Appendix A. Proof of no asymmetric equilibrium in Case 2 and Case 3

The asymmetric equilibrium means that (Invest, Not Invest) or (Not Invest, Invest) is the equilibrium.

B.1. Case 2

We firstly assume asymmetric equilibrium (Invest, Not Invest) exists. For this equilibrium to exist, D1 need to choose Invest as his best response and D2 need to choose Not Invest as his best response. If Invest is the best response for D1, then the following condition need to be satisfied:

$$\begin{cases} v - I > v(1 - p_{NI}) \\ v - I > v(1 - \frac{1}{2} p_{NI}) \end{cases}$$

Solve the above formula we get the condition that Invest is the best response of D1:

$$\frac{I}{v} < \frac{1}{2} p_{NI}$$

If Not Invest is the best response for D2, then the following condition need to be satisfied:

$$\begin{cases} v - I < v(1 - p_{NI}) \\ v - I < v\left(1 - \frac{1}{2} p_{NI}\right) \end{cases}$$

Solve the above formula we get the condition that Not Invest is the best response for D2:

$$\frac{I}{v} > p_{NI}$$

Therefore, together we get that, in order to have (Invest, Not Invest) as the equilibrium, the following condition (3) and (6) need to be satisfied:

$$\begin{cases} \frac{I}{v} < \frac{1}{2} p_{NI} \\ \frac{I}{v} > p_{NI} \end{cases}$$

The above condition (3) and condition (6) are impossible to satisfy at the same time. So, the asymmetric equilibrium (Invest, Not Invest) does not exist.

Using the similar process, we can prove that there is no asymmetric equilibrium (Not Invest, Invest) exits.

In conclusion, there is no asymmetric equilibrium that exists in case 2.

B.2. Case 3

We firstly assume asymmetric equilibrium (Not Invest, Invest) exists. For this equilibrium to exist, D1 need to choose Not Invest as his best response and D2 need to choose Invest as his best response. If Not Invest is the best response for D1, then the following condition need to be satisfied:

$$\begin{cases} v\left(1-\frac{1}{2}p_I\right)-I<v(1-p_{NI}) \\ v-I<v(1-\frac{1}{2}p_{NI}) \end{cases}$$

Solve the above formula we get the condition that Not Invest is the best response of D1:

$$\frac{I}{v}>p_{NI}-\frac{1}{2}p_I$$

If Invest is the best response for D2, then the following condition need to be satisfied:

$$\begin{cases} v\left(1-\frac{1}{2}p_I\right)-I>v(1-p_{NI}) \\ v-I>v(1-\frac{1}{2}p_{NI}) \end{cases}$$

Solve the above formula we get the condition that Invest is the best response for D2:

$$\frac{I}{v}<\frac{1}{2}p_{NI}$$

Therefore, together we get that, in order to have (Invest, Not Invest) as the equilibrium, the following condition (9) and (12) need to be satisfied:

$$\begin{cases} \frac{I}{v}>p_{NI}-\frac{1}{2}p_I \\ \frac{I}{v}<\frac{1}{2}p_{NI} \end{cases}$$

$$\because p_{NI}>p_I \therefore p_{NI}-\frac{1}{2}p_I>\frac{1}{2}p_{NI}$$

The above condition (9) and condition (12) are impossible to satisfy at the same time. So, the asymmetric equilibrium (Not Invest, Invest) does not exist.

Using the similar process, we can prove that there is no asymmetric equilibrium (Invest, Not Invest) exits.

In conclusion, there is no asymmetric equilibrium that exists in case 3.

Appendix B. Equilibrium with a monetary penalty

In case 2, the equilibrium strategies for the defenders are either (invest-invest) or (no invest – no invest):

In the invest-invest case, the attacker never attacks, so we don't have to consider the penalty situation.

In the no invest-no invest case, the attacker always attacks, and asks for a ransom. The affected defender needs to decide whether to pay the ransom or not.

D1 | D2	Invest	Not Invest
Invest	$v - I, v - I$	$v - I, v - p_{NI}(v - p_{pay(}v - r - f))$
Not Invest	$v - p_{NI}\left(v - p_{pay}(v - r - f)\right), v - I$	$v - \frac{1}{2}p_{NI}(v - p_{pay}(v - r - f))$

If $p_{pay} = 1$, which means $v - r - f \geq 0$, so $f \leq v - r$, the defender's utility become

$$v - \frac{1}{2}p_{NI}(v)$$

If $p_{pay} = 0$, which means $v - r - f < 0$, so $f > v - r$, the defender's utility become

$$\frac{1}{2}v$$

Since $p_{NI} \leq 1$, the affected defender should always pay the ransom if $f \leq v - r$.

Appendix C. MTurk Experiment instruction and quiz

Instruction-Background Information

This is an experiment in ransomware and protection investment. If you follow the instructions carefully and make good decisions, you may earn a considerable amount of money that will be paid to you in cash at the end of the experiment. You have already earned US$1 show-up fee for participating. You will earn *experimental dollars* during the experiments, and experimental dollars will be converted to US dollars at the end of the experiment with the following exchange rate.

1,000 experimental dollars = US$1

You will receive the show up fee ($1) and any additional earnings ONLY if you finish the experiment.

In this experiment, there are three players: Attacker, Defender 1, and Defender 2. During the experiment, you will be randomly assigned to be the Attacker, the Defender 1, or the Defender 2. You will play the same role for the entire experiment. In total, you are going to play 30 rounds. In the first round, you will be randomly paired up with other players to form a 1 attacker- 2 Defenders group to play the game. You will stay in the same group for the entire experiment.

Data Value

In each round, each Defender is given a "data value" of 100 experimental dollars. The defender will receive these 100 experimental dollars at the end of each period if this data value is not lost.

Ransomware Attack

In each round, the Attacker chooses one of three options: (a) attack Defender 1; (b) attack Defender 2; (c) do not attack. If the Attacker chooses to attack a Defender, he/she also decides a ransom to ask.

The Attacker's probability of being successful is 80%. A defender can reduce the probability to 30% by spending 30 experimental dollars to make a protection investment.

If the attack is successful, the affected Defender chooses whether to pay the ransom. If the Defender decides to pay, he/she will not lose his/her data value and the Attacker receives the ransom for the round. If the Defender decides NOT to pay the ransom, the Defender loses his/her data value and the Attacker receives nothing for the round.

If the Attacker decides not to attack, the Attacker receives a fixed payment of 40 experimental dollars for the round.

Protection Investment

Protection Investment can reduce the Attacker's probability of being successful from 80% to 30%. That is, if a Defender who made protection investment is attacked, the Attacker's probability of being successful is 30%. If a Defender who did not make protection investment is attacked, the Attacker's probability of being successful is 80%. If a Defender decides to invest, a cost of 30 experimental dollars will occur for the round.

You will be allowed to continue only if you pass the following quizzes.

Quiz Question 1

Pretending the following scenario happened for a particular round in the experiment:

Defender 1 decided not to make the protection investment.

Defender 2 decided to make the protection investment.

Attacker decided not to attack.

What is the experiment dollar payoff for the Defender 1? Answer: 100

What is the experiment dollar payoff for the Attacker? Answer: 40

Quiz Question 2

Pretending the following scenario happened for a particular round in the experiment:

Defender 1 decided not to make the protection investment.

Defender 2 decided to make the protection investment.

Attacker decided to attack Defender 1 and asked 55 as ransom.

It was a successful attack, and Defender 1 decided to pay the ransom.

What is the experiment dollar payoff for the Defender 1? Answer: 45

What is the experiment dollar payoff for the Attacker? Answer: 55

Quiz Question 3

Pretending the following scenario happened for a particular round in the experiment:

Defender 1 decided not to make the protection investment.

Defender 2 decided not to make the protection investment.

Attacker decided to attack Defender 2 and asked 60 as ransom.

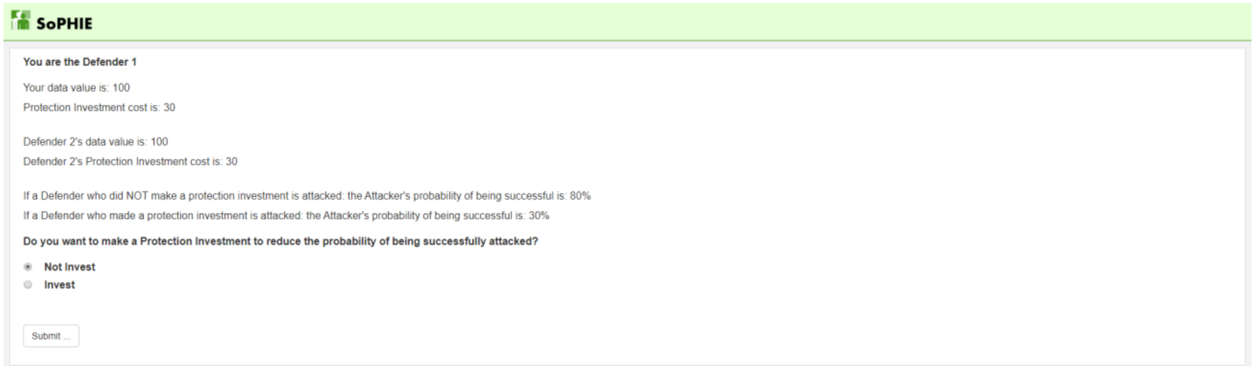
It was an unsuccessful attack.

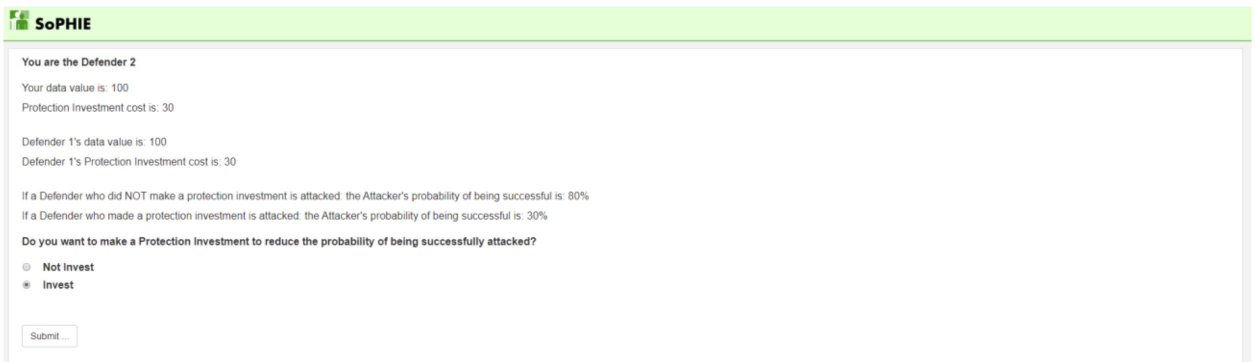
What is the experiment dollar payoff for the Defender 2? Answer: 100

What is the experiment dollar payoff for the Attacker? Answer: 0

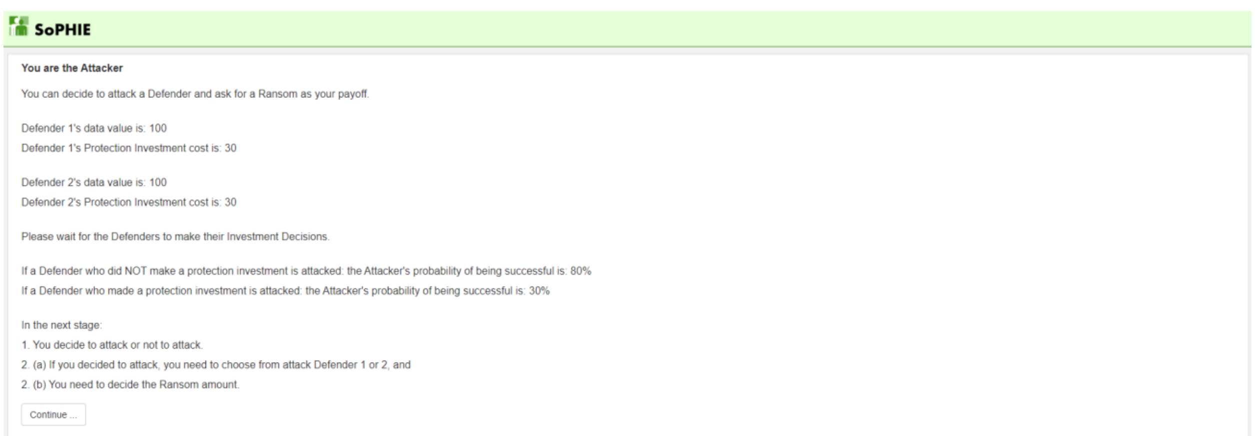
Appendix D. SoPHIE Screenshot (Baseline Treatment)

Defenders make investment decisions





Attacker waits for defenders



Attacker makes attack and ransom decisions

SoPHIE

You are the Attacker

Defender 1's data value is: 100

Defender 1's Protection Investment cost is: 30

Defender 1's Investment decision is: Not Invest

If you attack Defender 1, the probability of successful attack is: 80%

Defender 2's data value is: 100

Defender 2's Protection Investment cost is: 30

Defender 2's Investment decision is: Invest

If you attack Defender 2, the probability of successful attack is: 30%

If you decide not to attack, your payoff is 40

What is your attacking decision?*

- (•) Attack Defender 1
- ( ) Attack Defender 2
- ( ) Do not attack

If you decided to attack, how much Ransom to ask?* 60

Continue...

Affected defender makes payment decision

SoPHIE

You are the Defender 1

Your data value is: 100

Protection Investment cost is: 30

Your Investment decision is: Not Invest

Defender 2's data value is: 100

Defender 2's Protection Investment cost is: 30

Defender 2's Investment decision is: Not Invest

Who got attacked: You got attacked

Attacker successfully attacked you. If you do not pay the ransom, you lose your data value.
Ransom Amount is: 60

Do you want to pay the ransom?

- ( ) **Not Pay**
- (•) **Pay**

Submit ...

End of a round: show profit

SoPHIE

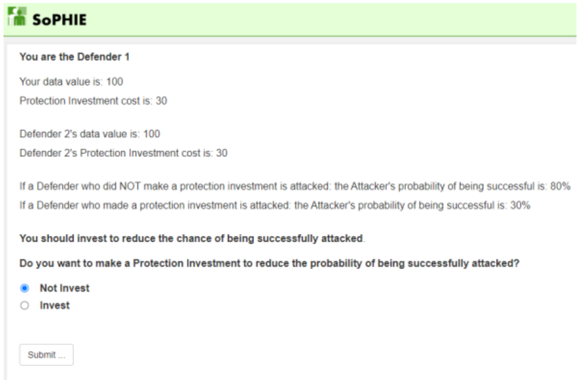
Results

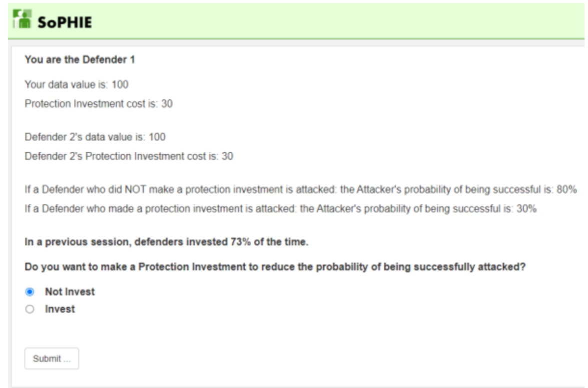
Round	Defender 1 Decision	Defender 2 Decision	Attacker Decision	Attack Outcome	Ransom Amount	Pay Decision		Defender 1 Payoff	Defender 2 Payoff	Attacker Payoff
1	Not Invested	Not Invested	Attacked D1	Succeeded	60	Paid		40	100	60
2	Invested	Invested	Not Attacked	N/A	N/A	N/A		70	70	40

Continue ...

Benefit Appeals: Should Invest **Descriptive Appeals: 73% Invest**

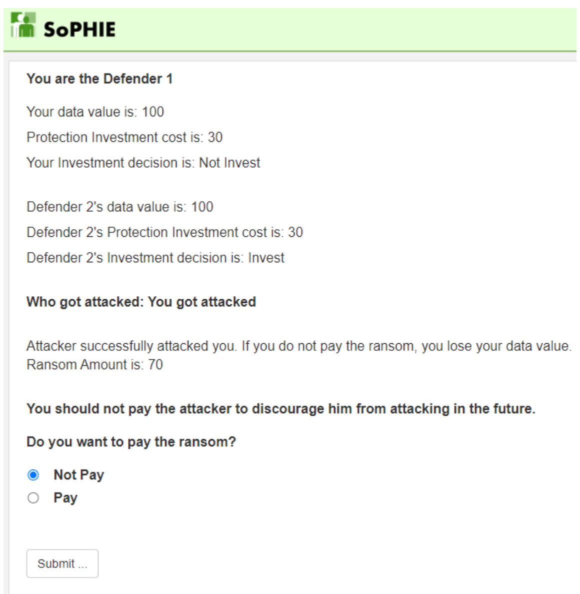
Defenders make investment decisions

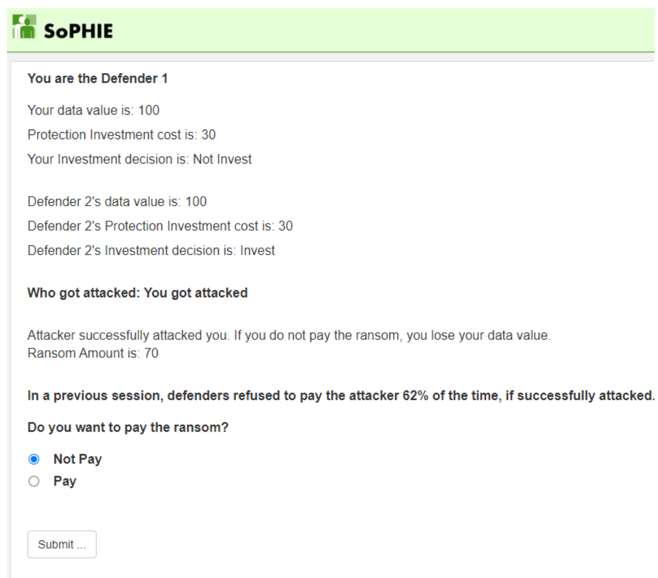




Benefit Appeals: Should Not Pay **Descriptive Appeals: 62% Not Pay**

Affected defender makes payment decision





Appendix E. Model Estimation with Benefit and Descriptive Appeals

Benefit Appeals	Should Invest		Should Not Pay	
Parameter	Estimate	p-value	Estimate	p-value
γ_p: payment bounded rationality	0.0313	0.000	0.0023	0.000
γ_r: ransom bounded rationality	0.1404	0.000	0.1270	0.000
γ_a: attack bounded rationality	0.0685	0.000	0.0453	0.000
γ_i: investment bounded rationality	0.0914	0.000	0.1153	0.000
α_m: anchoring past (self)	31.5578	0.000	25.0615	0.000
α_y: anchoring past (other)	6.9687	0.000	11.0368	0.000
α_f: pay fairness	0.3536	0.000	9.0885	0.000

Descriptive Appeals	73% Invest		62% Not Pay	
Parameter	Estimate	p-value	Estimate	p-value
γ_p: payment bounded rationality	0.0353	0.000	0.0186	0.000
γ_r: ransom bounded rationality	0.0833	0.000	0.1680	0.000
γ_a: attack bounded rationality	0.0883	0.000	0.0541	0.000
γ_i: investment bounded rationality	0.1299	0.000	0.1260	0.000
α_m: anchoring past (self)	19.2649	0.000	28.2915	0.000
α_y: anchoring past (other)	4.9174	0.000	5.4026	0.000
α_f: pay fairness	0.1011	0.014	0.7430	0.000

Appendix F.

Baseline Model with Risk Aversion Estimation Result

Parameter	Estimate	p-value
γ_p: payment bounded rationality	0.0156	0.000
γ_r: ransom bounded rationality	0.1384	0.000
γ_a: attack bounded rationality	0.0459	0.000
γ_i: investment bounded rationality	0.3545	0.000
α_m: anchoring past (self)	21.4969	0.000
α_y: anchoring past (other)	6.5884	0.003
α_f: pay fairness	0.4435	0.000
θ_r: risk aversion	0.0157	0.049

Appendix G. Alternative Interventions: Penalty and Chat

The first is a manipulation of direct incentives where we introduce a penalty for ransom payment. We expect a direct penalty to increase the utility of not-paying ransoms (i.e., increase the treatment utility of not-paying). According to the model analysis in Section 5, we anticipate a substantial decrease in ransoms, and all other security outcomes (investment rate, attack rate and payment rate) either remain unchanged or decrease slightly.

The second is a "chat" treatment where we manipulate social interactions between the two defenders. That is, we allow the defenders to have free-form communications before they make investment decisions. A comprehensive analysis of costless communication, or cheap talk, can be found in a number of studies (Farrell, 1987, Farrell and Gibbons, 1988, Rabin, 1990, Rabin, 1994). We posit that cheap talk is a means to deliver players' intentions and thereby improve coordination for both investment and ransom payment decisions. Communication allows defenders to exchange reasoning for their actions and influence each other.

We recruit and coordinate the penalty and chat treatments exactly identical as the baseline treatment that is described in Section 4.5. For the penalty treatment, we add the following sentence in addition to the baseline instruction to describe the penalty when a defender decides to pay the ransom, "If the Defender decides to pay, he/she will not lose his/her data value. However, a penalty of 15 will be incurred to him/her, and the Attacker receives the ransom for the round." For the chat treatment, we add the following paragraph in addition to the baseline instruction to describe the chatting process, "At the beginning of each round, Defenders can use a chat box to discuss possible choices and strategies for 30 seconds. These messages will be strictly private communication between Defenders."

Table G1 summarizes the descriptive statistics of the four security outcomes in the two treatment conditions, compared with those predicted by game theory and in baseline. Table G2 reports the results of Mann-Whitney tests comparing the security outcomes in the two treatment conditions with those in baseline. Table G3 presents the estimation of the base model for the two treatments separately.

Table G1 Summary Statistics – Penalty and Chat

Decision	Game Theory	Baseline	Penalty	Chat
Number of Groups	N/A	20	20	21
Investment Rate	100%	50.17% (28.87%)	60.08% (28.27)	57.38% (29.26%)
Attack Rate	0%	50.17% (40.82%)	38.00% (39.63%)	40.48% (41.07%)
Ransoms	100	66.14	54.30	65.81

	85 for penalty	(12.48)	(10.80)	(14.14)
Payment Rate	100%	49.61% (25.19%)	63.32% (26.01%)	54.76% (34.29%)

Notes:
1. Standard deviations are reported in parentheses.
2. All results are significantly (p-value < 0.01) different from the game theory predictions.
3. Investment rate and attack rate are reported in group average across all periods.
4. Ransoms are reported in group average across all periods conditioned on attack decision.
5. Payment rate is reported in group average across all periods conditioned on successful attacks.

Table G2 Decision Compare with Baseline

	Investment	Attack	Ransom	Payment
Baseline	50.17%	50.17%	66.14	50.39%
Penalty	60.08% (0.250)	38.00% (0.064)	54.30 (0.004)	63.32% (0.090)
Chat	57.38% (0.489)	40.48% (0.136)	65.81 (0.607)	54.76% (0.978)

Note: Mann-Whitney test, p-values are reported in parentheses

Table G3 Model Estimation with Penalty and Chat

	Penalty		**Chat**	
Parameter	Estimate	p-value	Estimate	p-value
γ_p: payment bounded rationality	0.0316	0.000	0.0197	0.000
γ_r: ransom bounded rationality	0.1381	0.000	0.1622	0.000
γ_a: attack bounded rationality	0.0565	0.000	0.0640	0.000
γ_i: investment bounded rationality	0.0877	0.000	0.1315	0.000
α_m: anchoring past (self)	23.5013	0.000	22.2947	0.000
α_y: anchoring past (other)	15.6999	0.000	8.6884	0.000
α_f: pay fairness	0.4297	0.001	0.5359	0.000
Δ_f: penalty	7.6221	0.349		
Δ_i: chat invest			9.6521	0.000
Δ_p: chat pay			15.6932	0.280

G.1 Penalty Treatment

We set the penalty of paying a ransom, in this treatment, to be 15% of the data value. We employ the model outline in Section 5.4 to estimate the impact of the intervention. Table G4 summarizes the estimation results. The penalty intervention significantly increases the utility of not-paying. The β_p parameter is positive and significant, with p-value < 0.001. Note that, in this case, the treatment utility of not-paying should be interpreted as the additional mental cost of paying a ransom, in addition to the actual monetary penalty included in the utility calculation explicitly as a cost in the model. This is consistent with the interpretation that the penalty itself, which is communicated as such to the subjects, is viewed with a negative connotation and further increases the utility of not-paying ransoms, on top of the actual monetary incentive. The penalty intervention has no impact on the utility of investing. The β_i parameter is not significantly different from 0 with p-value = 0.129. This is in-line with our expectation that a penalty of paying ransom has no direct impact on the utility of investing.

Table G4. Comparison between Penalty Treatment and Baseline

comparison to Base	Penalty
Parameter	Estimate (P-value)
γ_p: payment bounded rationality	0.0237 (0.000)
γ_r: ransom bounded rationality	0.1435 (0.000)
γ_a: attack bounded rationality	0.0511 (0.000)
γ_i: investment bounded rationality	0.0739 (0.000)
α_m: anchoring past (self)	29.6958 (0.000)
α_y: anchoring past (other)	13.8246 (0.000)
α_f: pay fairness	0.3030 (0.000)
β_i: treatment utility (invest)	2.0731 (0.129)
β_p: treatment utility (not-pay)	13.3767 (0.001)

Analyses in section 5.6.1 suggest that an increase in treatment utility of not-paying should be accompanying a decrease in ransoms. Indeed, as shown in Table G2, we find the ransom is significantly lower, with p-value < 0.01, than the baseline. The other security outcomes (investment rate, attack rate and payment rate) are not significantly different from the baseline, consistent with the model analysis.

G.2 Chat Treatment

We believe that communication has the potential to improve coordination and increase understanding of the decision problems. However, as opposed to the penalty intervention, there is less theoretical guidance of whether and how enabling communication will impact one or both of the treatment utilities. Note that all communication in this scenario is cheap-talk, as there is no enforcement mechanism for any intentions or suggestions that are communicated.

The model formulation in section 5.4 is designed to isolate the impact of a treatment intervention, compared to the baseline, which is adequate in all previous analysis. In this case, however, we also observe the actual *content* of the communication (i.e., text sent by subjects). We develop a variation of the behavioral model for treatment comparison to incorporate this additional information with the goal to determine whether the content is important in addition to the general ability to communicate for nudging a decision maker's utility to act.

In particular, we create a pair of new indicator variables y_i and y_p, where $y_i/y_p = 1$ if investment/payment is mentioned in the particular round by either defender 1 or defender 2. We opt to employ this simpler approach, as opposed to a full text mining analysis, because the messages are generally short, and we are only interested in the two decisions. We set these indicator variables by looking for specific, pre-determined, words such as "invest" and "investment" (see Appendix H for examples of chatting messages). We refer to these two variables as *chat content variables*. Please see Table G6 for the frequencies of these words mentioned. Recall that treatment utility formulations, defined in section 5.4 is as follows.

$$perceived\ investment\ cost = true\ cost - \alpha_{me}(x\, x_{me}) - \alpha_{you}(x\, x_{you}) - \beta_i z$$

$$u_p(x_{pj}) = (v - x_r - \alpha_f \max(r - [v - r - c_i x_{ij}],0))x_{pj} - c_i x_{ij} + \beta_p(1 - x_{pj})z$$

We re-formulate β_i and β_p to $(\beta_i + \Delta_i y_i)$ and $(\beta_p + \Delta_p y_p)$ respectively. Hence, the above equations become:

$$perceived\ investment\ cost = true\ cost - \alpha_{me}(x\, x_{me}) - \alpha_{you}(x\, x_{you}) - (\beta_i + \Delta_i y_i)z$$

$$u_p(x_{pj}) = (v - x_r - \alpha_f \max(r - [v - r - c_i x_{ij}],0))x_{pj} - c_i x_{ij} + (\beta_p + \Delta_p y_p)(1 - x_{pj})z$$

The interpretation is that the respective utilities are changed by β_i and β_p in the treatment (when z=1), and *further* changed by Δ_i and Δ_p if certain words are mentioned during the chat session. Please see Table G5 for the estimates.

We conclude that indeed communication has an impact on the utility of investing, but only if investment is discussed. β_i is not significant but Δ_i is highly significant with p-value < 0.001. We also find that investment rate, in the rounds where investment is discussed, is significantly higher than investment rate in the baseline (74.63% vs 50.17%) with p-value < 0.01 (please see Table G1 and G2 for the relevant statistics). Furthermore, this difference goes away if we compare the overall investment rate of the chat treatment with that of the baseline. This is strong evidence that communication matters, but only when the discussion is relevant.

Table G5. Comparison between Chat Treatment and Baseline

comparison to Base	Chat
Parameter	Estimate (P-value)
γ_p: payment bounded rationality	0.0177 (0.000)
γ_r: ransom bounded rationality	0.1675 (0.000)
γ_a: attack bounded rationality	0.0521 (0.000)
γ_i: investment bounded rationality	0.0840 (0.000)
α_m: anchoring past (self)	30.8140 (0.000)
α_y: anchoring past (other)	11.7297 (0.000)
α_f: pay fairness	0.4233 (0.000)
β_i: treatment utility (invest)	-0.4463 (0.735)
β_p: treatment utility (not-pay)	12.3924(0.023)
Δ_i: chat invest	14.8104 (0.000)
Δ_p: chat pay	10.0757 (0.583)

In addition, we find an increase in the utility of not-paying which does not depend on whether "paying" is discussed. β_p is significant with a p-value = 0.023 but Δ_p is not significant. We speculate that, as 75% of the infrequent ransom payment discussion occurred in the first half of the experiment, the impact of discussions, in this case, persisted throughout the session. Empirically, we find no evidence of a change in the ransom or ransom payment rate. That is not outside of our expectation as we observe a significant amount of decision noise.

Finally, we find that relevant discussions take place only in a minority of the rounds (Table G6). While communication seems to have improved investment in the right circumstances, these results, taken in total, point to a concern of this approach. Namely, there is a lack of control of what is discussed, even when discussion can improve decision-making.

Table G6. Message Frequency in Chat Treatment

Message Content	# of Messages	Percentage
Investment	134	21.54%
Ransom Payment	27	4.34%
None of them[1]	461	74.12%

In response to this treatment, we find the attacker neither lowers the ransom nor changes the attack compared with the baseline (Table G2). We also did not find a drop of payment rate. Neither

[1] "None of them" means the defenders mentioned neither investment nor ransom payment, but may be something else at the round.

do we detect an increase of investment rate overall. The results may be caused by the low number of rounds with relevant communications.

References

Farrell J. 1987. Cheap talk, coordination, and entry. *The RAND Journal of Economics*. **18**(1) 34-39.

Farrell J, Gibbons R. 1988. *Cheap talk, neologisms, and bargaining*. Massachusetts Institute of Technology, Cambridge, Mass.

Rabin M. 1990. Communication between rational agents. *Journal of Economic Theory*. **51**(1) 144-170.

Rabin M. 1994. *Incorporating behavioral assumptions into game theory*. Springer.

Appendix H. Examples of the Defenders' Chat Message from Chat Treatment

837;"";"1532571875895";"Pa.46";"I WOULD INVEST IF I WERE YOU"

839;"";"1532571928634";"Pa.46";"MAYBE THEY ONLY ATTACK IF WE DONT INVEST"

913;"";"1532572471523";"Pa.4";"They never attack if we both invest"

1003;"";"1532573056300";"Pa.4";"...maybe you should invest this time"

890;"";"1532572349488";"Pa.4";"I'll invest"

184;"";"1532122735864";"Pa.59";"they might attack if we choose not to defend though"

165;"";"1532122494616";"Pa.18";"I'm going to just keep investing"

187;"";"1532122798962";"Pa.59";"haha yeah i say we stick with investing"

We omitted sentences containing only one or two words, such as "Hi", "that's good", or any stopping words. We have a total of 622 lines of chat messages.

If a defender mentioned "inv", "invest", "investment", or "30, 70" we count it as an investment message. We have 134 lines of investment messages, which represent 21.54% of total messages, many of defenders were talking about "I will or will not to invest", and suggesting what the other defender should do, such as "you should invest".

If a defender mentioned "pay", "payment", "ransom", or "money", we count it as a ransom payment message. We have 27 lines ransom payment messages, which represent 4.34% of total messages.

If a defender did not mention any keyword listed above, such as "investment" or "ransom", we categorize it as a no key word message. We have 461 no key word messages, which represent 74.12% of total messages.