

**Verena Utikal**

# A fault confessed is half redressed - Confessions and Punishment

Research Paper Series  
Thurgau Institute of Economics and Department of Economics  
at the University of Konstanz

Member of

---

---

---

**thurgauwissenschaft**

[www.thurgau-wissenschaft.ch](http://www.thurgau-wissenschaft.ch)



**THURGAU INSTITUTE  
OF ECONOMICS**  
at the University of Konstanz

# **A fault confessed is half redressed**

## **- Confessions and Punishment\***

26 October 2010

Verena Utikal<sup>1</sup>

### **Abstract**

Confessions after failures are socially desirable. However, confessions also bear the risk of punishment. In a laboratory experiment I examine how confessions work. I analyze whether the willingness to punish harmful failures depends on how the harmed party has learned about the outcome. The harmed party can learn about the outcome via random detection or self-report by the performer. I find that confessions are a powerful instrument: Punishment for confessed failures is less likely than for randomly detected failures.

**JEL Classification:** C91, D82, D83.

**Keywords:** Cheap Talk, Confession, Experiment, Intentions.

---

\* I would like to thank Kate Bendrick, Lisa Bruttel, Gerald Eisenkopf, Urs Fischbacher, Simeon Schudy, Pascal Sulser, Irenaeus Wolff and the participants of a conference in Granada for helpful comments. Support from the Center of Psychoeconomics at the University of Konstanz is gratefully acknowledged.

<sup>1</sup> Department of Economics, University of Konstanz, 78457 Konstanz, Germany,  
verena.utikal@uni-konstanz.de

## 1 Introduction

Imagine you are a worker in a company. After a failure in performance, you have two possibilities. Either you can remain quiet hoping that nobody ever knows or you can confess your shortcoming which is of course socially desirable. All major theories of moral decision making (Freud (1940), Kohlberg (1980)) emphasize the centrality of truthfulness. Not confessing can be seen as a lie. On the other hand, in confessing you expose yourself to potentially ruthless punishment. So what should you do? Confess or remain quiet?

In general, there are two types of confessions. There are spontaneous confessions and confessions after accusation. Both types are “a detailed written or oral statement in which a person admits to having committed some transgression” (Kassin and Gudjonsson (2004)). In this paper I focus on spontaneous confessions. A spontaneous confession is the transfer of private information, up to now unknown to the receiver. To spontaneously confess a failure is transferring private information such that the receiver can give his best response to the situation. By this means a confession can prevent further costs for the receiver. However, decreasing the receiver’s future costs comes along with risk of punishment for the confessor. Nevertheless, there are reasons why confessors might be punished less than offenders who are randomly detected. First, offenders who do not confess value the other’s payoff less than they fear punishment. Confessing is therefore a kind action and confessors deserve less punishment. A second crucial argument why confessions can reduce punishment is ambiguity of intentions. People respond to the intentionality of an action and punish less if an offense was caused unintentionally (Brandts and Sola (2001), Offerman (2002), Charness (2004), Charness and Rabin (2005), Charness and Levine (2007), Blount (1995), and Falk et al. (2008), Nelson Jr. (2002)), Charness and Rabin (2002), Falk et al. (2003)). After offenses with ambiguous intentionality a confession is a cheap device to declare one’s noble intentions and to therefore reduce punishment. This reasoning is also part of taxation law. In many countries tax evaders who are detected through a (random) investigation not only have to balance their debts, but have to additionally pay a penalty. Tax evaders who voluntarily report their tax evasion do not have to pay this penalty<sup>2</sup>.

---

<sup>2</sup> A particular interesting procedure is tax amnesty. For an overview on international tax amnesties see Torgler (2003).

*How* an offense is revealed is therefore very relevant. Yet how people's punishment decisions differ with respect to spontaneous confession and random detection of offenses has not been examined. So far, existing psychological studies compare behavior after spontaneous confessions and confessions after accusation<sup>3</sup>. The comparison between a spontaneous confession and a random detection of an offense is missing in the literature. This is an important question since it describes the most common and basic situation. The decision between spontaneous confession and silence is indispensable for every offender. Only some offenders also have to decide whether to confess after accusation, but all of them have to choose whether to confess spontaneously or to risk random detection. Therefore the fundamental question is: How do spontaneous confession and random detection of an offense affect punishment?

To answer this question I conduct a laboratory experiment. There are two related strings of literature that deal with how reports of private information affect punishment and cooperation. In the strategic information literature people have private information about an outcome chosen by nature which they can report. The deception-analyzing literature uses pre-play announcements. My contribution is a more natural design. In my experiment a person (the agent) rather than nature decides the outcome. The outcome (failure or success) is unknown and crucial information for another person's (the principal's) best response, but is private information to the agent. I compare punishment after two possible detections of failure. The agent can report the failure to the principal and the failure can also be randomly detected. In the design the intentions leading to failures are ambiguous. A failure can have two reasons. Either it can happen because a person intended to fail and did not even try to succeed. Second, the failure can happen due to inability. Either the failure is revealed via confession or through random detection.

I find that confessions are a powerful instrument: Confessed faults are indeed significantly redressed. However, the results suggest that people do not partly forgive. After a report, it seems like people either stop punishing or they punish nevertheless and in fact to the same degree as without report. The rest of this paper is organized as follows. Section 2 summarizes the related literature. Section 3 explains the experimental design and procedures. Section 4 gives theoretical predictions and Section 5 presents the results. Section 6 concludes.

---

<sup>3</sup> See chapter 2 for an overview.

## 2 Related Literature

In the psychological literature there are several studies on confessions. They are concerned with behavior after spontaneous confessions and confessions after accusation. Several studies use vignettes (Darby and Schlenker (1982), Kerby and Johnson (2005), Takaku et al. (2001), Gold and Weiner (2000), Robinson et al. (1994)). Readers of the vignettes have to state how they perceive the character of the offender, or how they attribute responsibility for the act, forgiveness and hypothetical punishment. Usually confessions without a prior accusation and in ambiguous causal situations have strong beneficial effects on the perceptions of the offender's trustworthiness and on sympathy and forgiveness. Weiner et al. (1991) use a laboratory manipulation study to compare cooperation after spontaneous confession and confession after accusation. They find the strongest cooperation and greatest forgiveness towards spontaneous confessors. Accused confessors do not elicit greater forgiveness than deniers.

Numerous studies have indicated that apologies – just like confessions - can mitigate anger and increase forgiveness (Ohbuchi and Sato (1994), Girard et al. (2002), Ohtsubo and Watanabe (2009), Wada (1998), Scher and Darley (1997), Exline et al. (2007), McCullough et al. (1997), McCullough et al. (1998), Schmitt et al. (2004), Ohbuchi et al. (1989), Struthers et al. (2008), Schweitzer et al. (2006), Ho (2007), Fischbacher and Utikal (2010), Bottom et al. (2002)), although confessions seem to be more effective (Holtgraves (1989), Hale (1987), Felson and Ribner (1981)). Apologies and confessions are apparently similar concepts but differ in an important aspect. It is true that both apologies and confessions can help to reveal the intention behind the preceding offense. The victim usually does not know whether the harmdoer is a friendly person who harmed accidentally or whether the harmdoer intended to harm. However, the victim's punishment decision strongly depends on the offender's type. Punishment for intentional offenses is higher than for accidental harm. An apology or a confession can be an offender's only chance to inform the victim about his type. However there is one essential difference between apologies and confessions. An apology cannot undo the offense and it cannot repair the caused damage either. In contrast, by transferring private information a confession can prevent further harm.

Signals with respect to private information are also used in the economic literature on strategic information (Dickhaut et al. (1995), Cai and Wang (2006), Sánchez-Pagés and Vorsatz (2007), Sánchez-Pagés and Vorsatz (2009), Gneezy (2005), Hurkens and

Kartik (2009)). These experiments use strategic information games, introduced by Crawford and Sobel (1982). The games have the following structure: A sender with private information about a state of nature sends a message about this state to the receiver, who can decide about a following action, which is payoff relevant for both players and depends on the actual state of nature. The main theoretical prediction by Crawford and Sobel (1982) is that the less aligned the preferences of the sender and the receiver the less information revealed about the true state by the sender's message. Experimental evidence of the literature mentioned above supports this prediction and shows that senders tell the truth more often than in the sequential equilibrium.

Not confessing a failure can be considered as a lie. This paper can therefore well be integrated into studies analyzing deception through pre-play communication (Brandts and Charness (2003), Charness and Dufwenberg (2006), Gneezy (2005)). In these games in the first stage one player sends a message indicating an intended play, which is either favorable or unfavorable for the other player. In the second stage the players choose their actions. The sender may or may not play according to his message. In the third stage the receiver may reward or punish the sender. The results show that people are liar and lying averse. That means that people have a preference for truth telling - for themselves and also with respect to others. These findings are also supported by Hurkens and Kartik (2009), Charness and Dufwenberg (2006), Lundquist et al. (2009), Sánchez-Pagés and Vorsatz (2007), and Vanberg (2008). Sánchez-Pagés and Vorsatz (2009) show that people are even willing to forgo positive payoffs in order to remain silent.

For a comprehensive review on the psychology of false confessions in court proceedings see Kassin and Gudjonsson (2004). Horowitz (1956) analyzes the psychological situation of a person who confesses an act of which she has been accused. There is also a broad literature in educational sciences that discusses children's lying behavior and how to promote truth telling. For a review see Stouthamer-Loeber (1986). Lyon and Dorado (2008) and Wagland and Bussey (2005) analyze factors that facilitate truth telling for children.

### **3 Experimental Design and Procedure**

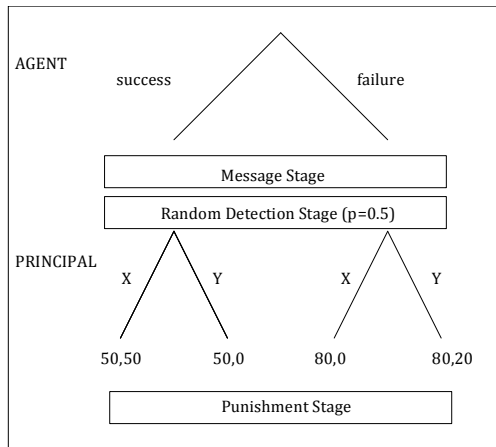
To analyze whether punishment for confessors and randomly detected offenders differs, I use a variation of the strategic information game introduced by Crawford and

Sobel (1982). My basic game is as follows. An agent can decide how to divide 100 points between himself and his principal. He can choose between an equal allocation and an unequal allocation, which is advantageous for the agent. However, the agent cannot just freely choose one allocation. He has to correctly answer a multiple choice question for the equal allocation. The unequal allocation results when he answers the question incorrectly. I used the same questions as Fischbacher and Utikal (2010) for their study on apologies. Questions were easy but not trivial. For example, I asked for the capital of Japan, giving Tokyo, Osaka, Yokohama and Kyoto as possible answers. By ensuring not all participants could perfectly solve the questions the unequal allocation could be the result of either inability or intentional unkindness. Ambiguity of intentions is necessary because after a failure we usually do not know whether the bad performer is a friendly person who tried to succeed and failed accidentally or whether the bad performer intended to fail. However, as explained in the introduction, people's punishment decisions strongly depend on the performer's type. Several studies show that punishment for intentional failures is higher than for accidental failures.

Following the answer, the agent receives his self-allocated share. The principal has to give an appropriate best response in order to receive his allocated share. The best response depends on the agent's performance but the chosen allocation is the agent's private information. It is possible that the principal does not learn which allocation was chosen. However the decision could be randomly detected. Also, the agent has the option to write a message to the principal. I am interested in whether agents use this message to self-report their performance. On the one hand a report on performance is a friendly move since it secures the principal a positive payoff. On the other hand reporting a failure leaves the agent exposed to punishment. I examine whether it makes a difference for the principal's behavior how he learns about the agent's performance. I analyze whether confessors are treated differently than non-confessors whose failures were randomly detected. Therefore, at the end of the game, the principal can punish the agent. In detail, the sequence of the game is the following. The sequence of action is also shown in Figure 1.

1. The agent receives a multiple-choice question. If the agent answers the question correctly, he receives 50 points. If he answers the question incorrectly, he receives 80 points.

2. The agent learns whether his answer was correct. The principal does not learn the agent's answer or payoff.
3. The agent can send an individual message to the principal.
4. Afterwards, a die determines whether the principal learns if the agent's answer was correct. The probability for random detection is  $p = 0.5$ .
5. The principal chooses  $X$  or  $Y$ . If the agent answered correctly, the principal receives 50 points by choosing  $X$  and 0 points by choosing  $Y$ . If the agent answered incorrectly, the principal receives 0 points by choosing  $X$  and 20 points by choosing  $Y$ . If the principal has no information about the agent's performance, the principal nevertheless has to choose between  $X$  and  $Y$ .
6. The last stage of the game is a punishment stage. Both players receive an additional 20 points in order to avoid negative payoffs. The principal can now deduct points from the agent. One deducted point costs 0.2 points. Punishment is restricted such that it cannot yield negative payoffs.



**Figure 1: The Game**

In order to assess the impact of a confession, I will focus on the situation where the agent failed to answer the question, but the principal could avoid a payoff of zero by giving the appropriate payoff-maximizing response  $Y$ . The players end up with the

unequal allocation (80, 20). The situation differs only in how the principal has learned about the failure. Assuming a failure of the agent and the best response  $Y$  by the principal the allocation (80, 20) can be reached via four different ways: First, the agent reports his failure. Second, the failure is randomly detected. Third, the agent reports his failure, and additionally the failure is randomly detected. Fourth, the principal does not learn the agent's performance but chooses  $Y$  nevertheless. In this last case, the principal of course does not know that the allocation (80, 20) is reached. I analyze punishment behavior for the three possibilities of knowingly reaching (80, 20). I also compare the results with punishment behavior after unknowingly reaching (80, 20).

With the objective of determining the consequences of confessing I conducted two treatments. The first treatment is the treatment *report option* as explained above. In the second treatment *no report option* the agent cannot send messages. I thereby control for people's behavior when confessing is not possible and determine the punishment benchmark for a failure.

To introduce ambiguity of intentionality we selected the questions such that not all participants could perfectly solve the questions. To control for the difficulty of the questions and to give the principals the opportunity to form an individual view on the difficulty of their agent's questions the principals receive their agents' multiple choice questions at the same time. If they answer this solo question correctly, they receive 5 points. I call the fraction of principals who are able to solve the multiple choice question the *solvability benchmark*. On average principals were able to solve 82% of the questions.

Before the experiment started, subjects were randomly assigned to their role as agent or principal. The experiment lasts 10 rounds. I used perfect stranger matching in order to avoid reputation effects and to keep the principals from receiving identical messages. I rule out spillover effects across and within sessions by giving every pair of players a different question.

The whole procedure was common knowledge. I conducted 4 sessions for every treatment (86 subjects in *report option*, 92 subjects in *no report option*) in the time from April to July 2010. All sessions were conducted at the LakeLab (TWI/University of Konstanz). A total number of 178 students from various disciplines took part in the experiment. Each subject sat at a randomly assigned PC terminal and was given a copy of

instructions.<sup>4</sup> A set of control questions was provided to ensure the understanding of the game. The experiment did not start until all subjects had answered all questions correctly. I ensured that no subject participated more than once in the experiment. Participants receive the income of all periods. One point translates into 0.02 euros. The experiment took about 60 minutes, average income of participants was 11.90 euros (\$14.92) plus a show-up fee of 2 euros (\$2.51). The games were programmed with z-Tree (Fischbacher (2007)). I recruited participants using the online recruiting system ORSEE (Greiner (2004)).

## 4 Predictions

Is the willingness to punish a harmful failure affected by how a person learns about that specific failure? Assuming selfish preferences, players do not differentiate between confessed and randomly detected failures. If players are selfish, they will not punish failures since punishment is costly. The agent will answer the question incorrectly which I assume the agent is able to do with certainty. This is plausible since the questions were easy and it was even easier to find at least one wrong answer. Assuming common knowledge of selfish preferences confessions do not make sense since the principal's behavior does not depend on preceding actions. He will choose  $Y$  in any case and not punish.

Theories that model non-selfish motives based on outcome-oriented preferences like Fehr and Schmidt (1999) or Bolton and Ockenfels (2000) also predict that confessions have no impact on punishment. These models assume that people have an additional component in the utility function which captures disutility from inequity. In these models, utility depends not only on the own payoff but also on the payoff of the other player. However, utility depends only on the final allocation, i.e. it does not depend on the procedure with which this outcome was reached. This implies that punishment should only depend on the outcome. In particular it does not depend on how the principal has learned about the outcome. Principals will not differentiate between confessed and randomly detected failures. In the framework of a linear model of Fehr and Schmidt (1999), confessions can nevertheless occur. An agent with  $\beta_i \geq 1/2$  prefers the equal split to all other unequal allocations and will therefore intend to answer the

---

<sup>4</sup> A translation of the instructions can be found in the appendix.

question correctly. If the agent fails in answering the question nevertheless, he prefers (80,20) over (80,0) which is why without punishment option he will confess his failure to have the principal give his best response. An agent with  $0 < \beta_i < 1/2$  prefers (80,20) to all other allocations, hence he will intentionally answer the question incorrectly and without punishment option confess his failure. With possible punishment all types of agents will confess their failure if their utility for (80,20) minus expected punishment exceeds their utility for (80,0) minus expected punishment.<sup>5</sup>

Also in the intention-based model of Levine (1998) agents might report their performance because they care for the principal's payoff. In this model players differ in how they value the other players' payoff. *Ceteris paribus*, players reward those with a high positive other-regarding concern and punish those with a low, negative other-regarding concern. An agent, who does not exploit the principal but answers correctly values the principal's payoff and hence does not deserve punishment. Successful agents will therefore want to make their performance known to the principal and give a report. Agents who fail to answer correctly are of two types. Either they have a high positive other-regarding concern and intended to answer correctly but failed or they have a low negative other-regarding concern and intended to answer incorrectly. Independent from their other-regarding concern, agents who failed face a tradeoff with respect to reporting performance. On the one hand, reporting a failure is a nice and friendly move since although the equal allocation was missed, it at least guarantees the principal a positive payoff. On the other hand, confessing is admitting failure and might therefore lead to punishment. To avoid punishment agents will therefore either remain silent or lie and report success hoping that the failure will not be randomly detected. As multiple studies show that people prefer remaining silent to telling a lie (Sánchez-Pagés and Vorsatz (2009), Sánchez-Pagés and Vorsatz (2007), Hurkens and Kartik (2009), Lundquist et al. (2009)), I expect that agents who failed will remain silent. In contrast, successful performers will report their performance. This leads to Prediction 1.

**Prediction 1** Successful agents report their performance more often than agents who failed.

---

<sup>5</sup> An agent with  $\beta_i = 0$  will answer incorrectly and will never confess without punishment option.

How will principals respond to reports of performance? There is no reason for an agent to report failure if the performance was actually successful. So principals should trust negative reports and choose *Y*. A report of success is less informative. After a report of success principals cannot be sure whether to choose *X* or *Y* since in order to prevent punishment agents might report success even though they answered the question incorrectly. Principals who do not receive a report will decide between *X* and *Y* according to their belief. If they have a stronger belief that the agent answered correctly, they will choose *X*. If have a stronger belief that the agent answered incorrectly, they will choose *Y*.

Apart from wanting to secure the principal a positive payoff, there is another motivation to confess a failure. Compared to randomly detected failures, a failure disclosed through confession might be punished less. The reasons are the following. First, not reporting a failure can be seen as lying. As several studies have shown that people are liar-averse and punish others for not being honest (Hurkens and Kartik (2009), Charness and Dufwenberg (2006), Lundquist et al. (2009), Sánchez-Pagés and Vorsatz (2007) and Sánchez-Pagés and Vorsatz (2009)), I expect punishment for offenders who remain silent to be more likely than for confessors. Second, in the Levine model (1998) mentioned above, principals punish those agents less who show their positive other regarding concern by reporting their performance and making sure that the principal will receive a positive payoff. Agents who do not care about the principal's payoff and do not confess their failure are punished more. This leads to Prediction 2.

**Prediction 2** Punishment for reported performance is less likely than for randomly detected performance.

In a laboratory experiment, Fischbacher and Utikal (2010) show that apologies after offenses with ambiguous intentionality are a social norm. Punishment of apologizers is as likely as for offenders who did not have the option to apologize by design. However, punishment of offenders who had the option to apologize but did not use it is more likely than for offenders with no apology option. When apologies are possible, people seem to expect offenders to apologize. Apologies and confessions differ in one important aspect. An apology cannot undo the offense and it cannot repair the caused damage either. In contrast to an apology by transferring private information a confession is able to prevent further harm. Nevertheless, apologies and confessions are

related concepts as both apologies and confessions can help to reveal the intention behind the preceding offense. Therefore I expect the same effect for confessions as summarized in Predictions 3 and 4.

**Prediction 3** If the failure is detected, punishment for bad performers without report option is less likely than for bad performers who had the option to report their failure but did not use it.

**Prediction 4** If the failure is detected, punishment for bad performers without report option is as likely as for bad performers who reported their failure.

## 5 Results

This section is divided into two parts. First, I focus on the agent's task performance and his reporting behavior. In the second part I present results on the principal's decisions following a report or a random detection.

### 5.1 The agent

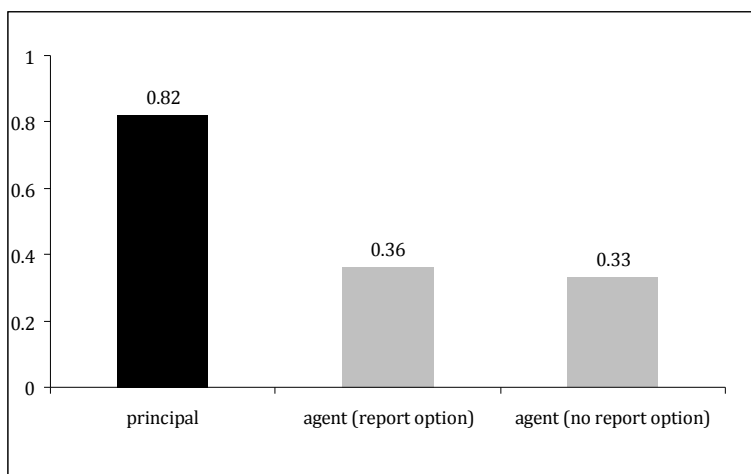


Figure 2: Successful task performance

In this section I present results on the agent's behavior. I start with the agent's performance in the multiple choice question. A correct answer is costly for the agent but

yields on average a higher payoff for the principal. The principal receives the same question and receives 5 points for a correct answer. I use the principal's average performance in order to create a solvability benchmark for the questions. Figure 2 shows that the principals are able to solve 82% of the questions. The agents solve 36% in *report option* and 33% in *no report option*. Not surprisingly the agents' performance is significantly below the principal's performance ( $p < 0.01$ ). This shows that people respond accordingly to incentives. Table 4 (column 1) presents the corresponding regression.

In *report option* agents could report their performance and therefore reduce punishment if they failed. Agents in *no report option* did not have this option. This can induce a different behavior in performance. However, I find that the agents' performance does not vary across treatments ( $p > 0.8$ ). Agents who have the option to report their failures do not perform worse than agents who cannot inform their principals. Table 4 (column 2) presents the corresponding regression. This result indicates that agents do not exploit an open communication structure. One might expect that more agents fail in *report option* since they can report their performance and thereby secure a positive payoff for the principal. However, this does not seem to be the case here.

In order to analyze the written messages, 20 additional subjects were recruited to independently sort messages into at least one of the 4 following categories: true REPORT, LIE, MERCY, and OTHER (small talk, jokes, etc.). In case of disunity between the raters, I applied the majority rule. Messages were only sorted into one particular category if the majority of raters decided to do so<sup>6</sup>. I present the frequency of the various categories in Table 1. Messages can appear in more than one category.

Category	Example	failure	success
No. of observations		274	154
NO MESSAGE	-	0.38	0.19
REPORT	My answer was incorrect/ correct.	0.43 <sup>7</sup>	0.66 <sup>8</sup>
LIE	My answer was incorrect/ correct.	0.05	0.00
MERCY	Please don't punish me!	0.08	0.17
OTHER		0.13	0.18

**Table 1: Percentage of message categories after failure and success with report option**

<sup>6</sup> Following this procedure two observations had to be taken out of the analysis because there was no majority for any category. All other messages were categorized.

<sup>7</sup> Within these 43% (119 messages) 9% (25 messages) were also categorized into *other* or *mercy*.

<sup>8</sup> Within these 66% (102 messages) 20% (32 messages) were also categorized into *other* or *mercy*.

**Result 1** Agents who failed remain quiet more often than successful agents. They also report their performance less often than successful agents.

**Evidence** The most common behavior after failure is to remain silent and to not send a message (38%). After successful performance only 19% of agents do not send a message. This difference is significant ( $p < 0.01$ ). The majority of successful performers reports their performance (66%). Failure significantly decreases the willingness to report the performance by 23%. This result confirms Prediction 1. I also find that successful performers are less likely to lie. Behavior in the message stage does not vary across periods. Table 4 (columns 3-5) shows the corresponding regressions.

## 5.2 The principal

The focus of this paper is on how confessions affect punishment. In order to conduct the analysis I distinguish between four situations. In the first situation the principal obtains no information about the agent's performance. In the other three situations the principal learns about the agent's performance. The first way the principal can learn about the agent's performance is that the agent reports it. Second, the agent's performance can be randomly detected or third, the agent reports and the performance is also randomly detected. Table 2 lists the number of observations of the four crucial situations.

	report option		no report option
no information	62		254
	RANDOM	REPORT	REPORT+RANDOM
learned failure	108	45	74
learned success	37	36	66

**Table 2: Number of observations**

I am going to assess two effects: First, does the channel through which the principal learns about the agent's performance influence the principal's following move? And second, how do confessions affect punishment behavior? In order to answer the first question, I focus on the fraction of principals not choosing the best response after they have learned the agent's performance, i.e. choosing  $Y$  after reported/randomly detected success and choosing  $X$  after reported/randomly detected failure (See Table 3). I find that people almost perfectly react to the information they receive. Without any

information on the agent's performance, 62% choose *Y* (38% choose *X*) in the *report option* treatment. In *report option*, after RANDOM detection of success, a REPORT of success, or REPORT+RANDOM the share of principals choosing *Y* decreases to 0%, 3% and 0% respectively.<sup>9</sup> After RANDOM detection of failure, or a REPORT of failure, or REPORT+RANDOM the share of principals choosing *X* significantly decreases to 2%, 4% and 0% respectively.<sup>10</sup> Compared to *report option*, principals with no information about the performance in the *no report option* treatment have a stronger belief that the agent was successful.<sup>11</sup> Here, without any information, 49% choose *Y* (51% choose *X*). After RANDOM detection of good performance the share of principals choosing *Y* decreases to 3%. After RANDOM detection of a failure the share of principals choosing *X* decreases to 3%.<sup>12</sup> To summarize, principals almost perfectly respond to random detection as well as to reports of performance. Thereby they do not distinguish between confessions and potential false reports of success.

	report option			no report option
choosing <i>Y</i> (choosing <i>X</i> ) with no information	0.62 (0.38)			0.49 (0.51)
	RANDOM	REPORT	REPORT+RANDOM	
<i>X</i> after learned failure	0.02	0.04	0.00	0.03
<i>Y</i> after learned success	0.00	0.03	0.00	0.03

**Table 3: Fraction of principals not choosing best response**

After the choice of *X* or *Y* the principal has to make his punishment decision. Let us first take a look at punishment behavior of principals who received no information about the agent's performance - neither through self-report nor random detection. Principals with no information who chose *X* punish with 22% probability. Punishment probability for principals with no information who chose *Y* is 34% and therefore significantly higher than for principals who chose *X*.<sup>13</sup> This behavior is very natural in that principals choose *Y* because they have a stronger belief that the agent failed. Following, if they believe the agent failed he deserves more punishment.

<sup>9</sup> Probit regression with standard errors clustered on session:  $p < 0.001$

<sup>10</sup> Probit regression with standard errors clustered on session:  $p < 0.001$

<sup>11</sup> Probit regression with standard errors clustered on session:  $p < 0.1$

<sup>12</sup> Probit regression with standard errors clustered on session:  $p < 0.001$

<sup>13</sup> Probit regression with standard errors clustered on session:  $p < 0.001$

Principals who do not have to form a belief about performance but chose  $Y$  because they know through random detection that the agent failed punish significantly more often. Table 5 (column 1) presents the results of the corresponding regression. Principals who received a self-report about the agent's successful performance and chose  $X$  are less likely to punish. Table 6 (column 1) presents the results of the corresponding regression.

I now focus on the second question: How do different channels of learning the agent's performance affect the principal's punishment decision? First, I examine the influence on punishment probability. I start with analyzing punishment probabilities after learned failure followed by punishment probabilities after success. Finally, I will discuss the impact on punishment intensity.

I begin with the analysis of how different ways of learning about the agent's failure affect punishment probability. From now on I will only include observations after learned failure and best response  $Y$ , and directly compare punishment decisions for the different channels of learning.

**Result 2** Punishment for reported failures is less likely than for randomly detected failures.

**Evidence** Figure 3 shows the punishment probability for the agent after the principal has learned the agent's performance and chosen the best response to this information ( $Y$  after failure,  $X$  after success). Punishment for agents who failed and reported is 19 percent less likely less likely than for agents who failed and were randomly detected ( $p < 0.01$ ). Table 5 (column 2) presents the results of the corresponding regression. The result confirms Prediction 2<sup>14</sup>.

In order to compare whether a report decreases punishment or not reporting increases punishment I conducted the additional *no report option* treatment. On the one hand there is no difference across treatments in punishment probability after random detection. Bad performers who had the option to report but did not use it are not punished more often than agents who failed and did not have the option to report. On

---

<sup>14</sup> It is worth mentioning that the principal's own performance affects punishment probability negatively. Principals who answered correctly punish significantly less often. Probit regression with standard errors clustered on session:  $p < 0.001$

the other hand though, I find that punishment after reporting a failure is 15 percent lower than after random detection in the *no report option* treatment. However, this difference is not significant ( $p=0.14$ ). Table 5 (column 4) presents the results of the corresponding regression. These results contradict Predictions 3 and 4. I conclude that a confession is not a social norm since punishment probability for not confessing a failure even though possible does not vary from punishment probability for the failure alone.

So far I have focused on two possible ways of detection. In the preceding analysis I compared random detection and detection via self-report. Our experimental design allows a third medium of detection: An agent might report his performance and be additionally randomly detected.

**Result 3** Random detection after a report on negative performance has already taken place increases punishment probability.

**Evidence** I find that punishment probability for agents who reported failure and were additionally randomly detected is higher than punishment probability after a report only ( $p<0.001$ ). Punishment probability after reported failure and additional random detection is as high as punishment probability after detection only. Table 5 (columns 2-3) presents the results of the corresponding regressions. This result is rather surprising. In our design the random detection took place after the message had already been sent and received. Therefore, whether the performance was randomly detected or not should not have an influence on punishment since the confession already happened. Nevertheless, a confession loses its effect when the offense is subsequently randomly detected. One possible reason is that after having learned the failure through random detection people might believe that they would have learned about the failure anyway and therefore do not value or appreciate the other's confession anymore. A very similar proceeding can also be found in the German jurisdiction of tax evasion. Tax evaders who self-report to the police do not usually pay a penalty (after having balanced their debts). However, tax evaders who self-report but whose tax honesty has already begun to be investigated cannot hope for exemption from punishment even if the tax evader had no information about the investigation. (§ 371

paragraph 2 Abgabenordnung<sup>15</sup>. For a commented edition see for example Pahlke and König (2009)).

Even after success, principals could be motivated to punish. Though the agent answered the question correctly, not reporting a positive performance weakens the agent's friendliness since non-reporters do not actively prevent the zero-payoff for the principal. In this case principals could punish agents for not ensuring them a positive payoff. This is exactly what I find.

**Result 4** Punishment for reported success is less likely than for randomly detected success.

**Evidence** Principals not only punish after failures, but also after successful performance of the agent. Even though one cannot compare the two values directly (since the related status quo allocations are different) I can state that successful agents are punished less often than agents who failed.<sup>16</sup> In the following analysis I focus on the situation where the agent answered correctly and the principal gave the best response  $X$ . As Figure 3 shows, 27% of randomly detected successful performances are punished. After reported success only 9% punish. After a report punishment probability is therefore 19 percent lower than after randomly detected successful performance ( $p < 0.01$ ). After not reporting a positive performance principals punish agents for not ensuring them a positive payoff. Agents who report their performance are punished with lower probability – even (and this is in contrast to the result on punishment after learned failures) when the performance is additionally randomly detected ( $p < 0.1$ ). Table 6 (columns 6-8) lists the corresponding regressions.

Until now, I have focused on punishment probabilities. However, the design also allows measurement of the effects of confessions on assigned punishment points. I am now going to discuss the impact of confessions on punishment intensity. Particularly interesting is the conditional punishment level, i.e. in case of punishment, how many punishment points do principals assign?

---

<sup>15</sup> Abgabenordnung from October 1, 2002 (BGBl. I S. 3866; 2003 I S. 61), with changes from July 30, 2009 (BGBl. I S. 2474)

<sup>16</sup> Probit regression with standard errors clustered on session:  $p < 0.001$ .

**Result 5** Report on performance only affects the event of punishment but not the degree of punishment.

**Evidence** Figure 4 shows the conditional punishment level after the principal has learned the agent's performance and chose the best response to this information ( $Y$  after failure,  $X$  after success). Using the dependent variable conditional punishment level instead of punishment probability, I cannot confirm Prediction 1. Punishment for reported failures is as high as for detected failures. It seems as if reports on performance can stop people from punishing. But if they punish nevertheless, the amount of punishment will be the same. Figure 4 presents the results and Table 6 the corresponding regressions for conditional punishment after failure and success.

To summarize, the income-maximizing strategy for the agent is to answer the question incorrectly and to report the failure: This way (ignoring the 20 additional points for each player in the punishment stage), an agent earns on average 62 points compared to 51 points after remaining silent. If the agent answers the question correctly, he earns on average 48 points after a report and 42 points after not reporting the performance.

## 6 Conclusion

To confess a failure is socially requested. All major theories of moral decision making (Freud (1940), Kohlberg (1980)) emphasize the centrality of truthfulness. Parents and teachers teach that confessing a failure is the right thing to do. There are also dozens of proverbs throughout the world recommending confession of failures rather than denial<sup>17</sup>. This paper analyzes whether these recommendations are advisable and appropriate from an economic point of view. Are confessors punished less than people who did not confess but were detected anyway?

In a laboratory experiment I find that our teachers, our parents and the proverbs are right. Punishment for reported failures is less likely than for randomly detected failures. Compared to randomly detected unreported failures, confessing your failure

---

<sup>17</sup> A fault confessed is half redressed. (Zulu Proverb)

Confessed faults are half-mended. (Scottish proverb)

Faillè avouée est à moitié pardonnée. (French proverb)

To confess a fault freely is the next thing to being innocent of it. (Publius Syrus)

makes punishment probability decrease. There is only one exception: If you confess your failure and afterwards the failure is also randomly detected, confessions do not help. Here, punishment is as likely as after random detection without confession. After having learned the failure through random detection people seem to believe that they would have learned about the failure anyway and therefore do not value or appreciate the other's confession anymore.

I find that compared to random detection punishment is less likely when a failure is confessed. Nevertheless, I cannot confirm the proverb '*A fault confessed is half redressed*'. Remarkably, a confession seems to only affect the event of punishment but not the degree of punishment. A confession can stop people from punishing. But if they punish nevertheless, the amount of punishment will be the same. People apparently either totally forgive or do not forgive at all. The proverb might look something more like:

A marked many faults confessed,  
For the lucky who eschew arrest  
are not half but fully redressed.

Finally I want to point out that my results also have important implications for failure management in companies. Receiving a report on the agent's performance assures the principal a positive payoff. On the other hand when agents have the option to report performance they might not even try to succeed because this is less costly for them and without feeling guilty they can report their failure and secure the principal a positive payoff. However, I do not find such behavior. Agents who have the option to report their failures do not perform worse than agents who cannot inform their principals. Managers who have the choice should therefore implement an open structure of communication, such that failures can easily be reported.

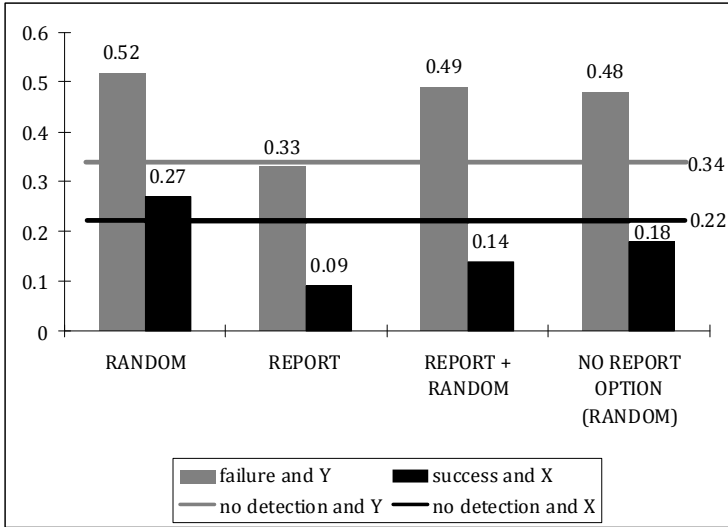


Figure 3: Punishment Probability after learned performance and best response

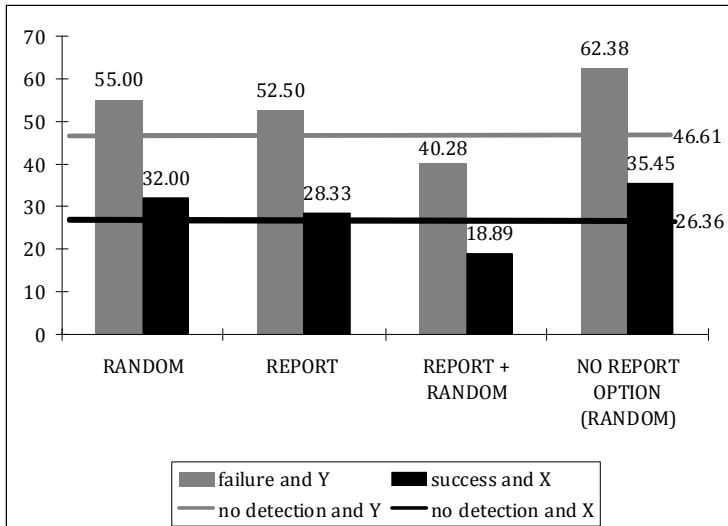


Figure 4: Conditional Punishment Level after learned performance and best response

	successful performance all	successful performance agents	no message agents in report option	report agents in report option	lie agents in report option
	(1)	(2)	(3)	(4)	(5)
no report option		-0.011 (.044)			
Principal	.371*** (.036)				
Success			-.200*** (.063)	.227** (.089)	-1.00*** (0.00)
Period	-.014*** (.003)	-.002 (.005)	-.014 (.016)	.006 (.013)	.002 (.003)
success in the previous period	.357*** (.039)	.451*** (.055)			
No of observations	1602	801	430	430	275
Wald $\chi^2$	334.26	57.78	10.32	68.98	0.38
Prob> $\chi^2$	0.0000	0.0000	0.0057	0.0000	0.5350
Pseudo R <sup>2</sup>	0.2634	0.1617	0.0403	0.0352	0.0011
No of Clusters	8	8	4	4	4

\*, p < 0.10, \*\*, p < 0.05, \*\*\*, p < 0.01.

**Table 4: Probit regressions, reporting marginal effects, with robust standard errors, standard errors in parentheses clustered on session**  
 Dependent variables: successful performance, no message, report and lie

	Punishment probability							
	After Y		After learned failure and best response Y		After X		After learned success and best response X	
	all	report option	report option	all	all	report option	report option	all
		RANDOM or	RANDOM or	RANDOM or	RANDOM or	RANDOM or	RANDOM or	RANDOM or
		REPORT or	REPORT or	REPORT	REPORT or	REPORT or	REPORT or	REPORT
		RANDOM and	RANDOM and		RANDOM and	RANDOM and	RANDOM and	REPORT
		REPORT	REPORT		REPORT	REPORT	REPORT	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RANDOM	.182* (.107)		.161* (.083)	.044 (.094)	.042 (.067)		.051 (.33)	-1.04 (.067)
REPORT	-.015 (.125)	-.192*** (.064)		-.151 (.098)	-.127** (.050)	-.145*** (.021)		.082 (.076)
REPORT + RANDOM	.151 (.106)	-.032 (.101)			-.081 (.061)	-.112* (.061)		
NO REPORT OPTION								
(RANDOM)	.139*** (.052)				-.039 (.042)			
No. of clusters	8	4	4	8	8	4	4	8
No. of observations	530	223	117	288	347	138	101	133
Wald $\chi^2$	37.44	9.84	3.68	9.17	27.10	21.61	2.65	17.15
Prob> $\chi^2$	0.0000	0.0073	0.0551	0.0102	0.0000	0.0001	0.1038	0.0002
Pseudo R <sup>2</sup>	0.0182	0.0155	0.0183	0.0119	0.0200	0.0401	0.0080	0.0346

\*: p < 0.10, \*\*: p < 0.05, \*\*\*: p < 0.01.

**Table 5: Probit regressions (reporting marginal effects) with robust standard errors, standard errors in parentheses clustered on session**  
 Dependent variable: punishment probability

Conditional Punishment Level						
	After Y		After learned failure and best response Y		After X	
	all	report option	all	REPORT or RANDOM and REPORT	all	REPORT or RANDOM and REPORT
	(1)	(2)	(3)	(4)	(5)	(6)
			REPORT or RANDOM and REPORT	RANDOM or REPORT	all	REPORT or RANDOM and REPORT
					REPORT or RANDOM and REPORT	REPORT or RANDOM and REPORT
RANDOM	10.55 (9.95)		-14.22 (14.94)	-8.24 (10.15)	5.64 (1.46)	-9.44 (12.81)
REPORT	7.56 (12.99)	-3.00 (15.67)		-11.31 (17.03)	1.97 (10.95)	
REPORT + RANDOM	-7.03 (9.38)	-17.65** (8.04)			-7.47 (8.61)	
NO REPORT OPTION (RANDOM)	18.70** (8.05)				9.09 (7.74)	
constant	48.51*** (6.40)	59.16*** (6.66)	55.51*** (12.73)	67.96*** (7.80)	26.26*** (5.49)	32.00*** (11.02)
No. of clusters	8	4	4	8	8	4
No. of observations	228	150	50	135	66	12
Wald $\chi^2$	16.87	6.63	0.91	1.22	4.77	0.54
Prob> $\chi^2$	0.0020	0.0363	0.3409	0.5435	0.3114	0.4609
Pseudo R <sup>2</sup>	0.0078	0.0058	0.0045	0.0016	0.0074	0.0053
Left-censored	0	0	0	0	0	0
Uncensored	185	86	44	102	66	12
Right-censored	43	19	6	33	0	0
Replications	488	497	500	489	432	479

\*: p < 0.10, \*\*: p < 0.05, \*\*\*: p < 0.01.

**Table 6: Tobit<sup>18</sup> regressions with robust standard errors, standard errors in parentheses clustered on session, bootstrapped standard errors**  
 Dependent variable: conditional punishment level

<sup>18</sup> Since punishment is restricted to 0  $s_{punishment} \leq 100$  I use a Tobit regression.

## 7 Appendix – Instructions of the experiment

Today you are participating in an economic experiment. By reading the following instructions carefully, you can – depending on your decisions – earn money in addition to the 2 euro show-up fee. Therefore it is important to read the instructions carefully. During the experiment it is not allowed to communicate with other participants. That is why we ask you not to talk with each other. If you have any questions, please take another look at the instructions. If you still have a question, please raise your hand. We will then come to you and answer your question in private. During the experiment we do not use euros, but points. All points you receive during the experiment will be changed into euros at the end of the experiment: 1 point = 0.01 euros. The following pages give you instructions on the course of the experiment. At the end of the instructions you will find some control questions that will help you to understand the experiment. The experiment will start as soon as all participants are familiar with the experiment.

**Summary:** In this experiment there are two types of players: player A and player B. At the beginning of the experiment you will see a screen indicating whether you are a player A or player B for the whole experiment. In every round a player A and a player B form a team. The experiment lasts for 10 rounds containing 6 stages. In stage 1 both players answer a question. Player A receives a team question. If he answers the team question correctly, he receives 50 points. If he answers the team question incorrectly, he receives 80 points. In stage 2 player A learns whether his answer was right or wrong. Player B, however, does not receive this information. In stage 3 player A can send a message to player B. In stage 4 a random mechanism determines whether player A learns if player A's answer was correct. In stage 5 player B chooses between two amounts of points. These amounts depend on whether player A answered the question in stage 1 correctly. In stage 6 both players receive an additional 20 points and player B can deduct points from player A. Then a new period begins. In every period new teams are formed. In every period the players interact with a player they have not interacted with before. At the end of the experiment you will receive a 2 euro show up fee and all points you received during the experiment: 1 point = 0.02 euros.

**The experiment:** At the beginning of every period every player A will be matched with a randomly determined player B.

Stage 1 Player A and player B receive a question. There are team questions and solo questions. Team questions: Every player A receives a team question. If he answers the question correctly, he receives 50 points. If he answers the question incorrectly, he receives 80 points. The correct answer to the team question influences player B in stage 5. How this exactly works will be explained later on. Solo question: Player B receives the same question as a solo question. If he answers the question correctly, he receives 5 points. If he answers the questions incorrectly, he receives 0 points.

Stage 2 After every team question player A learns whether he answered the question correctly and his corresponding points. Player B learns whether he answered his solo question correctly. However, player B does not learn whether player A answered the team question correctly.

Stage 3 After the team question player A has the option to write a message to player B.

Stage 4 After player B receives the message or learns that player A did not send him a message, a random mechanism determines whether player B learns if player A answered the team question correctly. This works as follows. Player 1 roles a six-sided die. If it shows a 1, 2, or 3, player B learns whether the answer was correct or incorrect. If the die shows a 4, 5, or 6, player B does not learn whether the answer to the team question was correct.

Stage 5 Next, player B chooses between two amounts of points. We call these amounts X and Y. X and Y depend on whether player A answered the team question correctly. If player A answered the team question correctly and player B chooses X, player B receives 50 points. If player A answered the team question correctly and player B chooses Y, player B receives 0 points. If player A answered the team question incorrectly and player B chooses X, player B receives 0 points. If player A answered the team question incorrectly and player B chooses Y, player B receives 20 points. This is clarified in the following tables.

Team question correct	Player B chooses X	Player B chooses Y
Player B receives	50	0

Team question incorrect

	Player B chooses X	Player B chooses Y
Player B receives	0	20

Player B chooses X or Y. Due to the previous events it might happen that player B does not know whether player A answered the team question correctly. Player B has to choose X and Y anyhow.

Stage 6 Next, both players receive an additional 20 points. Player B can deduct points from player A. By deducting points these points are erased. Deducting 5 points cost 1 point, deducting 1 point costs 0.2 points. Example: Assume, player A answered the team question correctly, player B chooses X. Player A's points:  $50+20$  additional points= $70$ ; Player B's points:  $50+20$  additional points= $70$ . Assume, player B deducts 50 points from player A. Player A's points:  $50+20-50=20$ ; Player B's points:  $50+20-(50/5)=50+20-10=60$ ; Player B can deduct up to 100 points but not more points than player A owns. For the previous example this means that if player B deducts more than 70 points, only 70 points will be deducted. The costs for player B are  $70/5=14$ . At the end of the period you will see a screen that shows your gains and losses from the period. Then the period is over. Every player A is matched with another player B. After 10 rounds you see a screen that shows your income form all periods.

## 8 Literature of Chapter 3

- Blount, Sally.** 1995. "When Social Outcomes Arent Fair - the Effect of Causal Attributions on Preferences." *Organizational Behavior and Human Decision Processes*, 63(2), 131-144.
- Bolton, Gary E. and Axel Ockenfels.** 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90 (1), 166-193.
- Bottom, William P.; Kevin Gibson; Steven E. Daniels and J. Keith Murnighan.** 2002. "When Talk Is Not Cheap: Substantive Penance and Expressions of Intent in Rebuilding Cooperation." *Organization Science*, 13(5), 497-513.
- Brandts, Jordi and Gary Charness.** 2003. "Truth or Consequences: An Experiment." *Management Science*, 49(1), 116-130.
- Brandts, Jordi and Carles Sola.** 2001. "Reference Points and Negative Reciprocity in Simple Sequential Games." *Games and Economic Behavior*, 36 (2), 138-157.
- Cai, Hongbin and Joseph Tao-Yi Wang.** 2006. "Overcommunication in Strategic Information Transmission Games." *Games and Economic Behavior*, 56(1), 7-36.
- Charness, Gary.** 2004. "Attribution and reciprocity in an experimental labor market." *Journal of Labor Economics*, 22(3), 665-688.
- Charness, Gary and Martin Dufwenberg.** 2006. "Promises and partnership." *Econometrica*, 74(6), 1579-1601.
- Charness, Gary and David I. Levine.** 2007. "Intention and stochastic outcomes: An experimental study." *Economic Journal*, 117(522), 1051-1072.
- Charness, Gary and Matthew Rabin.** 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117 (3), 817-869.
- Charness, Gary and Matthew Rabin.** 2005. "Expressed preferences and behavior in experimental games." *Games and Economic Behavior*, 53(2), 151-169.
- Crawford, Vincent P. and Joel Sobel.** 1982. "Strategic Information Transmission." *Econometrica*, 50(6), 1431-1451.
- Darby, Bruce W. and Barry R. Schlenker.** 1982. "Children's reactions to apologies." *Journal of Personality and Social Psychology*, 43(4), 742-753.
- Dickhaut, John W.; Kevin A. McCabe and Arijit Mukherji.** 1995. "An Experimental Study of Strategic Information Transmission." *Economic Theory*, 6(3), 389-403.

- Exline, Julie Juola; Lise Deshea and Virginia Todd Holeman.** 2007. "Is apology worth the risk? Predictors, outcomes, and ways to avoid regret." *Journal of Social & Clinical Psychology*, 26(4), 479-504.
- Falk, Armin; Ernst Fehr and Urs Fischbacher.** 2003. "On the nature of fair behavior." *Economic Inquiry*, 41(1), 20-26.
- Falk, Armin; Ernst Fehr and Urs Fischbacher.** 2008. "Testing Theories of Fairness-Intentions Matter." *Games and Economic Behavior*, 62(1), 287 - 303.
- Fehr, Ernst and Klaus M. Schmidt.** 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3), 817-868.
- Felson, Richard B. and Stephen A. Ribner.** 1981. "An attributional approach to accounts and sanctions for criminal violence." *Social Psychology Quarterly*, 44(2), 137-142.
- Fischbacher, Urs.** 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics*, 10(2), 171-178.
- Fischbacher, Urs and Verena Utikal.** 2010. "On the acceptance of apologies." *TWI Research Paper Series*.
- Freud, Sigmund.** 1940. "An outline of psychoanalysis." *Standard Edition*, 23, 141-207.
- Girard, Michelle; Etienne Mullet and Stacey Callahan.** 2002. "Mathematics of forgiveness." *American Journal of Psychology*, 115(3), 351-375.
- Gneezy, Uri.** 2005. "Deception: The Role of Consequences." *American Economic Review*, 95(1), 384-394.
- Gold, Gregg J. and Bernard Weiner.** 2000. "Remorse, confession, group identity, and expectancies about repeating a transgression." *Basic and Applied Social Psychology*, 22(4), 291-300.
- Greiner, Ben.** 2004. "The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics," University of Cologne, Department of Economics,
- Hale, Claudia L.** 1987. "A Comparison of Accounts: When is a Failure not a Failure?" *Journal of Language and Social Psychology*, 6(2), 117.
- Ho, Benjamin.** 2007. "Apologies as signals." *Cornell University Discussion Paper*.

- Holtgraves, Thomas.** 1989. "The form and function of remedial moves: Reported use, psychological reality and perceived effectiveness." *Journal of Language and Social Psychology*, 8(1), 1.
- Horowitz, Milton W.** 1956. "The psychology of confession." *The Journal of Criminal Law, Criminology, and Police Science*, 47(2), 197-204.
- Hurkens, Sjaak and Navin Kartik.** 2009. "Would I lie to you? On social preferences and lying aversion." *Experimental Economics*, 12(2), 180-192.
- Kassin, Saul M. and Gisli H. Gudjonsson.** 2004. "The psychology of confessions." *Psychological Science in the Public Interest*, 5(2), 33-67.
- Kerby, Jessica A. and Phillip Johnson.** 2005. "Students' reactions to cheating: An examination of the effects of confession on forgiveness and emotions." *Journal of College and Character*, 6(4).
- Kohlberg, Lawrence.** 1980. "The Psychology of Moral Development. The Nature and Validity of Moral Stages (Essays on Moral Development, Volume 2)." *New York: Harperscollins College Division.*
- Lundquist, Tobias; Tore Ellingsen; Erik Gribbe and Magnus Johannesson.** 2009. "The aversion to lying." *Journal of Economic Behavior and Organization*, 70(1), 81-92
- Lyon, Thomas D. and Joyce S. Dorado.** 2008. "Truth induction in young maltreated children: The effects of oath-taking and reassurance on true and false disclosures." *Child abuse & neglect*, 32(7), 738-748.
- McCullough, Michael E.; K. Chris Rachal; Steven J. Sandage; Everett L. Worthington, Jr.; Susan Wade Brown and Terry L. Hight.** 1998. "Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement." *Journal of Personality and Social Psychology*, 75(6), 1586-1603.
- McCullough, Michael E.; Everett L. Jr. Worthington and Kenneth C. Rachal.** 1997. "Interpersonal forgiving in close relationships." *Journal of Personality and Social Psychology*, 73(2), 321-336.
- Offerman, Theo.** 2002. "Hurting hurts more than helping helps." *European Economic Review*, 46(8), 1423 - 1437.
- Ohbuchi, Ken-ichi; Masuyo Kameda and Nariyuki Agarie.** 1989. "Apology as aggression control: Its role in mediating appraisal of and response to harm." *Journal of Personality and Social Psychology*, 56(2), 219-227.
- Ohbuchi, Ken-Ichi and Kobun Sato.** 1994. "Children's Reactions to Mitigating Accounts: Apologies, Excuses, and Intentionality of Harm." *Journal of Social Psychology*, 134(1), 5-17.

**Ohtsubo, Yohsuke and Esuka Watanabe.** 2009. "Do sincere apologies need to be costly? Test of a costly signaling model of apology." *Evolution and Human Behavior*, 30(2), 114-123.

**Pahlke, Armin and Ulrich König.** 2009. *Abgabenordnung*. München: Beck.

**Robinson, Dawn T.; Lynn Smith-Lovin and Olga Tsoudis.** 1994. "Heinous crime or unfortunate accident? The effects of remorse on responses to mock criminal confessions." *Social Forces*, 73(1), 175-190.

**Sánchez-Pagés, Santiago and Marc Vorsatz.** 2007. "An experimental study of truth-telling in a sender-receiver game." *Games and Economic Behavior*, 61(1), 86-112.

**Sánchez-Pagés, Santiago and Marc Vorsatz.** 2009. "Enjoy the silence: An experiment on truth-telling." *Experimental Economics*, 12(2), 220-241.

**Scher, Steven J. and John M. Darley.** 1997. "How Effective Are the Things People Say to Apologize? Effects of the Realization of the Apology Speech Act." *Journal of Psycholinguistic Research*, 26(1), 127-140.

**Schmitt, Manfred; Mario Gollwitzer; Nikolai Förster and Leo Montada.** 2004. "Effects of objective and subjective account components on forgiving." *The Journal of social psychology*, 144(5), 465-486.

**Schweitzer, Maurice E.; John C. Hershey and Eric T. Bradlow.** 2006. "Promises and lies: Restoring violated trust." *Organizational Behavior and Human Decision Processes*, 101(1), 1-19.

**Stouthamer-Loeber, Magda.** 1986. "Lying as a problem behavior in children: A review." *Clinical Psychology Review*, 6(4), 267-289.

**Struthers, C. Ward; Judy Eaton; Alexander G. Santelli; Melissa Uchiyama and Nicole Shirvani.** 2008. "The effects of attributions of intent and apology on forgiveness: When saying sorry may not help the story." *Journal of Experimental Social Psychology*, 44(4), 983-992.

**Takaku, Seiji; Bernard Weiner and Ken-Ichi Ohbuchi.** 2001. "A cross-cultural examination of the effects of apology and perspective taking on forgiveness." *Journal of Language and Social Psychology*, 20(1), 144-166.

**Tavuchis, Nicholas.** 1991. *Mea culpa: A sociology of apology and reconciliation*. Stanford University Press.

**Torgler, Benno.** 2003. "Tax morale: Theory and empirical analysis of tax compliance," Unpublished Ph.D. Thesis, University of Basel,

**Vanberg, Christoph.** 2008. "Why Do People Keep Their Promises? An Experimental Test of Two Explanations." *Econometrica*, 76(6), 1467-1480.

**Wada, Mitsuhiro.** 1998. "The Mitigative Effects of an Apology and Excuses in a Social Predicament." *Journal of Yamanashi Eiwa Junior College*, 32, 84-72.

**Wagland, Paul and Kay Bussey.** 2005. "Factors that Facilitate and Undermine Children's Beliefs About Truth Telling." *Law and Human Behavior*, 29(6), 639-655.

**Weiner, Bernard; Sandra Graham; Orli Peter and Mary Zmuidinas.** 1991. "Public confession and forgiveness." *Journal of Personality*, 59(2), 281-312.

THURGAU INSTITUTE  
OF ECONOMICS  
at the University of Konstanz

Hauptstr. 90  
CH-8280 Kreuzlingen 2

Telefon: +41 (0)71 677 05 10  
Telefax: +41 (0)71 677 05 11

[info@twi-kreuzlingen.ch](mailto:info@twi-kreuzlingen.ch)  
[www.twi-kreuzlingen.ch](http://www.twi-kreuzlingen.ch)