Online appendix: Strategic Discrimination in Hierarchies

Dominik Duell[*]and Dimitri Landa[†]

Monday 16th September, 2019

Online appendix

A Experimental design appendix

A.1 Treatments

We implement four treatments within the research agenda: the main – STRATEGIC – treatment features induced groups and the opportunity to reward the agent with a bonus (henceforth, referred to as the availability of the sanctioning device), following closely the model described above. The NON-IDENTITY treatment does not induce group identities, and the NON-STRATEGIC treatment induces identities but removes the sanctioning device. The NON-STRATEGIC/NON-IDENTITY treatment features neither induced identities nor the sanctioning device.

Our experiment included 202 subjects, 101 in the role of a principal and 101 in the role of an agent, generating 4040 subject-round observations in 11 sessions (see Table A.1).

Table A.1: Experimental treatments, number of subjects (N), and of subject-round observations (n)

	Identity	**No identity**
With sanctioning	STRATEGIC (N=110, n=2200)	NON-IDENTITY (N=38, n=760)
Without sanctioning	NON-STRATEGIC (N=40, n=800)	NON-STRATEGIC/NON-IDENTITY (N=14, n=280)

A.2 Setup

Sessions were carried out at the Center for Experimental Social Sciences/NYU. Each experimental session lasted 20 rounds with 14-22 participating subjects. Participants signed up via a web-based

[*] University of Essex

[†] New York University

recruitment system that draws on a large, pre-existing pool of potential subjects. Subjects were not recruited from the authors' courses. The recruitment system contains a filter that blocked subjects from participating in more than one session of a given experiment. The subject pool consists almost entirely of undergraduates from around the university. Subjects interacted anonymously via networked computers. The experiments were programmed and conducted with the software z-Tree (Fischbacher, 2007).

After giving informed consent according to standard human subjects protocols, subjects received written instructions that were subsequently read aloud in order to promote understanding and induce common knowledge of the experimental protocol. No deception was employed at any point in the experiment, in accordance with the long-standing norms of the lab in which the experiment was carried out. Before the principal-agent game stage commenced, subjects were asked three questions concerning their understanding of the payoff tables provided to them in the instructions. 90% of participating subjects answered those questions correctly. At the end of the experiment, an exit survey was conducted.

In communicating the game to the subjects we referred to type as "Special Number," to noise as "Random Bump," to outcome as the "Choice Outcome", to subjects in the role of agents as "Player 1," and to subjects in the role of principals as "Player 2"; the value generated by principal's decision whether to double type or effort in the outcome-function was termed "Increased Outcome." Subjects did not see agent's payoff function but received a table of all possible payoffs given type, effort, and noise, and the principal's reward decision, and in the instructions were told:

> "When you are participating in the role of Player 1, your payoff in a given round will depend on the *choice outcome* in that round (and so indirectly, on your *special number*, your *effort* level, and the realised *random bump*) but also directly on the chosen level of *effort* and on the decision of Player 2 you are matched with whether to give you a *bonus*."

A.3 Group identity inducement

At the beginning of each session of both the STRATEGIC and the NON-STRATEGIC treatments, subjects were shown 5 pairs of paintings, with one painting by Paul Klee paired with one by Vassily Kandinsky, and were asked which painting they prefer in each pair. Based on which painter a subject preferred in a majority of pairs, he/she was assigned to be a *Klee* or a *Kandinsky.*

The STRATEGIC treatment condition generated 55 Klees (subjects who preferred paintings by Paul Klee most of the time) and 55 Kandinskys (subjects who preferred those by Vassily Kandinsky most of the time). In the NON-STRATEGIC treatment there were 21 Klees and 19 Kandinskys.

Once identities were assigned, subjects participated in an activity aimed at strengthening the attachment to the new identities. In particular, they were given a quiz in which they were asked to identify the painter (Klee or Kandinsky) of five further paintings. In answering the question about each of those paintings, subjects gave initial guesses which were made available to other subjects in the same identity group before everyone was asked for their final answer. Subjects within a group received $1 if the majority of members of their group named the correct painter in the final answer. Additionally, they received another $1 when members of their group gave at least as many correct final answers on all five quizzes as members of the other group.

During the quiz, a majority of members in both groups gave correct answers in four out of five painting quizzes. Ultimately, all subjects received a payoff of $5 at this stage of the experiment. This positive group experience in a competitive environment is part of the intended group-identity strengthening; the experimenters intentionally selected paintings whose authors are moderately easy to identify. Subjects were told how many correct answers their group gave and were notified that members of their group “gave at least as many correct answers” as members of the other group. Members of both groups, Klees and Kandinskys, in all treatments performed approximately equally well.

A.4 Instructions

Introduction
During the following experiment, we require your complete undivided attention and ask that you follow instructions carefully. Please turn off your cell phones and, for the duration of the experiment, do not take actions that could distract you or other participants, including opening other applications on your computer, reading books, newspapers, and doing homework.

This is an experiment on group decision-making. In this experiment you will make a series of choices. At the end of the experiment, you will be paid depending on the specific choices that you made during the experiment and the choices made by other participants. If you follow the instructions and make appropriate decisions, you may make an appreciable amount of money.

This experiment has 3 parts. Your total earnings will be the sum of your payoffs in each part plus the show-up fee. We will start with a brief instruction period, followed by Part 1 of the experiment. After Part 1 is completed, we will pause to receive instructions for Part 2 and complete the session accordingly.

If you have questions during the instruction period, please raise your hand after I have completed reading the instructions, and your questions will be answered out loud so everyone can hear. Please restrict these questions to clarifications about the instructions only. If you have any questions after the paid session of the experiment has begun, raise your hand, and an experimenter will come and assist you. Apart from the questions directed to the experimenter, you are expressly asked to refrain from communicating with other participants in the experiment, including making public remarks or exclamations. Failure to comply with these instructions will result in the termination of your participation and the forfeiture of any compensation.

Part 1
In Part 1 of the experiment, everyone will be shown 5 pairs of paintings by two artists, Paul Klee and Wassily Kandinsky. You will be asked to choose which painting in each pair you prefer. You will then be classified as member of the "KLEEs" (or "a KLEE" as a shorthand) or member of the "KANDINSKYs" (or "a KANDINSKY" as a shorthand) based on which artist you prefer most and informed privately about your classification. Everyone's identity as a KLEE or as a KANDINSKY will stay fixed for the rest of the experiment (that is, in both Part 1 and Part 2 of the experiment).

You will then be asked to identify the painter (Klee or Kandinsky) of five other paintings. For each of those paintings, you will be asked to submit two answers: your initial guess and your final answer. After submitting your initial guess, you will have an opportunity to see the initial guesses of your fellow KLEEs if you are a KLEE, or of fellow KANDINSKYs if you are a KANDINSKY, and then also an opportunity to change your answer when you are submitting your final answer.

If you are a KLEE and a half or more of KLEEs give a correct final answer then, regardless of whether your own final answer was correct or incorrect, you and each of your fellow KLEEs will receive $1. Similarly, if you are a member of the KANDINSKYs and a half or more of KANDINSKYs give a correct final answer then, regardless of your own final answer, each of the KANDINSKYs, including you, will receive $1. However, if you are a KLEE and more than a half of KLEEs give an incorrect final answer, then, regardless of whether your own final answer was correct or incorrect, you and each of the KLEEs will receive $0. And similarly, if you are a KANDINSKY and the final answers from more than a half of KANDINSKYs were incorrect, then you and each of your fellow KANDINSKYs will receive $0 regardless of what answer he or a she gave personally.

In addition, if you and your fellow group members answer at least as many quiz questions correctly than members of the other group, you will receive an additional payoff of $1. That is, if you are a KLEE and you and your fellow KLEEs give more correct answers than the KANDINSKYs, you receive the additional payoff. If you are a KANDINSKY and you and your fellow KANDINSKYs give more correct answers than the KLEEs, you receive the additional payoff.

We will now run Part 1 of the experiment. After Part 2 has finished, we will give you instructions for Part 2.

Part 2
We will now move on to Part 2 of the experiment. Part 2 will consist of 20 different rounds. At the beginning of the first round, you will be randomly assigned a role of either Player 1 or Player 2. You will keep that role for the rest of Part 3 of the experiment. Throughout this part of the experiment, you will also retain your identity as a member of the KLEEs or a member of the KANDINSKYs, as assigned in Part 2 of the experiment.

Matched group
In each round, all participants in the experiment will be randomly matched into pairs, each consisting of one Player 1 and one Player 2. Because every participant will be randomly re-matched with other participants into a different group in each round of the experiment, the composition of matched pairs will vary from one round to the next. All of participants' interactions will take place anonymously through a computer terminal, so your true personal identity will never be revealed to others, and you will not know who precisely is in your pair in any round of the experiment. However, every time you are matched with another participant (Player 1 or Player 2), you will be told whether that participant is a member of the KLEEs or a member of the KANDINSKYs.

In each round, a member of the group who takes on the role of Player 1 in that round will be randomly assigned a number, which we will refer to as Player 1's *special number*. That number will be shown only to that participant and never to other participants in the experiment. You should know, however, that Player 1's *special number* is one of three possible numbers: 1, 2 or, 3, and is chosen by the computer for assigning to Player 1 so that each of these numbers is equally likely to be picked. In each round, Player 1 is assigned a new *special number*, which stays fixed until the round ends, at which point a new *special number* is assigned. As with all other players, her identity as a member of the KLEEs or a member of the KANDINSKYs does not change from one round to the next.

Choices within each round of the experiment
At the beginning of each round, in each group, the member who is designated as Player 1 will choose a number: 1, 2, or 3, which you can think of as Player 1's level of *effort*. Please note that, while Player 1's *effort* is her choice, Player 1's *special number* is not her choice, but is assigned to Player 1 by the computer. Player 1's choice of *effort* will help determine *the choice outcome* in that round. In particular, *the choice outcome* will be computed as follows:

the choice outcome = Player 1's effort + Player 1's special number + random bump,

where the possible values of the *random bump* are -1, 0, or 1, and any one of these three values will be possible and equally likely to occur.

For example, suppose that a given Player 1's *special number* is 2, he or she chooses a level of *effort* equal to 1, and the realised value of the *random bump* is -1. Then *the choice outcome* is 2 + 1 - 1 = 2.

After *the choice outcome* is computed, it will be shown to Player 2. However, Player 2 will not see Player 1's *special number* nor her choice of *effort* nor the realised value of the *random bump.*
After seeing *the choice outcome,* Player 2 will be given an opportunity to *increase* the outcome by doubling the contribution to outcome of either Player 1's *effort* or of her *special number* – whichever of those two Player 2 decides to increase. A new outcome will, then, be computed, based on the corresponding *choice outcome*, but now increased because of the doubled contribution of *effort* or

special number, as indicated by Player 2. We will refer to this new resulting outcome as *the increased outcome.*

For example, suppose that a given Player 1's *special number* is 2, he or she chooses a level of *effort* equal to 1, and the realised *random bump* is -1. Suppose, further, that Player 2 decides to increase the outcome by raising the contribution of *effort.* Then *the increased outcome* is 2 + [2(1)] - 1 = 3. (Note that the product in the square brackets [] is the newly increased value of *effort.*) If, in contrast, Player 2 decides to raise the contribution of Player 1's *special number,* then *the increased outcome* is [2(2)] + 1 - 1 = 4. (Note that the product in the square brackets [] is now the newly increased contribution of Player 1's *special number.*)

Of course, if Player 1 had chosen a level of *effort* equal to 3, instead, then, with her *special number* (2) and the realised *random bump* (-1), *the choice outcome* would be 1 + 3 - 1 = 3. If Player 2 had further chosen to increase the outcome by increasing the contribution of Player 1's *special number,* then *the increased outcome* would be 2(1) + 3 - 1 = 4. But if Player 2 had chosen to increase the contribution of Player 1's *effort,* then *the increased outcome* would be 1 + 2(3) - 1 = 6.

In addition to deciding how to increase the *choice outcome*, Player 2 also decides if she wants to give Player 1 a *bonus* - a special addition to Player 1's payoff in that round.

After *the increased outcome* is shown to Player 2 and Player 2's bonus decision is shown to Player 1, the round ends and the players proceed to the next round.

This completes the description of a single round of play. I will now describe how your payoff for the experiment will be calculated.

Payoffs

If you are participating in the role of Player 1, your payoff in a given round will depend on *the choice outcome* in that round (and so indirectly, on your *special number*, your *effort* level, and the realised *random bump*) but also directly on the chosen level of *effort* and on the decision of Player 2 you are matched with whether to give you a *bonus.*

Please look now at Table 1 on page 9 of these instructions. This table gives you the values of Player 1's payoffs for all possible values of your *special number*, your *effort* level, and the realised *random bump.* For your convenience we are reproducing a piece of this table in the text of these instructions. Please, turn back to page 6 of the instructions.

Special		Random			
Number	Effort	Bump	Outcome	Bonus	No Bonus
1	1	-1	1	**6.54**	**4.05**
		0	2	**8.44**	**6.54**
		1	3	**10.05**	**8.44**
	2	-1	2	**6.49**	**4.59**
		0	3	**8.10**	**6.49**
		1	4	**9.52**	**8.10**
	3	-1	3	**6.15**	**4.54**
		0	4	**7.57**	**6.15**
		1	5	**8.85**	**7.57**

Suppose, for example, that in a given round, your *special number* was 1, your *effort* was 2, and the *random bump* was -1. You can see in the table above that the resulting choice outcome is 2. Suppose that Player 2 decided not to give you a *bonus* this round. You will find your payoff for this example by finding *special number* equal to 1 in the left-most column, *effort* equal to 2 in the column second from the left, and *random bump* equal to -1 in the third column from the left. Then, you will see in the right-most column of this row of Table 1 that your payoff for that round will be $4.59.

Suppose, however, that you are considering a higher level of *effort*, say 3. If the random bump happens to be same, -1, then the outcome will be 3. If the Player 2 decides to give you a *bonus* in this case, then your payoff in this round can be found by locating *special number* equal to 1 in the left-most column, *effort* equal to 3 in the second column from the left, *random bump* equal to -1, and then looking at the second to last column of this row, which shows a payoff of $6.15.

To give you further assistance in visualizing your choices as Player 1, we will also provide you the relevant payoff information on the screen as you are making your *effort* choices. This information will be equivalent to what you see in Table 1. Please look now at page 8 of this handout, which reproduces a screenshot similar to what you will see each round. The screenshot shows a question that we will ask Player 1 as a part of his *effort* choice: "What minimal outcome do you think Player 2 will demand to give you a bonus?" Then, for a given such outcome that you are specifying, the screen will show you what payoffs you may get with what probabilities (corresponding to different random bumps) given different available choices of *effort*.

If you are participating in the role of Player 2, your payoff in a given round will be equal to *the increased outcome* you obtained in that round – that is, it will depend on *the choice outcome* produced by Player 1 you are matched with (and so on Player 1's *special number*, her choice of *effort*, and the realised *random bump*), as well as on your decision on how to increase it.

Please look now at Table 2 on page 10 of the instructions where you can see how Player 2's payoffs are computed from *the choice outcome* and Player 2's decision how to increase it. Now, for example, suppose that in a given round, Player 1's *special number* was 2, she chose a level of *effort* equal to 1, and the value of the *random bump* was -1. If you chose to increase the outcome by increasing *effort*, then your payoff in that round is

$$2 + [2 \times 1] - 1 = \$3$$

In contrast, if you chose to increase the outcome by increasing Player 1's *special number*, then your payoff in that round is

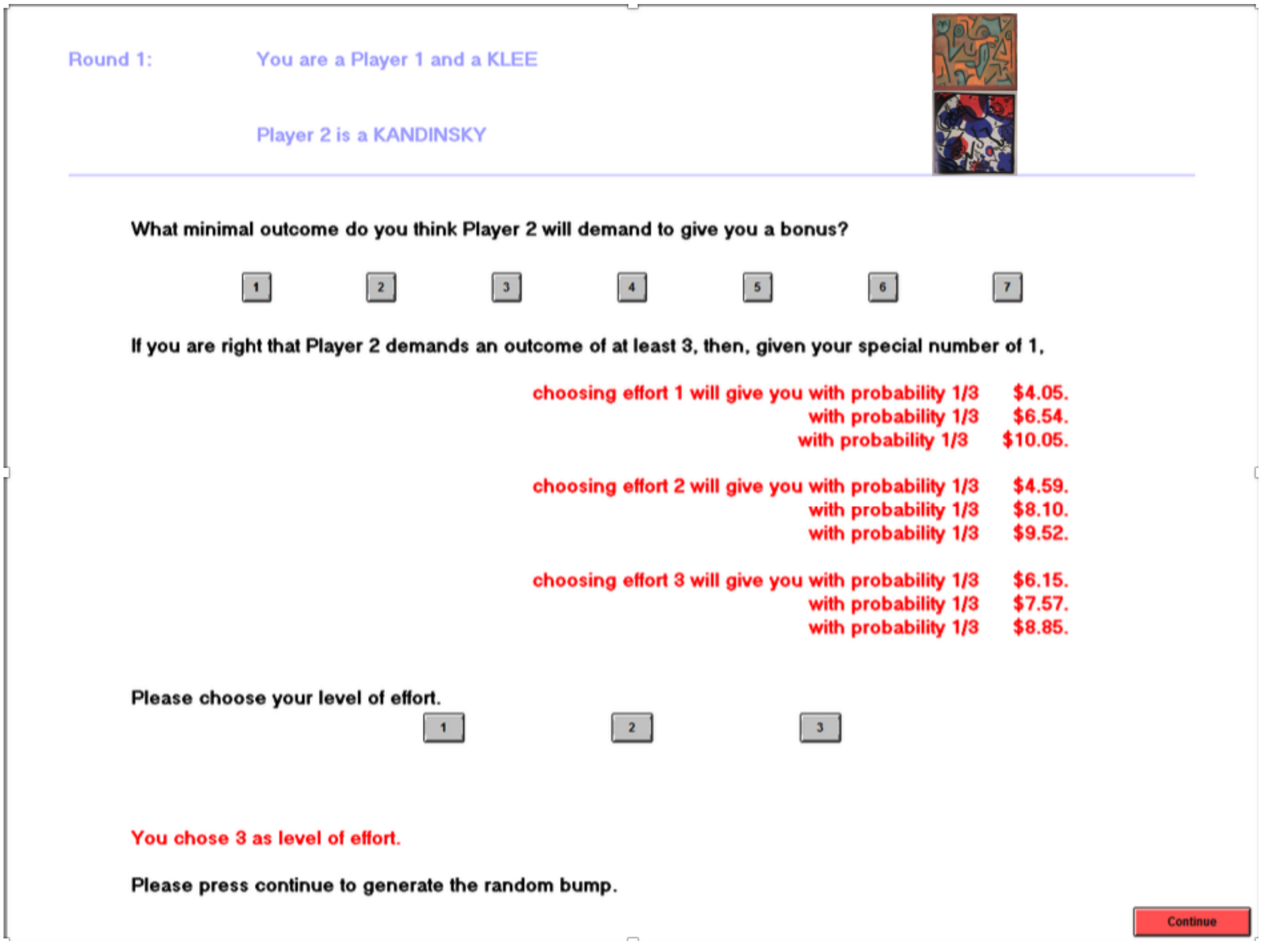
$$[2 \times 2] + 1 - 1 = \$4$$

You will see this by finding *special number* equal to 2 in the left-most column, *effort* equal to 1 in the second column from the left, and *random bump* equal to -1 in the third column from the left. The value in the same wow of the next column shows that the *the choice outcome* associated with this example is 2. The values in this row in the two columns on the right, then, tell you what *the increased outcome* and thus your payoff from this round as Player 2 will be. In case you decide to double *special number*, your payoff will be 4. In case you decide to increase *effort*, your payoff will be 3.

Again, your total payoff for the experiment will be the two highest round payoff from three randomly chosen rounds plus your payoffs from Part 1 of the experiment plus the show-up fee of $7.

If you have any questions, please ask them now.

Figure A.1: Screen shot of agents' belief elicitation and effort decision in the STRATEGIC treatment. Screen shot was embedded as Figure 1 on page 8 of the instructions given to subjects.



Round 1: You are a Player 1 and a KLEE

Player 2 is a KANDINSKY

What minimal outcome do you think Player 2 will demand to give you a bonus?

If you are right that Player 2 demands an outcome of at least 3, then, given your special number of 1,

choosing effort 1 will give you with probability 1/3 $4.05.
with probability 1/3 $6.54.
with probability 1/3 $10.05.

choosing effort 2 will give you with probability 1/3 $4.59.
with probability 1/3 $8.10.
with probability 1/3 $9.52.

choosing effort 3 will give you with probability 1/3 $6.15.
with probability 1/3 $7.57.
with probability 1/3 $8.85.

Please choose your level of effort.

You chose 3 as level of effort.

Please press continue to generate the random bump.

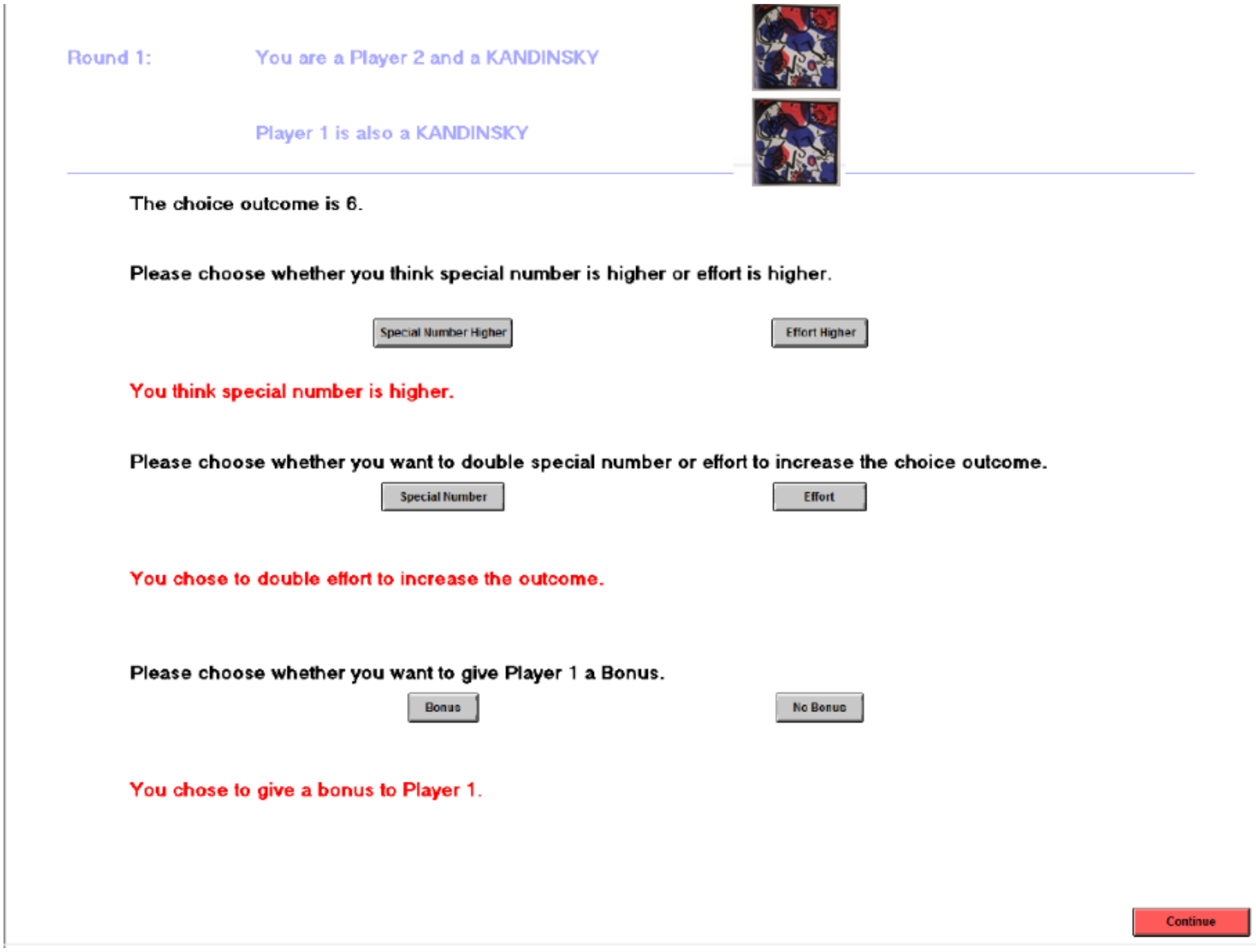
Table 1: Player 1's round payoff

Special Number	Effort	Random Bump	Outcome	Bonus	No Bonus
1	1	-1	1	**6.54**	**4.05**
		0	2	**8.44**	**6.54**
		1	3	**10.05**	**8.44**
	2	-1	2	**6.49**	**4.59**
		0	3	**8.10**	**6.49**
		1	4	**9.52**	**8.10**
	3	-1	3	**6.15**	**4.54**
		0	4	**7.57**	**6.15**
		1	5	**8.85**	**7.57**
2	1	-1	2	**8.44**	**6.54**
		0	3	**10.05**	**8.44**
		1	4	**11.47**	**10.05**
	2	-1	3	**8.10**	**6.49**
		0	4	**9.52**	**8.10**
		1	5	**10.80**	**9.52**
	3	-1	4	**7.57**	**6.15**
		0	5	**8.85**	**7.57**
		1	6	**10.02**	**8.85**
3	1	-1	3	**10.05**	**8.44**
		0	4	**11.47**	**10.05**
		1	5	**12.57**	**11.47**
	2	-1	4	**9.52**	**8.10**
		0	5	**10.80**	**9.52**
		1	6	**11.97**	**10.80**
	3	-1	5	**8.85**	**7.57**
		0	6	**10.02**	**8.85**
		1	7	**11.12**	**10.02**

Table 2: Player 2's round payoff

				Increased Outcome when	
Special Number	**Effort**	**Random Bump**	**Outcome**	**Special Number Doubled**	**Effort Doubled**
1	1	-1	**1**	**2**	**2**
		0	**2**	**3**	**3**
		1	**3**	**4**	**4**
	2	-1	**2**	**3**	**4**
		0	**3**	**4**	**5**
		1	**4**	**5**	**6**
	3	-1	**3**	**4**	**6**
		0	**4**	**5**	**7**
		1	**5**	**6**	**8**
2	1	-1	**2**	**4**	**3**
		0	**3**	**5**	**4**
		1	**4**	**6**	**5**
	2	-1	**3**	**5**	**5**
		0	**4**	**6**	**6**
		1	**5**	**7**	**7**
	3	-1	**4**	**6**	**7**
		0	**5**	**7**	**8**
		1	**6**	**8**	**9**
3	1	-1	**3**	**6**	**4**
		0	**4**	**7**	**5**
		1	**5**	**8**	**6**
	2	-1	**4**	**7**	**6**
		0	**5**	**8**	**7**
		1	**6**	**9**	**8**
	3	-1	**5**	**8**	**8**
		0	**6**	**9**	**9**
		1	**7**	**10**	**10**

Figure A.2: Screen shot of principals reward decision and attribution decision screen in the STRATEGIC treatment.



Round 1: You are a Player 2 and a KANDINSKY

Player 1 is also a KANDINSKY

The choice outcome is 6.

Please choose whether you think special number is higher or effort is higher.

You think special number is higher.

Please choose whether you want to double special number or effort to increase the choice outcome.

You chose to double effort to increase the outcome.

Please choose whether you want to give Player 1 a Bonus.

You chose to give a bonus to Player 1.

B Statistical appendix

B.1 Session statistics

Table B.1: Number of subjects and number of observations by treatment.

Treatment		# of subjects	# of observations
STRATEGIC	Klees	55	1100
	Kandinskys	55	1100
	Total	110	2200
NON-IDENTITY	Total	38	760
NON-STRATEGIC	Klees	21	420
	Kandinskys	19	380
	Total	40	800
NON-STRATEGIC/NON-IDENTITY	Total	14	280
		202	4040

The STRATEGIC treatment condition generated 55 Klees (subjects who preferred paintings by Paul Klee most of the time) and 55 Kandinskys (subjects who preferred those by Vassily Kandinsky most of the time). In the NON-STRATEGIC treatment there were 21 Klees and 19 Kandinskys. During the quiz, a majority of members in both groups gave correct answers in four out of five painting quizzes. Ultimately, all subjects received a payoff of $5 at this stage of the experiment. This positive group experience in a competitive environment is part of the intended group-identity strengthening; the experimenters intentionally selected paintings whose authors are moderately easy to identify. Subjects were told how many correct answers their group gave and were notified that members of their group “gave at least as many correct answers” as members of the other group.

B.2 Summary statistics

Table B.2: Means (standard deviation), minimum, and maximum values of type, effort, outcome, attribution decision (0 = attributed to type, 1 = attributed to effort), and reward decision (0 = not rewarded, 1 = rewarded) by treatment.

	STRATEGIC		**NON-IDENTITY**	**NON-STRATEGIC**			
Variable	In-group	Out-group		In-group	Out-group	Min	Max
Type	1.97 (.82)	2.01 (.80)	2.01 (.81)	2.00 (.79)	2.05 (.79)	1	3
Effort	1.79 (.78)	1.74 (.79)	1.76 (.84)	2.11 (.77)	2.17 (.76)	1	3
Expected demand	3.38 (1.2)	3.48 (1.3)	3.77 (1.4)			1	7
Outcome	3.70 (1.3)	3.68 (1.3)	3.81 (1.3)	4.03 (1.1)	4.14 (1.2)	1	7
Attribution	.555 (.50)	.534 (.50)	.455 (.50)	.66 (.48)	.57 (.50)	0	1
Reward	.594 (.49)	.483 (.50)	.605 (.49)	-	-	0	1

Table B.3: Rates of principals' decisions to award a bonus (*reward decision*) and principals' attribution of good/bad outcomes to effort (*attribution decision*). We do not observe reward decisions in the NON-STRATEGIC and NON-STRATEGIC/NON-IDENTITY treatments. The reward rate of non-incentivizing principals would be the average over the reward rate of those who always award a bonus (rate of 1) and those who never do so (rate of 0) and is therefore omitted.

Treatment	**Incentivizer**	**Outcome**	**Match**	***rewarded***	***attribution***
STRATEGIC	Yes	Bad	In-group	0.220 (0.415)	0.505 (0.501)
			Out-group	0.165 (0.372)	0.585 (0.494)
		Good	In-group	0.770 (0.422)	0.561 (0.497)
			Out-group	0.691 (0.463)	0.426 (0.496)
	No	-	In-group	-	0.618 (0.488)
		-	Out-group	-	0.573 (0.497)
NON-STRATEGIC	-	Bad	In-group	-	0.803 (0.497)
	-		Out-group	-	0.794 (0.401)
	-	Good	In-group	-	0.478 (0.408)
	-		Out-group	-	0.507 (0.503)
NON-IDENTITY	Yes	Bad	-	0.165 (0.373)	0.441 (0.498)
		Good	-	0.742 (0.440)	0.484 (0.502)
	No	-	-	-	0.450 (0.499)
NON-STRATEGIC /NON-IDENTITY	-	Bad	-	-	0.767 (0.427)
	-	Good	-	-	0.694 (0.466)

Table B.4: Means of agents' effort choices (*effort decision*) and agents' *expected demand belief.*

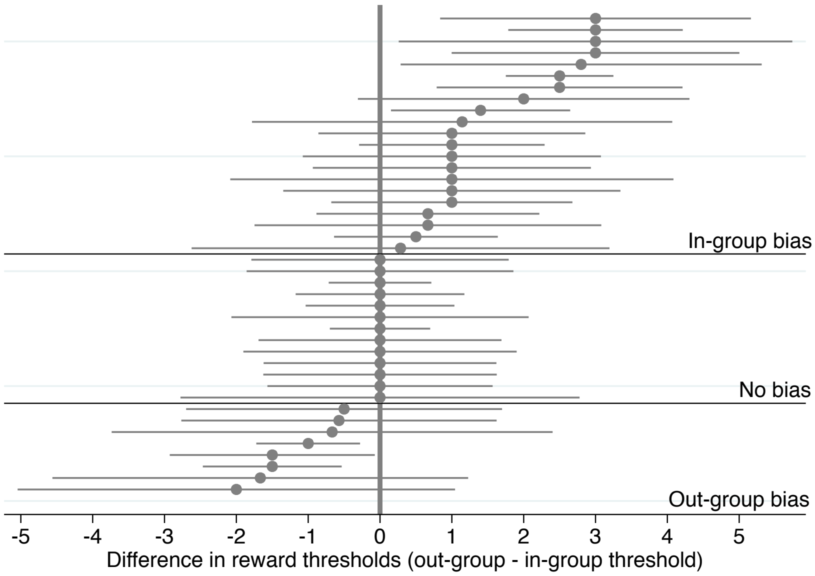
Treatment	**Match**	***type***	***effort***	***expected demand***
STRATEGIC	In-group	1	1.99 (0.100)	3.38 (1.21)
		2	1.74 (0.089)	
		3	1.62 (0.095)	
	Out-group	1	1.91 (0.097)	3.48 (1.31)
		2	1.73 (0.097)	
		3	1.58 (0.086)	
NON-STRATEGIC	In-group	1	2.69 (0.094)	
		2	2.05 (0.119)	
		3	1.61 (0.180)	
	Out-group	1	2.74 (0.092)	
		2	2.12 (0.116)	
		3	1.74 (0.154)	
NON-IDENTITY	-	1	2.05 (0.153)	3.77 (1.26)
	-	2	1.74 (0.151)	
	-	3	1.48 (0.162)	
NON-STRATEGIC **/NON-IDENTITY**	-	1	2.72 (0.)	
	-	2	2.09 (0.)	
	-	3	1.34 (0.)	

B.3 Robustness and further statistical analysis

B.3.1 Principals' choices and beliefs

We estimate the incentivizing principals' individual reward thresholds from choices in 20 rounds by each principal. Apart from the overall tendency towards in-group biased reward choices, we find that, at the individual-level, more principals can be characterized by a reward threshold that is significantly different from zero in the direction of in-group bias. In particular, 8 principals feature significant positive in-group bias in reward thresholds (see Figure B.1) while only 3 principals show significant negative in-group bias in reward thresholds.

Figure B.1: Distribution of incentivizing principals' reward thresholds in the STRATEGIC treatment. 95% confidence intervals estimated by an individual-level bootstrap of the threshold computation procedure as explained in Section 5.1.



50% of incentivizing principals are classified by our measure as demanding to see lower outcomes from the in-group than from the out-group agents (henceforth: as being "in-group biased in rewards"), while a significantly smaller share, 20%, are classified as demanding the reverse. We can reject the hypotheses that the prevalence of in-group biased principals in our sample is driven by chance; and that the distribution of statistically significant individual-level in-group bias is a chance overestimate, but cannot reject the corresponding hypothesis with respect to out-group bias.
Further, at the aggregate level, the distribution of in-group bias in reward thresholds is significantly skewed towards positive bias; the appropriate one-sample t-test ($p < .01$), sign-test ($p = .01$), and sign rank-test ($p = .01$) all support the interpretation that incentivizing principals are more likely to be significantly in-group bias in reward thresholds than not biased at all or biased towards the out-group. Running a randomization test that generates 10000 samples of 20 reward decisions of the 42 incentivizing principals (where reward decisions are modelled according to the regression estimates in for the STRATEGIC treatment shown in Table B.5), yields that we should expect at most 26 principals with positive in-group bias and no fewer than 8 principals with negative in-group bias in a world where reward choices are random and not contingent on group identity. More precisely, only in 1 out of 10000 random samples do we find that the number of out-group biased principals is 8 or fewer and, at the same time, the number out in-group bias principals is at least 21. Formulating a hypothesis test based on the implied simulated null distribution, we find that the observed pattern of in-group bias in reward thresholds cannot have occurred by chance ($p < .01$).

Table B.5: Principals' reward decisions regressed on outcome, in-group status of the matched agent, the interactions of those variables, and round of play in the STRATEGIC treatment.

VARIABLES	
outcome	0.30 (0.0)***
in-group	-0.02 (0.39)
in-group × *outcome*	0.13 (0.10)
round	-0.05 (0.01)***
constant	-0.63 (0.30)*
AIC	1436.70
BIC	1461.71
Log Likelihood	-713.35
Deviance	1426.70
Observations	1100
Subjects	55

Standard errors clustered by subject in parentheses

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

How should we think about a principal's bias more generally *across outcomes*? We cannot measure it for the incentivizing principals by simply comparing average attribution choices above and below the good outcome thresholds in in-group and out-group matches because the sets of good outcomes tend to be different in those matches (In contrast, because those sets are the same for the non-incentivizing principals, that comparison is the right measure of their group-specific bias.)

An incentivizing principal who is willing to reward in-group agents for lower outcomes than out-group agents may appear to be more likely to attribute good outcomes to effort in the in-group than in the out-group matches but may, in fact, be group-neutral in attribution at a fixed level of outcome. Avoiding the confounding effects of differences in good outcome thresholds by measuring in-group bias in attribution at the level of the individual principal would be problematic because, for a particular subject, a given good outcome in the in-group matches may not have an equivalent outcome in the out-group matches, and certainly does not have an equivalent bad outcome.

We get around this problem by making inferences based on the behavior in in-group and out-group matches of comparable principals, pooling together principals who show similar biases in reward decisions. We estimate principals' attribution choices in a regression framework to assess the robustness of results on principals' in-group bias in attribution in relation to their in-group bias in rewards at a given level of outcome.

Table B.6: Logistic regression of incentivizing principals' attribution decision on covariates in the STRATEGIC treatment.

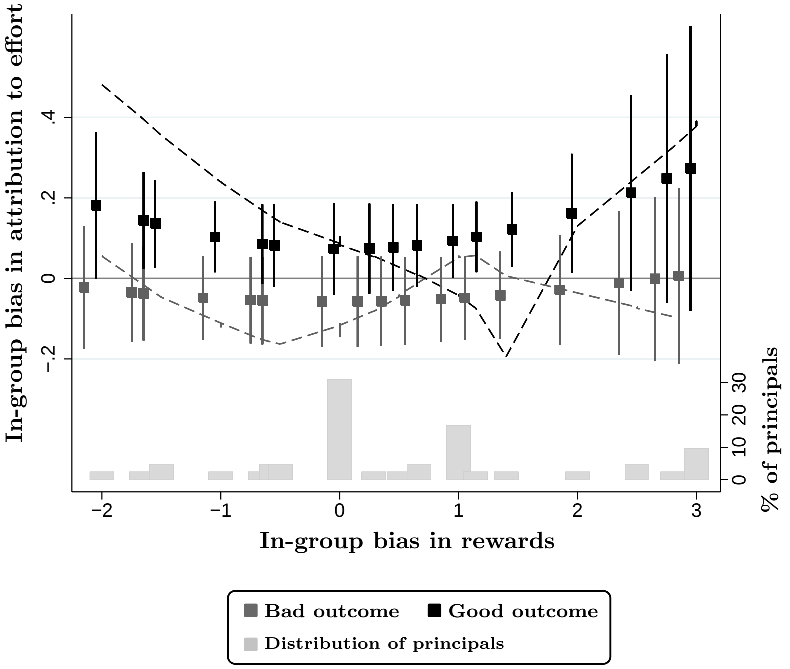
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
good outcome	-0.21 (0.14)	-0.22 (0.14)	-0.67 (0.21)**	-0.59 (0.21)**	-0.53 (0.22)*	-0.38 (0.25)	-0.33 (0.27)
in-group		0.08 (0.14)	-0.31 (0.19)	-0.25 (0.20)	-0.19 (0.21)	-0.23 (0.24)	-0.23 (0.24)
good.outcome × *in-group*			0.85 (0.28)**	0.71 (0.30)*	0.62 (0.31)*	0.53 (0.35)	0.53 (0.35)
in-group bias in rewards				0.09 (0.06)	0.15 (0.10)		
good outcome × *in-group bias in rewards*					-0.18 (0.18)		
in-group × *in-group bias in rewards*					-0.11 (0.16)		
good outcome × in-group × *in-group bias in rewards*					0.24 (0.24)		
in-group bias in rewards^2						0.10 (0.04)*	0.10 (0.04)*
good outcome × *in-group bias in rewards*^{2}						-0.12 (0.09)	-0.12 (0.09)
in-group × *in-group bias in rewards*^{2}						0.03 (0.09)	0.03 (0.09)
good outcome × in-group × *in-group bias in rewards*^{2}						0.07 (0.13)	0.07 (0.13)
outcome							-0.03 (0.07)
round	-0.04 (0.01)**	-0.04 (0.01)**	-0.03 (0.01)**	-0.03 (0.01)**	-0.03 (0.01)**	-0.04 (0.01)**	-0.04 (0.01)**
constant	0.59 (0.16)***	0.55 (0.18)**	0.72 (0.19)***	0.64 (0.19)***	0.59 (0.20)**	0.47 (0.21)*	0.56 (0.29)
AIC	1157.55	1159.24	1152.27	1151.57	1156.43	1146.35	1148.15
BIC	1171.75	1178.17	1175.94	1179.98	1199.03	1188.95	1195.48
Log Likelihood	-575.77	-575.62	-571.13	-569.79	-569.22	-564.18	-564.07
Deviance	1151.55	1151.24	1142.27	1139.57	1138.43	1128.35	1128.15
Observations	840	840	840	840	840	840	840
Subjects	42	42	42	42	42	42	42

Standard errors clustered by subject in parentheses

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

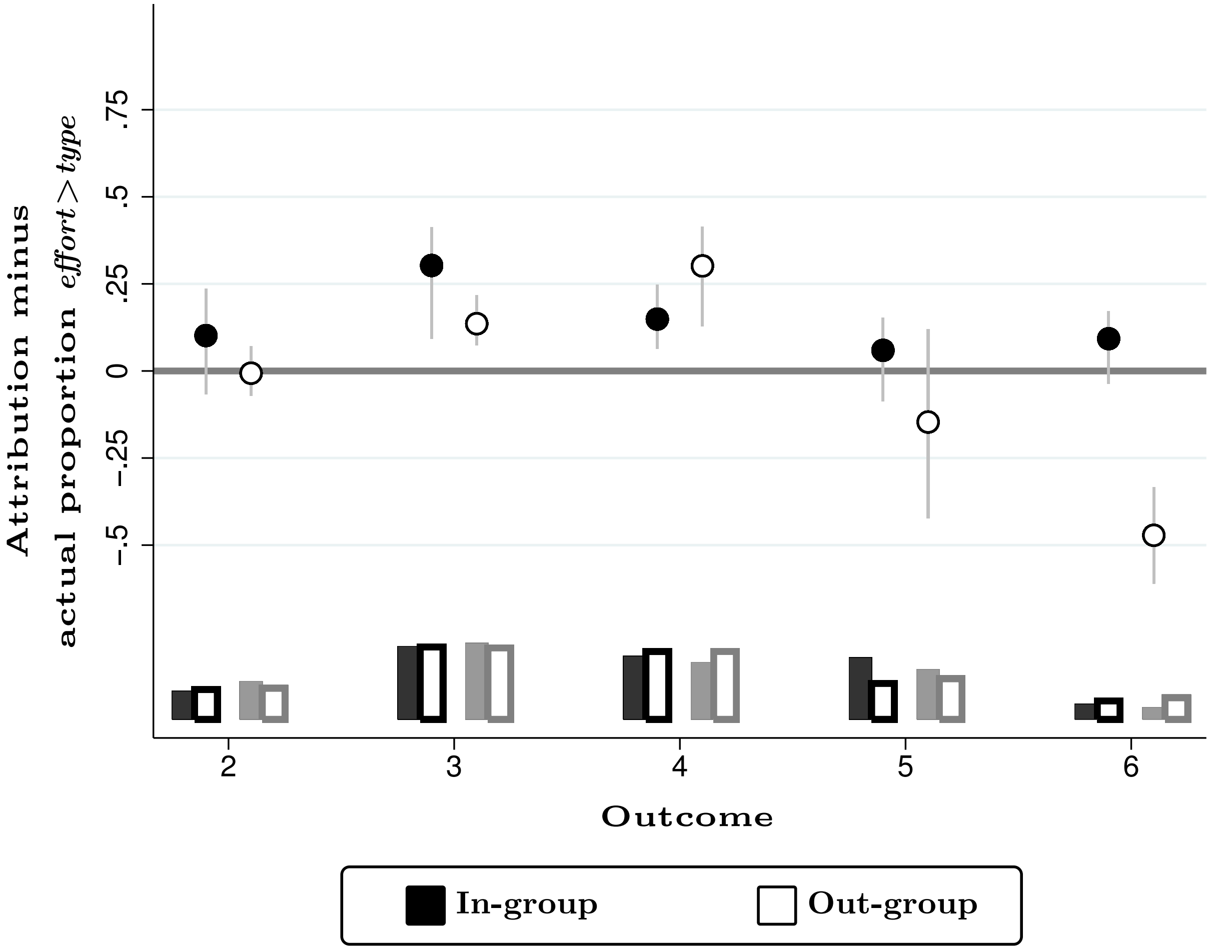
To this end, Figure B.2 is based on the results from a regression of attribution to effort on in-group status of the matched agent, whether an outcome is below or above the threshold (whether an outcome is a bad or good outcome) for each level of in-group bias in rewards, the particular outcome observed, as well as covariates. Based on the regression estimates we generate the marginal effect of in-group vs. out-group status of the agent on attribution (= in-group bias in attribution) of good and bad outcomes over principals in-group bias in rewards (markers). We also superimpose a curve of lowess estimates of the directly observed average of in-group bias in attribution for each level of principals' in-group bias in rewards for good and bad outcomes (dashed lines). Estimates are taken from Model 7 for incentivizing principals in Table B.6. Informed by an U-shaped curve drawn by the loess estimator of average in-group bias in attribution, we fit a model that includes the square of in-group bias in rewards.

Figure B.2: Average difference in the rates of attribution to effort between in-group and out-group matches (= in-group bias in attribution) over in-group bias in rewards of incentivizing principals in the STRATEGIC treatment. The loess curve of in-group bias in attribution is fitted at a given pair of above/below the threshold and in-group bias in rewards.



We also asked whether principals are right *within* identity matches? Here, we consider behavior within counter-factual principal-agent pairs that match on the actual (for the principals) and the expected (by the agents) reward threshold outcomes. At each level of outcome in a distinct identity match condition, we record the principal's *correctness in attribution within identity match.* Holding fixed the outcome, this quantity measures the difference in the proportion of observations for which the agents' effort levels were larger than their type and the proportion of observations where principals' correctly attribute those outcomes to effort. Figure B.3 plots the correctness measure where the value of 0 on the y-axes corresponds to the principals' always correctly guessing the ordering of type and effort for the given outcome levels. We show negative deviations (underestimation of agents' effort relative to type) and positive deviations (overestimation) from a correct guess.

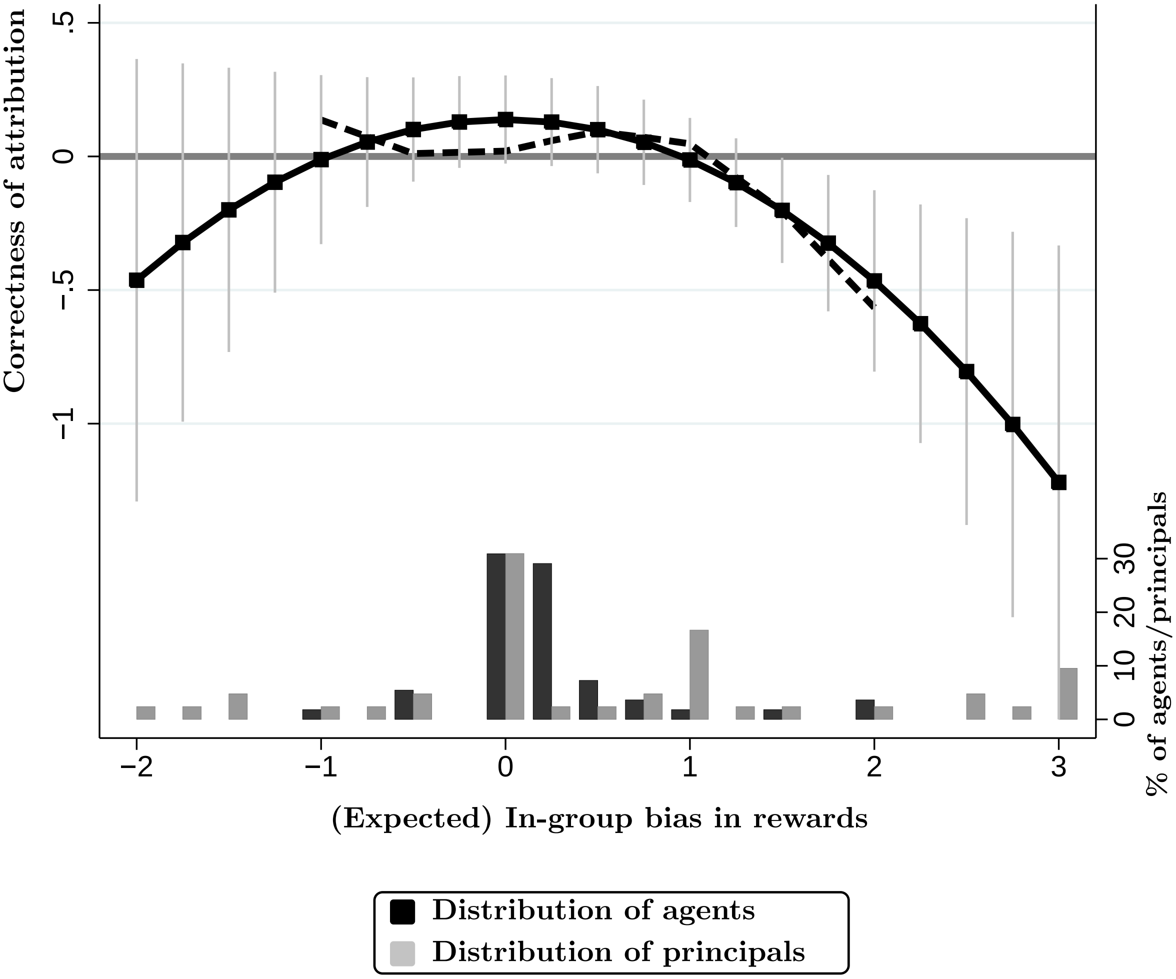
Figure B.3: *Correctness in attribution within identity match* at each level of outcome for the counterfactually matched incentivizing principals and agents with the corresponding expectations about in-group bias in rewards. Results are shown for agents/principals who demonstrate expected/actual in-group bias in rewards; the unit of this analysis is groups of agents/principals who show similar levels of (expected) in-group bias (values within the interval of 1 on the (expected) in-group bias in rewards-scale); 95% confidence bounds based on a subject-level clustered bootstrap.



Note, for pairs with (expected) in-group bias, the principal's attribution choice is closer to correct in in-group rather than out-group matches; the most systematic attribution mistakes are due to the under-attribution to effort in the higher than average range. Relating this back to our motivating example, Bob's interpretation of Alice's performance tends to under-appreciate Alice's effort. Relating to the discussion of the two effects on agents' choices we saw in the previous section, we may say that principals focus on the expected bias effect, and under-appreciate the implications of expected in-group bias in rewards on the manifestation of the expected demand effect.

Then, are principals right *across* identity matches? The values in the figure correspond to pairs of incentivizing principals and agents, matching principals' in-group bias in rewards and agents' expectations of in-group bias. The distance from zero on the vertical axis gives a measure of *correctness of attribution across identity matches.* It is computed as the difference between (1) the average difference between attribution to effort in in-group and out-group matches at a given outcome and (2) the difference between the proportions of observations with effort greater than type in in-group and out-group matches.

Figure B.4: *Correctness of attribution across identity matches* over in-group bias in rewards. Markers give the predicted correctness of attribution estimated from a regression of correctness of attribution on (expected) in-group bias in rewards and its squared value. The dashed line gives the lowess estimate from the raw value of correctness of attribution. The unit of analysis pairs matched groups of agents/principals who show similar levels of (expected) in-group bias (values within the interval of 1 on the (expected) in-group bias in reward scale); 95% confidence bounds based on a subject-level clustered bootstrap.



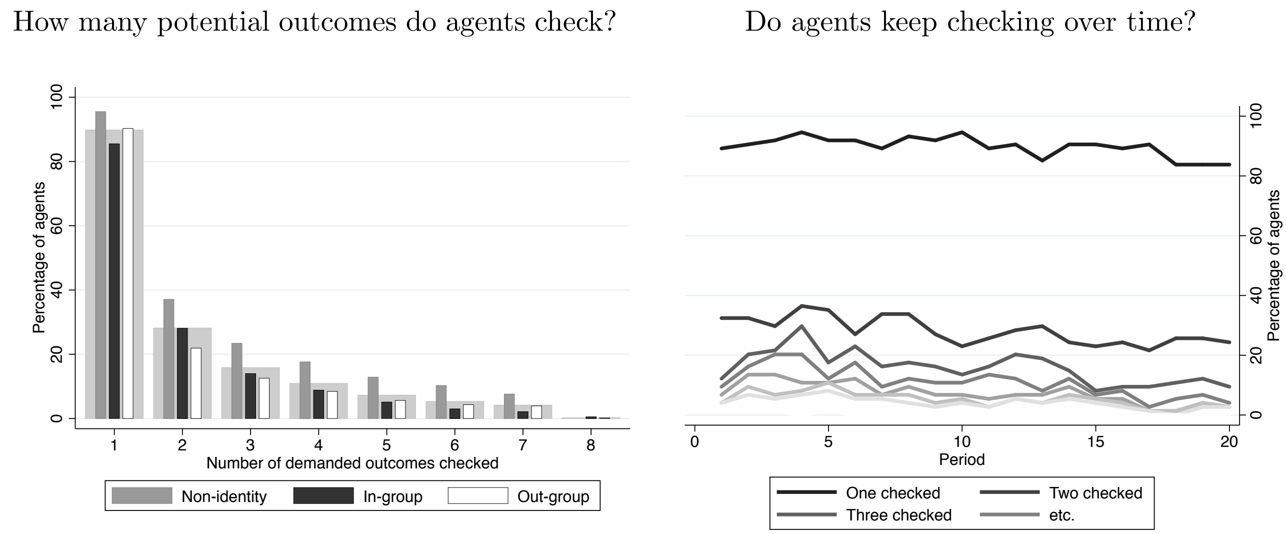
The evidence in the figure reinforces our interpretation. Principals who are in-group biased in rewards display the largest deviation from a correct guess about agents' in-group bias in effort, consistent with our conjecture of their failure to anticipate correctly the strength of the expected demand effect on out-group agents.

B.3.2 Agents' choices and beliefs

90% of agents check at least one minimal outcome they expected to be demanded by their matched principals; the willingness to check stays constant throughout all 20 periods of the experiment. 28%

of agents also investigate the payoff consequences of a second minimal outcome demanded and 16% a third value. In the modal case – in 26% of the agent-rounds – agents obtain information about payoffs for a minimally required outcome of 4, the next highest-frequency outcome value checked is 3 (22%). The distribution of checked outcomes is approximately normal, centered around 4.
Subjects in the role of an agent do not simply click through all potential outcomes. Most of them only check outcomes from the middle of the outcome range and tend to do so only once. If agents had clicked through all possible values of outcome, we would not be able to claim confidently they were checking the expected outcome that is most reasonable to them, given their match. Since agents are very specific in their expectation of the payoff information they want to obtain, and their behavior with respect to which expected outcome they check to obtain their potential payoffs does not change over the course of the experiment, their choices here indicate a targeted and reasoned attempt to learn payoffs at the expected outcome threshold. In short, agents' outcome-checking choices appear to elicit what they believe is the outcome principals are most likely to demand in order to reward.
Defining this measure as only the first click by an agent does not change the results of our analysis.

Figure B.5: Agents' inquiries of payoff consequences of expected demanded minimal outcomes


How many potential outcomes do agents check?

Do agents keep checking over time?

Table B.7: Agents' effort regressed on a treatment-dummy (STRATEGIC serves as base category), agent's type, in-group status of the matched principal (STRATEGIC vs NON-STRATEGIC model), the interactions of those variables, and round of play.

VARIABLES		
treatment	1.13	(0.20)***
type	-0.16	(0.04)***
treatment × *type*	-0.34	(0.10)**
in-group	0.11	(0.13)
treatment × *in-group*	-0.10	(0.19)
type × *in-group*	-0.03	(0.06)
treatment × *type* × *in-group*	-0.01	(0.09)
round	-0.01	(0.00)
constant	2.11	(0.12)***
R^2	0.14	
Observations	1500	
Subjects	75	
RMSE	0.74	

Standard errors clustered by subject

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table B.8: Regression of agents' effort on covariates for the STRATEGIC treatment. *Risk-aversion* is measured by the number of *safe choices* made in a (Holt and Laury, 2002)-list; 4 out of 55 agents with inconsistent choices moving through the list – switching back and forth between safe and risky option are excluded.

VARIABLES	(1)	(2)	(3)	(4)
in-group	0.07	-0.33	-0.09	-1.10
	(0.138)	(0.308)	(0.290)	(0.977)
type	-0.16***	-0.08	-0.07	-0.32
	(0.046)	(0.137)	(0.144)	(0.455)
expected demand		0.20**	0.29***	-0.27
		(0.075)	(0.083)	(0.199)
expected in-group bias			0.10	-0.57
			(0.215)	(0.952)
type × *expected demand*		-0.03	-0.04	0.06
		(0.035)	(0.036)	(0.105)
expected demand × *expected in-group bias*			-0.11**	0.16
			(0.048)	(0.262)
in-group × *expected demand*		0.17*	0.07	0.34
		(0.096)	(0.085)	(0.264)
in-group × *type*	-0.02	0.03	0.02	0.28
	(0.059)	(0.145)	(0.146)	(0.432)
in-group × *expected in-group bias*			-0.17	0.46
			(0.175)	(0.846)
in-group × *expected demand* × *type*		-0.03	-0.03	-0.10
		(0.042)	(0.043)	(0.115)
in-group × *expected demand* × *expected in-group bias*			0.12*	-0.22
			(0.067)	(0.273)
risk aversion				-0.31*
				(0.179)
in-group × *risk aversion*				0.25
				(0.179)
type × *risk aversion*				0.06
				(0.098)
expected demand × *risk aversion*				0.12***
				(0.041)
expected in-group bias × *risk aversion*				0.13
				(0.215)
type × *expected demand* × *risk aversion*				-0.02
				(0.023)
expected demand × *expected in-group bias* × *risk aversion*				-0.06
				(0.061)
in-group × *expected demand* × *risk aversion*				-0.07
				(0.048)
in-group × *type* × *risk aversion*				-0.07
				(0.078)
in-group × *expected in-group bias* × *risk aversion*				-0.15
				(0.169)
in-group × *expected demand* × *type* × *risk aversion*				0.02
				(0.021)
in-group × *expected demand* × *expected in-group bias* × *risk aversion*				0.09
				(0.059)
round	-0.00	-0.00	-0.00	-0.00
	(0.004)	(0.005)	(0.005)	(0.005)
Constant	2.11***	1.45***	1.21***	2.73***
	(0.131)	(0.258)	(0.263)	(0.877)
R-squared	0.033	0.152	0.180	0.240
Observations	1,020	949	949	949
Subjects	55	51	51	51

Standard errors clustered by subject

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

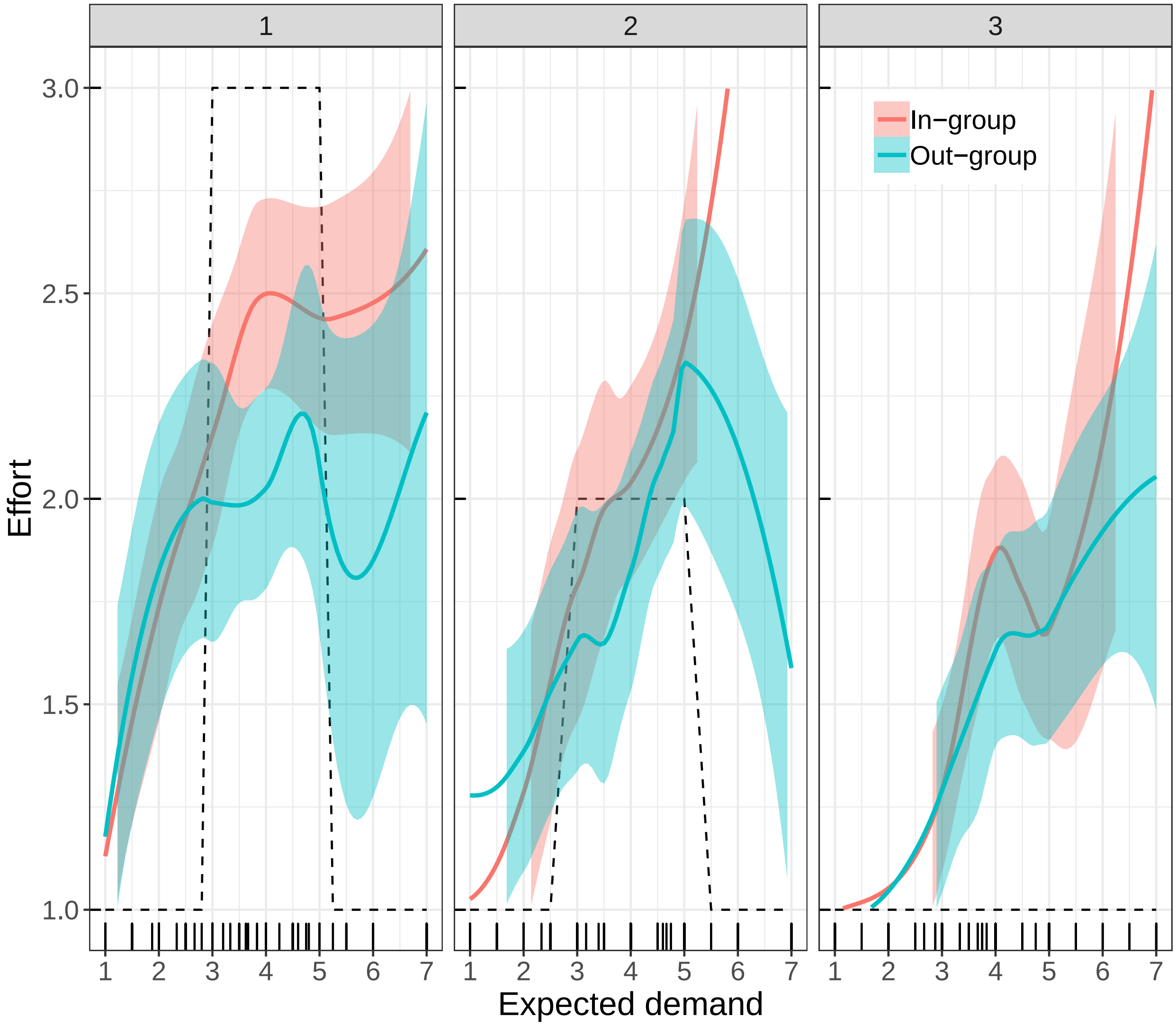
Table B.9: Marginal effects of **expected demand** on effort computed from a local regression equivalent to the regression specification reported in Table B.8 on ranges of expected bias (expected bias rounded to the nearest .5) for which we have enough observations.

Expected bias	AME	SE	z	p	lower	upper
-.5	0.0143	0.1911	0.0748	0.9403	-0.3602	0.3888
0	0.1945	0.0579	3.3609	0.0008	0.0811	0.3079
.5	0.2267	0.1007	2.2504	0.0244	0.0293	0.4241
1	0.0421	0.0629	0.6704	0.5026	-0.0811	0.1653

Table B.10: Marginal effects of **expected bias** on effort computed from a local regression equivalent to the regression specification reported in Table B.8 on ranges of expected demand (expected demand rounded to the nearest integer) for which we have enough observations.

Expected demand	AME	SE	z	p	lower	upper
2	-0.1838	0.3126	-0.5879	0.5566	-0.7964	0.4288
3	-0.0589	0.1845	-0.3191	0.7496	-0.4205	0.3028
4	0.1086	0.1997	0.5438	0.5866	-0.2829	0.5001
5	0.1453	0.1970	0.7376	0.4608	-0.2408	0.5315
6	2.4077	2.3350	1.0311	0.3025	-2.1688	6.9842

Figure B.6: Loess smoothed curve of effort over expected demand for agents who expect in-group bias in rewards of their matched principals plotted by in-group status.



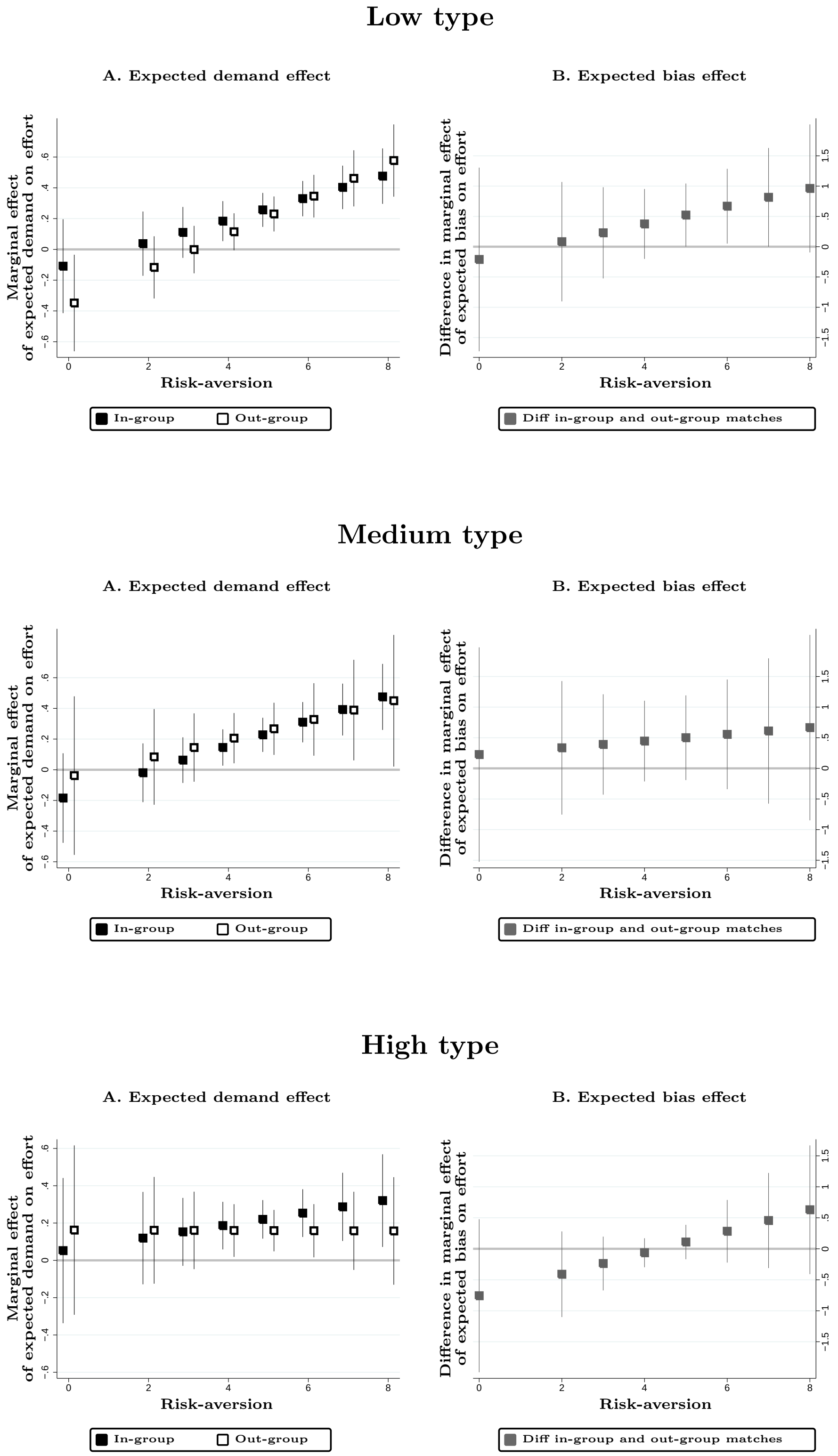
B.3.3 Agents' choices and risk preferences in the STRATEGIC treatment

Agents' choices are a comparison of a sure value (cost of investment) to an expected value of a lottery (outcome and reward, contingent on realizations of random variables). It would be reasonable to suppose that in making such choice, subjects will respond to the explicitly given payoffs in ways that track their personal unmodeled risk preferences, which would induce a variation in effort choice where our prediction of agent choices for a given type t — in particular, for the OCP equilibria — expects no variation relative to the differences in the expected retention threshold $z \in \{3, 4, 5\}$. However, we find no significant relationship between the agents' risk preference and their expected demand in either in- or out-group matches; the marginal effect of risk aversion on expected demand in both is not systematically different from zero $(-.01\,(-.14, .11)$ and $.07\,(-.07, .20)$, respectively). If agents' risk preferences have an effect on their behavior it is on the effort choices they make, not on their beliefs about the principals.

Indeed, we find that agents' risk preferences importantly condition how expected demands and ex-

pected in-group bias in principals' reward decisions relate to effort. The marginal effect of expected demands on effort increases systematically with agents' risk aversion; while the marginal effect of expected demand on effort is not significantly different from zero for risk-seeking subjects (3 safe choices or less), they are for subjects considered risk neutral (4 safe choices) or risk averse (more than 4 safe choices. Similarly, the expected bias effect, the marginal effect of agents' expectation of principals' bias in reward decisions also grows stronger with risk aversion: the difference in marginal effect of expected bias on effort is systematically larger in in- than out-group matches and increasingly so for more risk-averse agents, while there is no such difference for more risk-accepting agents. Marginal effects relating to the analysis of the effect of risk aversion on effort are estimated from the regression reported in Table B.8; also see Figure B.8
A plausible, if speculative, way of understanding the behavioral motivations behind this result is by conceiving of the agents as viewing the bonus as a reference payoff (Kahneman and Tversky, 1979) and seeking to insure themselves against losing it with investment into effort (Kőszegi and Rabin, 2007). Consistent with this interpretation, when the agents anticipate higher outcome demands, the more risk-averse among them react more strongly by investing more on the margin to meet those demands. However, if that payoff is too distant – too risky – the insurance premium may become too expensive to be worth purchasing, and so we should see the more risk-averse agents losing interest in it faster. Perhaps, the status of the bonus as a reference payoff itself becomes for the risk-averse agents less plausible when the risks associated with it become too great. Put somewhat differently, those agents for whom the gap between in-group and out-group expectations is high tend to regard the bonus in the out-group matches as a particularly distant prospect, accounting for the relationship we reported above. If this interpretation is right, the patterns reported should be most pronounced when agents have lower types. Indeed, that is the case: the effect of risk preference on the expected demand effect and on the expected bias effect is highest for type 1 agents. See Figure B.7.

Figure B.7: Marginal effect of expected demands on effort (= expected demand effect, Panel A) and difference in marginal effect of expected in-group bias on effort (= difference in expected bias effect, Panel B) over risk-aversion (number of safe choices in the Holt and Laury (2002)-list) by type. Marginal effects are estimated from Model 4 in B.8.

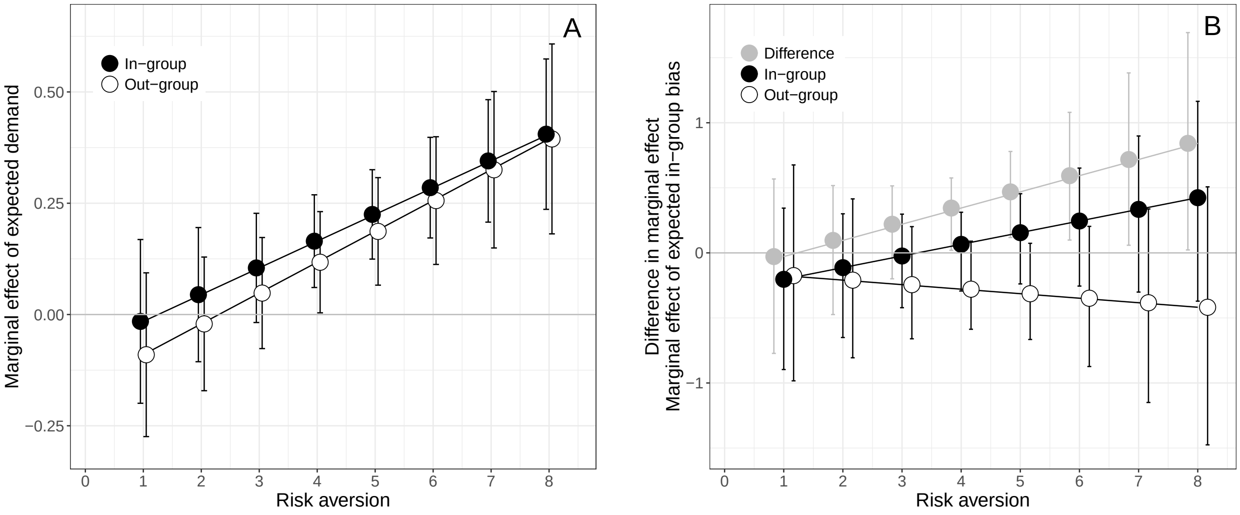


Low type

Medium type

High type

Figure B.8: Marginal effect of expected demands on effort (= expected demand effect, Panel A) as well as marginal effect (= expected bias effect, Panel B) by in-group status plotted over risk-aversion. We also show the difference in marginal effect of expected in-group bias on effort (Panel B). The effects are estimated from Model 4 in Table B.8



B.3.4 Average treatment effects: NON-STRATEGIC treatment

Table B.11: Logistic regression of attribution decision on indicators of treatment status, being classified as non-incentivizing principals, in-group status, and high (good) outcome as well as the interactions of those variables and round of play. For non-incentivizing principals and principals in the NON-STRATEGIC treatment, high outcomes are defined as those that are above > 4, in contrast to low outcomes (< 4. For incentivizing principals, good outcomes are defined as those that are above the principals individual reward threshold as defined in Section 5.1.

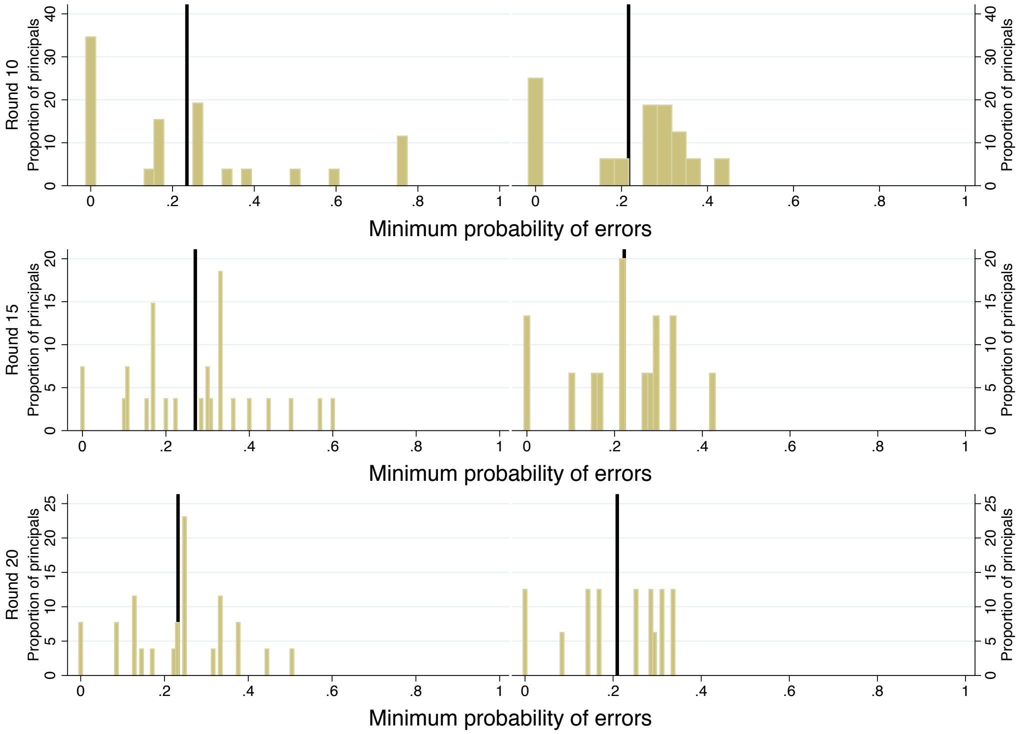
VARIABLES	
non-incentivizing	-0.43
	(0.398)
NON-STRATEGIC	0.87*
	(0.524)
in-group	-0.32
	(0.196)
non-incentivizing × *in-group*	0.89*
	(0.516)
NON-STRATEGIC × *in-group*	1.40*
	(0.740)
high (good) outcome	-0.66**
	(0.298)
non-incentivizing × *high (good) outcome*	1.69***
	(0.629)
NON-STRATEGIC × *high (good) outcome*	-1.16*
	(0.659)
in-group × *high (good) outcome*	0.85***
	(0.294)
non-incentivizing × *in-group* × *high (good) outcome*	-1.95**
	(0.819)
NON-STRATEGIC × *in-group* × *high (good) outcome*	-1.57*
	(0.833)
round	-0.02**
	(0.012)
Constant	0.60**
	(0.243)
Observations	1,229

Standard errors clustered by subject in parentheses
*** p<0.01, ** p<0.05, * p<0.1

B.3.5 History of play

Principals' reward and attribution decisions As expected, with increasing number of rounds played, the threshold in the outcome space above which incentivizing principals are willing to reward the agent improves with respect to minimizing committed categorization errors. Figure B.9 shows a decrease in the spread of the probability of errors associated with the error minimizing threshold computed for each principal (while the mean remains constant); in other words, the computation of principals' thresholds becomes more precise with round of play. There is an element of noise we seem unable to pick up with our definition of each individual principals' threshold; the categorization error associated with the threshold that minimizes errors lingers around a probability of .2 of committing a categorization error.

Figure B.9: Distribution of the probability of an error in categorizing reward decisions associated with the principal's reward threshold (= error minimizing threshold above which incentivizing principals are willing to reward the agent).



Looking at principals reward decisions in the aggregate, we do not find a relationship of experience of favourable treatment in general and in in-group and out-group in particular; we express favourable past experience in current round t as the average outcome in round 1 to $t-1$. Table B.12 shows no significant effect of experience on current reward choices; here we model reward decisions as a function of outcome, favourable past experience (overall and separated by in- and out-group), the in-group status of the matched agent (applicable in the comparison STRATEGIC and NON-STRATEGIC treatment), the interaction of those variables, and round of play.

Table B.12: Logistic regression of principals' reward decisions on outcome on average of past outcomes in the in- and out-group.

VARIABLES	
outcome	0.31***
	(0.085)
in-group	-0.38
	(1.702)
outcomes in the past in the in-group	0.19
	(0.308)
outcomes in the past in the out-group	-0.11
	(0.221)
in-group ×*outcome*	0.10
	(0.100)
in-group ×*outcomes in the past in the in-group*	0.06
	(0.275)
in-group ×*outcomes in the past in the out-group*	0.07
	(0.280)
Constant	-1.72
	(1.497)
Observations	1,083

Standard errors clustered by subject in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Modeling the attribution decisions of incentivizing principals as a function of outcome, the in-group status of the matched agent, and, similar to above, past outcome experience, shows that there is also no effect of a history of favourable experience with any agent, in-group agents, or out-group agents on the decision whether to attribute outcomes to effort. For our argument of the existence of strategic discrimination, behavior among incentivizing principals in the STRATEGIC treatment, because they accept to act in a strategic environment, and comparing those to principals in the NON-STRATEGIC treatment is the relevant counterfactual; Model (2) and (4) in Table B.13 gives the regression results for this comparison.

Table B.13: Logistic regression of principals' attribution decisions on outcome, in-group status of the matched principal, average of past outcomes in STRATEGIC, NON-STRATEGIC treatment, where the treatment-variable takes the STRATEGIC treatment as its base category, and in the STRATEGIC treatment on average of past outcomes in the in- and out-group separately; standard errors are computed based on clustering by subject. Model (2) and (4) exclude non-incentivizing principals in the STRATEGIC treatment from the analysis.

		STRATEGIC and NON-STRATEGIC			
VARIABLES	All treatments	(1)	(2)	(3)	(4)
NON-STRATEGIC	-0.61	1.37	1.38	0.83	0.69
	(1.612)	(2.923)	(2.978)	(2.455)	(2.503)
outcome	-0.02	-0.02	-0.10	-0.02	-0.10
	(0.071)	(0.082)	(0.091)	(0.085)	(0.104)
outcomes in the past	0.01	-0.01	0.06		
	(0.200)	(0.231)	(0.232)		
in-group		-0.16	-1.41	0.87	-0.36
		(1.218)	(1.084)	(1.208)	(1.319)
outcomes in the past in the in-group				-0.02	-0.12
				(0.251)	(0.278)
outcomes in the past in the out-group				0.11	0.23
				(0.156)	(0.137)
in-group × *outcome*		0.01	0.08	0.04	0.13
		(0.112)	(0.124)	(0.110)	(0.128)
in-group × *outcomes in the past*		0.04	0.27		
		(0.297)	(0.267)		
in-group × *outcomes in the past in the in-group*				-0.06	0.02
				(0.223)	(0.253)
in-group × *outcomes in the past in the out-group*				-0.21	-0.08
				(0.187)	(0.225)
NON-STRATEGIC × *outcome*	-0.46**	-0.44**	-0.36*	-0.54**	-0.46**
	(0.193)	(0.197)	(0.201)	(0.224)	(0.232)
NON-STRATEGIC × *outcomes in the past*	0.70*	0.16	0.08		
	(0.365)	(0.599)	(0.605)		
NON-STRATEGIC × *in-group*		-4.25	-3.04	-3.12	-1.93
		(4.072)	(4.060)	(4.126)	(4.192)
NON-STRATEGIC × *outcomes in the past in the in-group*				0.42	0.52
				(0.485)	(0.501)
NON-STRATEGIC × *outcomes in the past in the out-group*				-0.03	-0.15
				(0.317)	(0.309)
NON-STRATEGIC × *in-group* × *outcome*		-0.07	-0.15	-0.09	-0.18
		(0.251)	(0.257)	(0.230)	(0.239)
NON-STRATEGIC × *in-group* × *outcomes in the past*		1.18	0.97		
		(0.899)	(0.896)		
NON-STRATEGIC × *in-group* × *outcomes in the past in the in-group*				-0.25	-0.32
				(0.695)	(0.709)
NON-STRATEGIC × *in-group* × *outcomes in the past in the out-group*				1.18*	1.05
				(0.640)	(0.658)
round	-0.02*	-0.02*	-0.02*	-0.02	-0.02
	(0.010)	(0.011)	(0.012)	(0.014)	(0.016)
Constant	0.38	0.47	0.51	0.06	0.25
	(0.881)	(0.953)	(1.044)	(1.198)	(1.299)
Standard errors clustered by subject in parentheses					
Observations	1,995	1,634	1,330	1,412	1,161

*** p<0.01, ** p<0.05, * p<0.1

Learning, in the shape of forming beliefs about agents' behavior given past experience with agents' performance (outcomes), does not exist once experience with in-group vs out-group agents is introduced into the model (Model 5 and 6).

We do find that our result of an attribution bias for good outcomes by incentivizing principals in the STRATEGIC treatment fully emerges in the second half (round 11 to 20) of the experiment only. Figure B.10 and B.11 replicate Figure 2 from the main text and shows that the observed attribution bias for good outcomes by incentivizing principals exists in direction in the first half but is significantly different from zero only in the second half of the experiment.

Figure B.10: In-group bias in attribution to effort by outcome and treatment in the **first half** of the experiment (round 1 to 10). Gray marker in NON-STRATEGIC panel is conservative estimate.

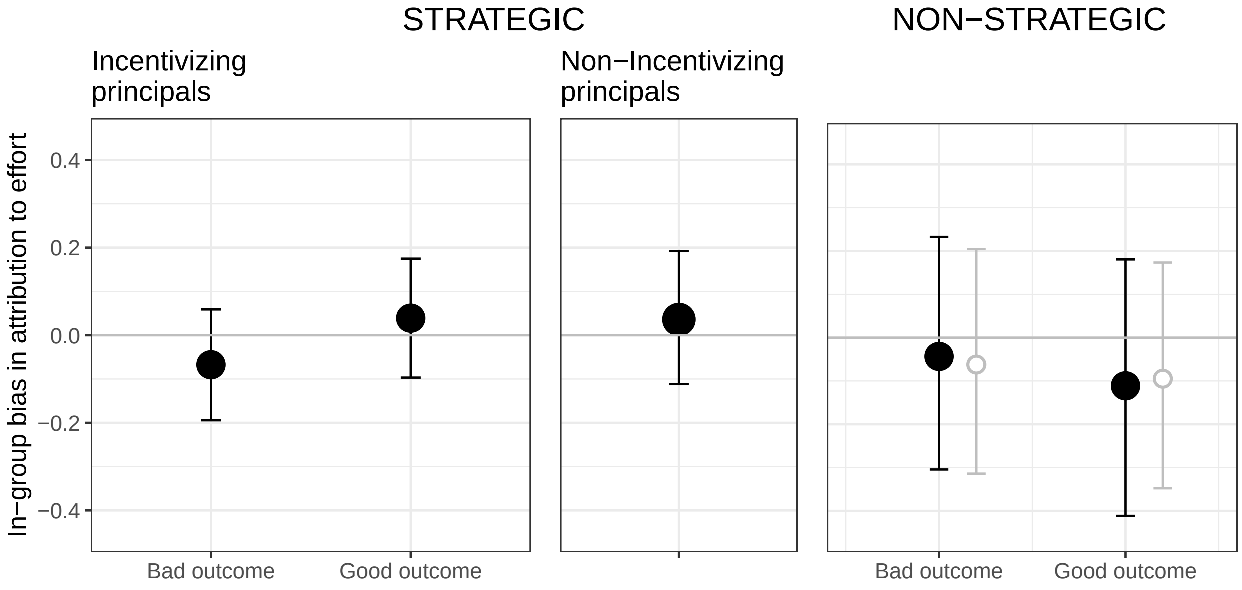
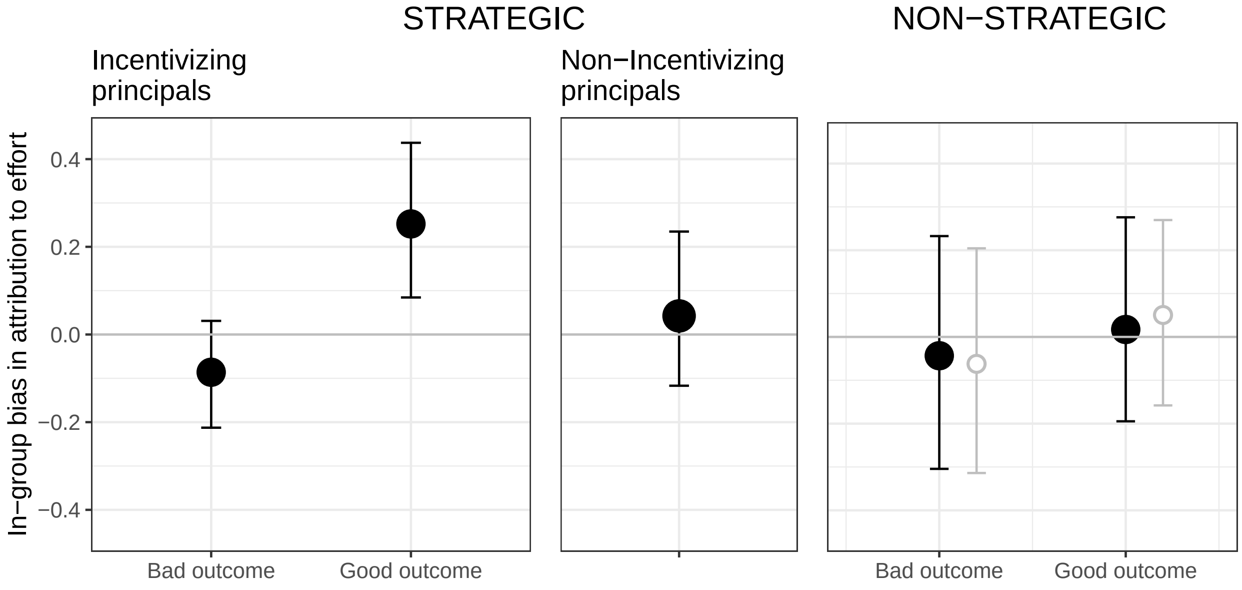


Figure B.11: In-group bias in attribution to effort by outcome and treatment in the **second half** of the experiment (round 11 to 20). Gray marker in NON-STRATEGIC panel is conservative estimate.



Agents' effort decisions and expected demand beliefs Elaborating on the effect of history of play on agents, we see that agents' beliefs do not respond to individual agents' experience with reward decisions of their matched principals. Table B.15 shows no significant effect of the past rate of being rewarded overall, in in-group, or in out-group matches on agents' current expectations of principals' demands. There is, however, an effect of favourable treatment as out-group agent in the past in terms of principals' reward decisions on agent's current effort choice in the Strategic treatment (Table B.14). In particular, in the STRATEGIC treatment, the marginal effect of an increase in the rate of reward in the in-group in past rounds raises effort of agents in in-group matches by $.46\,(.12, .81)$. A rise in receiving a reward in the out-group increases effort in in-group matches $(.41\,(.02, .80))$ and out-group matches $(.44\,(.03, .86))$; marginal effects are estimated from Model (2) in Table B.14). Given that this relationship seems not to be related to a positively updated belief about the likelihood of receiving a reward from principals in the current round, we do not think that this finding takes away from our interpretation of strategic discrimination.

We also do not see different patterns emerging with respect to expected demand and expected bias effect as reported in Figure 3 in the main text (See Figure B.12 and B.13.)

Table B.14: Regression of agents' effort on type, rate rewarded in the past in the in- and out-group separately, and in-group status of the matched principal in the STRATEGIC treatment.

VARIABLES	(1)	(2)
type	-0.19***	-0.18***
	(0.053)	(0.056)
rewarded in the past	0.32	
	(0.257)	
expected demand	0.14**	0.16***
	(0.054)	(0.051)
in-group	-0.20	-0.26
	(0.211)	(0.220)
rewarded in the past in the in-group		0.02
		(0.215)
rewarded in the past in the out-group		0.41**
		(0.199)
in-group × *type*	-0.06	-0.08
	(0.057)	(0.060)
in-group × *rewarded in the past*	0.20	
	(0.233)	
in-group × *expected demand*	0.07	0.05
	(0.058)	(0.052)
in-group × *rewarded in the past in the in-group*		0.45**
		(0.204)
in-group × *rewarded in the past in the out-group*		0.03
		(0.206)
round	0.00	0.00
	(0.006)	(0.006)
Constant	1.41***	1.24***
	(0.233)	(0.218)
Observations	1,164	1,032
R-squared	0.147	0.184

Standard errors clustered by subject in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table B.15: Regression of agents' effort on type, rate rewarded in the past in the in- and out-group separately, and in-group status of the matched principal in the STRATEGIC treatment; standard errors are computed based on clustering by subject.

VARIABLES	(1)	(2)
type	0.23***	0.22***
	(0.077)	(0.083)
rewarded in the past	0.48	
	(0.446)	
in-group	-0.27	-0.30
	(0.266)	(0.278)
rewarded in the past in the in-group		0.74
		(0.481)
rewarded in the past in the out-group		0.01
		(0.411)
in-group × *type*	0.08	0.13
	(0.118)	(0.133)
round	-0.00	0.00
	(0.011)	(0.014)
Constant	2.65***	2.44***
	(0.430)	(0.548)
Observations	1,164	1,032
R-squared	0.034	0.049

Standard errors clustered by subject in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure B.12: Predicted levels of effort plotted over expected in-group bias and expected demands for in-group matches (Panel A) and out-group matches (Panel B) in the **first half** of the experiment (round 1 to 10).

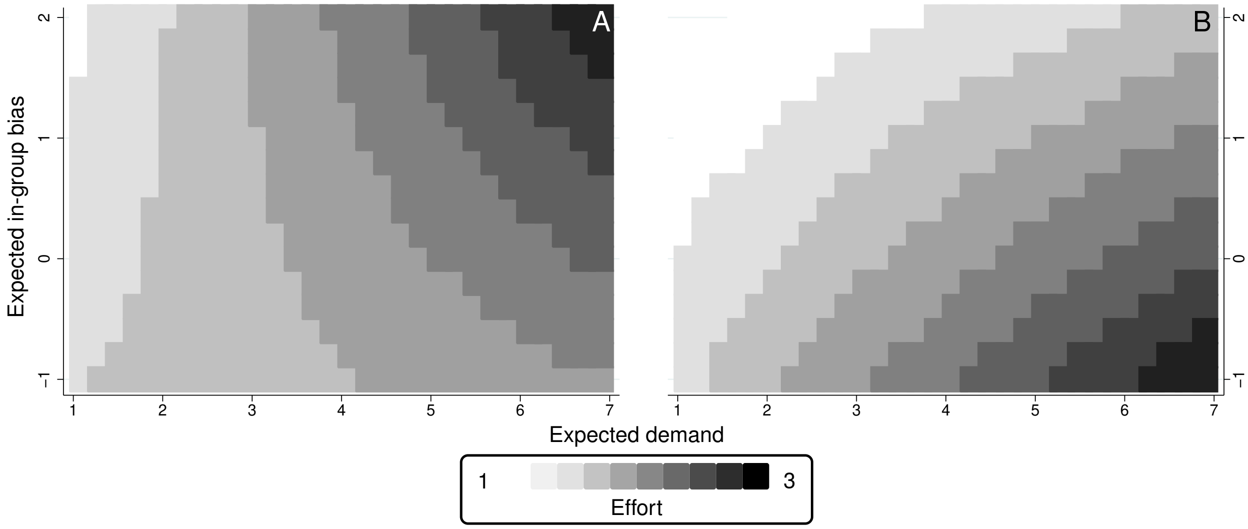
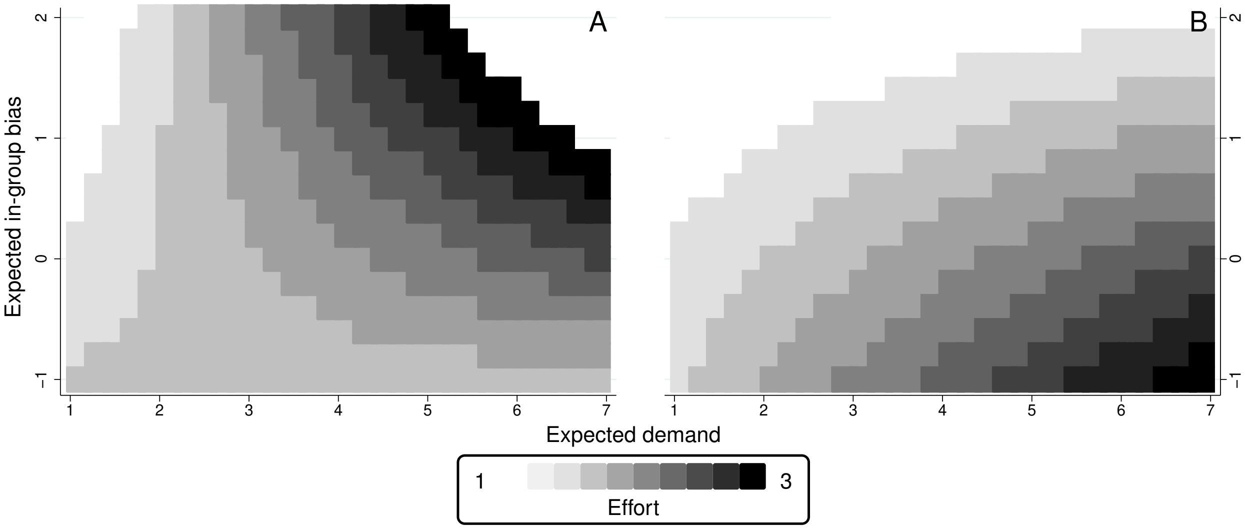


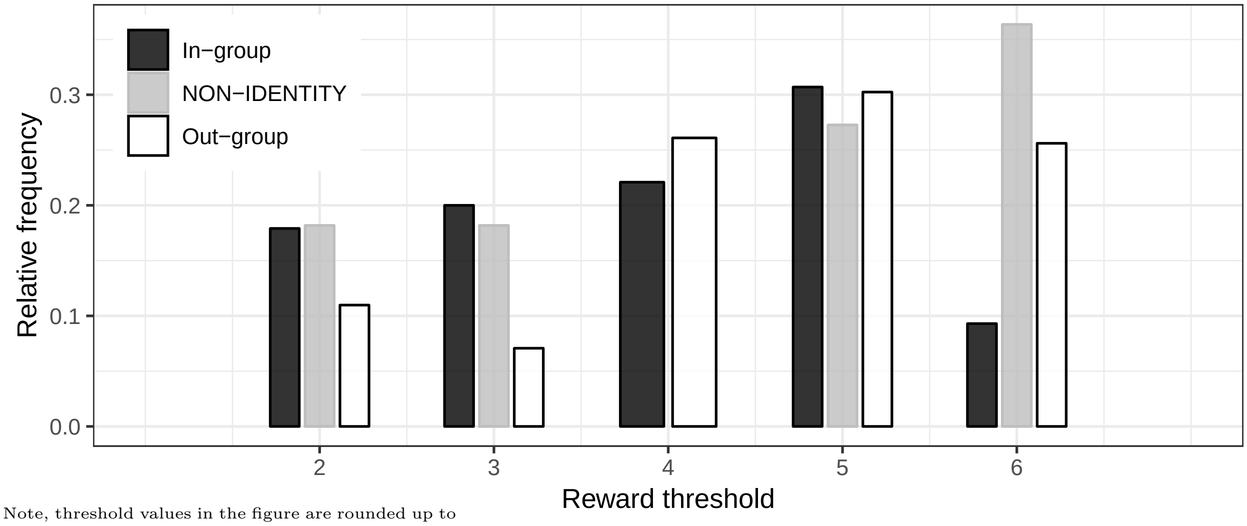
Figure B.13: Predicted levels of effort plotted over expected in-group bias and expected demands for in-group matches (Panel A) and out-group matches (Panel B) in the **second half** of the experiment (round 11 to 20).



B.4 NON-IDENTITY treatment

Incentivizing principals constitute 58% of the principals in the NON-IDENTITY treatment. Principals' reward decisions are significantly increasing in outcome. The marginal effect of outcome on awarding a bonus s $.06\,(.01,.12))$ The average principal-specific outcome threshold is 4.45. Incentivizing principals in the NON-IDENTITY treatment do not attribute to effort differently upon observing good and bad outcomes. The average share of reward decisions incorrectly classified by the error-minimizing threshold is .13 suggesting that principals' reward decisions are largely consistent with their inferred individual thresholds.

Figure B.14: Incentivizing principals' reward thresholds by in-group status and treatment.



Note, threshold values in the figure are rounded up to nearest integer and there were no thresholds at 4 in the NON-IDENTITY treatment.

B.5 Testing the power of incentives across strategic and non-strategic treatments

We compare the marginal effect of agents' type on their effort in STRATEGIC and NON-IDENTITY treatment as well as in NON-STRATEGIC and NON-STRATEGIC/NON-IDENTITY treatment. The marginal effect of type on effort for agents is $-.28\,(-.44, -.13)$ in the NON-IDENTITY treatment, $-.18\,(-.25, -.10)$ in the STRATEGIC treatment, $-.52\,(-.70, -.34)$ in the NON-STRATEGIC treatment, and $-.69\,(-.84, -.54)$ in the NON-STRATEGIC/NON-IDENTITY treatment. Comparing the first two numbers, we observe that as the agent type increases, the effort decreases, as predicted by our theoretical analysis. The similar conclusion holds for the second set of numbers. The effect of identity appears to be to slightly muffle the marginal effects, but the decrease is not statistically significant.

References

Fischbacher, Urs. 2007. “z-Tree: Zurich Toolbox for Ready-made Economic Experiments.” *Experimental Economic* 10(2):171–178.

Holt, Charles and Susan Laury. 2002. “Risk Aversion and Incentive Effects.” *American Economic Review* 92(5):1644–55.

Kahneman, Daniel and Amos Tversky. 1979. “Prospect Theory: An Analysis of Choices under Risk.” *Econometrica* 47(2):263–91.

Kőszegi, Botond and Matthew Rabin. 2007. “Reference-dependent risk attitudes.” *The American Economic Review* pp. 1047–1073.