A Theory

This Appendix is structured as follows: In Section A.1, the model is presented, and in Section A.2, we derive the equilibrium outcome for each treatment. The behavioral predictions of Section 4 are derived from the Propositions 1 - 4.

A.1 Model

The Game Played We consider a model played by two players, a manager and an employee. The manager is matched with an employee of loyalty type $l \in \{\underline{l}, \bar{l}\}$, where $\underline{l} < \bar{l}$ ($l = \underline{l}$: $\underline{l}$-employee, $l = \bar{l}$: $\bar{l}$-employee), which is private information of the employee. The manager's belief that she faces a loyal employee is $\Pr(l = \bar{l})$=$q(r)$. She has a common prior $\Pr(l = \underline{l}) = 1 - \alpha$ and $\Pr(l = \bar{l}) = \alpha$. In stage 1, the manager decides whether or not to embezzle $e \in \{0, 1\}$, which is observed by the employee. Then, the employee decides whether or not to report $r \in \{0, 1\}$, which is observed by the manager in treatments without anonymity (treatments B and I), but not in treatments with anonymity (treatments A and AI). In stage 3, the manager decides whether or not to cooperate with the employee $c \in \{0, 1\}$. If the manager cooperates, the employee decides to return $t \in [t_0, t_1]$, where $t_0 < t_1 < 1$, back to the manager. Otherwise, the game ends after stage 3.

Treatments

- In treatment B, the manager observes the reporting decision before deciding on cooperation and there is no minimum payment to the employee if the manager does not cooperate.
- In treatment I, the manager observes the reporting decision before deciding on cooperation and the employee must get at least x where $\underline{l} > x > 0$ when he reports and the manager does not cooperate.
- In treatment A, the manager does not observe the reporting decision before deciding on cooperation and there is no minimum payment to the employee if the manager does not cooperate.
- In treatment AI, the manager does not observe the reporting decision before deciding on cooperation. The employee must get at least x when he reports and the manager does not cooperate.

Payoffs All payoffs (monetary and non-monetary) are summarized in Table A.1. First, the payoff of the manager depends on whether or not she embezzles, whether or not the employee reports, whether or not she cooperates and which t is returned in case of cooperation. The manager's potential gain from embezzlement is the monetary payoff E. If the employee reports and there was embezzlement, the manager pays a net fine F. If

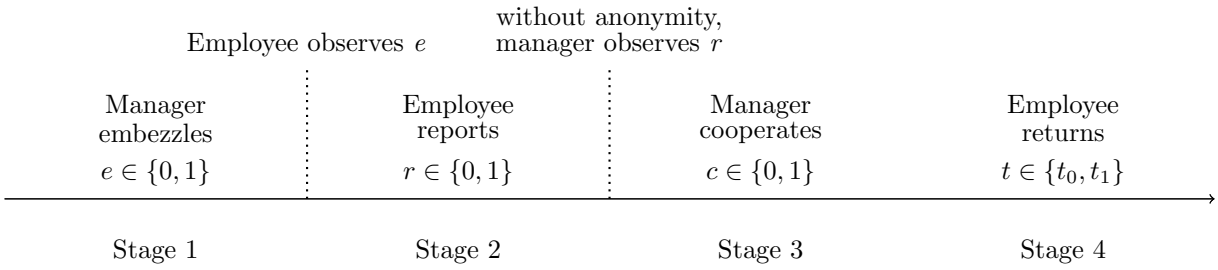

Figure A.1: Model

she chooses to cooperate, she pays an investment I to create a pie of size 1, which the employee then distributes between himself and the manager by sending either t_0 or t_1. In treatments with(out) immunity, the manager does (not) have to pay x to the employee if he has reported.

Second, the payoff of the employee depends on whether or not the manager embezzles, whether or not he reports, whether or not the manager cooperates and whether he returns t_1 or t_0 in case the manager cooperates. If the manager embezzles and the employee does not report, he faces a moral cost from undetected embezzlement $\delta = 1 - l$. Note that $\underline{l} < \bar{l} \implies \underline{\delta} > \bar{\delta}$. If the manager does not embezzle, but the employee does report, he faces a moral cost l. If the manager cooperates and the employee returns t_1 that leaves him with a payoff of $1 - t_1$. If he returns t_0 instead, he faces a moral cost l, since he did not reciprocate the cooperation decision properly.

A.2 Equilibrium Analysis

A.2.1 Preliminaries

When deriving my predictions, we focus on Perfect Bayesian Equilibria (PBE) in pure strategies (i.e., all players choose best responses given their beliefs and given the strategies of the other players, where beliefs are formed in accordance with Bayes' Rule whenever possible). More precisely, we focus on separating equilibria where an $\underline{l}(\bar{l})$-employee does (not) report when the manager chooses to embezzle. This captures the trade-off between the detection of embezzlement and signalling trustworthiness to support productive cooperation.

Assumption 1. *$\underline{l} < t_1 - t_0 < \bar{l}$, i.e. the disutility of an $\underline{l}(\bar{l})$-employee from returning the low amount t_0 is smaller (larger) than the monetary gain from returning low amount t_0 instead of high amount t_1.*

Assumption 2. *$t_0 < I < t_1$, i.e. cooperation pays off for the manager only if the employee returns the high amount t_1.*

Treatments without immunity

embezzlement	report	cooperation	return	employee	manager
0	0	0	n.a.	0	0
0	0	1	t_0	$1 - t_0 - l$	$t_0 - I$
0	0	1	t_1	$1 - t_1$	$t_1 - I$
0	1	0	n.a.	$-l$	0
0	1	1	t_0	$1 - t_0 - 2l$	$t_0 - I$
0	1	1	t_1	$1 - t_1 - l$	$t_1 - I$
1	0	0	n.a.	$-\delta$	E
1	0	1	t_0	$-\delta + 1 - t_0 - l$	$E + t_0 - I$
1	0	1	t_1	$-\delta + 1 - t_1$	$E + t_1 - I$
1	1	0	n.a.	0	$-F$
1	1	1	t_0	$1 - t_0 - l$	$-F + t_0 - I$
1	1	1	t_1	$1 - t_1$	$-F + t_1 - I$

Differences in treatments with immunity

embezzlement	report	cooperation	return	employee	manager
0	1	0	n.a.	$x - l$	$-x$
1	1	0	n.a.	x	$-F\ -x$

Table A.1: Theory Payoffs

Assumption 3. $t_1 < \bar{l}$, *i.e. the disutility from embezzlement must be smaller than the profit from cooperation for an* $\bar{l}$*-employee.*

Assumption 4. $t_0 < I - x$, *i.e. receiving the low amount* t_0 *does not pay off for the manager compared to paying the reward for reporting.*

Assumption 5. $t_1 + x < \bar{l}$, *i.e. the sum of the disutility from embezzlement and the reward for reporting must be smaller than the profit from cooperation for an* $\bar{l}$*-employee.*

A.2.2 Return behavior: Equilibrium outcome

Lemma 1. (Return behavior) *In every equilibrium in all treatments, if the manager chose* $c = 1$*, an* $\underline{l}$*-employee always chooses* $t = t_0$ *and an* $\bar{l}$*-employee always chooses* $t = t_1$*. That is,*

$$t^*(c,l) = \begin{cases} t_1 & \text{if } l = \bar{l} \text{ and } c = 1, \\ t_0 & \text{if } l = \underline{l} \text{ and } c = 1, \\ n.a. & \text{if } c = 0. \end{cases}$$

Proof. First, the employee can only choose a transfer t for $c = 1$, i.e. a pie of 1 is created. So assume that the utility of the employee in stage 4 is $U_4(t_0) = 1 - t_0 - l$ and $U_4(t_1) = 1 - t_1$. From $U_4(t_1) > U_4(t_0)$ follows $l > t_1 - t_0$. By assumption 1, a type $\bar{l}$ ($\underline{l}$) employee will choose t_1 (t_0). □

A.2.3 Treatment *Baseline*: Equilibrium Outcome

Lemma 2. (*Baseline*: Cooperation) *The manager always (never) cooperates if she chose to embezzle and the employee does not (does) send a report. If the manager didn't embezzle, she only cooperates if the share of loyal employees is high enough. That is,*

$$c^*(r,e,\alpha) = \begin{cases} 1 & \text{if } e = 1 \text{ and } r = 0, \\ 1 & \text{if } e = 0 \text{ and } r = 0 \text{ and } \alpha > \alpha', \\ 0 & else, \end{cases}$$

with $\alpha' := \frac{I - t_0}{t_1 - t_0}$.

Proof. While the profit for the manager from not cooperating in stage 3 is zero, the expected profit from cooperation for the manager in stage 3 is $\pi(c = 1) = \Pr(\bar{l}) \cdot t_1 + 1 - \Pr(\bar{l}) \cdot t_0 - I$. For cooperation to be profitable it must hold that $\pi(c = 1) > \pi(c = 0) \iff \Pr(\bar{l}) \cdot t_1 + (1 - \Pr(\bar{l})) \cdot t_0 > I$. First, we consider the case where the manager chose $e = 1$. In the candidate separating equilibrium, the reporting decision perfectly reveals the employee's type. That means $r = 1 \implies \Pr(\bar{l}) = q^*(1) = 0$ and $r = 0 \implies \Pr(\bar{l}) = q^*(0) = 1$. Therefore, by assumption 2, the manager will choose $c^*(r = 0) = 1$ and $c^*(r = 1) = 0$, since $r = 1 \implies \pi(c = 1) = t_0 - I < 0$ and $r = 0 \implies \pi(c = 1) = t_1 - I > 0$. If the manager chose $e = 0$, reporting cannot be profitable for any type. Therefore, the employee not reporting does not reveal the employee's type, such that $\Pr(\bar{l}) = q^*(0) = \alpha$. Therefore, the manager cooperates only if $\alpha > \alpha' := \frac{I - t_0}{t_1 - t_0}$. □

Lemma 3. (*Baseline*: Reporting) *An $\underline{l}$($\bar{l}$)-employee always (never) reports if the manager chooses to embezzle. Both types do not report if the manager does not embezzle.*

That is,

$$r^*(l,e) = \begin{cases} 1 & \textit{if } l = \underline{l} \textit{ and } e = 1, \\ 0 & \textit{if } l = \bar{l} \textit{ and } e = 1, \\ 0 & \textit{if } e = 0. \end{cases}$$

Proof. The employee anticipates subsequent cooperation and payment decisions, as well as beliefs by the manager. First, we consider the scenario where the manager chose $e = 1$ and faces an $\bar{l}$-employee. Since the $\bar{l}$-employee would choose t_1 in stage 4 if the manager cooperates, his utility in stage 2 is $\bar{U_2}(r = 1) = 0$ and $\bar{U_2}(r = 0) = -\bar{\delta} + 1 - t_1$. From $\bar{U_2}(r = 0) > \bar{U_2}(r = 1)$ follows $-\bar{\delta} + 1 - t_1 > 0 \iff \bar{\delta} = 1 - \bar{l} < 1 - t_1$ and therefore $\bar{l} > t_1$, which holds by assumption 3. Second, we consider the scenario where the manager chose $e = 1$ and faces an $\underline{l}$-employee. Since the $\underline{l}$-employee would choose t_0 in stage 4 if the manager cooperates, his utility in stage 2 is $\underline{U_2}(r = 1) = 0$ and $\underline{U_2}(r = 0) = -\underline{\delta} + 1 - t_0 - \underline{l}$ From $\underline{U_2}(r = 1) > \underline{U_2}(r = 0)$ follows $-\underline{\delta} + (1 - t_0) - \underline{l} < 0$ and therefore $0 > -t_0$. As we consider a scenario where loyalty costs only occur in the case of a false report for both types, neither type reports when there is no embezzlement. In consequence, a type $\underline{l}(\bar{l})$-employee will optimally choose (not) to report if the manager embezzles and neither type reports if the manager does not embezzle. □

Lemma 4. (*Baseline*: Embezzlement) *A manager only chooses to embezzle if the share of loyal employees α is sufficiently high. That is,*

$$e^*(\alpha) = \begin{cases} 1 & \textit{if } l = \alpha > \alpha' \textit{ and } \alpha > {\alpha_B}'', \\ 1 & \textit{if } l = \alpha < \alpha'' \textit{ and } \alpha > {\alpha_B}''', \\ 0 & \textit{else}, \end{cases}$$

with $\alpha' := \frac{I - t_0}{t_1 - t_0}$, ${\alpha_B}'' := \frac{t_0 - I + F}{t_0 - I + F + E}$ *and* ${\alpha_B}''' := \frac{F}{t_1 - I + F + E}$.

Proof. If the manager chooses $e = 0$, both types would not report and the decision about c depends on the size of α. If $\alpha > \alpha'$ the manager would cooperate and receive an expected payoff of $\pi(c = 1) = \alpha \cdot t_1 + (1 - \alpha) \cdot t_0 - I > 0$. If $\alpha < \alpha'$ the manager would get $\pi(c = 0) = 0$. Therefore, whether $e = 1$ is profitable depends as well on the size of α. Given $\alpha > \alpha'$, for embezzlement to be profitable it must hold that $\pi(e = 1) > \pi(e = 0) \iff \pi(c = 1) = \alpha \cdot (t_1 + E - I) + (1 - \alpha) \cdot (-F) > \alpha \cdot t_1 + (1 - \alpha) t_0 - I \iff \alpha > {\alpha_B}'' := \frac{t_0 - I + F}{t_0 - I + F + E}$. Given $\alpha < \alpha'$, for embezzlement to be profitable it must hold that $\pi(e = 1) > \pi(e = 0) \iff \pi(c = 1) = \alpha \cdot (t_1 + E - I) + (1 - \alpha) \cdot (-F) > 0 \iff \alpha > {\alpha_B}''' := \frac{F}{t_1 - I + F + E}$. □

Proposition 1. (*Baseline*: Equilibrium Outcome) *The Baseline treatment has the following equilibrium outcome: (i) An $\underline{l}(\bar{l})$-employee always (never) reports if the manager chooses to embezzle. (ii) Any employee does not report if the manager does not embezzle. (iii) A manager never cooperates if the employee sent a report. (iv) A manager cooperates if she embezzled and the employee did not send a report or if she didn't embezzle and the share of $\bar{l}$-employees α is larger than $\alpha' := \frac{I-t_0}{t_1-t_0}$. (v) A manager does only embezzle, if α is larger than ${\alpha_B}'' := \frac{t_0-I+F}{t_0-I+F+E}$ and if $\alpha > \alpha'$, or if α is larger than ${\alpha_B}''' := \frac{F}{t_1-I+F+E}$ and if $\alpha < \alpha'$. (vi) An employee of type $\bar{l}(\underline{l})$ always chooses $t = t_1$ $(t = t_0)$.*

A.2.4 Treatment *Immunity*: Equilibrium Outcome

Lemma 5. (*Immunity*: Cooperation) *The manager always (never) cooperates if she chose to embezzle and the employee does not (does) send a report. If the manager didn't embezzle she only cooperates if the share of loyal employees is high enough. That is,*

$$c^*(r,e,\alpha) = \begin{cases} 1 & \text{if } e = 1 \text{ and } r = 0, \\ 1 & \text{if } e = 0 \text{ and } r = 0 \text{ and } \alpha > \alpha', \\ 0 & else, \end{cases}$$

with $\alpha' := \frac{I-t_0}{t_1-t_0}$.

Proof. As in the baseline treatment, if the manager chose $e = 0$, reporting cannot be profitable for any type since $x < \underline{l} < \bar{l}$. Therefore, not reporting does not reveal the employee's type such that $\Pr(\bar{l}) = q^*(0) = \alpha$. Therefore, the manager cooperates only if $\alpha > \alpha' := \frac{I-t_0}{t_1-t_0}$. For $e = 1$, the cooperation decision has to be evaluated differently given the reporting decision of the employee, since reporting is incentivized. First, if the employee does not report, the scenario is identical to the baseline treatment. Second, if the employee does report, the payoff for the manager from not cooperating is now $\pi(c = 0) = -x$. The expected profit from cooperation for the manager in stage 3 is still $\pi(c = 1) = \Pr(\bar{l}) \cdot t_1 + (1 - \Pr(\bar{l})) \cdot t_0 - I$. For cooperation to be profitable it must hold that $\pi(c = 1) > \pi(c = 0) \iff \Pr(\bar{l}) \cdot t_1 + (1 - \Pr(\bar{l})) \cdot t_0 > I - x$. In the candidate separating equilibrium, the reporting decision perfectly reveals the employee's type. That means $r = 1 \implies \Pr(\bar{l}) = q^*(1) = 0$ and $r = 0 \implies \Pr(\bar{l}) = q^*(0) = 1$. Therefore, by assumption 4 the manager will choose $c^*(r = 0) = 1$ and $c^*(r = 1) = 0$, since $r = 0 \implies \pi(c = 1) = t_1 - I > 0$ and $r = 1 \implies \pi(c = 1) = t_0 - I + x < 0$ (which is harder to sustain compared to the baseline treatment). □

Lemma 6. (*Immunity*: Reporting) *An employee of type $\underline{l}$ ($\bar{l}$) always (never) reports if the manager chooses to embezzle. Both types do not report if the manager does not embezzle. That is,*

$$r^*(l) = \begin{cases} 1 & \text{if } l = \underline{l} \text{ and } e = 1, \\ 0 & \text{if } l = \bar{l} \text{ and } e = 1, \\ 0 & \text{if } e = 0. \end{cases}$$

Proof. Suppose assumption 4 holds: Again, we first consider the scenario where the manager chose $e = 1$ and faces an employee of type $l = \bar{l}$. Since the employee would choose t_1 in stage 4 if the manager cooperates, his utility in stage 2 is $\bar{U}_2(r = 1) = x$ and $\bar{U}_2(r = 0) = -\bar{\delta} + 1 - t_1$. From $\bar{U}_2(r = 0) > \bar{U}_2(r = 1)$ follows $-\bar{\delta} + 1 - t_1 > x \iff \bar{\delta} = 1 - \bar{l} < 1 - t_1 - x$ and therefore $\bar{l} > t_1 + x$, which holds by assumption 5. Second, we consider the scenario where the manager chose $e = 1$ and the employee is of type $l = \underline{l}$. Since the employee would choose t_0 in stage 4 if the manager cooperates, his utility in stage 2 is $\underline{U}_2(r = 1) = x$ and $\underline{U}_2(r = 0) = -\underline{\delta} + 1 - t_0 - \underline{l}$ From $\underline{U}_2(r = 1) > \underline{U}_2(r = 0)$ follows $-\underline{\delta} + 1 - t_0 - \underline{l} < x$ and therefore $x > -t_0$. As we consider a scenario where the reward x is smaller than the loyalty costs–which occur in case of a false report–for both types, neither type reports when there is no embezzlement. In consequence, a type $\underline{l}(\bar{l})$ employee will optimally choose (not) to report if the manager embezzles and neither type reports if the manager does not embezzle. □

Lemma 7. (*Immunity*: Embezzlement) *A manager only chooses to embezzle if the share of loyal employees α is sufficiently high. That is,*

$$e^*(\alpha) = \begin{cases} 1 & \text{if } l = \alpha > \alpha' \text{ and } \alpha > {\alpha_I}'', \\ 1 & \text{if } l = \alpha < \alpha'' \text{ and } \alpha > {\alpha_I}''', \\ 0 & \text{else}, \end{cases}$$

with $\alpha' := \frac{I - t_0}{t_1 - t_0}$, ${\alpha_I}'' := \frac{t_0 - I + F + x}{t_0 - I + F + E} + x$ and ${\alpha_I}''' := \frac{F + x}{t_1 - I + F + E + x}$.

Proof. If the manager chooses $e = 0$, both types would not report and the decision about c depends on the size of α. If $\alpha > \alpha'$, the manager would cooperate and receive an expected payoff of $\pi(c = 1) = \alpha \cdot t_1 + (1-\alpha) \cdot t_0 - I > 0$. If $\alpha < \alpha'$ the manager would get $\pi(c = 0) = 0$. Therefore, whether $e = 1$ is profitable depends on the size of α as well. Given $\alpha > \alpha'$, for embezzlement to be profitable it must hold that $\pi(e = 1) > \pi(e = 0) \iff \pi(c = 1) = \alpha \cdot (t_1 + E - I) + (1-\alpha) \cdot (-F - x) > \alpha \cdot t_1 + (1-\alpha) t_0 - I \iff \alpha > {\alpha_I}'' := \frac{t_0 - I + F + x}{t_0 - I + F + E + x}$. Given $\alpha < \alpha'$, for embezzlement to be profitable it must hold that $\pi(e = 1) > \pi(e = 0) \iff$

$\pi(c=1) = \alpha \cdot (t_1 + E - I) + (1-\alpha) \cdot (-F - x) > 0 \iff \alpha > \alpha_I''' := \frac{F+x}{t_1 - I + F + E + x}$. □

Proposition 2. (*Immunity*: Equilibrium Outcome) *The immunity treatment has the following equilibrium outcome: (i) An $\underline{l}(\bar{l})$-employee always (never) reports if the manager chooses to embezzle. (ii) Any employee does not report if the manager does not embezzle. (iii) A manager never cooperates if the employee sent a report. (iv) A manager cooperates if she embezzled and the employee did not send a report or if she didn't embezzle and the share of $\bar{l}$-employees α is larger than $\alpha' := \frac{I - t_0}{t_1 - t_0}$. (v) A manager does only embezzle, if α is larger than $\alpha_I'' := \frac{t_0 - I + F + x}{t_0 - I + F + E + x}$ and if $\alpha > \alpha'$, or if α is larger than $\alpha_I''' := \frac{F+x}{t_1 - I + F + E + x}$ and if $\alpha < \alpha'$. (vi) An employee of type $\bar{l}(\underline{l})$ always chooses $t = t_1$ $(t = t_0)$.*

A.2.5 Treatment *Anonymity*: Equilibrium Outcome

Lemma 8. (*Anonymity*: Cooperation) *The manager only cooperates if the share of loyal employees is high enough. That is,*

$$c^*(\alpha) = \begin{cases} 1 & \text{if } \alpha > \alpha', \\ 0 & else, \end{cases}$$

with $\alpha' := \frac{I - t_0}{t_1 - t_0}$.

Proof. In the anonymity treatment, the reporting decision does not convey any information about the type of the employee. The crucial condition for the cooperation decision of the manager is therefore: $\alpha \cdot t_1 + (1-\alpha) \cdot t_0 \lessgtr I$. First, we consider $\alpha < \alpha' := \frac{I - t_0}{t_1 - t_0}$. If the manager chose $e = 1$, both types will report the embezzlement, because they thereby avoid the disutility from undetected embezzlement without any other consequences. the belief of the manager is $\Pr(\bar{l}) = q(1) \implies \alpha \implies c^* = 0$. As before, if the manager chose $e = 0$, reporting cannot be profitable for any type since $x < \underline{l} < \bar{l}$. Therefore, the employee not reporting does not reveal his type, such that $\Pr(\bar{l}) = q(0) \implies \alpha \implies c^* = 0$. Second, we consider $\alpha > \alpha' := \frac{I - t_0}{t_1 - t_0}$. If the manager chose $e = 1$, both types will report the embezzlement, because they avoid the disutility from undetected embezzlement and do not affect the cooperation decision of the manager. That means, the belief of the manager is $\Pr(\bar{l}) = q(1) \implies \alpha \implies c^* = 1$. For $e = 0$, still, reporting cannot be profitable for any type. Therefore, the not reporting employee does not reveal his type, such that $\Pr(\bar{l}) = q(0) \implies \alpha \implies c^* = 1$. □

Lemma 9. (*Anonymity*: Reporting) *Any employee reports if the manager chooses to embezzle. Both types do not report if the manager does not embezzle. That is,*

$$r^* = \begin{cases} 1 & \text{if } e = 1, \\ 0 & \text{if } e = 0. \end{cases}$$

Proof. First, we consider $\alpha < \alpha' := \frac{I-t_0}{t_1-t_0}$. If the manager chose $e = 1$, the utility of the employee in stage 2 is $U_2(r = 1) = 0$ and $U_2(r = 0) = -\delta \implies r^*(l) = 1$. Second, we consider $\alpha > \alpha' := \frac{I-t_0}{t_1-t_0}$. If the manager chose $e = 1$, the utility of an $\underline{l}$-employee in stage 2 is $\underline{U}_2(r = 1) = 1-t_0-\underline{l}$ and $\underline{U}_2(r = 0) = -\underline{\delta}+1-t_0-\underline{l} \implies r^*(l) = 1$. For an $\bar{l}$-employee, the utility in stage 2 is $\bar{U}_2(r = 1) = 1 - t_1$ and $\bar{U}_2(r = 0) = -\bar{\delta} + 1 - t_1 \implies r^*(l) = 1$. As before, if the manager chose $e = 0$, reporting cannot be profitable for any type. In consequence, any employee will optimally choose (not) to report if the manager does (not) embezzle. □

Lemma 10. (*Anonymity*: Embezzlement) *A manager never embezzles. That is,* $e^* = 0$.

Proof. If the manager chooses $e = 1$, both types would report and she would make a loss for sure since her cooperation decision is independent of the reporting decision such that $\pi(e = 0) = \alpha \cdot t_1 + (1 - \alpha) \cdot t_0 - I > \alpha \cdot t_1 + (1 - \alpha) \cdot t_0 - I - F = \pi(e = 1) \iff 0 > -F \implies e^* = 0$. □

Proposition 3. (*Anonymity*: Equilibrium Outcome) *The anonymity treatment has the following equilibrium outcome: (i) Any employee does not report if the manager does not embezzle. (ii) A manager cooperates if the share of $\bar{l}$-employees α is larger than $\alpha' := \frac{I-t_0}{t_1-t_0}$. (iii) A manager never embezzles. (iv) An employee of type $\bar{l}(\underline{l})$ always chooses $t = t_1$ $(t = t_0)$.*

A.2.6 Treatment *Anonymity and Immunity*: Equilibrium Outcome

Note that the only difference between treatments "Anonymity" and "Anonymity and Immunity" is that, in the latter, reporting is rewarded in the case where cooperation does not take place. Since both types of employees report if and only if the manager embezzled, embezzlement is already completely deterred by the provision of anonymity. In consequence, both types do not report and the reward does not come into effect. It follows that the respective equilibrium outcomes are the same in both treatments:

Proposition 4. (*Anonymity and Immunity*: Equilibrium Outcome) *In the treatments "Anonymity" and "Anonymity and Immunity", the equilibrium outcomes coincide.*

B Supplementary Material

B.1 Translated Instructions

Welcome to today's experiment! If you read the following instructions carefully, you can earn a significant payment - depending on your decisions.

Please note, that from now on and during the whole experiment no communication is allowed. If you have any questions, please direct these at one of the experimenters. Neglecting these rules results in exclusion from this experiment and all payments.

All your decisions during this experiment will remain anonymous and cannot be related to you by either the experimenters nor the fellow subjects. Your earnings will be accounted in points. The points you acquire during this experiment will be exchanged for euro at the end. The exchange rate is: **10 points = 50 eurocent**.

General procedure:

There are **three roles** in this experiment: *Manager*, *employee* and *a third party*. These roles are assigned randomly. If you are drawn into the role *manager*, you'll maintain this role throughout the entire experiment. If you start with one of the other two roles, your role will be drawn randomly before each period. In each period you are part of a group consisting of exactly one manager, one employee and one third party. Also the group composition will result from a random draw in every period.

The experiment is divided into two parts consisting of multiple periods. Beneath you find the procedure of a period in part 1. For the second part, you'll receive instructions on your screen immediately before it starts.

Procedure of a period in part 1:

Every subject is endowed with 100 points. After the roles are assigned, the manager chooses between two alternatives (CIRCLE or TRIANGLE). CIRCLE has no payoff consequences for any member of the group. TRIANGLE represents violating the law, resulting in a gain (50 points) for the *manager*, and a loss (90 points) for the *third party*. Again, there are no consequences for the *employee*.

After the manager has made her choice about CIRLCE and TRIANGLE, the employee has to decide whether she wants to file a complaint. This decision is taken separately for both alternatives (complaint if CIRCLE was chosen; complaint if TRIANGLE was chosen). Filing a complaint causes costs for the manager in any case (10 points). If TRIANGLE has been chosen and a complaint has been filed, the manager has to pay an additional fine (60 points). The third party receives partial compensation for her damage

(80 points).

The table below displays all possible combinations of the decisions made by the manager and the employee as well as its respective payoffs for all group members.

Manager chooses alternative	*Employee* files a complaint	Payoffs		
		Manager	*Employee*	*Third Party*
Circle	No	0	0	0
Circle	Yes	-10	0	0
Triangle	No	50	0	-90
Triangle	Yes	-20	0	-10

Subsequently, all group members are informed about the chosen alternative[and whether there has been a complaint].

To conclude a period, the manager and the employee play an investment game. First, the manager chooses an amount x between -30 and 60 points. Negative numbers mean that points are taken from the employee. Positive numbers mean that points are sent to the employee. If the manager deducts points from the employee these points are transferred and the investment game ends. If the manager sends a positive amount to the employee, it will be multiplied by three. In this case, the employee chooses an amount y between 0 and $3 \cdot$ x which she would like to return to the manager. There are no consequences for the third party in the investment game.

Payoffs in the investment game:

Manager = - x + y points,

Employee = max(x, $3 \cdot$ x) - y points,

Third party = 0.

At the end of a period [all of the group members are informed whether there was a complaint] your surplus adds up from your **endowment** (100 points), **your revenue from the decisions made** (see table) and **your revenue from the investment game**.

Summary of a period in part 1

1. Manager chooses alternative CIRLCE or TRIANGLE (violation of law)
2. Employee decides upon reporting
3. Every member of a group learns about the chosen alternative [and the reporting

decision]

4. Manager and employee engage in an investment game

(5. Every member of a group learns about the reporting decision)

5./6. The surplus is computed

After you have completed the second part and a questionnaire, **one period** is drawn for payout. You'll receive the points you earned in that period converted according to the exchange rate plus 5 euro as show up fee.

Thank you for participating and good luck!

B.2 Control Questions

1. Do you keep your role through the entire experiment?
   ☐ Yes, always.
   ☐ No, my role is randomly drawn in each period.
   ☐ Yes, in case I am a manager. If I am an employee or the third party, it may change from period to period.
2. Do you have the same members in your group over several periods?
   ☐ No.
   ☐ Yes, in the second part of the experiment.
   ☐ Yes, always.
3. If the manager chooses TRIANGLE, ...
   ☐ she receives a profit and harms the employee as well as the third party.
   ☐ she does not receive a profit, but harms the employee as well as the third party.
   ☐ she receives a profit and harms the third party, but not the employee.
4. If the manager chooses CIRCLE and the employee files a report, ...
   ☐ all payoffs are unaffected.
   ☐ it causes a cost for the manager. Both the employee and the third party are not affected.
   ☐ it causes a cost for the manager. Both the employee and the third party receive a profit.
5. If the manager sends 30 points in the investment game, how many points does the employee receive?
   _____

B.3 Questionnaire

Demographics

1. How old are you? _____
2. What is your sex? ☐ Male ☐ Female

3. What are you studying? _____
4. How much work expericence do you have?
   (a) Internships (in month): _____
   (b) Full-time (in month): _____
   (c) Student jobs (in month): _____

Risk preferences

1. Imagine you had won 100,000 euros in a lottery. Almost immediately after you collect your winnings, you receive the following financial offer from a reputable bank, the conditions of which are as follows: There is the chance to double the money within two years. It is equally possible that you could lose half of the amount invested. What fraction would you choose to invest?
   ☐ 0 ☐ 20,000 ☐ 40,000 ☐ 60,000 ☐ 80,000 ☐ 100,000

Attitudes towards whistleblowing

1. What is your opinion with respect to the following claims?
   (a) A person should be supported in disclosing serious misbehavior, even if this requires disclosure of insider information.
   ☐ Strongly agree ☐ Agree ☐ No opinion ☐ Disagree ☐ Strongly disagree
   (b) A person should be supported in disclosing already mild misbehavior, even if this requires disclosure of insider information.
   ☐ Strongly agree ☐ Agree ☐ No opinion ☐ Disagree ☐ Strongly disagree
   (c) I would disclose serious misbehavior, even it would cause disadvantages for me.
   ☐ Strongly agree ☐ Agree ☐ No opinion ☐ Disagree ☐ Strongly disagree
   (d) I would disclose already mild misbehavior, even it would cause disadvantages for me.
   ☐ Strongly agree ☐ Agree ☐ No opinion ☐ Disagree ☐ Strongly disagree
   (e) If the chance is larger that misbehavior is detected it could be deterred.
   ☐ Strongly agree ☐ Agree ☐ No opinion ☐ Disagree ☐ Strongly disagree
2. In your opinion, how acceptable are the following actions?
   (a) Disclosing insider information about serious misbehavior by a person in authority of an organization.
   ☐ Very acceptable ☐ Acceptable ☐ Neither, nor ☐ Unacceptable ☐ Very unacceptable
   (b) Disclosing insider information about serious misbehavior by regular employees of an organization.
   ☐ Very acceptable ☐ Acceptable ☐ Neither, nor ☐ Unacceptable ☐ Very unacceptable

(c) Disclosing insider information about serious misbehavior by a friend or family member of an organization's member.
☐ Very acceptable ☐ Acceptable ☐ Neither, nor ☐ Unacceptable ☐ Very unacceptable

3. Imagine you had insider information about serious misbehavior in an organization you are a member of. How important was each of the following items for the decision to tell someone about it?
   (a) Persons in authority would support me.
   ☐ Very important ☐ Important ☐ Neither, nor ☐ Unimportant ☐ Very unimportant
   (b) I would be legally obliged to report.
   ☐ Very important ☐ Important ☐ Neither, nor ☐ Unimportant ☐ Very unimportant
   (c) Somebody would act to end the misbehavior.
   ☐ Very important ☐ Important ☐ Neither, nor ☐ Unimportant ☐ Very unimportant
   (d) Only people I choose would know my identity.
   ☐ Very important ☐ Important ☐ Neither, nor ☐ Unimportant ☐ Very unimportant
   (e) Apart from the people I contact, the information would remain confidential.
   ☐ Very important ☐ Important ☐ Neither, nor ☐ Unimportant ☐ Very unimportant
   (f) I would remain completely anonymous.
   ☐ Very important ☐ Important ☐ Neither, nor ☐ Unimportant ☐ Very unimportant

C Descriptive Statistics and Survey Responses

Table C.1 displays the average characteristics of the subjects cut by treatment and role.

	Anonymity		No Anonymity	
	Manager	Employee	Manager	Employee
characteristic				
age	24.0	24.2	24.8	25.9
female	0.79	0.63	0.77	0.57
risk	1.16	1.34	1.20	0.95
work experience	0.17	-0.15	0.01	0.06
attitude reporting	-0.15	0.09	-0.26	0.03
attitude disclosure	-0.04	0.10	-0.13	-0.02
attitude environment	0.00	-0.03	-0.06	0.03
No. of subjects	19	38	30	60

Notes: The table reports the average characteristics of the subjects per treatment and role. *risk* is measured on a scale from 0-5, where 5 is extremely risk-loving. *work experience* is a standardized measure of the answers to question 4 in the "demographics" section of the questionnaire, where a higher score represents more month of work experience. *attitude reporting* is a standardized measure of the answers to question 1 in the "attitudes towards whistleblowing" section of the questionnaire, where a higher score represents a stronger support for whistleblowing. *attitude disclosure* is a standardized measure of the answers to question 2 in the "attitudes towards whistleblowing" section of the questionnaire, where a higher score represents a greater appropriateness of disclosing insider information. *attitude reporting* is a standardized measure of the answers to question 3 in the "attitudes towards whistleblowing" section of the questionnaire, where a higher score represents a greater importance of the legal environment for the decision to become a whistleblower.

Table C.1: Average characteristics per role over treatments

D Regression Analysis

In this section, we present the regression results for the decisions on embezzlement, truthful and false reporting, and the return behavior, which are discussed in Section 5.

	Truthful reporting			False reporting		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment A	0.131**	0.143**	0.144**	0.0893	0.111*	0.111*
	(0.0591)	(0.0610)	(0.0610)	(0.0569)	(0.0586)	(0.0593)
Treatment I	0.167***	0.189***	0.194***	0.194***	0.187***	0.206***
	(0.0437)	(0.0466)	(0.0473)	(0.0463)	(0.0475)	(0.0500)
Treatment AI	0.189***	0.213***	0.218***	0.421***	0.412***	0.433***
	(0.0610)	(0.0629)	(0.0634)	(0.0661)	(0.0658)	(0.0674)
CooperatingManager(lag)		0.0373	0.0238		-0.0982**	-0.0873**
		(0.0394)	(0.0389)		(0.0431)	(0.0415)
Constant	0.704***	0.675***	0.630***	0.118***	0.148***	-0.00979
	(0.0464)	(0.0530)	(0.0678)	(0.0299)	(0.0374)	(0.0728)
Period FE	No	No	Yes	No	No	Yes
N	784	735	735	784	735	735
N_{groups}	98	98	98	98	98	98
R^2	0.0343	0.0426	0.0502	0.104	0.113	0.129

Notes: The table reports results from a random-effects GLS regression where N_{groups} is the number of individuals. Dependent variable: (1)-(3): willingness to report truthfully (0 or 1), (4)-(6): willingness to report falsely (0 or 1). Standard errors in parentheses are clustered on the individual level. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table D.1: Regression Analysis: Reporting

	Embezzlement		
	(1)	(2)	(3)
Treatment A	-0.0967	-0.0127	-0.0103
	(0.0855)	(0.102)	(0.106)
Treatment I	-0.171***	-0.0134	-0.0204
	(0.0501)	(0.0600)	(0.0688)
Treatment AI	-0.334***	-0.0657	-0.0735
	(0.0749)	(0.0890)	(0.0885)
Embezzlement(lag)		0.708***	0.738***
		(0.0815)	(0.0846)
Report(lag)		0.0721	0.102
		(0.0912)	(0.0962)
ReportedEmbezzlement(lag)		-0.613***	-0.655***
		(0.135)	(0.137)
LowReturn(lag)		-0.0204	-0.0447
		(0.0696)	(0.0701)
Constant	0.412***	0.243***	0.166
	(0.0673)	(0.0782)	(0.140)
Period FE	No	No	Yes
N	784	206	206
N_{groups}	49	44	44
R^2	0.0694	0.163	0.198

Notes: The table reports results from a random-effects GLS regression where N_{groups} is the number of individuals. Dependent variable: embezzlement decision of managers (0 or 1). Standard errors in parentheses are clustered on the individual level. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table D.2: Regression Analysis: Embezzlement

	Returning more than sent		
	(1)	(2)	(3)
Treatment A	-0.0407	-0.0313	-0.0417
	(0.103)	(0.102)	(0.103)
Treatment I	-0.124**	-0.101*	-0.0982
	(0.0603)	(0.0599)	(0.0651
Treatment AI	-0.214*	-0.173	-0.154
	(0.122)	(0.132)	(0.135)
Embezzlement		-0.0968	-0.0927
		(0.0857)	(0.0782)
FalseReport		-0.183**	-0.208**
		(0.0902)	(0.0940)
UnreportedEmbezzlement		-0.0121	-0.0612
		(0.128)	(0.127)
Constant	0.540***	0.586***	0.690***
	(0.0627)	(0.0689)	(0.0913)
Period FE	No	No	Yes
N	212	212	212
N_{groups}	87	87	87
R^2	0.00633	0.0255	0.0504

Notes: The table reports results from a random-effects GLS regression where N_{groups} is the number of individuals. Dependent variable: (1)-(3): employees return more than what the manager sent (0 or 1) . ReportedEmbezzlement, FalseReport, UnreportedEmbezzlement are all binary variables. Standard errors in parentheses are clustered on the individual level. Significance levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table D.3: Regression Analysis: Return Behavior