slopes of investments over time for the different groups show that the slope for ingroup partners,  $intercept = 5.63, b = 0.22, SE = 0.08, t(95) = 2.81, p = .006$ , differs significantly from the slope for outgroup partners,  $intercept = 1.75, b = 0.86, SE = 0.09, t(43) = 9.08, p < .001$ , comparison ingroup - outgroup,  $\chi^2(1) = 24.24, p < .001$ . Moreover, the slope for the

outgroup partners differed significantly from the control group,  $intercept = 4.07$ ,  $b = 0.39$ ,  $SE = 0.04$ ,  $t(42) = 3.74$ ,  $p < .001$ , comparison control – outgroup,  $\chi^2(1) = 12.55$ ,  $p < .001$ . The slopes of the ingroup and the control group were not significantly different from each other,  $\chi^2(1) = 1.70$ ,  $p = .192$ . These results partially confirm hypothesis 2. The left-hand side of Figure 2 shows, using a boxplot of per-subject coefficients, increased change in investments for the outgroup, over the control, which in turn was larger than for the ingroup. The visualised coefficients were retrieved from linear multilevel models of each phase, with the sign of coefficients reversed for the violation phase for magnitude comparison.

[Insert Figure 2 here]

*Figure 2.* Boxplots of coefficients of investment change over trials for the different phases of the Trust Game and for the different groups, for Study 1.

In summary, ingroup partners received the highest investments, but investments increased the least. Meanwhile, outgroup partners received very low investments initially, but investments increased the most across the trust-building phase (see Figure 1). Lastly, no significant main effect or interactions with ingroup identification were observed on the investments in the trust-building phase.

### Trust-violation phase

A second linear multilevel model was created with group, trial number, and ingroup identification predicting investments in Round 4 to Round 6 of the Trust Game (see Figure 1). The same random structure and planned contrasts for group were used. The fixed effects of this model explained 18% of the variance within the data ( $R^2_{fixed} = 0.18$ ). Adding the random effects to the model increased the amount of variance explained to 71% ( $R^2_{total} = 0.71$ ). In this model, only a significant main effect of trial number was observed,  $F(1, 129) =$

$76.07, p < .001, b = -0.81$ . This effect showed that investments significantly decreased during the trust violation phase, and thus again confirmed hypothesis 1. The non-significant Group x Trial interaction,  $F(2, 129) = 1.01, p = .368, b_1 = -0.07, t(129) = -1.09, p = .277, b_2 = 0.10, t(129) = 0.93, p = .355$ , indicates that the slope of change in investments during this phase did not differ between the conditions, providing no evidence for the second hypothesis during the trust-violation phase. No significant main effect or interactions with ingroup identification were observed on the investments in the trust-violation phase.

In summary, investments in novel partners decreased in the trust-violation phase following the untrustworthy behaviours of others. This was not more pronounced for the outgroup over the ingroup.

### Trust-recovery phase

The third linear multilevel model was performed with group, trial number, and ingroup identification predicting investments in round six to round ten of the game (see Figure 1). The same random structure and planned contrasts for group were used. The fixed effects of this model explained 12% of the variance within the data ( $R^2_{fixed} = 0.12$ ). Adding the random effects to the model increased the amount of variance explained to 80.8% ( $R^2_{total} = 0.81$ ). As in the trust-violation phase, only a significant main effect of trial number was observed,  $F(1, 128) = 28.71, p < .001, b = 0.29$ . This main effect indicates that investments generally increased during the trust recovery phase, also confirming hypothesis 1 in this phase of the game. The non-significant Group x Trial interaction,  $F(2, 128) = 0.96, p = .386, b_1 = 0.05, t(128) = 1.38, p = .169, b_2 = 0.00, t(129) = 0.04, p = .971$ , indicates that the slope of change in investments during this phase did not differ between the groups (see Figure 1), disconfirming hypothesis 4. No significant main effect or interactions with ingroup identification were observed on the investments in the trust-recovery phase.

In summary, overall investments increased during this phase. The prediction that such recovery would occur faster for the ingroup than the outgroup was not supported.

### **Comparisons of slopes over phases**

To examine hypothesis 3 regarding stronger generalisation after negative than positive interactions, changes in investments over time needed to be directly compared between phases. Moreover, differences in generalisation between groups could also be compared over the whole game (instead of only during a specific phase of the game). Figure 2 shows the slope coefficients of trial number (i.e. the amount of change in investments over time) for each of the groups during the trust building, trust violation, and trust recovery phase.

For a statistical comparison of these coefficients between phases of the game and between groups, trial number was regressed on investments separately for each participant and for each phase. This analysis created an intercept and coefficient for trial number for each participant. The coefficients for the trust violation phase were multiplied by -1 to allow for a direct comparison the magnitude of the average slopes between the phases without the problem of the sign of the slope (e.g. slope coefficients of +2 and -2 are equal in magnitude, but would be statistically different in the model).

A linear multilevel model was performed with the slope coefficient as dependent variable and phase (building, violation, recovery) and group as predictors. A per-subject random intercept was added to account for the repeated measures design. For the variable phase (building, violation, recovery), the following contrasts were implemented: the first contrast compared the building phase with the violation phase, the second contrast compared the violation phase with the recovery phase. The fixed effects of this model explained 10% of the variance within the data ( $R^2_{fixed} = 0.10$ ). Adding the random effects to the model increased the amount of variance explained to 27.4% ( $R^2_{total} = 0.27.4$ ).

Significant main effects of both phase,  $F(2, 264) = 16.99, p < .001, b_1 = 0.04, t(264) = 0.79, p = .430, b_2 = -0.23, t(264) = -4.61, p < .001$ , and group,  $F(2, 132) = 4.00, p = .021, b_1 = 0.05, t(132) = 1.65, p = .102, b_2 = -0.13, t(132) = -2.31, p = .022$ , were observed. Post-hoc tests revealed that the slopes were significantly larger in the violation phase than in the building and recovery phase (see Table 2), indicating a stronger change in behaviour during the violation phase, and the lowest change in investments during the recovery phase. This result confirms hypothesis 3. In addition, the average slope for the outgroup was significantly larger than the slope for the ingroup and control group (see Table 3), confirming hypothesis 2.

Table 2

*T-test comparisons of the average slopes of investment change during each of the phases of the Trust Game*

	Trust building	Trust violation	Trust recovery
Building	-		
Violation		$t(134) = -3.66, p < .001,$ $d = 0.31$	-
Recovery		$t(134) = 2.39, p = .018, \quad t(134) = 5.13, p < .001$ $d = 0.44 \quad d = 0.21$	-

*Note.* Comparisons were made between the column header, minus the row header (i.e. building - violation, building - recovery, violation - recovery).

Table 3

*T-test comparisons of the average slopes of investment change for each of the groups*

	Ingroup	Outgroup	Control
Ingroup	-		
Outgroup		-	
<b>t(86) = -2.23, p = .028,</b>			
<b>d = 0.47</b>			
<b>t(88) = 0.30, p = .765,    t(85) = 2.53, p = .013</b>			
Control			-
<b>d = 0.06</b>		<b>d = 0.53</b>	

*Note.* Comparisons were made between the column header, minus the row header (i.e. ingroup - outgroup, ingroup - control, outgroup - control).

### **Relation between expectations and investments**

The relation between expectations of return, as measured each round before the participant made their investment decision, and investments was examined. Firstly, the correlation between expectations and investments over the whole game indicated a positive relation,  $r = 0.47, p < .001$ . Next, for each phase of the game, a model was created with expectations, group, and trial number predicting investments. The same random structure and contrasts were used as in the previous models. In all phases, only a significant main effect of expectations on investments was found: trust-building,  $F(1, 291) = 25.38, p < .001, b = 0.04$ , trust-violation,  $F(1, 213) = 6.00, p = .015, b = 0.06$ , trust-recovery,  $F(1, 196) = 4.31, p = .039, b = 0.06$ . Expectations strongly predicted investments in each phase of the game, but there was no effect of group.

## Discussion

Study 1 examined how people generalise their experiences with group members in their trust behaviour towards novel individuals from the same group. Moreover, the role of group membership (ingroup vs outgroup vs control) and valence (positive vs negative) in member-to-member generalisation process was investigated. The main research question of this project focusses on the generalisation of experiences in the game in trust towards novel individuals. It was predicted that people use their previous experiences and adjust their trust in novel individuals after experiencing positive (trust reciprocation) or negative (trust violation) interactions with other members of the same group. This hypothesis was confirmed; previous interactions strongly influenced behaviour towards novel partners in all phases of the game. Investments increased after positive interactions with previous group members and decreased after negative interactions.

Moreover, ingroup favouritism and a dislike of the outgroup were visible in this study. As seen in Figure 1, participants initially expected and invested much higher amounts in ingroup partners than in control partners, and participants invested even lower amounts in partners from the political outgroup UKIP. A tendency continued for people to invest more in the ingroup than the outgroup, even though differences from the control group were not significant in the last two phases of the game.

Secondly, it was predicted that outgroup interactions should be generalised more readily than experiences with ingroup members, due to heightened homogeneity perceptions within outgroups compared to ingroups (Judd & Park, 1988; Ostrom & Sedikides, 1992; Park & Rothbart, 1982). However, this effect of group membership on generalisation was only confirmed for the trust-building phase. People started off strongly distrusting outgroup members, but after experiencing positive interactions increased their trust much more strongly towards novel outgroup members than towards novel ingroup members. This

enhanced generalisation of outgroup interactions was not observed in the violation- and recovery-phase of the Trust Game.

Thirdly, a general negativity effect was hypothesised. Negative experiences should have a stronger effect on trust behaviour towards other group members than positive experiences. This hypothesis was confirmed, as changes in investments were more pronounced during the trust-violation phase than the trust-building and -recovery phase. Lastly, it was hypothesised that trust recovery towards novel group members after experiencing a violation would be stronger for ingroup members than outgroup members, and people would be more reluctant to trust novel outgroup members than novel ingroup members. However, the results showed a different pattern. Even though overall investments increased less steeply in the recovery phase compared to the trust-building phase, no differences between the groups was observed. It seems that, after establishing general levels of trust for each group at the end of the trust-building phase, people only focussed on previous experiences in their investment decisions and were no longer influenced by the group membership of the partners.

In summary, the results from the first study show evidence for member-to-member generalisation in trust behaviour. People adjusted their behaviour towards novel group members after experiencing positive or negative interactions with other members of that group. The differences in slopes between the three groups in the trust-building phase demonstrate that group membership predicted how people adjusted following trustworthy behaviour. Specifically, people showed ingroup favouritism in their expectations and trust but adjusted their behaviour towards trustworthy outgroup members in the first few rounds of the game. This outgroup generalisation effect did not continue through in the other phases of the game. Moreover, negative experiences led to a stronger change in behaviour towards novel individuals than positive experiences. The second study aimed to replicate Study 1 in a larger

and more representative online sample where participants from different political backgrounds could be compared.

## Study 2

Study 2 was conducted using the online platform Mechanical Turk, which has been shown to produce reliable data from more representative samples than traditional undergraduate samples (Gosling, Sandy, John, & Potter, 2010; Buhrmester et al., 2011; Casler, Bickel, & Hackett, 2013; Hauser & Schwarz, 2016). This setting allowed for a wider sample in a different country than Study 1, including equal numbers of participants from both sides of the political spectrum to confirm effects are not specific to the particular nature of the outgroup used in Study 1. The design of Study 1 was adapted to this online platform. Political affiliation was again used as a target group, where participants indicated their political affiliation by choosing either the version of the study available for supporters of the Democratic Party or the Republican Party.

## Method

### Participants and design

The participant pool consisted of 254 MTurk workers<sup>6</sup> in the United States. The data of 22 participants was removed due to inconsistent responses related to the political affiliation and orientation.<sup>7</sup> The remaining 226 participants (43% female,  $M_{age} = 37.12$ ,  $SD_{age}$

---

<sup>6</sup> The aim was to double the sample size from Study 1 to Study 2 to accommodate comparisons between Democrat and Republican participants.

<sup>7</sup> All participants indicated their political affiliation, as well as voting intentions (the experiment was conducted before the 2016 Presidential Election) and political orientation (liberal to conservative) at the beginning of the

= 11.72) were all US citizens, and consisted of equal numbers of self-identified Democrats and Republicans. Participants received a payment of 0.75 US Dollars for their time and had a chance to win a monetary bonus based on average earnings in the Trust Game, converted to dollars (one token = 10 cents). Ten participants were selected at random to receive the bonus, with an average bonus amount of 1.20 dollars. The same design was utilised as in Study 1.

### Materials and procedure

The experiment was programmed using the online software program Qualtrics and was distributed via Amazon's Mechanical Turk. Informed consent was obtained from all individual participants included in this study. Separate versions of the experiment were created for Democratic and Republican voters and were launched separately on the MTurk website. After informed consent was obtained and the party screening was successful, participants completed the ingroup identification scale. Next, participants' feelings towards the outgroup (either Democrats or Republicans, depending on own affiliation) were measured using a feeling thermometer.

The Trust Game was identical to Study 1, with 10 rounds of the game divided into three phases, and participants played the game rounds with 10 different individuals from either their political ingroup, outgroup, or the control. Symbols of the Republican and Democratic Party were used as images for the ingroup and outgroup partners (depending on chosen affiliation), and the same "affiliation unknown" image was used for control partners as in Study 1. After completing the 10 rounds of the game, participants completed the outgroup attitudes scale and demographics questions. Participants were informed that the

---

experiment. For 22 participants, voting intentions and ratings of political orientation (liberal to conservative) did not match with selected affiliation (e.g. self-identified Democrats intending to vote for Donald Trump and identifying as very conservative).

winners of the bonus amount based on average earnings in the game would be notified later on, after data collection was completed. A written debriefing was presented to participants before they submitted their answers. This study was approved by the appropriate institutional research ethics committee before data collection.

### Data analysis

The same analyses were performed as in Study 1. Additionally, the effect of political affiliation was explored as a predictor of investments in the game.

## Results

### Initial trust bias

An ANOVA on investments in the first round of the Trust Game showed a significant effect of group,  $F(2, 223) = 4.00, p = .020, \eta_p^2 = 0.04$ . Post-hoc independent *t*-tests confirmed that initial investments in the ingroup were significantly higher than initial investments in the outgroup and the control group (see Table 4 for descriptive statistics and t-test statistics). The outgroup and the control group did not differ significantly in their initial investments. The second ANOVA on first round expectations of return confirmed this pattern, showing a significant effect of group,  $F(2, 223) = 4.13, p = .017, \eta_p^2 = 0.04$  (see Table 4).

Table 4

*Mean and standard deviations for initial investments (0 – 10 tokens) and expectations of return in percentage of the received investment (0 to 100%) in partners from the different groups, and t-test comparisons of initial investments between groups*

<b>Investments</b>				
	Mean (SD)	1	2	3
<b>1. Ingroup</b> <i>(N = 73)</i>	6.32 (3.35)	-		
<b>2. Outgroup</b> <i>(N = 73)</i>	4.92 (3.69)	$t(142) = 2.40,$ $p = .018, d = 0.40$	-	
<b>3. Control</b> <i>(N = 80)</i>	4.98 (3.69)	$t(147) = 2.54,$ $p = .012, d = 0.41$	$t(142) = 0.10,$ $p = .919, d = 0.00$	-
<b>Expectations</b>				
	Mean (SD)	1	2	3
<b>1. Ingroup</b>	45.81 (22.53)	-		
<b>2. Outgroup</b>	35.89 (24.85)	$t(142) = 2.53,$ $p = .013, d = 0.42$	-	
<b>3. Control</b>	36.21 (24.11)	$t(150) = 2.55,$ $p = .012, d = 0.41$	$t(148) = 0.08,$ $p = .935, d = 0.01$	-

[Insert Figure 3 here]

*Figure 3.* Mean investments in different group partners for each round of the Trust Game for Study 2. Error bars indicate 95% confidence intervals. The dotted lines indicate the phases of the game (building, violation, recovery).

### Trust-building phase

The first model examined the effect of group, trial number, and average mean-centred ingroup identification score on investments in the trust-building phase of the game (Round 1 to Round 4, see Figure 3). The same fixed and random effects structure was used as in Study 1. The fixed effects of this model explained 6.9% of the variance within the data ( $R^2_{fixed} = 0.07$ ). Adding the random effects to the model increased the amount of variance explained to 89% ( $R^2_{total} = 0.89$ ).

In this model, significant main effects of group,  $F(2, 220) = 3.09, p = .047, b_1 = 0.27, t(220) = 1.61, p = .109, b_2 = 0.56, t(220) = 1.90, p = .059$ , and trial number,  $F(1, 220) = 17.29, p < .001, b = 0.18$ , were observed. The positive coefficient for the main effect of trial indicates that people increased their investments during the trust-building phase, confirming hypothesis 1. Post-hoc multiple comparisons based on the model show that the ingroup received higher investments than the outgroup,  $t(220) = 2.61, p = .027, d = 0.35$ , and higher investments than the control group,  $t(220) = -2.48, p = .037, d = 0.33$ . The outgroup and the control group did not differ in the received investments during the trust building phase,  $t(220) = 0.19, p = .981, d = 0.03$ . The least-square means as predicted from the model for each group are:  $M_{control} = 5.25 (SE = 0.35, 95\% CI [4.55, 5.95])$ ,  $M_{ingroup} = 6.52 (SE = 0.37, 95\% CI [5.79, 7.25])$ , and  $M_{outgroup} = 5.16 (SE = 0.37, 95\% CI [4.42, 5.88])$ , respectively. No differences were found in how the investments changed over trials, as the non-significant Group x Trial interaction indicates,  $F(2, 220) = 0.88, p = .414, b_1 = -0.03, t(220) = -0.98, p = .328, b_2 = 0.05, t(220) = 0.90, p = .369$ . This non-significant interaction disconfirms the

second hypothesis.

A significant Group x Identification interaction was observed,  $F(2, 220) = 4.37, p = .014, b_1 = 0.02, t(220) = 0.15, p = .878, b_2 = 0.66, t(220) = 2.96, p = .003$  (see Figure 4). This interaction indicates that the relation between ingroup identification and investments in the trust building phase is positive for the ingroup,  $intercept = 6.03, b = 0.74, SE = 0.32, t(71) = 2.32, p = .023$ , and negative for the outgroup,  $intercept = 4.90, b = -0.58, SE = 0.35, t(71) = -1.67, p = .099$ . No relation between identification and investments was observed for the control,  $intercept = 4.66, b = 0.02, SE = 0.29, t(81) = 0.07, p = .944$ .

Once again ingroup partners received the highest investments in this trust-building phase; the outgroup did not differ from the control. Ingroup favouritism was particularly pronounced for those high in ingroup identification (as seen in Figure 4), which led to increased investments for the ingroup, and decreased investments for the outgroup. Unlike Study 1, there were no between-group differences in how much investments increased.

[Insert Figure 4 here]

*Figure 4.* Relationship between ingroup identification and investments in the trust-building phase of the Trust Game for Study 2. A scatterplot is presented with separate regression lines for each group. Shaded areas around regression lines indicate 96% confidence intervals.

### Trust-violation phase

For the trust-violation phase, a second linear multilevel model was created with group, trial number, and ingroup identification predicting investments in Round 4 to Round 6 of the Trust Game (see Figure 3). The same random effects were included as in the trust-building phase. The fixed effects of this model explained 6.6% of the variance within the data ( $R^2_{fixed} = 0.07$ ). Adding the random effects to the model increased the amount of variance explained to 80.6% ( $R^2_{total} = 0.81$ ).

In this model, a significant main effect of trial number was observed,  $F(1, 220) = 8.66, p = .004, b = -0.51$ , which shows that investments generally decreased during the violation phase. This main effect again confirms hypothesis 1. The main effect of group was marginally significant,  $F(2, 220) = 2.58, p = .078, b_1 = 0.23, t(220) = 0.65, p = .519, b_2 = 1.36, t(220) = 2.18, p = .031$ . Post-hoc multiple comparisons based on the model show that the ingroup received higher investments than the outgroup,  $t(220) = 2.94, p = .010, d = 0.40$ , and higher investments than the control group,  $t(220) = -2.20, p = .074, d = 0.30$ . The outgroup and the control group did not differ in the received investments during the trust-violation phase,  $t(220) = 0.81, p = .696, d = 0.11$ . The least-square means as predicted from the model for each group are:  $M_{control} = 5.14$  ( $SE = 0.35$ , 95% CI [4.45, 5.83]),  $M_{ingroup} = 6.25$  ( $SE = 0.37$ , 95% CI [5.53, 6.98]), and  $M_{outgroup} = 4.72$  ( $SE = 0.37$ , 95% CI [4.00, 5.45]), respectively. No differences were found in how the investments changed over trials, as the non-significant Group x Trial interaction indicates,  $F(2, 220) = 0.55, p = .576, b_1 = -0.02, t(220) = -0.33, p = .744, b_2 = -0.12, t(220) = -1.00, p = .318$ .

In addition, the Group x Identification interaction was again significant,  $F(2, 220) = 3.72, p = .026, b_1 = 0.49, t(220) = 1.77, p = .077, b_2 = 0.99, t(220) = 2.12, p = .035$ . This interaction indicated the same results as in the trust-building phase. The 3-way Group x Trial x Identification interaction was marginally significant,  $F(2, 220) = 2.50, p = .084, b_1 = -0.10, t(220) = -1.88, p = .062, b_2 = -0.11, t(220) = -1.26, p = .208$ . To explore the three-way interaction in more depth, a median-split factor of ingroup identification was created ( $Mdn = 4.71, N_{low} = 105, N_{high} = 121$ ). Adding this factor to the model instead of the continuous scores produced a significant three-way interaction,  $F(2, 220) = 3.54, p = .031, b_1 = -0.33, t(220) = -2.50, p = .013, b_2 = -0.22, t(220) = -0.93, p = .352$ , see Figure 5 and Table 5.

Separate models were created for participants that indicated high and low identification with the ingroup. For highly identified participants, investments in the ingroup

decreased more strongly than investments in the control group,  $\chi^2(1) = 5.80, p = .048$  (see Table 5). This Group x Trial interaction examines hypothesis 2, but the result indicates an effect opposite to the hypothesised direction. For low identifiers, no differences were observed between the slopes of investments in the violation phase.

As in Study 1, the trust-violation phase decreased investments overall. The tendency for higher investments in the ingroup than the outgroup or the control group continued in this phase. Overall, the magnitude of decrease did not differ between ingroups and outgroups. However, participants reporting high ingroup identification showed more decrease in ingroup investments than for the control group.

[Insert Figure 5 here]

*Figure 5.* Relationship between ingroup identification and investments over trials in the trust-violation phase for Study 2. Regression lines for each group are presented and separate plots for high and low identifiers. Shaded areas around regression lines indicate 96% confidence intervals.

*Table 5*

Intercepts, unstandardized coefficients, and standard errors of the slopes of investments in partners from different groups, presented separately for high and low identified participants.

	Low identification			High identification		
	Intercept	b	SE	Intercept	b	SE
<b>Ingroup</b>	6.19	-0.14	0.22	12.10	-1.05***	0.24
<b>Outgroup</b>	6.14	-0.17	0.29	7.41	-0.63**	0.19
<b>Control</b>	7.84	-0.63**	0.23	7.14	-0.32	0.21

*Note.* The notations on the coefficient indicate whether the effect of trial number on investments is significantly different from zero.

\*\*\*  $p < .001$ , \*\*  $p < .01$ .

### **Trust recovery phase**

A third linear multilevel model was created with group, trial number, and ingroup identification predicting investments in the trust recovery phase (round 6 to 10, see Figure 3). The same random effects were again included. The fixed effects of this model explained 3.8% of the variance within the data ( $R^2_{fixed} = 0.04$ ). Adding the random effects to the model increased the amount of variance explained to 85.8% ( $R^2_{total} = 0.86$ ). In this model, only a significant main effect of trial number was observed,  $F(1, 220) = 19.81, p < .001, b = 0.18$ . This main effect indicates that investments generally increased during the trust-recovery phase, again confirming hypothesis 1. No significant main effect of group,  $F(2, 220) = 0.59, p = .556, b_1 = 0.11, t(220) = 0.42, p = .678, b_2 = 0.49, t(220) = 1.00, p = .318$ , or Group x Trial interaction,  $F(2, 220) = 0.05, p = .949, b_1 = 0.005, t(220) = 0.18, p = .860, b_2 = 0.014, t(220) = 0.27, p = .787$ , was observed during the trust-recovery phase. The non-significant disconfirms hypothesis 4 for Study 2. This phase showed that investments increased over time, with no between-group differences.

### **Comparisons of slopes over phases**

To examine hypothesis 3 regarding stronger generalisation after negative than positive interactions, changes in investments over time were again directly compared between phases and between groups. Figure 6 below shows the slope coefficients of trial number (i.e. the amount of change over time) for each of the groups during the trust building, trust violation, and trust recovery phase. As in Study 1, trial number was regressed on investments separately for each participant and for each phase to create an intercept and coefficient for trial number for each participant. Coefficients for the trust violation phase were multiplied by -1 to make

the numbers positive and allow for a direct comparison of the magnitude of the average slopes between the phases.

[Insert Figure 6 here]

*Figure 6.* Boxplots of coefficients of investment change over trials for the different phases of the Trust Game and for the different groups, for Study 2.

A linear multilevel model was created with the slope coefficient as dependent variable and the phase (building, violation, recovery) and group as predictors. A per-subject random intercept was added to account for the repeated measures design. The same contrasts were implemented for the variable phase (building, violation, recovery) as in Study 1. The fixed effects of this model explained 3% of the variance within the data ( $R^2_{fixed} = 0.03$ ). Adding the random effects to the model increased the amount of variance explained to 42.4% ( $R^2_{total} = 0.42$ ).

In this model, only a significant main effect of phase,  $F(2, 446) = 15.03, p < .001, b_1 = 0.11, t(446) = 2.77, p = .006, b_2 = -0.11, t(446) = -2.71, p = .007$ , was observed. The significance of both contrasts indicates that the slope of the violation phase was significantly steeper than both the building and recovery phase, confirming hypothesis 3. No main effect of group was observed on the investment coefficients over trials,  $F(2, 223) = 0.52, p = .597, b_1 = -0.00, t(223) = -0.02, p = .987, b_2 = 0.06, t(223) = 1.02, p = .310$ , disconfirming hypothesis 2. Moreover, the Group x Phase interaction was also non-significant,  $F(2, 446) = 0.71, p = .584, b_{11} = 0.03, t(446) = 1.03, p = .305, b_{21} = 0.01, t(446) = 0.20, p = .840, b_{12} = 0.01, t(446) = 0.287, p = .774, b_{22} = -0.05, t(446) = -0.95, p = .345$ .

### Differences between Democrats and Republicans

For the three phases of the Trust Game, it was explored whether Democrat and Republican participants differed in their behaviour. For each phase, a linear multilevel model was created with group, trial number, and political affiliation (Democrat, Republican) predicting investments. The same random structure and contrasts were used as in the previous models. A marginally significant Trial x Political affiliation interaction was observed in the trust-building phase,  $F(1, 220) = 3.21, p = .074, b = -0.15$ , and in the trust-violation phase,  $F(1, 220) = 5.60, p = .019, b = 0.45$ . For both these phases, the slope of investments over trials was somewhat steeper for Democrat participants,  $b_{Building} = 0.26, SE = 0.06, t(110) = 4.04, p < .001, b_{Violation} = -0.74, SE = 0.13, t(110) = -5.86, p < .001$ , than Republican participants,  $b_{Building} = 0.10, SE = 0.06, t(110) = 1.85, p = .068, b_{Violation} = -0.29, SE = 0.14, t(110) = -2.00, p = .048$ . No effects of political affiliation were observed in the trust recovery phase, Trial x Political affiliation,  $F(1, 220) = 0.44, p = .509, b = -0.05$ .

### **Relation between expectations and investments**

The relation between expectations of return and investments was again examined. Firstly, the correlation between expectations and investments over the whole game indicated a positive relation,  $r = 0.65, p < .001$ . Next, for each phase of the game, a model was created with expectations, group, and trial number predicting investments. The same random structure and contrasts were used as in the previous models. In all phases, only a significant main effect of expectations on investments was found: trust-building,  $F(1, 442) = 60.56, p < .001, b = 0.04$ , trust-violation,  $F(1, 281) = 26.44, p < .001, b = 0.09$ , trust-recovery,  $F(1, 463) = 48.00, p < .001, b = 0.08$ . Expectations strongly predicted investments in each phase of the game, but there were no effects of group.

### **Discussion**

Study 2 examined member-to-member generalisation in trust behaviour in a larger and more representative sample, to examine the role of ingroup identification in moderating generalisation processes. Moreover, Study 2 contained an equal number of participants from opposing political groups; this was not controlled for in Study 1. Four main findings were replicated from Study 1. Firstly, a clear effect of previous game experiences was observed, again confirming the first hypothesis. Participants increased and decreased their investments in new, unknown group members based on the trustworthiness of other group members. Secondly, ingroup favouritism in trust decisions was observed in the game. Both Democrats and Republicans showed higher trust when interacting with partners from their own political group more than when interacting with partners from the political outgroup. Thirdly, investments changed more strongly after negative interactions in the trust-violation phase than after positive interactions in the trust-building and trust-recovery phase, again confirming hypothesis 3. Lastly, Study 2 again showed no differentiation between groups on trust recovery after violation, indicating that people were equally likely to trust new ingroup as outgroup members after experiencing trust violations and trust reciprocation.

Unlike Study 1, an effect of ingroup identification was observed. People who highly identified with their political ingroup showed the strongest levels of ingroup favouritism, while people with less ingroup identification showed outgroup favouritism. The differences between the two studies in effects of ingroup identification could be due to the oppositional nature of American political parties, where identification with one party almost automatically implies opposition to the other party (Iyengar & Westwood, 2015). However, in the multiple party system in the UK, most political parties are not explicitly constructed in opposition to UKIP. People who might not strongly identify with one party, may particularly be in opposition to UKIP, and therefore identification with these other parties might be a weaker predictor of behaviour.

When examining the effect of group membership, differences were only found at initial investment levels (i.e. an initial ingroup bias) and during the violation phase of the Trust Game. Participants increased their investments at similar rates for all groups during the trust-building phase and the trust-recovery phase. However, when participants encountered negative behaviour from other partners in the violation phase, group information became more important. Surprisingly, the results only showed an effect of group membership during the violation phase for participants who highly identified with their political ingroup. High ingroup identifiers showed a stronger decrease in investments for the ingroup than for the control group (i.e. unknown affiliation). Changes in investments in outgroup partners did not significantly differ from the other groups. For low identifying participants, no differences were found between investments for partners from the different conditions. Thus, the second hypothesis about stronger outgroup generalisation than ingroup generalisation was not confirmed.

To summarise, Study 2 again showed that 1) people use previous experiences with other group members to inform their current decisions to trust, and 2) member-to-member generalisation in trust behaviour is stronger for negative interactions than positive interactions. Moreover, people who highly identify with the ingroup favour their ingroup in trust decisions, and show stronger generalisation of ingroup violations of trust than violations from the outgroup or a control where group membership is unknown.

### General discussion

The aim of this research was to examine how people use their previous experience with in- and outgroup members to inform their decisions to trust new, unknown individuals from the same group, a process we describe as member-to-member generalisation. Intergroup contact research demonstrates that positive and negative experiences with individual outgroup members correspondingly change attitudes towards their group as a whole (e.g. Brown & Hewstone, 2005; Pettigrew, Tropp, Wagner, & Christ, 2011). To examine whether actual behaviour towards novel individual group members is also influenced by such contact encounters a paradigm was designed where people interacted through a series of Trust Games with different members from the ingroup, outgroup, or a control (with no group information). We predicted that people generalise their experiences with group members in their trust towards novel individuals from that same group, and that this generalisation is enhanced for outgroup members compared to ingroup members, and for negative compared to positive interactions.

Both experiments showed an effect of previous interactions on investments in unknown partners, indicating member-to-member generalisation. People were not just guided by their group-based biases in decisions to trust, nor were they completely individuating the novel individual. Rather, they adjusted their behaviour towards novel individuals of the same group based on past experiences. The amount of adjustment depended at least partially on the individual's group membership. Between-group differences in generalisation were observed in the trust building phase in Study 1, and group identification moderated responses to trust-building and trust-violation in Study 2. Both studies consistently showed stronger generalisation of negative experiences than positive experiences, confirming hypotheses based on the negativity bias (Baumeister et al., 2001; Fazio et al., 2004; Rozin & Royzman, 2001; Shook et al., 2007). In the context of the Trust Game, reducing investments in

untrustworthy partners is more important to maintaining a profit than increasing investments in trustworthy partners, as the first option prevents losses, while the second option only increases gains. Furthermore, the hypothesised stronger recovery of trust for ingroup partners than outgroup partners was not confirmed in either study. After experiencing negative interactions with ingroup or outgroup partners, participants increased their trust again during the recovery phase.

Although a strict tit-for-tat strategy is illogical for novel partners, participants appear to be influenced more by their immediate previous experience (trustworthiness in the previous round), than by experiences that were further in the past (trust violation a few rounds back). However, increases in trust were weaker in the trust-recovery phase than in the trust-building phase. This indicates that people had become more cautious in trusting novel group partners after experiencing violations of trust.

Interestingly, the two experiments showed different findings regarding the effect of group membership on member-to-member generalisation. Study 1 found that initial dislike of the outgroup led to a stronger increase in trust towards novel outgroup members during the trust-building phase of the game. After that, similar levels of trust towards novel ingroup and outgroup members was established. Study 2 found that people who highly identified with the ingroup showed the highest trust in ingroup partners. However, the effect of group on generalisation of trust only occurred in the trust-violation phase for highly identified participants. People who strongly cared about their political ingroup, and therefore highly trusted their ingroup, decreased their trust in novel ingroup members more quickly than towards outgroup or control group members. We propose that these two contrasting findings can be explained together through *expectancy-violation* effects.

### **Expectancy-violation in trust behaviour**

The stronger changes in trust behaviour towards untrustworthy ingroup partners and trustworthy outgroup partners could be explained through the lens of an expectancy-violation effect. Research indicates that people respond more strongly to unexpected information than expected information. Exposure to group members behaving in a stereotype-incongruent manner leads to stronger affective arousal and more extreme evaluations of the group members, compared to stereotype-congruent behaving group members (Bartholow, Fabiani, Gratton, & Bettencourt, 2001; Bettencourt, Dill, Greathouse, Charlton, & Mulholland, 1997; Kernahan, Bartholow, & Bettencourt, 2000).

Moreover, behaviour that is incongruent with expectations or person impressions receives higher levels of cognitive processing and leads to better memory than congruent behaviour (Hamilton & Sherman, 1996; Stangor & McMillan, 1992). This effect occurs because a violation of expectations has a stronger information value and requires a change in response, whereas a confirmation of expectations conveys a signal to continue unchanged. Our results suggest that member-to-member generalisation in the game was not just driven by group membership or valence of the previous interaction, but also by how the previous experience related to expectations and general impressions of the group.

When people showed strong distrust of the outgroup (Study 1) or were highly identified with the ingroup (Study 2), they changed their behaviour towards novel partners more after experiencing incongruent interactions. Together, these two experiments with the adapted Trust Game paradigm offer some evidence that expectancy-violation effects can transfer from one group member to the next. This may be a potential driving force of generalisation of previous experiences in informing decisions to trust novel group members

### Limitations and future directions

Three limitations of this research need to be addressed. Firstly, a between-subject design was adopted to manipulate group membership of the partners. The choice of this design was a practical consideration, as a fully random order within-subject design would not allow for analysis of the effect of direct previous positive or negative experiences with a certain group on current decisions to trust and might create spill-over effects between groups. A blocked within-subject design would most likely lead participants to notice the pattern of trustworthy and untrustworthy behaviour that the partners displayed. However, a between-subjects design is limited in that it cannot establish intergroup bias on an individual level.

Secondly, high levels of individual variation in investment behaviour (i.e. majority of explained variance arising from random effects for participants) and inconsistency between studies of the effect of group membership create some difficulty in drawing conclusions about differences in ingroup and outgroup member-to-member generalisation. The observed variance in effects could be influenced by the large scope of potential investment decisions in the Trust Game (0-10) and could potentially be minimised by using a different type of cooperative game with more limited behavioural options and strategies. Restricting the response options to a binary Trust Game (i.e. trust vs not-trust, reciprocate vs not-reciprocate) will reduce individual variance in behaviour, and might produce more stable effects of group membership. However, a binary Trust Game cannot capture incremental changes in trust over time, as was demonstrated in the current research.

Thirdly, while the Trust Game paradigm provides a controlled setting in which to study social interactions and quantify trust, the paradigm also lacks external validity (Winking & Mizer, 2013). Therefore, it can be difficult to draw conclusions about daily life experiences of intergroup interactions. Moreover, as the Trust Game was played with virtual players, there was no physical interaction where people could see each other or communicate

with each other. Recent studies have shown that real, physical interactions in economic games can have rather different results than virtual interactions (e.g. Bhogal, Galbraith, & Manktelow, 2016).

Future research should seek to replicate the observed effects and explore further the role of expectancy-violation in these effects. These two experiments were the first to use this paradigm to examine member-to-member generalisation of trust behaviour. Interesting findings regarding generalisation, favouritism, and negativity-biases were consistently observed in the two studies. The first step for future work is to examine the robustness of this effect in large-scale replications, utilising different target groups and samples. Moreover, future work should examine member-to-member generalisation in a more natural and realistic setting. This could potentially be achieved by using different types of tasks and interactions, using face-to-face paradigms, such as cooperative tasks with ingroup or outgroup participants or confederates.

Furthermore, the effects of group membership that were observed in the two studies suggest an effect of expectancy-violation in the generalisation of trust behaviour towards novel group members. Expectations of trustworthiness of partners correlated strongly with investment behaviour in the game, making this explanation plausible. However, for a systematic test of the influence of expectancy-violation, expectations about the partners and the groups should be measured and manipulated prior to the game for a direct comparison. In general, we look forward to future research that will examine this important but under-researched type of generalisation, and how it applies to the formation and change of prosocial behaviour.

## Conclusions

This research examined how previous interactions with ingroup or outgroup members can influence trust behaviour towards novel individuals from the same group, thereby extending the intergroup contact literature beyond attitudes. The findings from two experiments show evidence for member-to-member generalisation; people are strongly influenced by their previous experiences with group members in forming decisions to trust novel group members. This effect is particularly strong for negative interactions, indicating a negativity bias. Moreover, violations of expectations and beliefs about the groups could be a driving force of the level of generalisation. People responded more strongly to negative interactions with trusted ingroup partners, or to positive interactions with distrusted outgroup partners. Future research should seek to replicate these findings in more natural settings, and explore further the role of expectancy-violation.

## Compliance with Ethical Standards

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## References

- Allport, G. W. (1954). *The nature of prejudice*. Reading, United Kingdom: Addison-Wesley.
- Axelrod, R., & Hamilton, W. (1981). The evolution of cooperation. *Science*, 211, 1390–1396.  
<https://doi.org/10.1126/science.7466396>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.  
<https://doi.org/10.1016/j.jml.2007.12.005>
- Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, 100, 603–617. <https://doi.org/10.1348/000712608X377117>
- Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin*, 139, 1090–1112. <https://doi.org/10.1037/a0030939>
- Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, 140, 1556–1581. <https://doi.org/10.1037/a0037737>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. <https://doi.org/10.1016/J.JML.2012.11.001>
- Bartholow, B. D., Fabiani, M., Gratton, G., & Bettencourt, B. A. (2001). A psychophysiological examination of cognitive processing of and affective responses to social expectancy violations. *Psychological Science*, 12, 197–204.  
<https://doi.org/10.1111/1467-9280.00336>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>

- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10, 122–142. <https://doi.org/10.1006/game.1995.1027>
- Bettencourt, B. A., Dill, K. E., Greathouse, S. A., Charlton, K., & Mulholland, A. (1997). Evaluations of ingroup and outgroup members: The role of category-based expectancy violation. *Journal of Experimental Social Psychology*, 33, 244–275. <https://doi.org/10.1006/jesp.1996.1323>
- Bhogal, M. S., Galbraith, N., & Manktelow, K (2016). Physical attractiveness and altruism in two modified dictator games. *Basic and Applied Social Psychology*, 38, 212-222. <https://doi.org/10.1080/01973533.2016.1199382>
- Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, 55, 429–444. <https://doi.org/10.1111/0022-4537.00126>
- Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 37, pp. 255–343). San Diego, CA, US: Elsevier Academic Press. [https://doi.org/10.1016/S0065-2601\(05\)37005-5](https://doi.org/10.1016/S0065-2601(05)37005-5)
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3–5. <https://doi.org/10.1177/1745691610393980>
- Camerer, C. (2003). *Behavioral game theory*. New York, NY, US: Russell Sage Foundation.
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29, 2156–2160. <https://doi.org/10.1016/J.CHB.2013.05.009>

- Cinnirella, M. (1997). Towards a European identity? Interactions between the national and European social identities manifested by university students in Britain and Italy. *British Journal of Social Psychology*, 36, 19–31. <https://doi.org/10.1111/j.2044-8309.1997.tb01116.x>
- Crawford, M. T., Sherman, S. J., & Hamilton, D. L. (2002). Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology*, 83, 1076–1094. <https://doi.org/10.1037/0022-3514.83.5.1076>
- Davies, K., Tropp, L. R., Aron, A., Pettigrew, T. F., & Wright, S. C. (2011). Cross-group friendships and intergroup attitudes: A meta-analytic review. *Personality and Social Psychology Review*, 15, 332–351. <https://doi.org/10.1177/1088868311411103>
- Deegan, M. P., Hehman, E., Gaertner, S. L., & Dovidio, J. F. (2015). Positive expectations encourage generalization from a positive intergroup interaction to outgroup attitudes. *Personality and Social Psychology Bulletin*, 41, 52–65.  
<https://doi.org/10.1177/0146167214556240>
- Dovidio, J. F., Gaertner, S. L., Kawakami, K., & Hodson, G. (2002). Why can't we just get along? Interpersonal biases and interracial distrust. In *Cultural Diversity and Ethnic Minority Psychology* (Vol. 8, pp. 88–102). <https://doi.org/10.1037/1099-9809.8.2.88>
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62–68.  
<https://doi.org/10.1037/0022-3514.82.1.62>
- Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology*, 87, 293–311.
- Fehr, J., & Sassenberg, K. (2009). Intended and unintended consequences of internal

- motivation to behave nonprejudiced: The case of benevolent discrimination. *European Journal of Social Psychology*, 39, 1093–1108. <https://doi.org/10.1002/ejsp.620>
- Finseraas, H., Hanson, T., Johnsen, Å. A., Kotsadam, A., & Torsvik, G. (2019). Trust, ethnic diversity, and personal contact: A field experiment. *Journal of Public Economics*, 173, 72–84.
- Graf, S., Paolini, S., & Rubin, M. (2014). Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and contact prevalence in five Central European countries. *European Journal of Social Psychology*, 44, 536–547. <https://doi.org/10.1002/ejsp.2052>
- Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, 52, 15–23.  
<https://doi.org/10.1016/J.JESP.2013.12.001>
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, 103, 336–355. <https://doi.org/10.1037/0033-295X.103.2.336>
- Haslam, S. A., Oakes, P. J., Turner, J. C., & McGarty, C. (1995). Social categorization and group homogeneity: Changes in the perceived applicability of stereotype content as a function of comparative context and trait favourableness. *British Journal of Social Psychology*, 34, 139–160. <https://doi.org/10.1111/j.2044-8309.1995.tb01054.x>
- Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, 48, 400–407. <https://doi.org/10.3758/s13428-015-0578-z>
- Herek, G. M. (1986). The instrumentality of attitudes: Toward a neofunctional

- theory. *Journal of Social Issues*, 42, 99–114. <https://doi.org/10.1111/j.1540-4560.1986.tb00227.x>
- Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, 59, 690–707. <https://doi.org/10.1111/ajps.12152>
- Judd, C. M., & Park, B. (1988). Out-group homogeneity: Judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology*, 54, 778–788. <https://doi.org/10.1037/0022-3514.54.5.778>
- Kernahan, C., Bartholow, B. D., & Bettencourt, B. A. (2000). Effects of category-based expectancy violation on affect-related evaluations: Toward a comprehensive model. *Basic and Applied Social Psychology*, 22, 85–100. [https://doi.org/10.1207/S15324834BASP2202\\_2](https://doi.org/10.1207/S15324834BASP2202_2)
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, 308, 78–83. <https://doi.org/10.1126/science.1108062>
- Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, 45, 152–168. <https://doi.org/10.1002/ejsp.2079>
- Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, 32, 991–1022. <https://doi.org/10.1177/0149206306294405>
- Marques, J. M., Yzerbyt, V. Y., & Leyens, J.-P. (1988). The “Black Sheep Effect”: Extremity of judgments towards ingroup members as a function of group identification. *European*

- Journal of Social Psychology, 18*, 1–16. <https://doi.org/10.1002/ejsp.2420180102>
- Meleady, R., & Seger, C. R. (2017). Imagined contact encourages prosocial behavior towards outgroup members. *Group Processes & Intergroup Relations, 20*, 447–464.  
<https://doi.org/10.1177/1368430215612225>
- Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin, 112*, 536–552. <https://doi.org/10.1037/0033-2909.112.3.536>
- Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology, 42*, 1051–1068.  
<https://doi.org/10.1037/0022-3514.42.6.1051>
- Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65–85.  
<https://doi.org/10.1146/annurev.psych.49.1.65>
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*, 751–783.  
<https://doi.org/10.1037/0022-3514.90.5.751>
- Pettigrew, T. F., Tropp, L. R., Wagner, U., & Christ, O. (2011). Recent advances in intergroup contact theory. *International Journal of Intercultural Relations, 35*, 271–280.  
<https://doi.org/10.1016/j.ijintrel.2011.03.001>
- R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs

- immediately; explicit attitude generalization takes time. *Psychological Science*, 19, 249–254. <https://doi.org/10.1111/j.1467-9280.2008.02076.x>
- Ratliff, K. A., & Nosek, B. A. (2011). Negativity and outgroup biases in attitude formation and transfer. *Personality and Social Psychology Bulletin*, 37, 1692–1703. <https://doi.org/10.1177/0146167211420168>
- Riek, B. M., Mania, E. W., Gaertner, S. L., McDonald, S. A., & Lamoreaux, M. J. (2010). Does a common ingroup identity reduce intergroup threat? *Group Processes & Intergroup Relations*, 13, 403–423. <https://doi.org/10.1177/1368430209346701>
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23, 393 - 404. <https://doi.org/10.5465/AMR.1998.926617>
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5, 296–320. [https://doi.org/10.1207/S15327957PSPR0504\\_2](https://doi.org/10.1207/S15327957PSPR0504_2)
- RStudio Team. (2015). RStudio: Integrated development environment for R. Boston, MA: RStudio, Inc. Retrieved from <http://www.rstudio.com/>
- Sheeran, P. (2002). Intention–behaviour relations: A conceptual and empirical review. In M. Hewstone & W. Stroebe (Eds.), *European review of social psychology* (Vol. 12, pp. 1–36). Chichester, UK: Wiley
- Shook, N. J., Fazio, R. H., & Eiser, J. R. (2007). Attitude generalization: Similarity, valence, and extremity. *Journal of Experimental Social Psychology*, 43, 641–647. <https://doi.org/10.1016/J.JESP.2006.06.005>
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression

- formation: A review of explanations. *Psychological Bulletin*, 105, 131–142.  
<https://doi.org/10.1037/0033-2909.105.1.131>
- Smith, E. R., & Henry, S. (1996). An in-group becomes part of the self: Response time evidence. *Personality and Social Psychology Bulletin*, 22, 635–642.  
<https://doi.org/10.1177/0146167296226008>
- Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, 111, 42–61. <https://doi.org/10.1037/0033-2909.111.1.42>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–37). Monterey, CA: Brooks-Cole.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology*, 46(1), 35-57. <https://doi.org/10.1086/406755>
- Van Oudenhoven, J. P., Groenewoud, J. T., & Hewstone, M. (1996). Cooperation, ethnic salience and generalization of interethnic attitudes. *European Journal of Social Psychology*, 26, 649–661. [https://doi.org/10.1002/\(SICI\)1099-0992\(199607\)26:4<649::AID-EJSP780>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-0992(199607)26:4<649::AID-EJSP780>3.0.CO;2-T)
- van 't Wout, M. & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgements in social decision-making. *Cognition*, 108, 796 - 803.  
<https://doi.org/10.1016/j.cognition.2008.07.002>
- Vermue, M., Seger, C. R., & Sanfey, A. G. (2018). Group-based biases influence learning about individual trustworthiness. *Journal of Experimental Social Psychology*, 77, 36–49.  
<https://doi.org/10.1016/j.jesp.2018.04.005>

Wilson, R. K., & Eckel, C. C. (2006). Judging a Book by its Cover: Beauty and Expectations in the Trust Game. *Political Research Quarterly*, 59, 189–202.

<https://doi.org/10.1177/106591290605900202>

Winking, J., & Mizer, N. (2013). Natural-field dictator game shows no altruistic giving. *Evolution and Human Behavior*, 34, 288-293.

<https://doi.org/10.1016/j.evolhumbehav.2013.04.002>.