C A G E

Anticipation of Discrimination, Misperceptions, and Trust: Application to Affective Polarization

CAGE working paper no. 738

December 2024
(Revised April 2025)

Devesh Rustagi,
Matthias Schief

WARWICK
THE UNIVERSITY OF WARWICK

UKRI Economic and Social Research Council

Anticipation of Discrimination, Misperceptions, and Trust: Application to Affective Polarization *

Devesh Rustagi† Matthias Schief‡

Abstract

Does anticipation of discrimination, beliefs individuals have about the discriminatory behavior of others toward them, undermine trust and cooperation? We develop a new design to isolate the role of anticipation of discrimination in cooperation dilemmas using a trust game. We capture the effect of anticipation on trust as the double difference between the amount transferred by trustors to outgroup vs. ingroup trustees when their own identity is revealed vs. concealed. We apply our design in the context of affective polarization in the UK using a large representative sample. We find that anticipation of discrimination undermines inter-partisan trust and cooperation by the same magnitude as the combined effect of taste-based and statistical discrimination. However, this anticipation is misperceived because trustees rarely discriminate along partisan lines, resulting in cooperation failure. Our design can be used to study anticipation of discrimination across different societal divisions including gender, ethnicity, religion, and caste.

JEL: C91, C93, J15, D72, Z13

Keywords: discrimination, anticipation, misperception, trust game, cooperation dilemma, affective polarization, United Kingdom

*We are especially thankful to Pedro Dal Bo, Stefano Caria, Oded Galor, Vesa-Matti Heikkuri, Stelios Michalopoulos, Jesse Shapiro, Bertil Tunggoden, Noam Yuchtman, David Weil, and Basit Zafar for extensive discussions and feedback on the paper. We are thankful to seminar participants at Brown University, MIT Sloan, Michigan Ann Arbor, University of Nottingham, and University of Warwick. Devesh Rustagi acknowledges funding for this project from Goethe University, Frankfurt. Our study was approved by the Review Board of Goethe University Frankfurt.

†University of Warwick. E-mail: devesh.rustagi@warwick.ac.uk

‡OECD. E-mail: matthias.schief@gmail.com. The views expressed in this paper are solely those of the authors and should not be interpreted as reflecting those of the Organisation for Economic Cooperation and Development.

I. Introduction

Many economic interactions require trust between individuals from different groups.[1] A large body of evidence shows that trust can be undermined because of *taste-based* (Becker, 1957) and *statistical discrimination* (Arrow, 1970; Phelps, 1972). However, we know little about the role of another factor which can also undermine trust – *anticipation of discrimination*. Individuals anticipate discrimination if they believe that others will engage in taste-based or statistical discrimination toward them. A classical example concerns minorities, who might be reluctant to invest in human capital if they anticipate discrimination from employers. Other examples include a reluctance to migrate despite better opportunities or the unwillingness to cooperate to provide public goods in anticipation of defection (Fehr and Gächter, 2000). As is evident from these examples, anticipation can result in self-fulfilling prophecies, which can undermine trust by reinforcing taste-based and statistical discrimination (Arrow, 1973; Coate and Loury, 1993). Anticipation of discrimination can also lead to coordination failures if people misperceive others to engage in taste-based or statistical discrimination.

There is hardly any empirical evidence on the importance of anticipation of discrimination for trust. This is because the anticipation of discrimination is often confounded with other motives that affect trust at the same time. For example, people might be reluctant to migrate or provide public goods not only because they anticipate discrimination but also because they themselves engage in taste-based and statistical discrimination. Separating the effect of anticipation from these confounding motives is challenging.

In this paper, we first develop a new experimental design to overcome these challenges and highlight the role anticipation of discrimination plays in the erosion of trust. Second, we investigate whether or not the anticipation of discrimination is misperceived (see Bursztyn and Yang, 2022; Alesina, Miano and Stantcheva, 2023). We apply our design to shed new light on the poorly understood motives behind an imperative global problem – affective polarization, which is defined as the hostility between members of opposing political camps that transcends questions of politics (Iyengar et al., 2019). Recent studies document widespread affective polarization in

[1] Putnam, Leonardi and Nanetti (1993); Platteau (2000); Fehr (2009); Tabellini (2010); Alesina and Giuliano (2015); Enke (2024)

many countries (Boxell, Gentzkow and Shapiro, 2020) and suspect that it plays a crucial role in the erosion of trust (see Guriev and Papaioannou, 2022).

Our study takes place in the UK, which has witnessed high levels of affective polarization, especially since Brexit (Hobolt, Leeper and Tilley, 2021). We recruit a large sample of about 1600 participants via Facebook and ensure that these participants have similar socio-economic characteristics as the adult population of the UK.[2]

Our experimental design is based on trust, which is a vital ingredient for economic exchange and development (Arrow, 1972; Fukuyama, 1996; Knack and Keefer, 1997; Algan and Cahuc, 2010; Putnam, 2000). We use a trust game, which is the cornerstone of cooperation dilemmas and for measuring trust as a cultural trait (Berg, Dickhaut and McCabe, 1995; Glaeser et al., 2000; Guiso, Sapienza and Zingales, 2006; Fehr, 2009; Alesina and Giuliano, 2015). The game has two players, a trustor and a trustee. We give both players the same endowment at the beginning of the game. Crucially, the players take decisions sequentially: the trustor decides first to send any amount between zero and her endowment to the trustee, which is tripled by the experimenter. The trustee then decides whether and how much to transfer back to the trustor. The total payoffs are maximized if the trustor sends her entire endowment in the expectation that the trustee reciprocates. However, if the trustee does not, the trustor loses money. This creates a cooperation dilemma.

The trust game has three implications for isolating the role of anticipation from other confounding motives: (a) since the trustor moves first, her decision could reflect anticipation of discrimination, as well as taste-based and statistical discrimination; (b) since the trustee is the second mover, she can condition her decision on the trustor's transfer. This implies that the trustee's decision reflects taste-based discrimination, but not statistical discrimination or anticipation of discrimination; (c) since the trustee does not need to anticipate what the trustor anticipates and the trustor knows that, the game rules out higher-order beliefs from playing a role. Thus, in our design, we need to separate anticipation of discrimination from other motives only for the trustor, whereas the behavior of trustees can be used to study whether anticipation of discrimination by trustors is misperceived.[3]

[2]Our sample size is similar to that of Bursztyn, Egorov and Fiorin (2020) who recruit 1600 participants for their online experiment.

[3]In the trust game, the trustee's decision may be affected by guilt aversion, which is defined as "striving to live up to others' expectations" (Charness and Dufwenberg, 2006), but the evidence is mixed (Ellingsen et al., 2010). Guilt aversion is expected to play a role in experiments with com-

To separate the different motives underlying a trustor's behavior, we augment the trust game as follows. We randomly match individuals with an ingroup or outgroup person, so that individuals are aware that there is a 50 % chance of matching with an outgroup person. Then we experimentally vary the information structure of the game. We flip a coin for each individual within a match to determine whether her own identity is revealed or concealed to the other person, such that the outcomes of the coin flip are common knowledge. This generates four key treatments: outgroup-revealed, ingroup-revealed, outgroup-concealed, and ingroup-concealed.

We take the *first* difference (D_1) between the amount sent by the trustors to outgroup vs. ingroup trustees, when the trustor's own identity is *revealed*. The first difference captures for the trustor the importance of taste-based and statistical discrimination, as well as anticipation of discrimination, as the trustor's own identity is revealed. We then take the *second* difference (D_2) between the amount sent by the trustors to outgroup vs. ingroup trustees, when the trustor's own identity is *concealed*. The second difference captures for the trustor the importance of taste-based and statistical discrimination, but cannot be influenced by anticipation of discrimination, as the trustor's own identity is concealed. Since the trustor's taste-based and statistical discrimination should not depend on whether her own identity is revealed or concealed, a double difference between D_1 and D_2 allows us to estimate the importance of anticipation of discrimination over and above taste-based and statistical discrimination. In addition, we elicit a trustor's expectation of the amount they think the trustee will return. This allows us to go beyond the reduced-form evidence and confirm whether anticipation of discrimination indeed undermines trust.

To investigate whether anticipation of discrimination is misperceived, we test if trustees display taste-based discrimination. In the standard trust game, researchers observe only one decision of trustees that is in response to a specific transfer by the trustor. To obtain a richer characterization of trustees' preferences, we implement the trust game using the strategy method (Selten, 1965). This allows us to elicit a trustee's conditional response function, that is, the amount the trustee sends back to the trustor for all possible amounts that a trustor could have sent. We then test

munication, which is not part of our design. In any case, to the extent that the trustees experience stronger guilt aversion when matched with an ingroup trustor than with an outgroup trustor, it can be considered part of the trustee's taste for discrimination. Similarly, if trustors believe that trustees are less guilt averse when interacting with an outgroup trustor, then this belief is anticipation of discrimination.

for taste-based discrimination by comparing conditional response functions across ingroup and outgroup treatments.

We apply our design in the context of affective polarization in the UK. The study was conducted in 2019, when Boris Johnson was the prime minister of the UK. We used self-reported support for Boris Johnson to classify participants into two groups: "supporters" and "opponents" of Boris Johnson. The participants then took part in the trust game explained above.

We find an economically large and statistically significant effect of anticipation of discrimination over and above taste-based and statistical discrimination. The magnitude of the effect turns out to be as large as the combined effect of taste-based and statistical discrimination. These results arise because trustors anticipate significantly higher defection from outgroup trustees but higher cooperation from ingroup trustees when their own identity is revealed. Crucially, in our context, this anticipation is misperceived, as trustees rarely discriminate between ingroup and outgroup individuals. In fact, the vast majority of trustees in our experiment are perfect reciprocators irrespective of whether they are matched with an ingroup or outgroup trustor. We find similar patterns when we use data on self-reported feelings from the post-experimental survey. These results highlight the importance of (misperceived) anticipation of discrimination in cooperation failures, leaving societies worse off.

Our paper contributes to several strands of literature. First, we complement the literature on trust and cooperation. Previous studies provide evidence only on the importance of taste-based and statistical discrimination (see reviews by Lane, 2016; Charness and Chen, 2020). We highlight the role of anticipation of discrimination over and above taste-based and statistical discrimination by using a new experimental design. A recent study by Ortiz (2023) examines anticipation of discrimination between Muslims and Christians in a Nigerian city using multiple dictator games and a simultaneous move coordination game. The study deduces anticipation of discrimination by comparing choices in the dictator games to choices in the simultaneous move coordination game. However, this design cannot isolate the role of anticipation of discrimination from the role of statistical discrimination, as well as higher-order beliefs. Our design circumvents these concerns by leveraging a sequential trust game. Moreover, since our design is based on a single game, it is easy to implement across different contexts.

Second, our paper contributes to the literature on misperceptions (Bursztyn and

Yang, 2022). Previous studies have documented misperception of social norms concerning women's participation in labor markets (Bursztyn, González and Yanagizawa-Drott, 2020) and wrong facts about immigration (Alesina, Miano and Stantcheva, 2023). We document that misperceptions can lead people to expect discrimination when there is none, leading to avoidable cooperation failures.

Third, our paper adds to the literature on affective polarization. Many studies have examined the patterns of affective polarization (Boxell, Gentzkow and Shapiro, 2020; Guriev and Papaioannou, 2022) but few have examined the underlying motives (Iyengar et al., 2019). Bursztyn, Egorov and Fiorin (2020) study the change in social norms related to xenophobia in the wake of the election of Donald Trump. Dimant (2023) randomly pairs individuals with supporters and opponents of Donald Trump in a public goods game and then records the change in behavior to study affective polarization. Our paper complements these findings by leveraging experimentally induced variation in the information structure of the game to separate the effect of anticipation of discrimination from taste-based and statistical discrimination.

Fourth, our paper also relates to the nascent literature in behavioral and labor economics that presents mixed evidence on the role of anticipation of discrimination. Charness et al. (2020) use a laboratory experiment to show that female students are less likely to reveal their avatar in math-related tasks because they expect discrimination. Similarly, Lepage, Li and Zafar (2022) find that when offered a choice between revealing the actual grade or pass/fail status, female candidates are more likely to reveal their actual grade in anticipation of discrimination. Studies by Aksoy, Chadd and Koh (2023) and Angeli, Matavelli and Secco (2023) report limited evidence on anticipation of discrimination in revealing LGBTQ+ identity and residential location, respectively. However, in these studies one cannot separate anticipation of discrimination from baseline aversion to revealing one's own identity to anyone regardless of who the interaction partner is. Furthermore, a common feature across all of these studies is that they consider settings in which only one party can discriminate. However, in many social interactions, such as cooperation, there is no such clean separation. Our study offers a design to isolate the role of anticipation of discrimination in these settings.

The paper is organized as follows. Section II describes the experimental design, and Section III our conceptual framework and strategy. Section IV presents the results from the application of our design to the context of affective polarization in

the UK, and Section V offers concluding remarks.

II. Experimental Design

Our design is based on a two-player sequential game, which offers several advantages over a two-player simultaneous move game. Table 1 summarizes the motives behind each player's decisions in both these games. In the simultaneous move game, the decisions by player 1 (P1) and player 2 (P2) reflect anticipation of discrimination, as well as taste-based discrimination, statistical discrimination, and anticipation of each other's anticipation. In contrast, in the sequential move game, the anticipation of each other's anticipation no longer plays a role. We need to isolate the role of anticipation of discrimination by Player 1 from her taste-based and statistical discrimination, whereas for Player 2 only her taste-based discrimination plays a role. We achieve this through a trust game, which we describe below.

Table 1: Motives Underlying Decisions

Simultaneous Move Game		Sequential Move Game	
P1's decision motives	P2's decision motives	P1's decision motives	P2's decision motives
Anticipation of **taste & statistical** discrimination by P2	Anticipation of **taste & statistical** discrimination by P1	Anticipation of of **taste-based** discrimination by P2	
Taste-based discrimination	Taste-based discrimination	Taste-based discrimination	Taste-based discrimination
Statistical discrimination	Statistical discrimination	Statistical discrimination	
Anticipation of P2's Anticipation	Anticipation of P1's Anticipation		

Notes: The table shows motives behind Player 1 (P1) and Player 2 (P2) decision in a two-player simultaneous move and a two-player sequential move game. The motives are listed in the columns. As an example, in the first column, P1's decision has four motives.

We describe our experimental design, which relies on the trust game and random assignment of participants to treatments.

II.A. Trust Game

There are two players in a trust game. Following the convention in the literature, we label the first player as "trustor" and the second player as "trustee". Both players receive the same endowment X at the start of the game.[4] The players take decisions sequentially. First, the trustor decides to send an amount t between 0 and X to the trustee. Any amount sent by the trustor is tripled by the experimenter. The trustor then decides to send any amount between 0 and $X + 3t$ back to the trustee. Each player takes two decisions in the following order:

- *Conditional* transfer as a trustee for each possible transfer by the trustor. We implement this using the strategy method (Selten, 1965).
- *Unconditional* transfer as a trustor.

We randomly decide ex post which one of these two decisions is payoff relevant. Since the participants are not aware ex ante which decision will be chosen, both are payoff relevant. The payoffs from the game are as follows:

$$\pi_{trustor} = X - t + \text{amount received from the trustee}$$
$$\pi_{trustee} = X + 3t - \text{amount sent back to the trustor}$$

In this game, the total payoffs are maximized if the trustor sends her entire endowment in the expectation that the trustee reciprocates. However, if the trustee refuses to do so then the trustor loses money. This creates a cooperation dilemma.

We additionally elicit beliefs of the trustors by asking them to guess the conditional transfer of the trustee in response to different possible amounts sent by the trustor. For each correct guess we pay individuals an additional amount.

II.B. Treatments

We use two-stage randomization to assign participants to treatments. First, we randomly pair participants with an ingroup or outgroup person, such that the participants are aware that there is a 50 percent chance of being paired with an outgroup individual. Second, for each individual within a pair, we flip a coin to determine

[4] In the original trust game only the trustee is endowed, resulting in kindness as a plausible motive behind transfers to trustees. Our design follows Cox (2004) to rule this out.

whether her own identity is revealed or concealed to the other individual, such that the outcomes of the coin flip are common knowledge.[5] The random assignment to ingroup or outgroup together with random assignment to the information structure of the game generates six treatments.

Table 2: Treatments

	Information Structure		Trustee is
	Trustor's identity	Trustee's identity	
1	Revealed	Revealed	In-group
2	Revealed	Revealed	Out-group
3	Concealed	Revealed	In-group
4	Concealed	Revealed	Out-group
5	Revealed	Concealed	Anonymous
6	Concealed	Concealed	Anonymous

Notes: The table shows six treatments. Our focus is on treatments 1-4.

III. Conceptual Framework

We use our framework first to assess the importance of anticipation of discrimination for transfers by trustors. We then assess whether this anticipation is misperceived by studying the transfers by trustees.

III.A. Anticipation of Discrimination

We focus on treatments 1-4 in which the identity of the trustee is revealed to the trustor. In these treatments, all three motives behind discrimination could affect a trustor's decision. We exclude treatments 5 and 6 from this analysis because in these treatments the identity of the trustee is concealed, implying that trustors cannot discriminate on the basis of any of the three motives under consideration.

Table 3 lists the different possible motives behind a trustor's decision in each of the four treatments. In Treatments 1-2, the trustor's own identity is revealed, so her decision may reflect taste-based discrimination, statistical discrimination, as well as

[5]To achieve a balance across treatments, we used the "evenly present elements" feature in Qualtrics built-in randomizer.

anticipation of discrimination. However, in Treatments 3-4, the trustor's own identity is concealed, which means anticipation of discrimination is ruled out, but taste-based and statistical discrimination still play a role.

Table 3: Conceptual Framework: Behavior of Trustors

	Information Structure		Trustee is	Motives
	Trustor's identity	Trustee's identity		
1	Revealed	Revealed	In-group	Taste, Statistical, Anticipation
2	Revealed	Revealed	Out-group	Taste, Statistical, Anticipation
3	Concealed	Revealed	In-group	Taste, Statistical
4	Concealed	Revealed	Out-group	Taste, Statistical

Notes: The table shows plausible motives behind trustor's behavior in the four main treatments.

We separate the effect of anticipation of discrimination from taste-based and statistical discrimination by computing three differences:

D_1 is the difference in transfers between Treatment 2 and Treatment 1, that is, the difference in transfers to outgroup vs ingroup individuals when the trustor's own identity is *revealed*. In this case, the trustor's choice of transfer reflects taste-based and statistical discrimination, as well as anticipation of discrimination.

$$D_1 = E\left[Transfer \mid \underbrace{outgroup,\ Revealed}_{\text{Treatment 2}}\right] - E\left[Transfer \mid \underbrace{ingroup,\ Revealed}_{\text{Treatment 1}}\right] \quad (1)$$

D_2 is the difference in transfers between Treatment 4 and Treatment 3, that is, the difference in transfers to outgroup vs ingroup individuals when the trustor's own identity is *concealed*. In this case, the trustor's choice of transfer may reflect taste-based and statistical discrimination, but anticipation of discrimination is no longer possible.

$$D_2 = E\left[Transfer \mid \underbrace{outgroup,\ Concealed}_{\text{Treatment 4}}\right] - E\left[Transfer \mid \underbrace{ingroup,\ Concealed}_{\text{Treatment 3}}\right] \quad (2)$$

We argue that the double difference, $D_3 = D_1 - D_2$, captures the role of anticipation

of discrimination. This is because revealing the trustor's identity is likely to change anticipation of discrimination, but unlikely to cause any change in the trustor's taste-based and statistical discrimination. To the extent trustors anticipate discrimination against them by outgroup individuals, we expect D_3 to be negative. This means that an estimate of D_3 allows us to capture the extent to which anticipation of discrimination undermines trust and cooperation over and above taste-based and statistical discrimination.

We go beyond a reduced-form interpretation and investigate how beliefs over transfers from the trustees differ across the four treatments. We also conduct similar analysis using data from a post-experimental survey on feelings towards ingroup and outgroup members. We discuss these in detail in Section IV.C.

Econometrically, we estimate the effect of anticipation of discrimination using the following linear specification:

$$y_{ij} = \alpha_0 + \beta_1 outgroup_j + \beta_2 Revealed_i + \beta_3 (outgroup_j * Revealed_i) + \mathbf{X'}_i \delta + \epsilon_i \quad (3)$$

where y_{ij} is the transfer made by trustor i to trustee j. The omitted category is transfer to *ingroup* when trustor's identity is *concealed*. With respect to the omitted category, β_1 captures the difference in transfer when the trustee is *outgroup* and the trustor's identity is concealed. β_2 captures the difference in transfer when the trustee is *ingroup* and the trustor's identity is *revealed*. β_3 captures the marginal effect when the trustee is *outgroup* and the trustor's identity is *revealed*. The coefficient of interest is β_3, which captures the effect of anticipation of discrimination and is the same as D_3. β_1 captures the role of taste-based and statistical discrimination, and is the same as D_2. $\mathbf{X}$ is a vector of individual characteristics that are expected to affect behavior in the trust game. These include generalized trust in strangers, gender, age, education, income, geographic location, religion, and political party.

III.B. Is Anticipation of Discrimination Misperceived?

We use the conditional decisions of the trustees from the strategy method to elicit their taste for discrimination. Previous studies show that the responses in the strategy method are not biased, relative to the responses in a simple sequential game (Brandts and Charness, 2000; Fischbacher and Gächter, 2010). We focus on treatments in which the identity of the trustor is revealed to the trustee. Ashraf, Bohnet and Piankov

(2006) find that the transfer by trustee's depend on the amount sent by the trustor. Accordingly, we infer taste-based discrimination by studying differences in conditional responses across the relevant treatments, as follows:

$$D_4 = E\left[\textit{Amount returned} \mid t, \textit{outgroup}\right] - E\left[\textit{Amount returned} \mid t, \textit{ingroup}\right] \quad (4)$$

D_4 captures the differences in conditional responses when matched with an outgroup vs. ingroup trustor, and reveals whether the trustee has a taste for discrimination or not. We estimate D_4 separately for each possible transfer (t) by the trustor. If estimates of D_4 turn out to be economically small for all levels of transfers, then anticipation of discrimination is misperceived.

IV. Application: Affective Polarization in the UK

Our design is not specific to any particular domain but can be applied to different contexts. We illustrate our design in the context of affective polarization. Recent studies have documented affective polarization in many countries (Boxell, Gentzkow and Shapiro, 2020; Guriev and Papaioannou, 2022). However, we know little about the motives that underlie affective polarization. For instance, Iyengar et al. (2019) writes:

> "There has been little to no research identifying the mechanisms underlying affective polarization. On the one hand, distaste for opposing partisans could be couched in raw, reflexive emotion. [On the other hand,] the aversion to economic transactions with opposing partisans may stem not from a visceral emotional response but from a perception that opponents are untrustworthy."

Motivated by these gaps, we conducted our study in the UK, which witnessed strong trends in affective polarization in the aftermath of Brexit. Our focus is on affective polarization among political leaders and their supporters. Following Guriev and Papaioannou (2022) who emphasize a strong relationship between trust and affective polarization, we conduct the trust game between supporters and opponents of Boris Johnson, a polarizing figure in British politics and a prominent proponent of

Brexit. Our study took place in 2019, when Johnson was the Prime Minister of the UK.

IV.A. Study Design

Our study proceeds as follows. First, we recruited 1560 participants from the UK via Facebook. These individuals were assigned to six treatments using stratified randomization based on gender, age, and geographical location. This ensures balance between treatments along these variables. Table A.1 in Appendix A shows that treatments are also balanced across other variables, such as income, education, religion, and political party.

Second, we invited these participants to participate in a trust game (see Appendix B for procedures and instructions). In the game, both the players received an endowment of £24 each, which was more than twice the minimum hourly wage in 2019. The participants made decisions in the role of both the trustor (first mover) and the trustee (second mover). As the trustor, participants could send any amount between £0 and £24 to the trustee in increments of £4. Similarly, the trustee could send back any amount to the trustor in increments of £4.

Third, we gauged an individuals' support or opposition to Boris Johnson using an adaptation of a standard question used in the literature on affective polarization (Druckman and Levendusky, 2019; Boxell, Gentzkow and Shapiro, 2020). The exact question we asked is: "How do you feel personally about Prime Minister Boris Johnson. Please rate on a scale from 0 to 100, where 0 means that you strongly oppose Boris Johnson and 100 means that you strongly support Boris Johnson." We used this information to sort the participants into two political camps. Individuals whose rating was above 50 were classified as "supporters", while individuals whose rating was below 50 were classified as "opponents".[6]

Finally, we performed a two-stage randomization to assign participants to our six treatments using the protocol described in Section II.B (see Appendix B for visualization).

Before participants took part in the game, we tested for their game comprehension. Participants were allowed to proceed only once they have understood the game and

[6]We excluded participants who rated their support at exactly 50 and could therefore not be assigned to either of the political camps. This was the case for 6.9% of the potential participants.

answered the control questions correctly. Overall, 5 percent of the participants were randomly selected for payment (participants were informed of this at the beginning). After the game, we implemented a post-experimental survey to collect data on basic socio-demographic characteristics, as well as data on standard questions on affective polarization, such as feelings towards political ingroup and outgroup members. On average, the entire study lasted 15-20 minutes and the selected participants earned £36 (three times the minimum hourly wage).

IV.B. Results

We start by documenting descriptive results on transfers by trustors in each of the four treatments in Figure 1.[7] When the trustor's identity is concealed, the transfer to ingroup and outgroup trustees are £12.14 and £10.08 respectively, a raw difference of £2. This difference reflects both taste-based and statistical discrimination by the trustors. However, when trustor's identity is revealed, the transfer to ingroup trustee increases but that to the outgroup trustee declines, and now the difference is £3.6. The double difference of £1.6 suggests that anticipation of discrimination by trustors plays an important role over and above taste-based and statistical discrimination.

We test these results econometrically using equation (3) and report the results in Table 4, which provides estimates of β_1, β_2, and β_3. Column 1 controls only for the stated support for Johnson, as this is our basis of classifying individuals into ingroup or outgroup. Relative to the omitted category, three findings are noteworthy. First, when trustor's identity is concealed, the transfer to outgroup trustee declines by nearly £2 and this difference is statistically significant at the 1-percent level. Second, when trustor's identity is revealed, the transfer to ingroup trustee increases by over £1.2 and the difference is statistically significant at the 10-percent level. Finally, the interaction term, which captures the effect of anticipation of discrimination, shows that when trustor's identity is revealed, the transfer to outgroup trustee additionally declines by close to £1.7 pounds and the difference is statistically significant ($p = 0.056$).

[7]For results on transfers in the remaining two treatments, see Figure A.1.

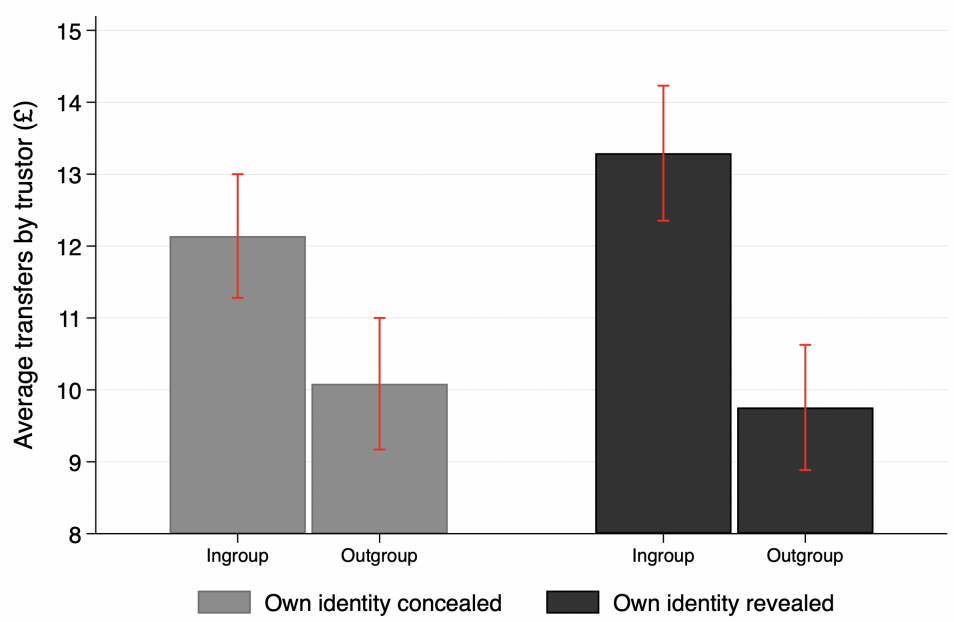

Figure 1: Transfers by Trustors (£)

Notes: The figure plots transfers by trustors in the four main treatments, whichappear from left to right as: Treatments 3, 4, 1, and 2.

Table 4: Transfer by Trustors and Anticipation of Discrimination

	Transfer by Trustor (£)				
	Johnson support	Trust in strangers	Stratification units FE	Income & education	Religion & politics
	(1)	(2)	(3)	(4)	(5)
Outgroup (Concealed)	-1.975	-1.670	-1.710	-1.814	-1.787
	(0.641)	(0.633)	(0.634)	(0.630)	(0.631)
Revealed (Ingroup)	1.244	1.514	1.543	1.536	1.535
	(0.645)	(0.630)	(0.634)	(0.629)	(0.630)
Outgroup×Revealed	-1.666	-2.049	-2.053	-1.907	-1.961
	(0.916)	(0.897)	(0.889)	(0.886)	(0.888)
Baseline Mean			12.14		
R^2	0.03	0.07	0.10	0.11	0.12
N	1536	1536	1536	1536	1536

Notes: OLS estimates with robust standard errors in parentheses. Baseline mean refers to transfer to ingroup members when trustor's identity is concealed (Treatment 3). We progressively include control variables. Column 1 controls for support for Johnson, column 2 additionally for a survey measure of trust in strangers, column 3 for stratification unit fixed effects (area, gender, and age-groups), column 4 for income and education, and column 5 for religion and political party.

We progressively include control variables that are likely to predict transfers by

trustors in columns 2-5. In column 2, we include a general tendency to trust strangers, which enters with a positive coefficient that is statistically significant at the 1-percent level. The coefficient on the interaction term increases slightly in absolute magnitude but its standard error declines, such that it is now statistically significant at the 5-percent level. In column 3, we introduce fixed effects for our stratification units: gender, age, and indicator for ten geographical regions of the UK. This does not lead to any change in the coefficient on the interaction term, which retains its magnitude and significance. In column 4, we control for income and education. These variables enter with positive coefficients that are statistically significant, still the coefficient on the interaction term remains robust in magnitude and retains its statistical significance. Similar results are obtained when we introduce controls for religion and politics in column 5. Our results hold when we carry out a randomization inference test with 5000 replications (p-value < 0.001).

In the model with the full set of controls (column 5), the magnitude of the coefficient on the interaction term implies that anticipation of discrimination leads to a decline in transfer by close to £2 pounds, which is 16 percent of the transfer in the benchmark category and thus economically large. Importantly, the marginal effect of anticipation of discrimination is similar in magnitude to the combined effect of taste-based and statistical discrimination (coef. 1.787, s.e. 0.631). This implies that, in our context, anticipation of discrimination is as important as other motives behind discrimination.

IV.C. Beliefs

The above results offer reduced-form evidence on the importance of anticipation of discrimination. We now go beyond and provide further evidence in support of this result by using data on beliefs from the experiment, as well as the post-experimental survey.

Experiment.– We use data on trustors' beliefs over expected transfers from the trustees. We define anticipation of defection by trustors as the expectation that sending any positive amount to the trustee will result in a payoff that is lower than the original endowment of the trustor.[8]

[8] As an example, if the trustor sends £12 but expects to receive less than £12 back, then she is better off sending nothing.

Figure 2 shows the share of trustors expecting defection in each of the four treatments. When the trustor's identity is concealed, the share of trustors expecting defection is 13 percent when the trustee is ingroup but 18 percent when the trustee is outgroup. However, this difference of 5 percentage points is not statistically significant. In contrast, when the trustor's identity is revealed, the share expecting defection declines when the trustee is ingroup but rises steeply when the trustee is outgroup. Now, the difference turns out to be 14 percentage points and is statistically significant at the 1-percent level. The double difference of 9 percentage points is economically large and statistically significant. It provides strong support to the reduced-form result on anticipation of discrimination.

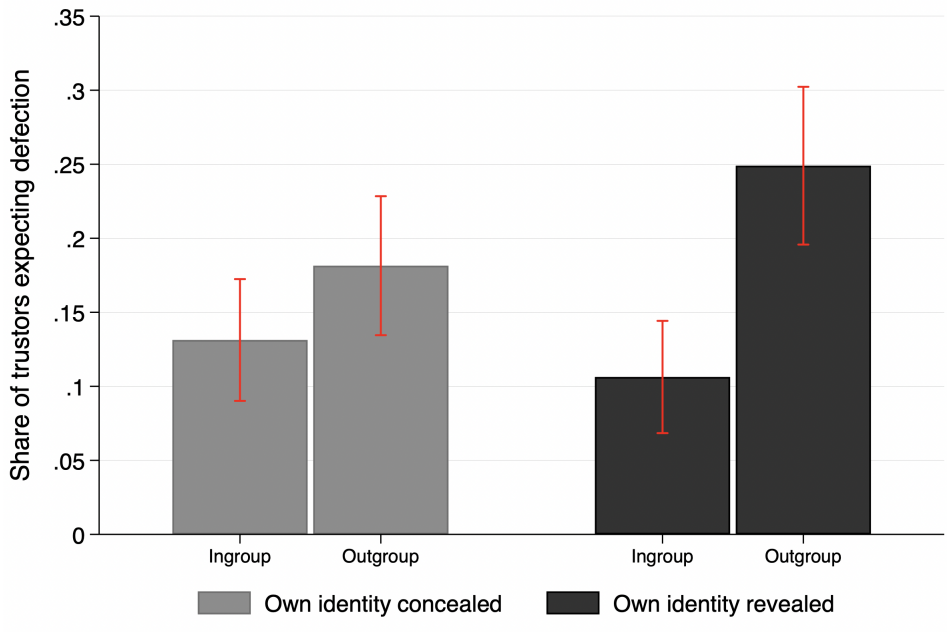

Figure 2: Trustor's Belief over Defection by Trustees

Notes: The figure shows share of defections expected by trustors in the four treatments, which appear from left to right as Treatments 3, 4, 1, and 2.

Results in Table 5 confirm these findings. Regardless of the specification we choose, the coefficient on *Outgroup* × *Revealed* turns out to be close to 0.09 in magnitude and is always statistically significant.

Table 5: Trustor's Anticipation of Defection from Trustees

	Anticipation of Defection				
	Johnson support	Trust in strangers	Stratification units FE	Income & education	Religion & politics
	(1)	(2)	(3)	(4)	(5)
Outgroup (Concealed)	0.053	0.042	0.041	0.042	0.046
	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)
Revealed (Ingroup)	-0.022	-0.031	-0.032	-0.032	-0.030
	(0.029)	(0.028)	(0.029)	(0.029)	(0.028)
Outgroup×Revealed	0.086	0.099	0.101	0.100	0.094
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
Baseline Mean			0.13		
R^2	0.02	0.04	0.06	0.06	0.7
N	1536	1536	1536	1536	1536

Notes: OLS estimates with robust standard errors in parentheses. Baseline mean refers to expected defection when trustor's identity is concealed and trustee is ingroup (Treatment 3). We progressively include control variables. Column 1 controls for support for Johnson, column 2 additionally for a survey measure of trust in strangers, column 3 for stratification unit fixed effects (area, gender, and age-groups), column 4 for income and education, and column 5 for religion and political party.

Post-Experimental Survey.– We use standard questions from the literature on affective polarization to elicit participants' beliefs about how *other people* feel towards members of the participants' political camp. Specifically, we asked the following questions:

- "In your opinion, how does a typical person who supports (opposes) Boris Johnson feel about other people who also support (oppose) Boris Johnson?" (ingroup political camp)
- "In your opinion, how does a typical person who supports (opposes) Boris Johnson feel about other people who oppose (support) Boris Johnson?" (outgroup political camp)

In both questions, we asked respondents to rate on a scale of 0-100, where 0 implies strong dislike and 100 implies strong liking.

Figure 3 shows the distribution of the responses to these questions, where we pool data on the beliefs of supporters and opponents of Boris Johnson. Consistent with the findings on beliefs in the trust game, the participants believe that people have on average warm feelings toward members of their own political camp (an average score of 66), whereas beliefs about the feelings of outgroup members are much more pessimistic

(an average score of 33). These differences are statistically significant at the 1-percent level and lend further support to the results on anticipation of discrimination.

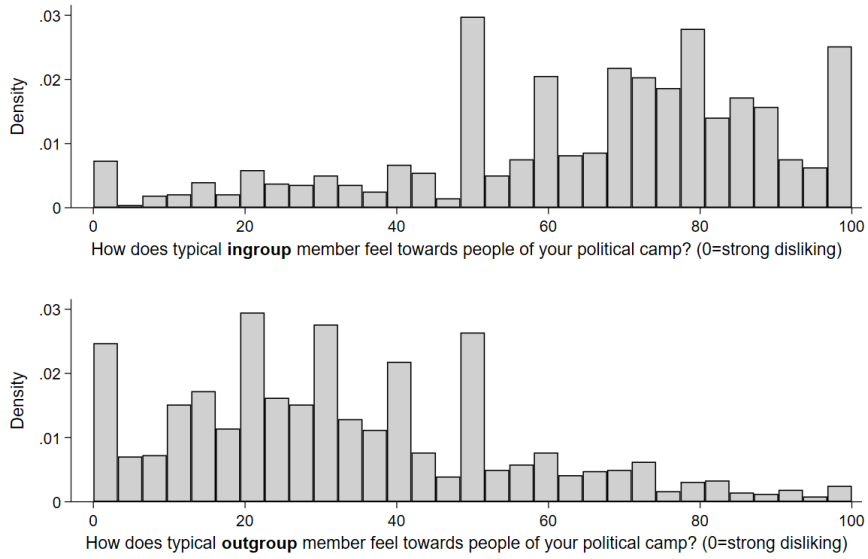

Figure 3: Beliefs about feelings of others towards people of respondent's political camp

Notes: The figure shows a histogram of self-reported beliefs from the post-experimental survey. The top panel shows the distribution of beliefs about the feelings ingroup members have for people of their own political camp (pooled across beliefs of supporters towards other supporters, and beliefs of opponents towards other opponents). The bottom panel does the same using beliefs about the feelings that outgroup members have for people of the opposite political camp (pooled across beliefs of supporters towards opponents, and beliefs about opponents towards supporters of Boris Johnson).

IV.D. Is Anticipation of Discrimination Misperceived?

Thus far, our results reveal a robust negative effect of anticipation of discrimination on transfers by trustors to trustees, which are further backed by data on beliefs from the game and the post-experimental survey. But is this anticipation warranted or is it misperceived? To test this, we use data from both the experiment and the survey.

Experiment.– As outlined in Section III, we start by comparing conditional transfers by trustees to trustors, when the identity of the trustor is revealed as ingroup or outgroup. Figure 4 reports the results by plotting the average conditional transfers by trustees in response to the transfers made by ingroup trustors (blue line) and outgroup trustors (red line). The black dotted line indicates perfect reciprocity – the

amount that the trustee would need to return for the final payoffs to be equalized.

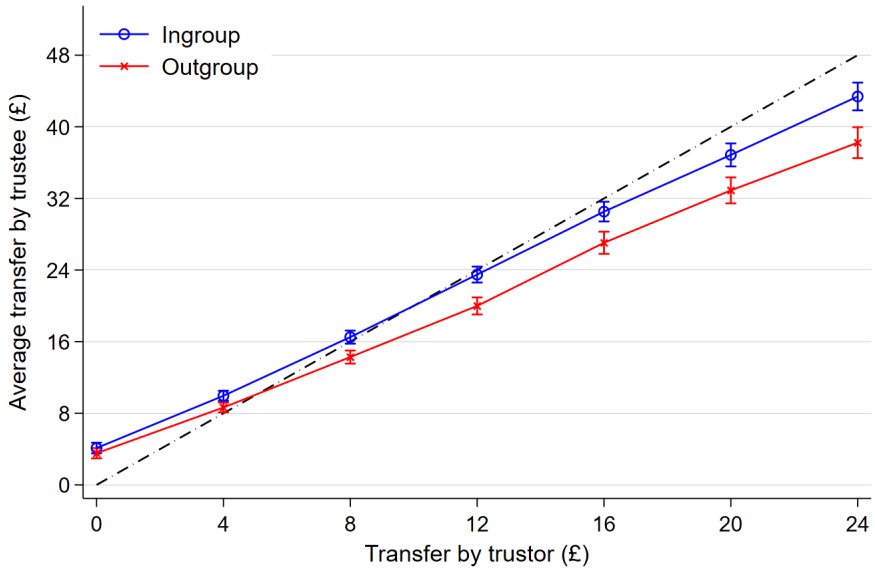

Figure 4: Conditional Transfer by Trustees

Notes: The figure shows conditional transfers by the trustees in response to every possible transfer by the trustor. The blue line indicates conditional transfers to ingroup trustors, whereas the red line indicates conditional transfers to outgroup trustors. The dashed line indicates payoff equalizing transfers by trustees to trustors.

It is evident from the figure that regardless of ingroup or outgroup identity, the conditional transfers by trustees closely follow the dotted line. This means that the trustees return similar amounts to trustors regardless of whether they are ingroup or outgroup. Although the amount returned to ingroup trustors is slightly larger than the amount returned to outgroup trustors, the difference is not economically large enough to justify anticipation of discrimination. In fact, the conditional response to both the ingroup trustors and the outgroup trustors is monotonically upward sloping, suggesting that the trustors would have maximized their expected final payoffs by transferring their entire endowments regardless of whether the trustee is ingroup or outgroup.

We analyze these patterns econometrically in Table A.2, where we regress the amount sent back by trustees on each level of the amount received from trustors and interact it with an outgroup indicator. In line with the results in Figure 4, we find that as the amount sent by trustors increases, the trustees reciprocate by sending a

larger amount when the trustor is ingroup. When the trustor is outgroup, the amount sent back by the trustees also increases, but slightly less than when matched with an ingroup trustor. As an example, when the trustor sends her entire endowment of £24, trustees on average reciprocate by sending £43 to an ingroup trustor and £38 to an outgroup trustor. These results suggest that despite this small difference, a risk-neutral trustor is always better off by sending everything to the trustee regardless of their identity.

We dissect this result further by looking at (i) the share of trustees who defect, and (ii) conditional transfers by trustees who do not defect. A trustee defects if she returns less that what the trustor had sent, resulting in the trustor having a final payoff lower than her original endowment. Figure 5 reports the results. Although the share of defectors rises when the other person is out-group, the magnitude of the difference (4 percentage points) is small. Furthermore, when we restrict the sample to trustees who do not defect, we find no difference in the average amount sent back by trustees to ingroup vs. outgroup trustors. These results suggest that the findings in Figure 4 are primarily at the extensive rather than at the intensive-margin.

Post-Experimental Survey.– We use a standard question to elicit participants' feelings toward people from the opposing political camp. We ask supporters (opponents) the following question: "How do you feel personally about people who oppose (support) Boris Johnson? Please answer on a scale from 0 to 100, where 0 means strong dislike and 100 means strong liking."[9] Figure 6 shows that feelings towards outgroup members are not extreme but moderate, with the most common score being 50. This result is in line with the experimental findings and bolsters the confidence in our findings that anticipation of discrimination is to a large extent misperceived.

[9]This question elicits the respondents' own partisan feelings rather than their beliefs about other people's partisan feelings documented in Figure 3.

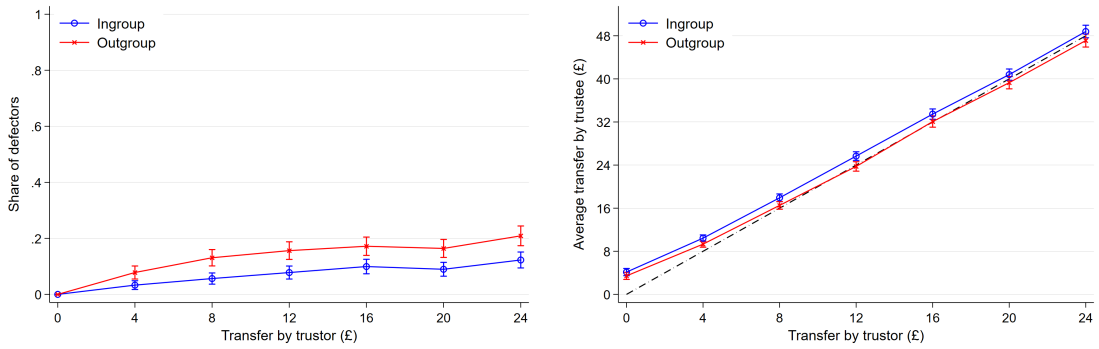

Figure 5: Share of Free Riders and Conditional Transfer by Trustees

Notes: The left figure shows the share of trustees who defect. The right figure shows the conditional transfers by trustees who do not defect in response to every possible transfer by the trustor. The blue line in the left figure indicates the share of trustees defecting when the trustor is ingroup, whereas in the right figure it indicates conditional transfers to ingroup trustors. The red line in the left figure indicates the share of trustees defecting when the trustor is outgroup, whereas in the right figure it indicates conditional transfers to outgroup trustors. The dashed line in the right figure indicates payoff equalizing transfers by trustees to trustors.

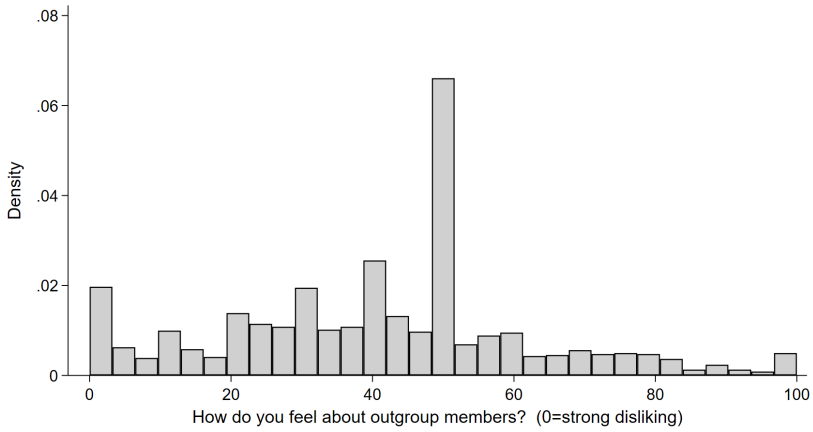

Figure 6: Reported feelings towards outgroup members

Notes: The figure shows a histogram of reported feelings towards outgroup members in the post-experimental survey. For the supporters of Johnson, we asked about their feelings towards the opponents of Johnson, and vice versa. We pooled data across feelings of supporters towards opponents, and feelings of opponents towards supporters of Boris Johnson.

IV.E. Most Polarized Sample

We lend further credibility to our results by showing that anticipation of discrimination widens in the most polarized sample in which people express extreme support (score of 100) or opposition (sore of 0) to Boris Johnson. Table A.3 reports the results and shows that the anticipation of discrimination increases from £ -2 in the full sample to £-4 in the most polarized sample (see column 5). Furthermore, the results in Figure A.2 and columns 3-4 of Table A.2 suggest that this anticipation is also largely misperceived.

V. Conclusions

Many economic activities require generalized trust and cooperation between ingroup and outgroup members. Several studies have documented that taste-based and statistical discrimination can hamper trust and cooperation. However, we know little about the extent to which anticipation of discrimination plays a role in this context. Progress in this area has been hampered by the difficulties in separating the effect of anticipation of discrimination from that of taste-based and statistical discrimination. This paper offers a new experimental design to fill this gap. It relies on a trust game in which players decide sequentially on their transfers.

To separate the motives underlying transfers by the trustors, we randomly vary the identity of the other player, as well as the information structure of the game. We achieve this by concealing the identity of the trustor in some treatments and revealing it in the other treatments. We then take the double difference between the transfers sent by trustors to ingroup vs. outgroup trustees when the trustors own identity is revealed vs. concealed. Since anticipation is possible only when the trustors own identity is revealed, the double difference allows us to estimate the effect of anticipation of discrimination over and above other motives. To test whether this anticipation is misperceived or justified, we elicit the preferences of trustees using the strategy method.

We apply this design in the context of affective polarization, which is on the rise in many countries but motives underlining it are poorly understood. Our study takes place in the UK which provides an ideal setting because of strong affective polarization in the aftermath of Brexit. We play the trust game between the opponents

and supporters of Boris Johnson. We find that anticipation of discrimination undermines inter-partisan trust and cooperation. The magnitude of the effect is as large as the combined effect of taste-based and statistical discrimination. This is further confirmed by data on beliefs that show trustors expect significantly higher discrimination from outgroup trustees when their own identity is revealed. However, we find that anticipation of discrimination is misperceived, as the vast majority of trustees do not discriminate by political identity. This results in cooperation failure.

Our design can be applied in many different contexts to study anticipation of discrimination along the lines of gender, caste, ethnicity, religion, and politics in different countries.

References

Aksoy, Billur, Ian Chadd, and Boon Han Koh. 2023. "Sexual identity, gender, and anticipated discrimination in prosocial behavior." *European Economic Review*, 154: 104427.

Alesina, Alberto, and Paola Giuliano. 2015. "Culture and Institutions." *Journal of Economic Literature*, 53(4): 898–944.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. 2023. "Immigration and redistribution." *The Review of Economic Studies*, 90(1): 1–39.

Algan, Yann, and Pierre Cahuc. 2010. "Inherited trust and growth." *American Economic Review*, 100(5): 2060–2092.

Angeli, Deivis, Ieda Matavelli, and Fernando Secco. 2023. "Expected Discrimination and Job Search." *Working Paper*.

Arrow, Kenneth. 1970. "Political and economic evaluation of social effects and externalities." In *The Analysis of Public Output*. 1–30. NBER.

Arrow, Kenneth J. 1972. "Gifts and exchanges." *Philosophy & Public Affairs*, 343–362.

Arrow, Kenneth J. 1973. "The Theory of Discrimination, Discrimination in Labor Markets, Ashenfelter, O. and A. Rees eds., 3-33."

Ashraf, Nava, Iris Bohnet, and Nikita Piankov. 2006. "Decomposing trust and trustworthiness." *Experimental economics*, 9: 193–208.

Becker, Gary S. 1957. *The economics of discrimination.* University of Chicago press.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, reciprocity, and social history." *Games and economic behavior*, 10(1): 122–142.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro. 2020. "Cross-country trends in affective polarization." National Bureau of Economic Research.

Brandts, Jordi, and Gary Charness. 2000. "Hot vs. cold: Sequential responses and preference stability in experimental games." *Experimental Economics*, 2: 227–238.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott. 2020. "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia." *Am. Econ. Rev.*, 110(10): 2997–3029.

Bursztyn, Leonardo, and David Y Yang. 2022. "Misperceptions about others." *Annual Review of Economics*, 14: 425–452.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin. 2020. "From Extreme to Mainstream: The Erosion of Social Norms." *Am. Econ. Rev.*, 110(11): 3522–3548.

Charness, Gary, and Martin Dufwenberg. 2006. "Promises and partnership." *Econometrica*, 74(6): 1579–1601.

Charness, Gary, and Yan Chen. 2020. "Social identity, group behavior, and teams." *Annual Review of Economics*, 12(1): 691–713.

Charness, Gary, Ramón Cobo-Reyes, Simone Meraglia, and Ángela Sánchez. 2020. "Anticipated discrimination, choices, and performance: Experimental evidence." *European Economic Review*, 127: 103473.

Coate, Stephen, and Glenn C Loury. 1993. "Will affirmative-action policies eliminate negative stereotypes?" *The American Economic Review*, 1220–1240.

Cox, James C. 2004. "How to identify trust and reciprocity." *Games and economic behavior*, 46(2): 260–281.

Dimant, Eugen. 2023. "Hate trumps love: The impact of political polarization on social preferences." *Management Science*.

Druckman, James N, and Matthew S Levendusky. 2019. "What do we measure when we measure affective polarization?" *Public opinion quarterly*, 83(1): 114–122.

Ellingsen, Tore, Magnus Johannesson, Sigve Tjøtta, and Gaute Torsvik. 2010. "Testing guilt aversion." *Games and Economic Behavior*, 68(1): 95–107.

Enke, Benjamin. 2024. "Moral boundaries." *Annual Review of Economics*, 16(1): 133–157.

Fehr, Ernst. 2009. "On the economics and biology of trust." *Journal of the european economic association*, 7(2-3): 235–266.

Fehr, Ernst, and Simon Gächter. 2000. "Fairness and retaliation: The economics of reciprocity." *Journal of economic perspectives*, 14(3): 159–182.

Fischbacher, Urs, and Simon Gächter. 2010. "Social preferences, beliefs, and the dynamics of free riding in public goods experiments." *American Economic Review*, 100(1): 541–56.

Fukuyama, Francis. 1996. *Trust: The social virtues and the creation of prosperity.* Simon and Schuster.

Glaeser, Edward L, David I Laibson, Jose A Scheinkman, and Christine L Soutter. 2000. "Measuring trust." *The quarterly journal of economics*, 115(3): 811–846.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does culture affect economic outcomes?" *Journal of Economic perspectives*, 20(2): 23–48.

Guriev, Sergei, and Elias Papaioannou. 2022. "The political economy of populism." *Journal of Economic Literature*, 60(3): 753–832.

Hobolt, Sara B, Thomas J Leeper, and James Tilley. 2021. "Divided by the vote: Affective polarization in the wake of the Brexit referendum." *British Journal of Political Science*, 51(4): 1476–1493.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J Westwood. 2019. "The origins and consequences of affective polarization in the United States." *Annual Review of Political Science*, 22: 129–146.

Knack, Stephen, and Philip Keefer. 1997. "Does social capital have an economic payoff? A cross-country investigation." *The Quarterly journal of economics*, 112(4): 1251–1288.

Lane, Tom. 2016. "Discrimination in the laboratory: A meta-analysis of economics experiments." *European Economic Review*, 90: 375–402.

Lepage, Louis-Pierre, Xiaomeng Li, and Basit Zafar. 2022. "Anticipated gender discrimination and grade disclosure." National Bureau of Economic Research.

Ortiz, Miguel. 2023. "Hate, fear and intergroup Conflict: Experimental evidence from Nigeria." *Working paper*.

Phelps, Edmund S. 1972. "The statistical theory of racism and sexism." *American Economic Review*, 62(4): 659–661.

Platteau, Jean-Philippe. 2000. *Institutions, social norms and economic development.* Routledge.

Putnam, Robert D. 2000. "Bowling alone: The collapse and revival of American community." *Simon Schuster*.

Putnam, Robert D, Robert Leonardi, and Raffaella Y Nanetti. 1993. *Making democracy work: Civic traditions in modern Italy.* Princeton university press.

Selten, Reinhard. 1965. "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes." *Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie*.

Tabellini, Guido. 2010. "Culture and institutions: economic development in the regions of Europe." *Journal of the European Economic Association*, 8(4): 677–716.

ONLINE APPENDIX

Anticipation of Discrimination and Misperceptions in Cooperation Dilemmas

Devesh Rustagi Matthias Schief

Appendix A

I. Summary statistics and balance check

Balance Check

Table A.1 shows that the treatments are balanced across a number of additional covariates that we did not target with our stratified sample protocol (age, gender, region).

Table A.1: Balance Table

	Treatments							
	1	2	3	4	5	6	Mean	Prob>F
Income	1.587	1.621	1.595	1.668	1.641	1.563	1.613	0.365
	(0.568)	(0.562)	(0.592)	(0.608)	(0.596)	(0.599)	(0.588)	
Education	0.516	0.462	0.541	0.560	0.533	0.488	0.517	0.253
	(0.501)	(0.500)	(0.499)	(0.497)	(0.500)	(0.501)	(0.500)	
Religion	0.480	0.522	0.498	0.444	0.502	0.528	0.495	0.445
	(0.501)	(0.501)	(0.501)	(0.498)	(0.501)	(0.500)	(0.500)	
Politics	0.386	0.375	0.351	0.421	0.382	0.321	0.373	0.288
	(0.488)	(0.485)	(0.478)	(0.495)	(0.487)	(0.468)	(0.484)	
Trusting	0.378	0.407	0.471	0.367	0.425	0.401	0.408	0.194
	(0.486)	(0.492)	(0.500)	(0.483)	(0.495)	(0.491)	(0.492)	
N	254	253	259	259	259	252	1536	

Notes: Columns 1-6 shows the means and standard deviations for the variables listed in the rows across the six treatments. Column 7 shows the mean in the full sample. Income is measured as an index with 1 indicating lower income, 2 indicating middle income, and 3 indicating upper income. Education is an indicator variable taking a value of 1 if the respondent has a university degree. Religion is an indicator variable taking a value of 1 if the respondent reports not being religious and 0 otherwise. Politics is an indicator variable taking a value of 1 if the respondent is a voter of the Labour Party, 0 otherwise. Trusting is an indicator variable taking a value of 1 if the respondent thinks that "most people can be trusted", and 0 otherwise. The last column shows the p-value of an F-test that tests for differences in means across treatments.

II. Results

Transfers by Trustors

Figure A.1 shows transfers by trustors in Treatments 5 and 6, that is, when their own identity is revealed but that of the trustee is concealed (Treatment 5) and when their own identity as well as that of the trustee are concealed (Treatment 6).

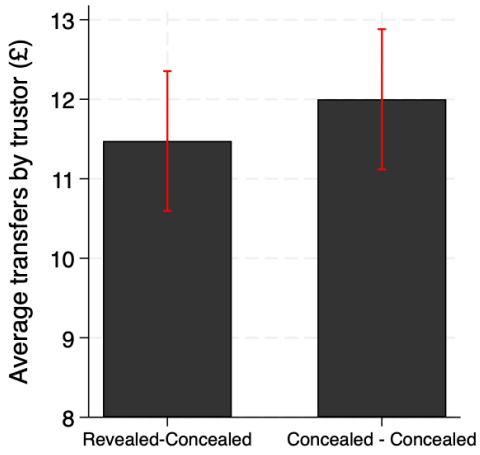

Figure A.1: Transfers by Trustors (£)

Notes: The figure plots transfers by trustors in treatments 5 and 6.

Misperceptions in Anticipation of Discrimination

Table A.2 documents misperception by showing the differences in amount sent by trustees for every possible transfer by the ingroup vs. outgroup trustor in game. The coefficients on amount sent back to outgroup trustors though statistically significant are very small in magnitude to warrant strong anticipation of discrimination.

Table A.2: Anticipation of Discrimination and Misperception

	Conditional Response of Trustees by Identity (£)			
	Full Sample		Most Polarized Sample	
	No controls	All controls	No controls	All controls
	(1)	(2)	(3)	(4)
In-group				
Decision 4	5.840	5.840	5.493	5.493
	(0.256)	(0.256)	(0.487)	(0.490)
Decision 8	12.390	12.390	12.028	12.028
	(0.334)	(0.335)	(0.677)	(0.681)
Decision 12	19.376	19.376	19.634	19.634
	(0.431)	(0.431)	(0.871)	(0.876)
Decision 16	26.409	26.409	27.211	27.211
	(0.540)	(0.541)	(1.060)	(1.066)
Decision 20	32.741	32.741	34.648	34.648
	(0.636)	(0.637)	(1.211)	(1.218)
Decision 24	39.267	39.267	41.380	41.380
	(0.785)	(0.787)	(1.480)	(1.489)
Out-group				
Decision 0	-0.578	-0.529	-1.599	-1.056
	(0.423)	(0.443)	(0.801)	(0.944)
Decision 4	-0.731	-0.731	0.101	0.101
	(0.396)	(0.397)	(0.759)	(0.764)
Decision 8	-1.648	-1.648	-1.953	-1.953
	(0.490)	(0.490)	(0.987)	(0.993)
Decision 12	-2.923	-2.923	-3.664	-3.664
	(0.658)	(0.659)	(1.293)	(1.301)
Decision 16	-2.902	-2.902	-3.662	-3.662
	(0.816)	(0.817)	(1.640)	(1.651)
Decision 20	-3.381	-3.381	-6.979	-6.979
	(0.974)	(0.976)	(1.918)	(1.930)
Decision 24	-4.580	-4.580	-7.275	-7.275
	(1.184)	(1.186)	(2.347)	(2.362)
Baseline Mean		4.12		
R^2	0.51	0.53	0.51	0.54
N	7175	7175	1925	1925

Notes: OLS estimates with robust standard errors clustered on the individual in parentheses. Baseline mean refers to the amount sent back by trustees when the in-group trustor sends 0. We progressively include control variables. Column 1 controls for support for Johnson, column 2 additionally for a survey measure of trust in strangers, column 3 for stratification unit fixed effects (area, gender, and age-groups), column 4 for income and education, and column 5 for religion and political party.

Results Using the Most Polarized Sample

Table A.3: Transfer by Trustors and Anticipation of Discrimination
Most Polarized Sample

	Transfer by Trustor (£)				
	No controls	Trust in strangers	Stratification units FE	Income & education	Religion, politics & ethnicity
	(1)	(2)	(3)	(4)	(5)
Out-group (Concealed)	-2.836	-1.983	-1.733	-1.984	-1.845
	(1.400)	(1.365)	(1.407)	(1.366)	(1.360)
In-group (Revealed)	2.668	2.868	3.020	3.046	3.049
	(1.270)	(1.197)	(1.234)	(1.237)	(1.249)
Out-group×Revealed	-3.341	-3.839	-4.247	-3.954	-4.147
	(1.906)	(1.823)	(1.821)	(1.800)	(1.815)
Baseline Mean			12.77		
R^2	0.07	0.13	0.22	0.24	0.25
N	405	405	404	404	404

Notes: OLS estimates with robust standard errors in parentheses. Baseline mean refers to expected defection when trustor's identity is concealed and trustee is in-group (Treatment 3). We progressively include control variables. Column 1 controls for support for Johnson, column 2 additionally for a survey measure of trust in strangers, column 3 for stratification unit fixed effects (area, gender, and age-groups), column 4 for income and education, and column 5 for religion and political party.

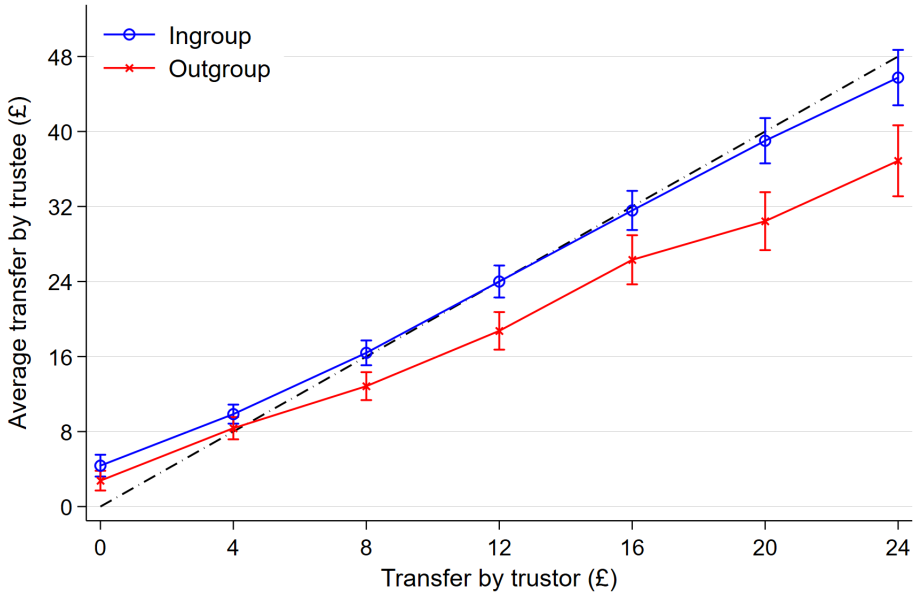

Figure A.2: Conditional Transfer by Trustees in the Most Polarized Sample

Notes: The figure shows conditional transfers by the trustees in response to every possible transfer by the trustor in the most polarized sample. The blue line indicates conditional transfers to ingroup trustors, whereas the red line indicates conditional transfers to outgroup trustors. The dotted line indicates payoff equalizing transfers by trustees to trustors.

Appendix B Experimental Instructions

I. Introduction

In this study, we will match you with another person in the United Kingdom who has also decided to participate in the study. You do not know anything about this person. Similarly, the other person knows nothing about you. You will interact with the other person. Your earnings will depend on the decisions made by you and the other person. Therefore, please carefully read the instructions on the next page.

Decision Situation

We will now introduce you to the basic decision situation. You and the other person will interact in the roles of player A and player B. At the start of the study, each player has £24 in the pocket. This money is provided by us.

Player A moves first and has the opportunity to send all, some, or none of this money to Player B. Any amount sent to Player B will be tripled. Example: if Player A decides to send £20, it will be tripled to £60 when it reaches Player B.

Player B can keep the money or decide to send some money back to Player A. The amount sent back to Player A is not tripled.

Note: Player A must decide how much money to send to Player B without knowing whether Player B will send money back. In contrast, Player B can take Player A's decision into account when deciding how much to send back.

Earnings

Earnings of Player A
£24 in pocket - amount sent to Player B + amount received from Player B

Earnings of Player B
£24 in pocket + 3 x amount sent by Player A - amount sent to Player A

Examples

Example I: Player A sends £0 to Player B. Player B sends £0 back to Player A. Hence, Player A earns £24 and Player B also earns £24

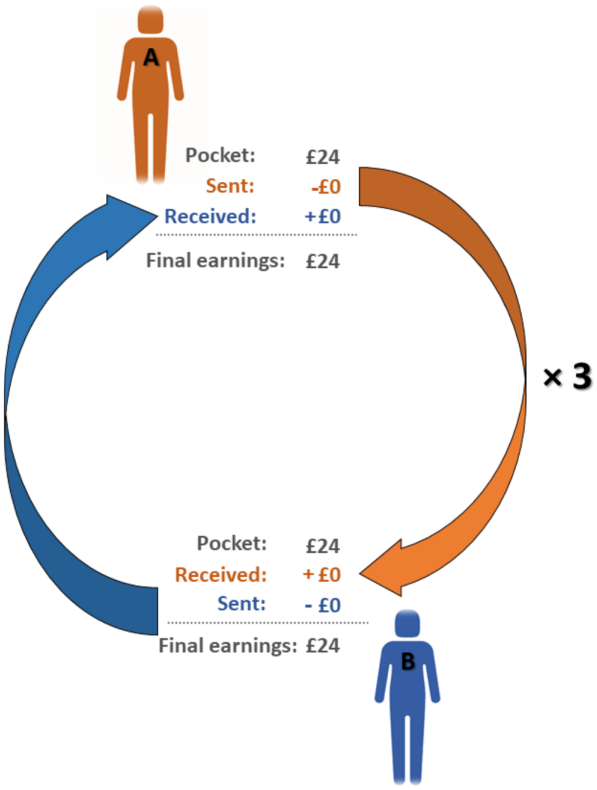

Figure A.3: Example I

Example II: Player A sends £20 to Player B. Player B sends £40 back to Player A. Hence, Player A earns £44 and Player B also earns £44.

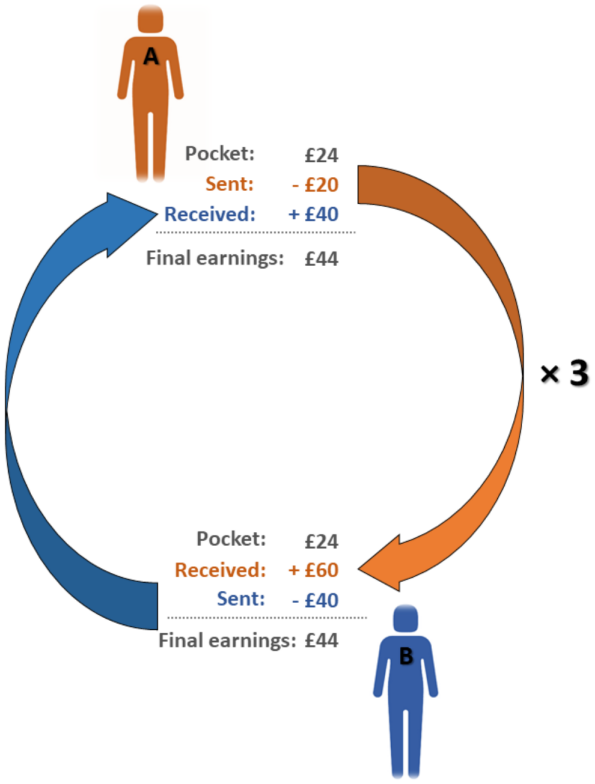

Figure A.4: Example II

Example III: Player A sends £20 to Player B. Player B sends £0 back to Player A. Hence, Player A earns £4 and Player B earns £84.

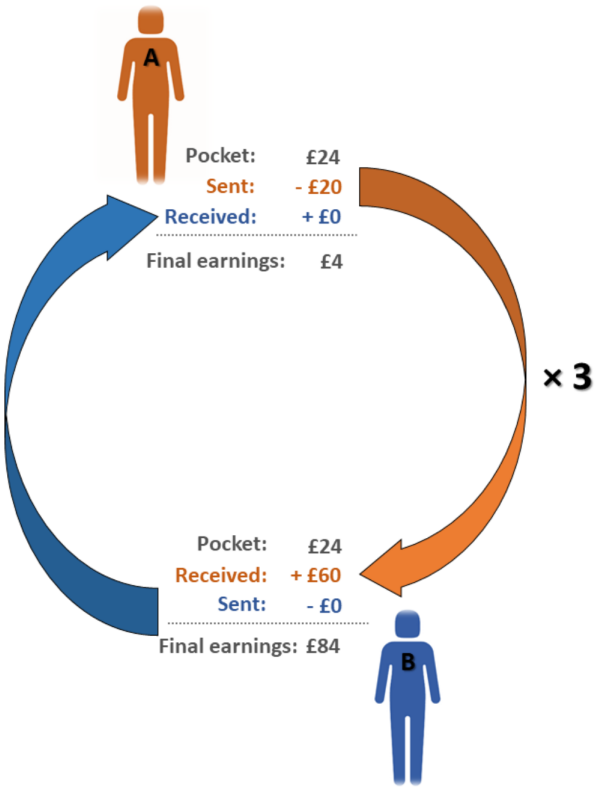

Figure A.5: Example III

Assignment of Roles

You will take decisions both in the role of Player A and Player B. When you decide in the role of Player A, the other person will decide in the role of Player B and vice versa. We will pay you for your decisions in either the role of Player A or Player B. We will toss a coin to decide on this. Please take all decisions carefully.

Game comprehension questions

To make sure that you have understood the decision situation, please answer the questions below.

Case 1: You are Player A and the other person is Player B. Out of the £24 in your pocket, you send £8 to the other person. After tripling, how much does the other person receive from you?

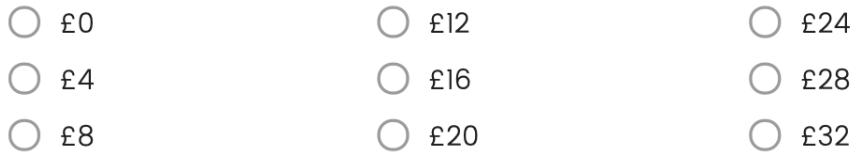

Figure A.6

The other person decides to send £8 back, what will be your and the other person's final earnings?

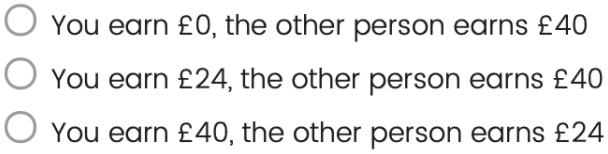

Figure A.7

Case 2: You are Player B and the other person is Player A. The other person sends £24 to you. After tripling, you receive £72. What will be your final earnings if you decide to send £48 back to the other person?

Figure A.8

What will be your final earnings if you decide to send £0 back to the other person?

Figure A.9

Introduction to Matching

Congratulation, you are ready to take your decisions!

We have now matched you with another person

Before you take your decisions, there is an opportunity for both you and the other person to learn one thing about each other: Whether you support or oppose Prime Minister Boris Johnson. There is a 50% chance that you are matched with someone who supports Boris Johnson and a 50% chance that you are matched with someone who opposes Boris Johnson.

We will toss a coin to decide on whether the other person learns about your political opinion.

- If you get Heads, we will **tell** the other person whether you support or oppose Boris Johnson.
- If you get Tails, we will **not tell** the other person whether you support or oppose Boris Johnson.

Similarly, we will toss another coin to decide on whether you learn about the other person's view on Boris Johnson.

Treatments

[Note for the reader: After this, we tossed coins to assign participants to one of the treatments. Here we are reproducing only one treatment as an example for the sake of illustration.]

We tossed coins and both got Heads!

- We can tell you that the other person opposes Boris Johnson.
- We have TOLD the other person that you support Boris Johnson.

You and the other person are both aware of this situation (visualised below).

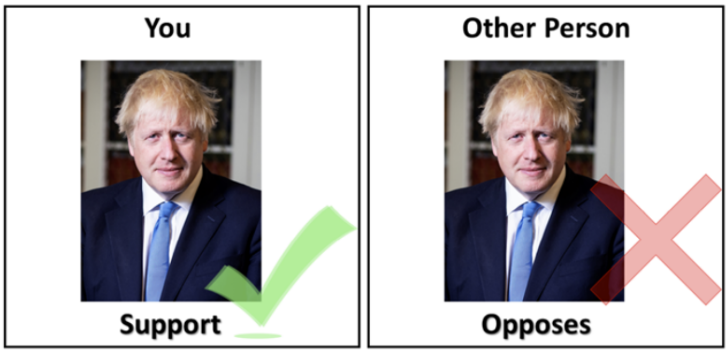

Figure A.10: SCREEN SHOT OF ONE OF THE TREATMENTS

ACTUAL DECISIONS

Decision as Player B

You will take this decision in the role of **Player B**. The other person is in the role of Player A.

You do NOT know the other person's views on Boris Johnson. Similarly, the other person does NOT know your views on Boris Johnson.

As of now we do not know the amount that the other person has sent to you. Therefore, we request you to take a decision for each possible amount sent by the other person. We will display the earnings associated with your decisions (in blue) when you click on the options below.

If the other person sends **£0**, how much will you send back?

Figure A.11

You earn:
Other person earns:

If the other person sends £**4** to you, it is tripled to £12. Now, the other person has £20 and you have £36. How much would you like to send back to the other person

Figure A.12

You earn:
Other person earns:

If the other person sends £**8** to you, it is tripled to £24. Now, the other person has £16 and you have £48. How much would you like to send back to the other person?

Figure A.13

You earn:
Other person earns:

If the other person sends £**12** to you, it is tripled to £36. Now, the other person has £12 and you have £60. How much would you like to send back to the other person?

Figure A.14

You earn:
Other person earns:

If the other person sends £**16** to you, it is tripled to £48. Now, the other person has £8 and you have £72. How much would you like to send back to the other person?

- ○ £0
- ○ £4
- ○ £8
- ○ £12
- ○ £16
- ○ £20
- ○ £24
- ○ £28
- ○ £32
- ○ £36
- ○ £40
- ○ £44
- ○ £48
- ○ £52
- ○ £56
- ○ £60
- ○ £64
- ○ £68
- ○ £72

Figure A.15

You earn:
Other person earns:

If the other person sends £**20** to you, it is tripled to £60. Now, the other person has £4 and you have £84. How much would you like to send back to the other person?

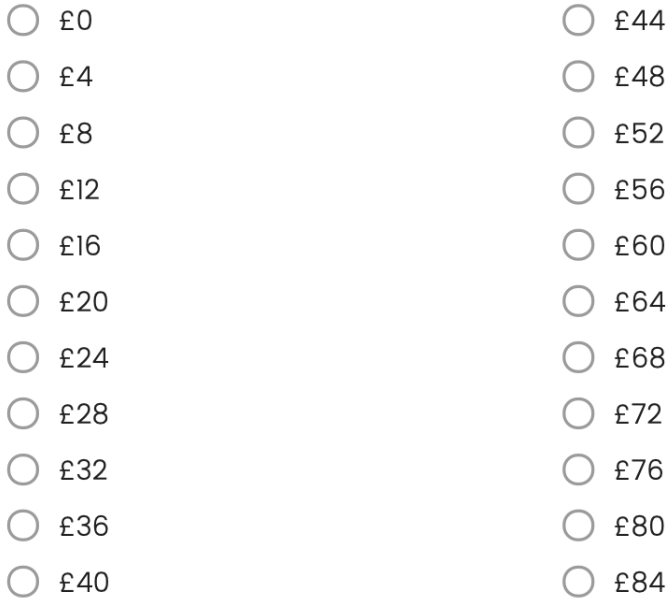

Figure A.16

You earn:
Other person earns:

If the other person sends £**24** to you, it is tripled to £72. Now, the other person has £0 and you have £96. How much would you like to send back to the other person?

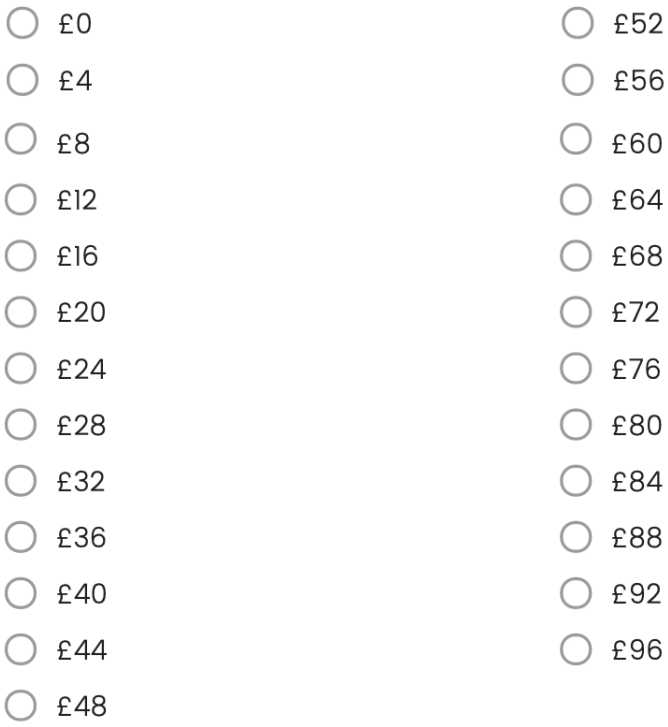

Figure A.17

You earn:
Other person earns:

Decision as Player A

You will take this decision in the role of **Player A**. The other person is now in the role of Player B.

You do NOT know the other person's views on Boris Johnson. Similarly, the other person does NOT know. your views on Boris Johnson.

Out of the £24 in your pocket, how many pound would you like to send to the other person?

£0 £4 £8 £12 £16 £20 £24

Now make a guess and increase your earnings

We have requested the other person in the role of Player B to decide on how much money to send back to you for each possible amount that you could have sent as Player A. Can you guess how much the other person will send back to you? For each correct guess we will increase your final earnings by 10 percentage points.
Out of £24 in your pocket, if you send £**0**, how much do you think the other person will send back?

- ○ £0 (You and the other person each earn £24)
- ○ £4 - £24 (You earn between £28 and £48, the other person earns between £0 and £20)

Figure A.18

Out of £24 in your pocket, if you send £**12**, how much do you think the other person will send back?

- ○ £0 - £8 (You earn between £12 and £20, and the other person earns between £52 and £60)
- ○ £12 - £20 (You earn between £24 and £32, and the other person earns between £40 and £48)
- ○ £24 (You and the other person both earn £36)
- ○ £28 - £60 (you earn between £40 and £72, and the other person earns between £0 and £32)

Figure A.19

Out of £24 in your pocket, if you send £**24**, how much do you think the other person will send back?

- ○ £0 - £20 (You earn £0 and £20, and the other person earns between £76 and £96)
- ○ £24 - £44 (You earn between £24 and £44, and the other person earns between £52 and £72)
- ○ £48 (You and the other person each earn £48)
- ○ £52 - £96 (You earn between £52 and £96, and the other person earns between £0 and £44)

Figure A.20

If you would like to revise your "Decision as Player A" at the top of this page, please feel free to do so now.