Take a chance:

Trust-building across identity groups

(Online Appendix)*

Yoshiko M. Herrera†

and

Andrew H. Kydd‡

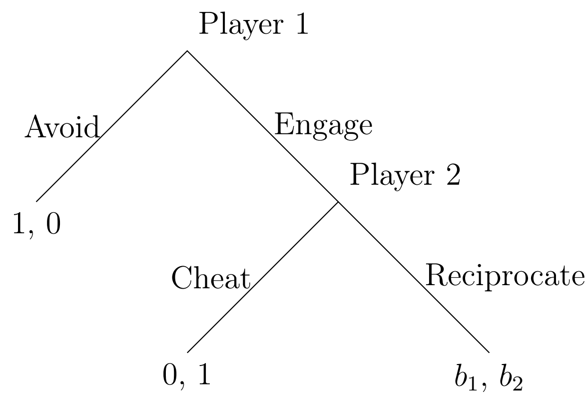
April 22, 2021

1 Introduction

The building block of our model is illustrated in Figure 1. Player 1 has an opportunity to avoid or engage with player 2. If player 1 engages, player 2 can reciprocate player 1's engagement or cheat. If player 1 avoids player 2, the individual payoffs are 1 and 0, (leaving

* *Journal of Peace Research*, Vol. ?? No. ?? pp. ??-??.
†Department of Political Science, University of Wisconsin, yherrera@wisc.edu.
‡Department of Political Science, University of Wisconsin, kydd@wisc.edu.

Figure 1: The Trust Game

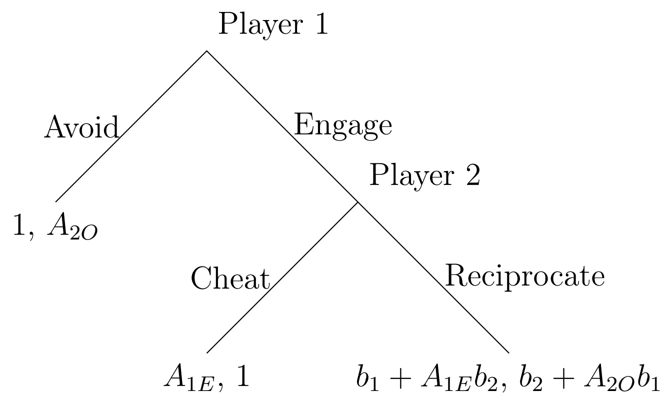


aside other-regarding preferences which we discuss below.) If player 1 engages and player 2 cheats, the payoffs are reversed, 0 and 1. If player 1 engages and player 2 reciprocates, each side gets a benefit, b_1 and b_2. To make the game an interesting representation of the trust problem, we assume both players prefer reciprocated engagement to avoidance, but player 2 might prefer to exploit.

Assumption 1 *Both players prefer reciprocated engagement to avoidance,* $b_1 > 1$*,* $b_2 > 0$ *(leaving aside social preferences).*

To model trust building between identity groups over time we need to extend the basic trust game to consider identity groups, multiple rounds of play, social preferences, and uncertainty about trustworthiness. To represent identity groups we increase the number of players to four and divide them into two identity groups, the *odds* (players 1 and 3) and the *evens* (players 2 and 4). To examine learning over time we consider a two round model. In the first round, player 1 and player 2 play the trust game and in the second round, players 3 and 4 play the game again. For now, we assume that players 3 and 4 observe the first round

Figure 2: The Trust Game with Social Preferences



directly, so they know what players 1 and 2 did.

Third, we need to endow the players with other regarding preferences, allowing them to like or dislike other players based on their group membership. We posit for each player *attachment parameters* for members of their own group and for members of the outgroup. For instance, player 1's attachment for members of their own group, the odds, is denoted by A_{1O} and their attachment to members of the even group is A_{1E}. The attachment parameters modify the payoffs as shown in Figure 2.

We assume that individuals like their own group, so that, for instance, A_{1O} is positive. We also assume that attachment parameters are less than 1 (and greater than -1), so that players do not derive more utility from other players than from their own material payoffs. Individuals may like or dislike the other group, so that A_{1E} could be positive or negative. We assume that the players' attachment parameters for their own group and the outgroup are independent of each other, so ingroup love does not automatically imply outgroup hostility. We also assume that they are fixed.

Assumption 2 *The attachment parameters for one's own group are bounded by 0 and 1,* A_{1O}, A_{2E}, A_{3O}, $A_{4E} \in [0, 1)$. *The attachment parameters for the other group are bounded by -1 and 1.* A_{1E}, A_{2O}, A_{3E}, $A_{4O} \in (-1, 1)$.

We posit two types for the even players, trustworthy and untrustworthy, where trustworthy types reciprocate and untrustworthy types cheat. The types are differentiated by their attachment parameters for the odd players. Let the even players' attachment parameters for the odd players take on one of two possible values, A^t_{EO} for the trustworthy type and A^u_{EO} for the untrustworthy type.

Assumption 3 A_{2O} *and* A_{4O} *can take on one of two values,* A^u_{EO} *and* A^t_{EO} *where* $-1 < A^u_{EO} < A^t_{EO} < 1$.

To model social learning, we assume that the even players' types are correlated, but not perfectly. This implies that if you see player 2 reciprocate, it increases your belief that player 4 will reciprocate, whereas if you see player 2 cheat, it makes you think player 4 is more likely to cheat as well. Formally, let the outgroup attachment parameters for players 2 and 4 be independently drawn from a Bernoulli distribution such that there is a t chance that the player is trustworthy, and has an attachment parameter equal to A^t_{EO} and a $1 - t$ chance that they are untrustworthy, and has an attachment parameter of A^u_{EO}. The probability t is therefore the true probability that an even player is trustworthy. We assume that the even players know their own type, and the true probability that their fellow even players are trustworthy.

Odd players, however, do not know the true likelihood that even players are trustworthy. (Aside from the even players' attachment for the odd players, and the true likelihood that the even players are trustworthy, all other parameters in the game are common knowledge.) So that learning may take place, let the true likelihood that an even player is trustworthy, t, be distributed according to a β density with mean $\frac{\nu}{\nu+\omega}$ where ν and ω are integers greater than 0. When the game begins, the prior probability held by players 1 and 3 that an even player is trustworthy is equal to the mean of the distribution.

$$t^p = \frac{\nu}{\nu + \omega} \tag{1}$$

If 2 reciprocates in the first round, the posterior belief increases to

$$t'(R) = \frac{\nu + 1}{\nu + 1 + \omega}. \tag{2}$$

Conversely, if player 2 cheats, then the posterior belief falls to

$$t'(C) = \frac{\nu}{\nu + \omega + 1}. \tag{3}$$

The level of trust is higher after observing reciprocation than after observing cheating.

$$t'(C) < t^p < t'(R) \tag{4}$$

2 The Even Players' Choices

To ensure that the trustworthy type of player 4 reciprocates and the untrustworthy type cheats, their attachment parameters must straddle the thresholds where reciprocating and cheating yield equal payoffs, derived as follows.

If player 4 cheats they get 1 and if they reciprocate they get $b_4 + A_{4O}b_3$. These are equal for the following level of attachment by player 4 for the odd players.

$$A^*_{4O} \equiv \frac{1 - b_4}{b_3} \tag{5}$$

In considering player 2, we need to consider whether the first round affects behavior in the second round. If not, then player 2 need only consider the first round payoffs. Player 2 gets 1 from cheating and $b_2 + A_{2O}b_1$ from reciprocating, so the critical value is similar to that for player 4.

$$A^*_{2O} \equiv \frac{1 - b_2}{b_1} \tag{6}$$

If the second round players will condition their behavior on what happened in the first round, the condition is slightly more complicated. In equilibrium, the only way player 3 will condition their behavior on player 2 is to engage if 2 reciprocates and avoid if 2 cheats. The reverse pattern is impossible because then the type of player 2 with the lower attachment for the odd group would reciprocate when the one with the higher attachment cheats, which is not incentive compatible.

With this in mind, if player 2 cheats they will get 1 in the first round. Player 3 will then avoid player 4 in the second round so player 2 will get A_{2O} in the second round. Player 2's total payoff from cheating is therefore $1+A_{2O}$. If player 2 reciprocates, they will get $b_2+A_{2O}b_1$ in the first round. Player 3 will engage, and so player 4 will reciprocate with probability t and cheat with probability $1-t$.[1] The payoff for reciprocating is therefore $b_2 + A_{2O}b_1$ from the first round and $t(A_{2E}b_4 + A_{2O}b_3) + (1-t)A_{2E}$ from the second. Reciprocation gives the same payoff as cheating if the following holds,

$$\begin{aligned}
\text{Payoff for Cheating} &= \text{Payoff for Reciprocating} \\
1 + A_{2O} &= b_2 + A_{2O}b_1 + t(A_{2E}b_4 + A_{2O}b_3) + (1-t)A_{2E} \\
1 - b_2 - tA_{2E}b_4 - (1-t)A_{2E} &= A_{2O}(b_1 + tb_3 - 1) \\
1 - b_2 - A_{2E}(1 + t(b_4 - 1)) &= A_{2O}(b_1 + tb_3 - 1)
\end{aligned}$$

which gives the following value of A_{2O}.

$$A^{\dagger}_{2O} \equiv \frac{1 - b_2 - A_{2E}(1 + t(b_4 - 1))}{b_1 + tb_3 - 1} \tag{7}$$

The following assumption sums up what is needed for the even players' behavior.

Assumption 4 *The trustworthy types of the even players will reciprocate while the untrustworthy types will cheat,* $-1 < A^u_{EO} < A^*_{4O}, A^*_{2O}, A^{\dagger}_{2O} < A^t_{EO} < 1$.

[1] We assume that player 2 knows the true distribution of types in the even group so its belief that player 4 is the low hostility type is t.

Table 1: Possible Strategy Profiles for the Odd Players

			After 1 Avoids	After 1 Engages and 2 Reciprocates	After 1 Engages and 2 Cheats
	Name	Player 1		Player 3	
1	1E, 3EEE	Engage	Engage	Engage	Engage
2	1E, 3EEA	Engage	Engage	Engage	Avoid
3	1E, 3EAE	Engage	Engage	Avoid	Engage
4	1E, 3EAA	Engage	Engage	Avoid	Avoid
5	1E, 3AEE	Engage	Avoid	Engage	Engage
6	1E, 3AEA	Engage	Avoid	Engage	Avoid
7	1E, 3AAE	Engage	Avoid	Avoid	Engage
8	1E, 3AAA	Engage	Avoid	Avoid	Avoid
9	1A, 3EEE	Avoid	Engage	Engage	Engage
10	1A, 3EEA	Avoid	Engage	Engage	Avoid
11	1A, 3EAE	Avoid	Engage	Avoid	Engage
12	1A, 3EAA	Avoid	Engage	Avoid	Avoid
13	1A, 3AEE	Avoid	Avoid	Engage	Engage
14	1A, 3AEA	Avoid	Avoid	Engage	Avoid
15	1A, 3AAE	Avoid	Avoid	Avoid	Engage
16	1A, 3AAA	Avoid	Avoid	Avoid	Avoid

3 Equilibria in the Model

We solve the model for Perfect Bayesian Equilibria, in which actors choose the best strategy given their beliefs and update their beliefs in accordance with Bayes' rule wherever possible.

The behavior of the even players is determined by the assumptions made above. There remains the behavior of the odd players to account for. There are sixteen possible strategy profiles for the odd players to consider, as shown in Table 1. Player 1 chooses to avoid or engage, then player 3 must have a choice for what to do if player 1 avoids, player 1 engages and player 2 reciprocates, and player 1 engages and player 2 cheats.

Consider player 3's choice. Let t' be player 3's belief about the likelihood of facing a trustworthy player 4 when it must choose. Player 3 gets 1 for avoiding player 4. If player 3 engages, there is a t' chance that player 4 reciprocates and a $1-t'$ chance that player 4 cheats, for a payoff of $t'(b_3 + A_{3E}b_4) + (1-t')A_{3E}$. Player 3 compares the following payoffs.

$$\begin{aligned} \text{Payoff for Avoiding} &= \text{Payoff for Engaging} \\ 1 &= t'(b_3 + A_{3E}b_4) + (1-t')A_{3E} \end{aligned}$$

These payoffs are equal for the following threshold for player 3's attachment to the even players.

$$A^*_{3E}(t') = \frac{1 - t'b_3}{1 - t'(1-b_4)} \tag{8}$$

There are three possible levels of belief that player 3 may have, ordered as follows.

$$t'(C) < t^p < t'(R)$$

$A^*_{3E}(t')$ is decreasing in t', that is, the more trusting player 3 is, the less attached to the outgroup they need to be to cooperate. Therefore there are three different values for the engagement threshold as a function of player 3's beliefs, as follows.

$$A^*_{3E}(t'(R)) < A^*_{3E}(t^p)) < A^*_{3E}(t'(C)) \tag{9}$$

If $A_{3E} < A^*_{3E}(t'(R))$, then player 3 will not engage even after observing player 2 reciprocate. If $A_{3E} \in \{A^*_{3E}(t'(R)), A^*_{3E}(t^p))\}$ then player 3 will engage only if they see player 2 reciprocate. If $A_{3E} \in \{A^*_{3E}(t^p), A^*_{3E}(t'(C)),)\}$, then player 3 will engage unless they see player 2 cheat. Finally, if $A_{3E} > A^*_{3E}(t'(C))$, they will engage even if they see player 2 cheat.

These considerations serve to eliminate some of the rows in Table 1. Those highlighted in red feature player 3 avoiding after player 1 engages and 2 reciprocates, and engaging after player 1 engages and player 2 cheats, which is not possible. Those in pink feature player 3 engaging after player 2 cheats and avoiding when player 1 avoids, when player 3 has its prior beliefs, which is also not possible. Finally, the rows in yellow have player 3 avoiding after player 2 reciprocates and engaging with no new information after player 1 avoids. All these patterns are incompatible with the previous conditions on player 3's beliefs and strategies. We consider the remaining eight cases below.

3.1 Profile 1

In this case player 1 and 3 engage no matter what. The condition for player 3 is the following.

$$A^*_{3E}(t'(C)) \leq A_{3E} \tag{10}$$

For player 1, avoidance gives a payoff of 1, and engaging gives a payoff of $t^p(b_1 + A_{1E}b_2) + (1 - t^p)(A_{1E})$. The payoff in the second round is the same either way and so does not affect the calculation. We set these equal to each other and solve for a threshold level.

$$\begin{aligned}\text{Payoff for Avoiding} &= \text{Payoff for Engaging} \\ 1 &= t^p(b_1 + A_{1E}b_2) + (1 - t^p)(A_{1E}) \\ A_{1E}^* &= \frac{1 - t^p b_1}{1 - t^p(1 - b_2)}\end{aligned}$$

Player 1 will choose to engage if the following holds.

$$A_{1E}^* \leq A_{1E} \tag{11}$$

3.2 Profile 2

Here player 1 engages and player 3 engages unless player 2 cheats. The condition for player 3 is the following.

$$A_{3E}^*(t^p) \leq A_{3E} \leq A_{3E}^*(t'(C)) \tag{12}$$

For player 1, avoidance gives 1 in the first round. Avoidance would send the game off the equilibrium path, but I assume that player 3's beliefs about player 4 remain unchanged by player 1's behavior. In the second round, player 3 will engage and the payoff for player 1 is therefore $t^p(A_{1O}b_3 + A_{1E}b_4) + (1 - t^p)A_{1E}$. If player 1 engages, player 2 may reciprocate or cheat, for a first round payoff of $t^p(b_1 + A_{1E}b_2) + (1 - t^p)A_{1E}$. If player 2 reciprocates, then player 3 will engage, and player 1 will get $t'(R)(A_{1O}b_3 + A_{1E}b_4) + (1 - t'(R))A_{1E}$. If player 2 cheats, player 3 will avoid, and player 1 will get A_{1O}. We set the payoffs equal to solve for the threshold level of A_{1E}.

$$\text{Payoff for Avoiding} = \text{Payoff for Engaging}$$

$$\begin{aligned} 1 + t^p(A_{1O}b_3 + A_{1E}b_4) + (1-t^p)A_{1E} &= t^p[b_1 + A_{1E}b_2 + t'(R)(A_{1O}b_3 + A_{1E}b_4) + \\ &\quad (1-t'(R))A_{1E}] + (1-t^p)(A_{1E} + A_{1O}) \end{aligned}$$

$$\begin{aligned} 1 + t^p(A_{1O}b_3 + A_{1E}b_4) &= t^p[b_1 + A_{1E}b_2 + t'(R)(A_{1O}b_3 + A_{1E}b_4) + (1-t'(R))A_{1E}] \\ &\quad +(1-t^p)A_{1O} \end{aligned}$$

$$\begin{aligned} 1 + t^pA_{1O}b_3 - t^pb_1 - t^pt'(R)A_{1O}b_3 - (1-t^p)A_{1O} &= -t^pA_{1E}b_4 + t^p[A_{1E}b_2 + t'(R)A_{1E}b_4 \\ &\quad +(1-t'(R))A_{1E}] \end{aligned}$$

$$\begin{aligned} A^{\dagger}_{1E} &= \frac{1 - t^pb_1 + A_{1O}(t^pb_3 - t^pt'(R)b_3 - (1-t^p))}{t^p[-b_4 + b_2 + t'(R)b_4 + 1 - t'(R)]} \\ A^{\dagger}_{1E} &= \frac{1 - t^pb_1 + A_{1O}(t^pb_3(1-t'(R)) - (1-t^p))}{t^pb_2 + t^p(1-b_4)(1-t'(R))} \end{aligned}$$

For the equilibrium to hold, it must be the case that the following holds.

$$A_{1E}^{\dagger} \leq A_{1E} \tag{13}$$

3.3 Profile 6

This is the reassurance equilibrium. The condition for player 3 is the following.

$$A_{3E}^{*}(t'(R)) \leq A_{3E} \leq A_{3E}^{*}(t^{p}) \tag{14}$$

If player 1 avoids player 2 in the first round, they get a payoff of 1 and in the second round player 3 will avoid player 4 as well, for a payoff of A_{1O}. Conversely, if player 1 engages in the first round, there is a t^p chance that player 2 reciprocates, which will give a payoff of $b_1 + A_{1E}b_2$ for the first round. In the second round, player 3 will also engage which will give player 1 a payoff of $t'(R)(A_{1O}b_3 + A_{1E}b_4) + (1 - t'(R))A_{1E}$. On the other hand, there is a $1 - t^p$ chance that player 2 will cheat, giving a first round payoff of A_{1E} and causing player 3 to avoid player 4, giving A_{1O} in the second round. The payoff for engaging is therefore the same as in the previous case.

$$t^{p}[b_1 + A_{1E}b_2 + t'(R)(A_{1O}b_3 + A_{1E}b_4) + (1 - t'(R))A_{1E}] + (1 - t^{p})(A_{1E} + A_{1O})$$

We set the payoff for avoiding equal to that for engaging and solve for the threshold.

$$\begin{aligned}
\text{Payoff for Avoiding} &= \text{Payoff for Engaging} \\
1 + A_{1O} &= t^p[b_1 + A_{1E}b_2 + t'(R)(A_{1O}b_3 + A_{1E}b_4) \\
&\quad +(1 - t'(R))A_{1E}] + (1 - t^p)(A_{1E} + A_{1O}) \\
1 - A_{1E} &= t^p[b_1 + A_{1E}b_2 + t'(R)(A_{1O}b_3 + A_{1E}b_4) \\
&\quad +(1 - t'(R))A_{1E} - (A_{1E} + A_{1O})] \\
1 - t^p(b_1 - A_{1O}(t'(R)b_3 - 1)) &= A_{1E} + t^p A_{1E}(b_2 + t'(R)b_4 - t'(R))
\end{aligned}$$

The relation can be solved for a condition on player 1's attachment to the even players.

$$A^{\ddagger}_{1E} = \frac{1 - t^p(b_1 + A_{1O}(1 - t'(R)b_3))}{1 - t^p(t'(R)(1 - b_4) - b_2)} \tag{15}$$

The following condition must hold.

$$A^{\ddagger}_{1E} \leq A_{1E} \tag{16}$$

3.4 Profile 8

Here player 1 engages and player 3 avoids no matter what. The condition for player 3 is the following.

$$A_{3E} \leq A^*_{3E}(t'(R)) \tag{17}$$

For player 1, avoiding gives a payoff of $1 + A_{1O}$ while engaging gives a payoff of $t^p(b_1 + A_{1E}b_2) + (1 - t^p)A_{1E} + A_{1O}$. We set these equal to each other and solve for the threshold.

$$\begin{aligned}
\text{Payoff for Avoiding} &= \text{Payoff for Engaging} \\
1 + A_{1O} &= t^p(b_1 + A_{1E}b_2) + (1 - t^p)A_{1E} + A_{1O} \\
1 - t^p b_1 &= A_{1E}(t^p b_2 + 1 - t^p) \\
A^*_{1E} &= \frac{1 - t^p b_1}{1 - t^p(1 - b_2)}
\end{aligned}$$

This is the same threshold as in Profile 1, and the equilibrium is possible if the following holds.

$$A^*_{1E} \leq A_{1E} \tag{18}$$

Now we turn to the profiles in which player 1 avoids.

3.5 Profile 9

In this case player 1 avoids and player 3 engages no matter what. The condition for player 3 is the same as in profile 1.

$$A^*_{3E}(t'(C)) \leq A_{3E} \tag{19}$$

The condition for player 1 is the opposite of that in profile 1.

$$A_{1E} \leq A^*_{1E} \tag{20}$$

3.6 Profile 10

Here player 1 avoids and player 3 engages unless player 2 cheats. The condition for player 3 is the same as in profile 2.

$$A^*_{3E}(t^p) \leq A_{3E} \leq A^*_{3E}(t'(C)) \tag{21}$$

The condition for player 1 is the opposite of that in profile 2.

$$A_{1E} \leq A^\dagger_{1E} \tag{22}$$

3.7 Profile 14

Here player 1 avoids and player 3 engages only if player 2 reciprocates. The condition for player 3 is the same as profile 6.

$$A^*_{3E}(t'(R)) \leq A_{3E} \leq A^*_{3E}(t^p) \tag{23}$$

The condition for player 1 is the opposite of profile 6.

$$A_{1E} \leq A^{\ddagger}_{1E} \tag{24}$$

3.8 Profile 16

Finally, in profile 16 both player 1 and 2 avoid in all circumstances. The conditions are as follows.

$$A_{3E} \leq A^*_{3E}(t'(R)) \tag{25}$$

$$A_{1E} \leq A^*_{1E} \tag{26}$$

4 The Model with Communication

Consider a variant of the previous game in which players 3 and 4 do not observe what happened in the first round, but player 1 can communicate to player 3 about whether player

2 reciprocated or not. That is, after the first round is over, player 1 can say "$2R$" meaning that player 2 reciprocated, or "$2C$" to indicate that 2 cheated. The game is otherwise identical. When can player 1 be relied on to tell the truth to player 3 about what happened in the first round?

Note, this question is only relevant if player 1 is willing to engage in the first place, and if player 3 is willing in principle to condition their behavior on what player 2 does. The two possibilities are profiles 2 and 6.

If player 1 says that player 2 cheated, player 3 will avoid player 4, giving player 1 a payoff of A_{1O} for the second round. If player 1 says player 2 reciprocated, player 3 will engage, and player 4 will reciprocate if trustworthy, which results in a payoff of $t'(A_{1O}b_3 + A_{1E}b_4) + (1 - t')A_{1E}$ for player 1. Player 1 must prefer to say that player 2 cheated when they did and reciprocated when they did, so the following conditions must hold.

$$
\begin{aligned}
A_{1O} &> t'(C)(A_{1O}b_3 + A_{1E}b_4) + (1 - t'(C))A_{1E} \\
A_{1O} &< t'(R)(A_{1O}b_3 + A_{1E}b_4) + (1 - t'(R))A_{1E}
\end{aligned}
$$

These conditions can be solved for a constraint on player 1's attachment for the outgroup as a function of their attachment for the ingroup. To be honest when player 2 cheated, the following must hold.

$$
\begin{aligned}
A_{1O} &> t'(C)(A_{1O}b_3 + A_{1E}b_4) + (1 - t'(C))A_{1E} \\
A_{1O} - t'(C)A_{1O}b_3 &> t'(C)(A_{1E}b_4) + (1 - t'(C))A_{1E}
\end{aligned}
$$

$$
\begin{aligned}
A_{1O}(1 - t'(C)b_3) &> A_{1E}(t'(C)b_4 + 1 - t'(C)) \\
A_{1O}\frac{1 - t'(C)b_3}{1 - t'(C)(1 - b_4)} &> A_{1E}
\end{aligned}
$$

$$
A_{1E} \leq A_{1O}A^*_{3E}(t'(C)) \tag{27}
$$

To be honest when player 2 reciprocated, it must be the case that the following holds.

$$
A_{1O}A^*_{3E}(t'(R)) \leq A_{1E} \tag{28}
$$

5 The Equilibria Illustrated

The equilibria are illustrated in Figure 3. The horizontal axis is player 1's attachment for the outgroup, A_{1E}, and the vertical axis is player 3's attachment to the outgroup, A_{3E}. Both range from -1 to 1, where higher values signify being more friendly to the outgroup. The parameters underlying the cutpoints are shown in Table 2.

The labels for the equilibria first list player 1's strategy, Avoid or Engage, and then Player 3's strategy, in the case where player 1 avoids, where player 1 engages and Player 2 reciprocates, and where player 1 engages and player 2 cheats.

In the top right corner is equilibrium 1E, 3EEE, where player 1 engages and player 3 engages no matter what happens in the first round. This is pattern 1 in Table 1. It is

Table 2: The Parameters and Cutpoints

Parameter	Value
b_1	2
b_2	0.5
b_3	2
b_4	0.5
ν	1
ω	1
t^p	0.5
$t'(R)$	0.66666
$t'(C)$	0.33333
A_{1O}	0.5

Cutpoint	Value
$A^*_{3E}(t'(R))$	-0.49997
$A^*_{3E}(t^p)$	0.0
$A^*_{3E}(t'(C))$	0.39998
A^*_{1E}	0.0
$A^\dagger_{1E}$	-0.24998
$A^\ddagger_{1E}$	0.0769

feasible when both player 1 and player 3 are positive towards the out group. Moving down is equilibrium 1E, 3EEA, where player 1 engages and player 3 engages unless observing player 2 cheat. In this case player 3 is willing to engage based on their prior beliefs, and so does not need reassurance.

Moving down once more, we have equilibrium 1E, 3AEA, where player 1 engages and player 3 engages only if observing player 2 reciprocate. Here player 3 needs reassurance to be willing to engage, and a successful interaction in round 1 provides that reassurance. Moving down to the lower right hand corner, is equilibrium 1E, 3AAA, where player 1 engages, but player 3 avoids even if they see player 2 reciprocate. In this case, player 3 is too negative towards the even group to cooperate, even after getting good information about them.

On the left hand side are the corresponding equilibria where player 1 avoids.

In the center, between the dotted lines, is the region in which player 1 can honestly communicate with player 3 about what happened in the first round.

Figure 3: The Equilibria in the Game