Anticipated Third-Party Rewards
Experimental Evidence on Social Control and Cooperation

Lilia Wasserka-Zhurakhovska

This contribution honors Christoph Engel's pioneering research on rule-following, legitimacy, and the institutional foundations of cooperation. Building on his experimental law and economics agenda, I investigate how anticipated third-party rewards influence trust and reciprocity in strategic interaction. The laboratory design combines a trust game with a subsequent helping game, in which Observers reward Trustees based on their behavior. Treatments vary depending on whether Investors and Trustees are informed in advance about the presence of Observers. Consistent with Engel's findings on conditional rule-following and the motivating power of institutions, the results show that Observers reward Trustees more generously when return transfers are higher, and that Trustees return more when they anticipate being observed. However, this strategic adjustment does not lead to higher investments by Investors, highlighting both the reach and the limits of anticipated indirect reciprocity. The study thus illuminates the subtle mechanisms through which institutions shape cooperation, while paying tribute to Engel's lifelong commitment to understanding how norms, roles, and third parties sustain social order.

A. Introduction

Christoph Engel has repeatedly asked how fragile social order can nonetheless persist when individuals face clear incentives to act selfishly. In his reflections on law as an "impossible discipline" (Engel, 2024), he stresses that rules and institutions do not operate mechanically. Their effectiveness depends on how they are perceived, interpreted, and legitimized. This line of questioning has shaped his experimental law-and-economics research: individuals are neither unconditional egoists nor simple altruists, but respond systematically to institutional and social cues that frame their interactions.

The present study takes up this challenge by examining whether the mere anticipation of institutional observation can alter behavior in a trust game. Specifically, I investigate a two-stage design in which two players interact in an investment game, while a third party (Observer) subsequently decides on rewards for the trustee, conditional on the trustee's behavior. My treatments vary whether trustees are aware from the outset that they may later be observed and rewarded (Announced treatment) or whether this possibility remains undisclosed until after their decisions (Unannounced treatment).

By embedding this mechanism into a controlled laboratory setting, the study connects directly to Engel's concern with conditional rule following, legitimacy, and the subtle role of institutional signals. Trustees behave more generously when they anticipate third-party evaluation and reward, suggesting that even minimal institutional cues are sufficient to trigger more cooperative behavior. Yet investors do not respond with higher levels of trust. This asymmetry highlights both the reach and the limits of anticipated indirect reciprocity: institutional interventions may strongly affect immediate reciprocity but fade as interactions become more indirect and cognitively demanding.

In this way, my findings resonate with Engel's broader research program. They echo his emphasis on conditionality, legitimacy, and the role of institutions in sustaining cooperation, while also illustrating the boundaries of such mechanisms. The study thus exemplifies Engel's call to accept complexity rather than simplifying it away and contributes to his intellectual legacy by advancing my understanding of how institutions shape behavior in layered social interactions.

B. Related Literature

Engel's research has repeatedly demonstrated that individuals condition their behavior on cues about norms and the expected actions of others. Rule compliance has been shown to depend on perceived compliance in one's reference group (Engel and Desmet, 2021). Similarly, even minimal social information has been shown to trigger peer effects in rule violation (Engel, 2023). These insights resonate strongly with my finding that trustees adapt when the existence of a monitoring third party is announced in advance. Even a minimal institutional cue suffices to trigger more cooperative behavior.

The institutional framing of decisions is another hallmark of Engel's research. The simple assignment of a judicial office led participants to enforce norms more strongly, suggesting that institutional roles activate a sense of duty (Engel and Zhurakhovska, 2017). My design parallels this mechanism: trustees act more generously once they recognize the evaluative role of a third party, even when the observer has no material stake. Similarly, legitimacy has been shown to derive both from outcomes and from the opportunity to participate in decision-making (Engel, Mittone and Morreale, 2024). The anticipatory effect of a third party in my study can be seen as a minimal form of participatory legitimacy.

my own work has likewise engaged with the institutional determinants of cooperation and fairness. Participation procedures increase subsequent kindness toward authorities, irrespective of outcomes (Kleine, Langenbach and Zhurakhovska, 2017). Similarly, impartial decision-makers have been shown to be swayed by stakeholder communication, though in asymmetric ways (Kleine, Langenbach and Zhurakhovska, 2016). Other studies emphasize how dishonesty and norm violations spread in distinct ways across genders (Böhm, Goerg and Wasserka-Zhurakhovska, 2023) and how leaders face ethical dilemmas that interact with gender differences in dishonesty (Grosch, Müller, Rau and Wasserka-Zhurakhovska, forthcoming). These contributions highlight the power of procedural and contextual signals to affect compliance and reciprocity, much as the presence of a third party in the current experiment alters trustee behavior.

Engel's research has also emphasized the importance of fairness perceptions and their fragility in social dilemmas. Rule-following declined once income heterogeneity was revealed, underscoring the role of fairness expectations (Engel, Mittone and Morreale, 2020). In my experiment, third-party rewards can be interpreted as fairness signals that amplify cooperative choices among trustees. Yet, in line with evidence showing that deciding on behalf of others does not mitigate selfishness (Engel and Cerrone, 2019), my investors remain unresponsive. This may be because the chain of reasoning is too indirect, or because the fairness signal fades with distance from the immediate decision context.

Finally, the broader literature on third-party enforcement underscores both the promise and the limits of institutional interventions. Norm enforcement has been identified as a cornerstone of large-scale cooperation (Fehr and Schurtenberger, 2018), and both direct and indirect sanctions have been shown to sustain cooperation (Balafoutas, Nikiforakis and Rockenbach, 2016). Third-party punishment can operate as

a costly signal of trustworthiness (Jordan et al., 2016), while theories of third-party reward and punishment highlight their role as mechanisms of norm enforcement (Chang et al., 2022). My findings add nuance to this literature by showing that anticipation of third-party reward shifts trustee behavior but does not necessarily cascade into increased trust. This asymmetry highlights a boundary condition for institutional interventions: they may bolster compliance at one stage of interaction while leaving upstream trust decisions unaffected.

C. Experimental Design

Each subject was randomly assigned to one of three roles: Investor/Observer (Player A), Trustee (Player B), or Outsider (Player C). The study combined a one-shot trust game with a subsequent one-shot helping game. The trust game builds on Berg, Dickhaut and McCabe (1995), while the helping game is a variant of the dictator game first introduced by Nowak and Sigmund (1998) and later adapted to study indirect reciprocity and strategic reputation building (Engelmann and Fischbacher, 2009). Both stages relied on the strategy method (Selten, 1967), requiring participants to make conditional choices for every possible scenario rather than only for the realized case.

In the trust game, both Player A (Investor) and Player B (Trustee) received an endowment of 100 Taler. The Investor could transfer up to 60 Taler, choosing any multiple of ten from the set $X \in \{0, 10, 20, 30, 40, 50, 60\}$, to Trustee. The transferred amount was then tripled by the experimenter. The Trustee, having received $3 * X$, then chose a return transfer $Y \in \{0, 1, 2, 3\}$ times the original investment X, using the strategy method. Thus, for every possible transfer by Investor, Trustee had to indicate whether they would return nothing, the transfer itself, double the transfer, or triple the transfer. In case of zero investment, no return was possible. The Outsiders did not take any decision in the trust game, nor were they informed about the decisions of the other players. Outsiders received a fixed payoff of 100 Taler in the trust game, identical to what Investors and Trustees earned if no transfer had been made.

In the helping game, each Player A was matched with a Player B from another group. In this game Player A became an Observer. Importantly, Observers did not evaluate the Trustee from their own trust game. Instead, they observed the complete strategy of an Investor–Trustee interaction

from a different group (i.e., the investment and the corresponding return transfer) and then decided how much to transfer. This ensured that Observers' transfers could not be interpreted as direct reciprocity toward their own partner, but only as third-party reward. Observers were endowed with 100 Taler and could transfer any amount between 0 and 100 to the Trustee (Player B). Transfers were tripled by the experimenter, so the Trustee received $3*T$. Observers' transfer decisions were elicited using the strategy method: they specified transfers for Trustees conditional on each possible history from the first stage (i.e., every combination of Investor transfer X and Trustee return decision Y). Additionally, Observers were asked how much they would transfer to a Passive Trustee (who had received 0 in the trust game) and to an Outsider.

Two treatments were implemented. In the Unannounced treatment, Investors and Trustees were not aware of the existence of Observers when making their decisions in the trust game. Only afterward were Observers introduced in the helping game. In the Announced treatment, by contrast, Investors and Trustees were informed ex ante that their trust game behavior could be observed by a third party, who might later reward Trustees in the helping game. For Observers, the two treatments were strategically identical: they always observed the Trustee's behavior and decided on transfers. However, they did know whether their existence was announced beforehand, which in principle could have influenced their rewarding behavior.

The experiment was run at the University of Bonn. It was programmed and conducted with z-Tree (Fischbacher, 2007). Four sessions with 96 subjects were held, resulting in 22 independent observations per treatment for the roles of Investor/Observer and Trustee. Not every group was assigned an Outsider, and while the instructions avoided explicitly stating that each group had one, they were carefully worded to prevent any deception in this regard. Subjects were recruited via ORSEE (Greiner, 2015). None had participated in prior trust or dictator/helping games. The subject pool consisted mainly of students: 19 from economics, 20 from law, and the remainder from other disciplines; 46 percent were female. Average earnings were 12.52 Euro (including a 4 Euro show-up fee), corresponding to about $ 15–16 depending on the exchange rate at the time. Sessions lasted about 70 minutes. Control questions ensured comprehension before the start, and participants filled out a short post-experiment questionnaire including demographics.

D. Behavioral Predictions

Anticipation functions similarly to backward induction. First, one must form expectations about the choices of Observers. Then, Trustees' reactions to these choices can be predicted. In the helping game, assuming individuals aim solely to maximize their own payoffs, the theoretical prediction is that no transfers will occur. A self-interested Trustee has no monetary incentive to return anything, regardless of the information available about the helping game. Similarly, a rational, self-interested Investor anticipates this and chooses not to invest in any treatment. Under the assumption of pure payoff maximization and common knowledge, the unique Nash equilibrium predicts zero transfers across all games and treatments.

However, empirical evidence contradicts this prediction. In helping games, positive transfers are observed (Forsythe et al., 1994), and in trust games, transfers occur in both directions (Berg, Dickhaut, and McCabe, 1995). The former is often attributed to social preferences such as the "warm glow" effect (Andreoni, 1990), while the latter is explained by strong direct reciprocity. Models by Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Levine (1998), and Rabin (1993) suggest that individuals derive utility from rewarding kind actions and punishing unkind ones. What counts as "kind" varies across models.

Levine's (1998) model is particularly relevant here, as it is not limited to two-player interactions. It assumes that individuals derive utility not only from their own payoffs but also from rewarding others based on their perceived altruism. In this framework, a person's utility increases when they reward someone who is seen as altruistic. In the present experiment, Observers do not need to form independent beliefs about Trustees' preferences; they can infer them from Trustees' return transfers to Investors. If some Observers are altruistic and value others' altruism, then:

> **Prediction 1a:** The higher Trustees' return transfers in the trust game, the more help will be transferred by Observers.

The second key question is whether Observers evaluate Trustees' altruism differently in the Unannounced and Announced treatments. Levine's model is a signaling model, where credible signals of trustworthiness should be rewarded. In the Announced treatment, high return transfers may signal trustworthiness, but they may also be interpreted as strate-

gic actions aimed at gaining rewards. Therefore, Observers may perceive these actions as less genuine.

Falk, Fehr, and Fischbacher (2008) show that intentions matter: actions driven by intrinsic motives are rewarded more than those driven by strategic ones. This suggests that Observers' transfers should be higher in the Unannounced treatment, where Trustees' return transfers cannot be strategically motivated. Similarly, Stanca, Brunni, and Corazzini (2009) find stronger indirect reciprocity when strategic motivations are ruled out.

> **Prediction 1b:** For any given level of Trustees' return transfers, Observers' transfers are lower in the Announced treatment than in the Unannounced treatment.

Following Levine's model, I can also predict Trustees' behavior:

> **Prediction 2a:** Trustees in the trust game make higher return transfers when they receive higher investments.

In the Announced treatment, Trustees know that Observers may reward them based on their behavior. If Trustees believe that some Observers value altruism and are willing to reward it, they may increase their return transfers to signal trustworthiness. If the expected reward outweighs the cost of the transfer, Trustees have an incentive to behave more generously.

> **Prediction 2b:** Trustees in the trust game make higher return transfers in the Announced treatment than in the Unannounced treatment.

Finally, Costa-Gomes, Huck, and Weizsäcker (2014) find that Investors' optimism about return transfers correlates positively with their investment levels. If Investors anticipate that Trustees will behave more generously in the Announced treatment, they may invest more.

> **Prediction 3:** Investors make higher investments in the Announced treatment than in the Unannounced treatment.

E. Results

We now turn to the experimental results. This section is structured along the behavioral predictions derived above. I begin by analyzing Observers' transfers in the helping game, followed by Trustees' behavior in the trust game, and finally Investors' behavior. Throughout, I use both non-parametric tests and regression analysis to evaluate treatment effects.

I. Observers' Transfers

Prediction 1a stated that Observers transfer more the higher the Trustees' relative return transfers are in the trust game. Figure 1 illustrates mean Observers' transfers across all possible combinations of investments and Trustees' relative return transfers, separately for the Unannounced and the Announced treatment. The pattern clearly supports Prediction 1a: the higher the relative return transfer, the more Observers transfer. This holds across both treatments.

A Spearman rank correlation confirms this result: in the Unannounced treatment, Observers' transfers and Trustees' relative return transfers are positively correlated ($\rho = 0.44$, $p < 0.01$), and similarly in the Announced treatment ($\rho = 0.53$, $p < 0.01$). These findings provide strong support for Prediction 1a.

Turning to Prediction 1b, I expected that Observers would transfer less in the Announced treatment for any given level of Trustees' relative return transfer. Yet the regression results in Table 1 show a positive and significant interaction between return transfers and the Announced treatment. This indicates that while baseline transfers are lower in Announced, the sensitivity of Observers to Trustees' generosity is actually stronger when Trustees could anticipate observation. In other words, Observers reward high-return Trustees more intensely if their actions were taken under announced observation. This contrasts with prior findings (e.g., Stanca, Brunni, and Corazzini, 2009) and suggests that factors beyond perceived intentions, such as the salience of reputational incentives, shape strong indirect reciprocity.

Taken together, these results strongly support Predictions 1a but not 1b. Observers reward Trustees more when they return more to Investors, but they do so even more when Trustees could anticipate that their behavior would be observed.

Table 1: Explaining Observers' Transfers – Comparison of Unannounced and Announced Treatment

	Model 1	Model 2	Model 3	Model 4
Announced Treatment	10.05	5.66	9.87	–1.24
	(10.59)	(10.38)	(8.52)	(9.43)
Relative Return Transfers	16.43***	16.43***	16.41***	14.82***
	(.72)	(.72)	(.72)	(1.00)
Investment	.12***	.12***	.12***	.044
	(.04)	(.04)	(.04)	(.06)
*Announced Treatment *Relative Return Transfer*				2.94**
				(1.37)
*Announced Treatment *Investment*				.14*
				(.08)
Own Investment		.03	–.23	–.23
		(.26)	(.23)	(.26)
Own Profit in TG		.36**	.12	.12
		(.17)	(.15)	(.15)
Transfer to Passive Trustee			.34	.71
			(.47)	(.45)
Transfer to Outsider			.32	.32
			(.45)	(.47)
Constant	–47.22***	–82.30***	–62.71***	–56.65***
	(7.99)	(20.41)	(17.12)	(17.25)
N	1056	1056	1056	1056
N of group	44	44	44	44
P model	< .001	< .001	< .001	< .001
Wald Chi2	528.48	530.67	539.19	550.51

Notes: Random effects Tobit regressions. Standard errors are presented in parentheses. Observations are clustered on individual level. The Announced dummy equals 1 for all observations of the Announced treatment, relative return transfers controls for the relative return of a trustee (Y) for a given investment, investment controls for the investment the trustee has received, own investment is the investment the helper has transfers in the trust game himself to her trustee, Announced* relative return transfer and Announced *investment are interaction terms, own profit in TG controls for the helper's profit from part 1 of the experiment. Transfer to passive trustee and Transfer to outsider are the levels transferred to passive players. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively. Left-censored = 577; right-censored = 14.

Figure 1: Mean Observers' transfers by Trustees' return transfers and treatment (Announced vs. Unannounced)
Notes: The upper figure displays Observers' transfers in the Unannounced treatment and the lower figure in the Announced treatment. On the x-axis, the exact condition is displayed, i.e., one can see if the recipient is either an Observer or a passive Trustee (no investment and thus no opportunity for a return transfer) or an active Trustee who has received an investment of X and has made a relative return transfer of Y (Y=0: Trustees keep full transfer; Y=1: Trustees return transfer and keep rest; Y=2: equal split; Y=3: full return). On the y-axis, mean Observers' transfers are displayed for the particular situation. Standard errors are indicated.

II. Trustees' Transfers

Prediction 2a stated that Trustees make higher return transfers the more they receive from Investors. Figure 2 shows mean relative return transfers for each level of investment. In the Unannounced treatment, Trustees' transfers increase with investment, consistent with Prediction 2a. The correlation between investments and return transfers is positive and significant (Spearman $\rho = 0.15$, $p < 0.05$).

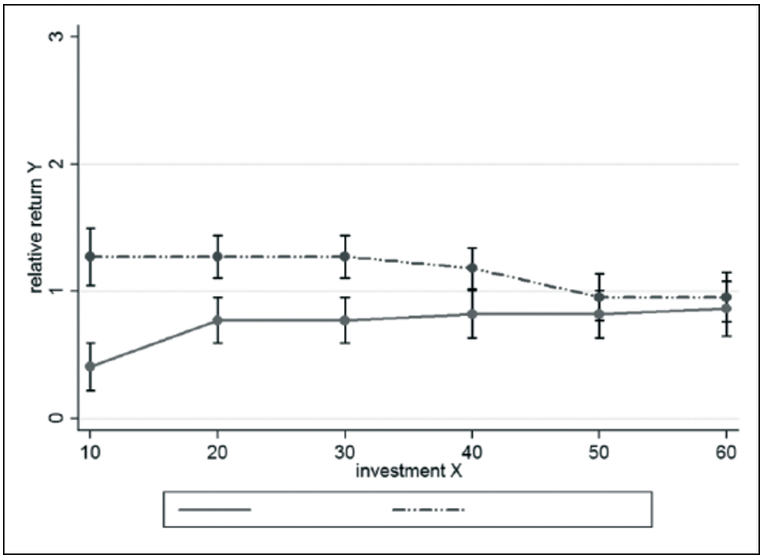

Figure 2: Trustees' relative return transfers by investment and treatment
Notes: On the x-axis, the investment the Trustee has received is depicted; on the y-axis, mean relative return transfers are displayed. The figure compares the Unannounced treatment and Announces treatments. Standard errors are indicated.

For the Announced treatment, however, the relationship is weaker and even negative ($\rho = -0.15$, $p < 0.10$), suggesting that Trustees do not condition as strongly positively on the level of investment when they anticipate being observed.

Prediction 2b stated that Trustees would make higher return transfers in the Announced treatment than in the Unannounced treatment. Figure 2 provides some support: average return transfers are indeed higher in the Announced treatment. A Mann–Whitney U test confirms that the difference is significant ($p < 0.05$).

Ordered Probit regressions with random effects (Table 2) further support this interpretation. The treatment dummy for Announced is positive and highly significant across all specifications, indicating that Trustees return more when they know Observers may reward them. At the same time, the interaction term between investment and the Announced treatment is negative and significant, consistent with the flatter slope observed in Figure 2. Trustees return more overall, but their responsiveness to investment is reduced when they can anticipate being observed.

Overall, Predictions 2a and 2b are supported: Trustees return more when they receive higher investments, and they return more in the Announced treatment, although their conditioning on investment is weaker. This pattern also indicates that trustees' behavior in this setting departs from the standard reciprocity usually observed in trust games, and is instead primarily driven by strategic concerns when anticipating potential rewards from Observers.

III. Investors' Transfers

Finally, Prediction 3 stated that Investors would invest more in the Announced treatment than in the Unannounced treatment. However, this prediction is not supported. Statistically, the investments do not differ between treatments (Mann–Whitney rank-sum: $p > 0.1$). The cumulative distribution functions in Figure 3 suggest some visual differences in the distribution, but these are not statistically significant. Similarly, parametric tests using ordered probit regressions (not reported here) confirm that

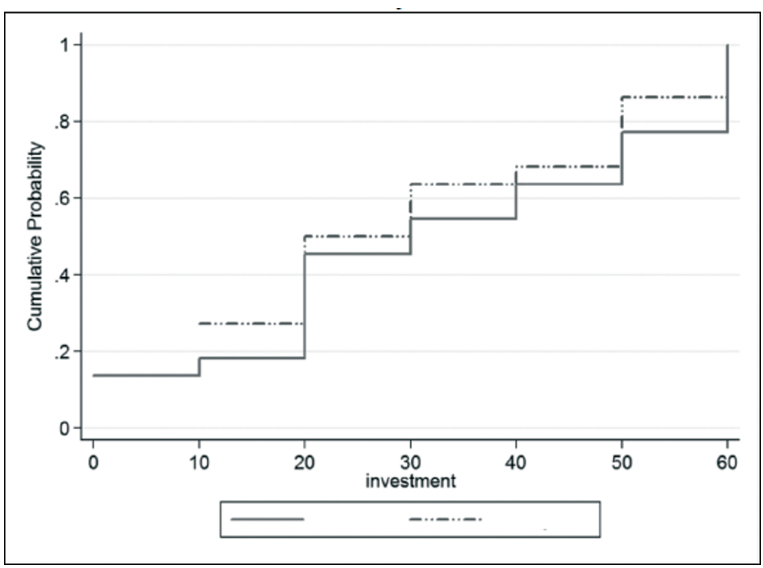

Figure 3: Cumulative distributions of Investors' investments across treatments
Notes: The graph shows an empirical cumulative distribution function of the investments for both treatments. On the x-axis, the investment is depicted; on the y-axis, the estimated probability for each investment is displayed.

Table 2: Explaining Trustees' Transfers – Comparison of Unannounced and Announced Treatment

	Model 1	Model 2	Model 3	Model 4
Announced Treatment	2.45***	2.45***	4.16***	4.20***
	(.27)	(.27)	(.50)	(.52)
Investment	–.00	–.00	.02***	.02***
	(.00)	(.00)	(.01)	(.01)
*Announced Treatment *Investment*			–.04***	–.05***
			(.01)	(.01)
*Announced Treatment *Female*				–1.11**
				(.44)
Cut 1 Constant	.29	.25	1.13***	1.06***
	(.18)	(.25)	(.34)	(.33)
Cut 2 Constant	2.29***	2.25***	3.28***	3.31***
	(.21)	(.27)	(.38)	(.39)
Cut 3 Constant	4.54***	4.50***	5.65***	5.51***
	(.39)	(.41)	(.52)	(.50)
Rho Constant	.86***	.86***	.87***	.86***
	(.02)	(.02)	(.02)	(.02)
N	264	264	264	264
N of groups	44	44	44	44
P model	0.033	0.101	< .001	< .001
LR Chi2	4.52	4.58	23.73	22.74

Notes: Random effects ordered Probit regressions. Standard errors are presented in parentheses. Observations are clustered on individual level. The Announced dummy equals 1 for all observations of the Announced treatment, investment controls for the investment (X) the trustee has received, the gender dummy equals 1 for all female observations. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

treatment has no statistically significant effect, and no gender–treatment interaction is found. Thus, Investors' behavior remains unaffected by the anticipation of third-party observation.

This pattern suggests that Investors are somewhat sensitive to the strategic implications of observation but less so than Trustees or Observers. One possible explanation, in line with Camerer, Ho, and Chong (2004), is that anticipating higher-order reasoning about indirect reciprocity is cognitively demanding, leading to weaker treatment effects on Investors' behavior.

F. Discussion

This study sets out to investigate how anticipated indirect reciprocity shapes trust, trustworthiness, and third-party helping. Building on a two-stage design with an investment game followed by a helping game, I compared two conditions: one in which the presence of Observers was announced to Investors and Trustees ex ante (Announced treatment), and one in which their presence remained unannounced until after the trust game (Unannounced treatment).

The behavioral predictions derived from models of social preferences and indirect reciprocity received substantial support. Consistent with Prediction 1a, I find a strong positive relationship between Trustees' return transfers and Observers' subsequent transfers. This result aligns with Levine's (1998) framework, where individuals derive utility from rewarding others perceived as altruistic. The Tobit regressions reported in Table 1 confirm that Observers' transfers increase systematically with the generosity of Trustees, both across and within treatments.

Prediction 1b, which anticipated lower transfers by Observers in the Announced treatment, is not supported. Instead, Observers reward generous Trustees even more strongly when Trustees act under announced observation. In other words, Observers do not penalize Trustees for behaving strategically, but still give up part of their own endowment to reward them. This suggests that strong indirect reciprocity operates even when cooperative behavior may be strategically motivated.

Turning to Trustees, I find support for both Predictions 2a and 2b. Trustees return more when investments increase, consistent with conditional reciprocity in the Unannounced treatment. Moreover, average return transfers are significantly higher in the Announced condition, as

confirmed by the Mann–Whitney tests and the ordered probit regressions in Table 2. At the same time, the negative interaction between investment and the Announced treatment shows that Trustees' responsiveness to investment is reduced once they anticipate third-party evaluation and potential reward. Overall, this pattern suggests that higher return transfers in the Announced treatment are driven less by reciprocity and more by strategic signaling aimed at attracting Observer rewards.

Finally, Prediction 3 receives less robust support. Investors' transfers are not systematically higher in the Announced treatment compared to the Unannounced treatment. While Figure 3 shows some indication of higher investments when Observers' presence was announced, the effect is not statistically significant. One explanation is that higher-order reasoning about Observers' potential transfers and Trustees' incentives proved too demanding for Investors, echoing Camerer, Ho, and Chong (2004). Alternatively, the attenuation of treatment effects across multiple layers of interaction may have limited Investors' responsiveness.

Taken together, these findings demonstrate that anticipated indirect reciprocity powerfully shapes behavior, but its effects are concentrated at the level of Trustees. Observers reward generosity; Trustees, in turn, strategically increase reciprocity when under observation. Investors, however, remain comparatively unaffected. This layered pattern underscores the importance of institutional signals and expectations of monitoring but also highlights the limits of such mechanisms in complex multi-stage interactions.

My results connect directly to Engel's broader research program. First, they speak to his long-standing interest in conditional rule following: Trustees adapt their behavior when institutionalized observation changes the perceived "rules of the game." Second, they resonate with his emphasis on legitimacy and third-party control: Observers effectively act as informal enforcers. Finally, they illustrate the importance of experimental methods for disentangling the subtle interplay of motives in institutional settings.

Of course, some limitations must be acknowledged. The study relied on a one-shot, laboratory setting with anonymity and strategy method elicitation. While these features increase control and internal validity, they may attenuate or amplify certain effects compared to repeated or naturalistic settings. Future research could explore how repeated interactions, reputation building, or heterogeneous observer roles affect the dynamics documented here. In making these points, I also follow Engel's broader in-

vitation to embrace the law as an "impossible discipline" (Engel, 2024). My experiment exemplifies his call to combine rigorous empirical methods with sensitivity to institutional context, and to accept complexity rather than simplifying it away. In this sense, the study not only extends the experimental law-and-economics toolkit, but also pays tribute to Engel's intellectual stance of taking risks in search of deeper insight.

G. Conclusion

This study provides evidence that anticipated indirect reciprocity can alter trust game behavior in systematic ways. Observers reward Trustees' generosity but penalize behavior that may appear strategically motivated. Trustees respond strategically to announced observation by increasing return transfers, while Investors remain comparatively unaffected.

The findings highlight both the power and the limits of anticipated indirect reciprocity. Strong effects emerge when the link between action and evaluation is immediate and transparent, as for Trustees, but weaken when reasoning requires higher-order inference, as for Investors. These insights contribute to understanding how institutions relying on external evaluation and monitoring may shape cooperative behavior, even in the absence of formal enforcement mechanisms.

For Engel's broader research agenda, the results underscore the role of conditional rule following and legitimacy in sustaining cooperation. They show how institutionalized observation can change incentives, but also how actors distinguish between genuine and strategic behavior. In this sense, the experiment speaks directly to Engel's interest in the foundations of social order: institutions are effective not only because they alter payoffs, but also because they shape perceptions of legitimacy and authenticity.

From a policy perspective, these findings suggest that institutional designs relying on third-party evaluation must carefully balance transparency with credibility. If evaluation mechanisms make actions appear overly strategic, their legitimacy may be undermined and their effectiveness reduced.

In closing, the study illustrates the value of combining formal models of social preferences with careful experimental design to capture the complex interplay of motives in social interactions. It is a contribution not only to the experimental literature on trust and reciprocity but also to

the intellectual legacy of Christoph Engel, whose work has consistently illuminated the mechanisms that underlie social order. In celebrating Christoph Engel's 70th birthday, this contribution aims to echo his commitment to risk-taking, interdisciplinarity, and the relentless pursuit of understanding how institutions shape behavior. If my results help sharpen the puzzle of cooperation under observation, they are indebted to the scholarly path Engel has charted.

References

Andreoni, J. (1990) 'Impure altruism and donations to public goods: a theory of warm-glow giving', *The Economic Journal*, 100(401), pp. 464–477.

Berg, J., Dickhaut, J. and McCabe, K. (1995) 'Trust, reciprocity, and social history', *Games and Economic Behavior*, 10(1), pp. 122–142.

Böhm, K.L., Goerg, S.J. and Wasserka-Zhurakhovska, L. (2023) 'How does unethical behavior spread? Gender matters!', *CESifo Working Paper*, No. 10314.

Camerer, C.F., Ho, T.-H. and Chong, J.-K. (2004) 'A cognitive hierarchy model of games', *The Quarterly Journal of Economics*, 119(3), pp. 861–898.

Cerrone, C. and Engel, C. (2019) 'Deciding on behalf of others does not mitigate selfishness: an experiment', *Economics Letters*, 183, p. 108616.

Costa-Gomes, M., Huck, S. and Weizsäcker, G. (2014) 'Beliefs and actions in the trust game: Creating instrumental variables to estimate the causal effect', *Games and Economic Behavior*, 88, pp. 298–309.

Dufwenberg, M. and Kirchsteiger, G. (2004) 'A theory of sequential reciprocity', *Games and Economic Behavior*, 47(2), pp. 268–298.

Engel, C. (2023) 'How little does it take to trigger a peer effect? An experiment on crime as conditional rule violation', *Journal of Research in Crime and Delinquency*, 60(4), pp. 455–492.

Engel, C., (2024). 'The Law–An Impossible Discipline,' *Review of Law & Economics*, 20(2), pp. 287–322.

Engel, C. and Desmet, P. (2021) 'People are conditional rule followers', *Journal of Economic Psychology*, 85, p. 102367.

Engel, C., Mittone, L. and Morreale, A. (2020) 'Tax morale and fairness in conflict: an experiment', *Journal of Economic Psychology*, 81, p. 102314.

Engel, C., Mittone, L. and Morreale, A. (2024) 'Outcomes or participation? Experimentally testing competing sources of legitimacy for taxation', *Economic Inquiry*, 62(2), pp. 563–583.

Engel, C. and Wasserka-Zhurakhovska, L. (2018) 'Do explicit reasons make legal intervention more effective? An experimental study', *MPI Collective Goods Preprint*, No. 2013/16.

Engel, C. and Zhurakhovska, L. (2014) 'Conditional cooperation with negative externalities: an experiment', *Journal of Economic Behavior & Organization*, 108, pp. 252–260.

Engel, C. and Zhurakhovska, L. (2016) 'When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives', *Applied Economics Letters*, 23(16), pp. 1157–1161.

Engel, C. and Zhurakhovska, L. (2017) 'You are in charge: Experimentally testing the motivating power of holding a judicial office', *The Journal of Legal Studies*, 46(1), pp. 1–50.

Engelmann, D. and Fischbacher, U. (2009) 'Indirect reciprocity and strategic reputation building in an experimental helping game', *Games and Economic Behavior*, 67(2), pp. 399–407.

Falk, A. and Fischbacher, U. (2006) 'A theory of reciprocity', *Games and Economic Behavior*, 54(2), pp. 293–315.

Falk, A., Fehr, E. and Fischbacher, U. (2008) 'Testing theories of fairness – intentions matter', *Games and Economic Behavior*, 62(1), pp. 287–303.

Forsythe, R., Horowitz, J.L., Savin, N.E. and Sefton, M. (1994) 'Fairness in simple bargaining experiments', *Games and Economic Behavior*, 6(3), pp. 347–369.

Fischbacher, U. (2007) 'z-Tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics*, 10(2), pp. 171–178.

Greiner, B. (2015) 'Subject pool recruitment procedures: organizing experiments with ORSEE', *Journal of the Economic Science Association*, 1(1), pp. 114–125.

Grosch, K., Müller, S., Rau S. and Wasserka-Zhurakhovska, L. (forthcoming) 'Ethical dilemmas in leadership: gender differences in dishonesty under responsibility', *The Leadership Quarterly*.

Kleine, M., Langenbach, P. and Zhurakhovska, L. (2016) 'Fairness and persuasion: how stakeholder communication affects impartial decision making', *Economics Letters*, 141, pp. 173–176.

Kleine, M., Langenbach, P. and Zhurakhovska, L. (2017) 'How voice shapes reactions to impartial decision-makers: an experiment on participation procedures', *Journal of Economic Behavior & Organization*, 143, pp. 241–253.

Levine, D.K. (1998) 'Modeling altruism and spitefulness in experiments', *Review of Economic Dynamics*, 1(3), pp. 593–622.

Nowak, M.A. and Sigmund, K. (1998) 'Evolution of indirect reciprocity by image scoring', *Nature*, 393, pp. 573–577.

Rabin, M. (1993) 'Incorporating fairness into game theory and economics', *American Economic Review*, 83(5), pp. 1281–1302.

Selten, R. (1967) 'Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments', in Sauermann, E. (ed.) *Beiträge zur experimentellen Wirtschaftsforschung*. Tübingen: Mohr, pp. 136–168.

Stanca, L., Bruni, L. and Corazzini, L. (2009) 'Testing theories of reciprocity: Do motivations matter?', *Journal of Economic Behavior & Organization*, 71(2), pp. 233–245.