Risk and Temptation: A Meta-Study on Social Dilemma Games*

Friederike Mengel †

University of Essex & Maastricht University

November 6, 2014

Abstract

This paper reports the results of a meta-study of 89 prisoner's dilemma experiments comprising more than 3000 participants across 6 countries. We organize existing evidence and explain seemingly contradictory results in the existing literature by focusing on two dimensions of the dilemma: "risk" (the percentage loss of unilaterally cooperating against a defector) and "temptation" (the percentage gain of unilaterally defecting against a cooperator). We find in particular, that (i) risk best explains the variation in cooperation rates across random matching ("stranger") and one-shot treatments, while (ii) temptation best explains the variation in repeated ("partner") interactions, (iii) consistently with reputational models temptation is more effectively contained the longer the time horizon of the game, (iv) there is more cooperation in partner than stranger if and only if risk is high and temptation low and (v) women are more cooperative than the average man if risk is low and less cooperative if risk is high, but there are no gender differences on average. These results can be useful for discriminating between competing theories of why people cooperate. For policy making, it is important to understand which dimension of the dilemma (risk or temptation) is best targeted via interventions to yield improved cooperation.

JEL Codes: C72, C90, D01, D70.

Keywords: *Prisoner's dilemma, Cooperation, Meta-study, Experiments, Game Theory;*

*I thank Jim Andreoni, Matt Embrey, David Hugh Jones, Willemien Kets, Stephanie Rosenkranz, Sigrid Suetens and seminar participants at Fort Lauderdale (ESA North America 2014), JRC Ispra and Maastricht (BEE-Lab meeting) for helpful comments, Sara Godoy for assistance in running the experiments and Jim Andreoni, Douglas DeJong, Rosemarie Nagel, Hans Theo Normann, Ryan Oprea and Lilia Zhurakhovska for providing data.

†Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ. *e-mail*: fr.mengel@gmail.com

1 Introduction

The prisoner's dilemma has been studied for over 50 years in areas as diverse as Biology, Economics, Political Science, Physics, Psychology or Sociology as a workhorse to understand civic behaviour or why people cooperate in social dilemma situations. Different theories of human cooperation have been proposed (Dresher (1961); Kreps et al. (1982); Axelrod (1984); Hamerstein (2003); Fischbacher et al. (2001)) and different factors favourable or unfavourable to cooperation have been studied in laboratory experiments (Andreoni and Miller (1993); Andreoni and Varian (1999); Bereby-Meyer and Roth (2006); Grimm and Mengel (2009); Friedman and Oprea (2012)). While there is agreement in the literature on some matters, the evidence is mixed on a number of other issues such as the role of the matching protocol (Andreoni and Croson (2008)), the importance of backward induction (Embrey et al. (2014)) or gender differences (Croson and Gneezy (2009)).

In this paper we try to organize existing evidence and explain seemingly contradictory results in the existing literature by focusing on two dimensions of the dilemma: "risk" and "temptation". We define "temptation" (TEMPT) as the percentage gain when unilaterally defecting against a cooperator and "risk" (RISK) as the percentage loss of unilaterally cooperating against a defector.[1] Both are reasons why cooperation may fail: if a player is confident that her opponent cooperates, then she will be *tempted* to defect herself and reap the highest possible payoff unilaterally. If a player, however, is unsure about her opponent's action then she will hesitate to cooperate, because she faces the *risk* that her opponent defects leaving her with the lowest possible payoff.

Our analysis is based on a meta-study of 89 prisoner's dilemma experiments with more than 3000 participants across 6 countries. Studies include exogenous random matching ("stranger") and fixed matching ("partner") treatments with 2x2 prisoner's dilemma games played for finitely many rounds without pre-play communication. Average cooperation rates across these 89 studies range from 0.04 to 0.84 with a mean of about 0.38.

We find that RISK best explains variation in cooperation rates across random matching ("stranger") and one-shot treatments, where people have no prior experience with their matches and hence face large uncertainty. Neither temptation, nor efficiency nor any of ten other indices used in previous literature can explain this variation once RISK is controlled for. This is consistent with a simple model of behaviour, where people reduce uncertainty by using pessimistic beliefs.[2]

Results are different in the partner (repeated game) condition. Here risk plays a minor role and it is mostly our measure of temptation (TEMPT) that can explain variation in cooperation rates. However, RISK as well as some other indices from the literature play a role as well in the partner setting. Consistently with reputational models (Kreps et al. (1982)) temptation is more effectively curbed (less detrimental to cooperation rates) the longer the time horizon.

In terms of the comparison between partner and stranger settings we find that there is more cooperation in partner than stranger if and only if risk is high (above median) and temptation is low (below median). Women are more cooperative than the average man if and only if risk is low and less cooperative if and only if it is high, but there are no gender differences on average across the studies considered. The fact that both these comparisons are mediated by the RISK and TEMPT measures can explain why previous literature (usually relying on one set of parameters) has found such mixed results (see Andreoni and Croson (2008) or Croson and Gneezy (2009) for surveys).

Disentangling the role of risk and temptation in social dilemma situations can help decide in

[1] If there is no temptation, then the game is a stag hunt game and if there is no risk it is a chicken game. We will also control for possible efficiency gains using a variable we call EFF . If RISK , TEMPT and EFF are all zero then then the game is a trivial game where all payoffs are equal.

[2] One model that produces such pessimistic beliefs is e.g. the multiple prior model of ambiguity aversion by Gilboa and Schmeidler (1989).

which of two games higher cooperation rates can be expected, which can in turn inform the design of mechanisms and policy. If e.g. two game forms Γ_1 and Γ_2 present the same level of efficiency, but Γ_1 has more risk and less temptation than Γ_2, then our results suggest that Γ_2 may be more conducive to high cooperation rates than Γ_1 in Stranger settings, while the reverse would be true in Partner settings. It can also help discriminate between theories of cooperation. Our results show, for example, that theories focused on risk, such as e.g. theories of conditional cooperation, should have good chances to explain behaviour in random matching situations. Those theories assume that agents are intrinsically motivated to cooperate as long as others do so as well (Fischbacher et al. (2001); Mengel (2007)). On the other hand, theories focused on creating incentives for cooperation (or avoiding temptation), should have better chances in Partner settings. Results can also help policy-makers to understand which aspect of the dilemma is best targeted to design more effective interventions. For example, interventions focused on reducing strategic uncertainty should be more effective in Stranger settings, where risk explains most of the variation in cooperation rates.[3,4]

The question of how indices derived from payoff parameters can predict cooperation rates in the prisoner's dilemma has attracted a lot of research in the late 1970-ies and 1980-ies. Much of this research is summarized in Murnighan and Roth (1983). The focus of this literature was not so much to distinguish risk and temptation, but rather to find one index, which typically mixes risk, temptation and efficiency, that is able to "summarize" incentives in the prisoner's dilemma. In Section 5 we compare the performance of our measures of risk and temptation with the indices proposed in this literature in terms of explaining variation in cooperation rates. We find that, particularly in the Stranger condition, the separation between motives works much better with our measure of risk outperforming all other indices. In a more recent study, Schmidt et al. (2001) also compare the impact of payoff parameters in six different experimental games using two "stranger" designs with exogenous and endogenous matching (where players can choose their match based on the history of play) and one "partner" setting with endogenous matching. They focus on three indices, called "greed", "fear" and "cooperator's gain", which correspond to the numerators of our indices `TEMPT` , `RISK` and `EFF`.[5] Variation in these parameters is low, though, with `RISK` ranging from 0.5 to 0.71 and `TEMPT` ranging from 0.1 to 0.29. They find that cooperation rates correlate with all three indices, but do not find systematic differences across their different designs. Other authors have studied similar notions in public good or trust games. Dawes and Thaler (1988) try to disentangle "greed" and "fear" in 7-player public good games. Their treatment manipulation for greed is different from what we call temptation. Snijders and Keren (1999) study the importance of risk and temptation by varying payoff parameters in (one-shot) trust games. They find that potential losses for a truster are important in determining behaviour, which seems to indicate that risk might play an important role in this game.

Two previous meta-studies on the prisoner's dilemma have focused on language and communication. Sally (1995) focuses on early experiments (1958-1992) and concludes that decisions are "usually inconsistent with a model of pure self interest" and that language used in the instructions seems to encourage cooperation. Balliet (2010) finds that pre-play (especially face-to-face) communication increases cooperation. To the extent that we interpret communication as reducing strategic uncertainty this is consistent with our results. In an ongoing meta study Embrey et al. (2014) ask whether

[3] To the extent that communication reduces strategic uncertainty, this could explain for example why pre-play communication is often found to increase cooperation rates in these settings (see e.g. the meta study of Balliet (2010) or Ledyard (1995) and Chaudhury (2011) for surveys on the related public good game).

[4] Examples of interventions designed to reduce temptation could include taxation schemes, with very progressive schemes better suited to curb temptation compared to linear schemes.

[5] In other words they measure absolute differences instead of percentage gains. This works in their setting, because they study ceteris paribus comparisons between six particular games, but is clearly unsuitable for comparing games where payoffs are on a different scale.

behaviour in the finitely repeated prisoner's dilemma is consistent with backward induction. They find that the mean time to first defection is predicted well by a "basin of attraction" index which combines elements of risk and temptation.[6]

Of course our paper is also related to all the studies contained in our data set, which are listed in Table 12 in Appendix A. Each of these studies have different research questions, but they have in common that they study a prisoner's dilemma satisfying the conditions we outline in the next section.

The paper is organized as follows. In Section 2 we explain how we collected our data and discuss our measures of risk and temptation. In Section 3 we provide some theoretical background. In Section 4 we present our main results. In Section 5 we discuss how they compare to the indices used by Murnighan and Roth (1983). Section 6 looks at dynamics and Section 7 exploits questionnaire data using subsets of of our full data set for which this additional information is available. Section 8 concludes. Additional tables and figures can be found in an Appendix.

2 Procedures

We study laboratory experiments on prisoner's dilemma games. In total our data set comprises 89 studies involving more than 3000 participants in experiments conducted in Germany, Japan, the Netherlands, Spain, the UK and the USA. Our data set contains 50 studies from the existing literature published between 1967 and 2013. Our criteria for inclusion are that (i) a (two player) 2x2 prisoner's dilemma is studied (no public good game or similar), (ii) the game is either one-shot or finitely repeated (not indefinitely repeated), (iii) matching is either "Partner", i.e. finitely repeated or "Stranger", i.e. random rematching (no networks etc.), (iv) matching is exogenous, (v) there is no pre-play communication nor other stages, such as punishment, in the game, (vi) choices are incentivized, (vii) the cooperation rate is either reported, data were made available to check it or it could be reasonably inferred from a graph and (viii) the study has been published before or in 2013.[7] We found studies via a keyword search on google scholar and via an e-mail to the ESA discussion group on January 6, 2014. All the studies are listed in Table 12 in Appendix A. Note that often the main treatments of a study don't satisfy our criteria listed above, but there is a control treatment that does.

In addition, we conducted our own lab experiments to cover a larger parameter space (17 studies). In our own experiments we conducted 10 period repeated prisoner's dilemma games in either a Partner or a Stranger setting. We also conducted some one-shot studies on Amazon Mechanical Turk, AMT (22 studies).[8] Lab Studies were conducted between December 2013 - January 2014 at EssexLab at the University of Essex. AMT studies were conducted between November 2013 - January 2014.

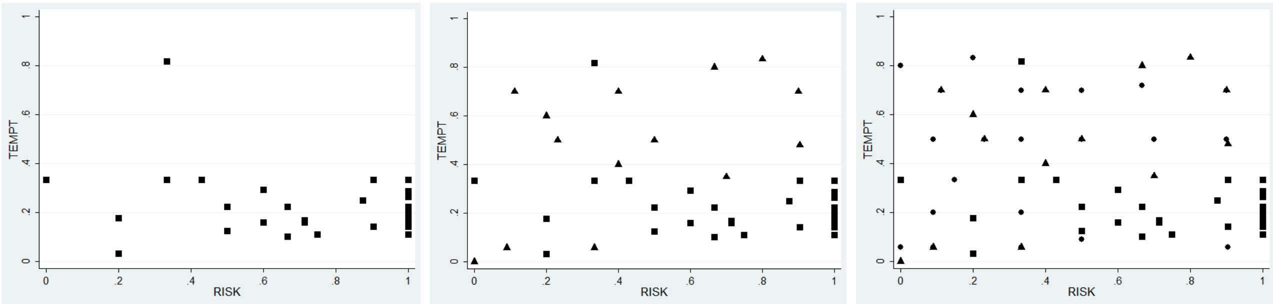
Figure 1 illustrates the variation in our two variables of interest: `RISK` and `TEMPT` for existing studies (Figure 1(a)), after adding our own lab studies (Figure 1(b)) and after adding both our own lab studies and studies on AMT (Figure 1(c)). Figure (1(a)) illustrates that the variation, especially in terms of the variable `TEMPT` in the existing literature is not very high. With one exception all studies have values of `TEMPT` lower than 0.4 and values of `RISK` higher than 0.2. We added our own

[6]There are also some meta-studies on the related public good game. Croson and Marks (2000) focus on threshold public goods and show that higher step returns (analogous to marginal per capita returns in the linear public good game) lead to more cooperation. Zelmer (2003) focuses on the linear public good game and finds that returns as well as framing, communication, partner matchings and the use of children as subjects had a positive effect on cooperation.

[7]We do allow for studies in which beliefs were elicited and we allow for asymmetric prisoner's dilemma games as well. Robustness checks show that excluding those would not affect the overall results (see Table 17 in Appendix B.3).

[8]In Appendix A we provide some more details about this platform and the characteristics of its users.

lab studies to increase the variation in our parameters of interest. To select payoff parameters for our own studies we partitioned the RISK-TEMPT space ($[0,1] \times [0,1]$) into squares of size $(0.2)^2$ and added our own studies s.t. there is at least one study in each element of the partition. Within these constraints we then selected payoffs arbitrarily preferring "easy numbers" (such as 200) to "difficult numbers" (such as 197.38965). Last, we randomly allocated these games to the lab (Partner or Stranger) or AMT (one-shot). We also conducted some studies in all three conditions: partner, stranger and one-shot. The parameter values used in our own studies are summarized in Tables 10 and 11 in Appendix A. Balancing tests, where we check whether the distributions of RISK and TEMPT are balanced across the partner and stranger settings can be found in Table 13 in Appendix B.2. We will show all our results both including and excluding the AMT studies.



(a) Existing Studies (b) With Own Lab Studies (c) All Studies

Figure 1: Variation in RISK and TEMPT parameters in the existing literature, including our own lab studies and including all studies (existing literature, own lab studies and own AMT studies).

We now proceed to describing our three key variables RISK, TEMPT and EFF defined on the payoff parameters of a symmetric prisoner's dilemma game such as the one in Table 1, which shows row player payoffs.

	C	D
C	a	b
D	c	d

Table 1: Prisoner's dilemma game with payoff parameters $c > a > d > b$.

To define RISK and TEMPT , first note that the expected payoff difference between cooperating and defecting $\Delta\pi_i^e = \pi_i^e(C) - \pi_i^e(D)$ can be written as

$$\begin{aligned} \Delta\pi_i^e &= p_{ij}a + (1-p_{ij})b - p_{ij}c - (1-p_{ij})d \\ &= -p_{ij}\underbrace{(c-a)}_{\text{Temptation}} - (1-p_{ij})\underbrace{(d-b)}_{\text{Risk}}, \end{aligned} \tag{1}$$

where p_{ij} is the subjective probability with which i believes that j will cooperate.

Temptation (TEMPT) We measure the extent of temptation the game presents by the percentage gain when unilaterally defecting against a cooperator, i.e. by TEMPT$= \left|\frac{c-a}{a+c}\right|$, which ranges between 0 and 1. The reason to use percentage gains as opposed to an absolute measure as in equation (1) is that we will compare many different studies conducted in different countries with different payoff scales and different exchange rates from experimental to local currency. The percentage measure puts all these studies on the same scale. Note that if there is no temptation (i.e. if $a = c$ or TEMPT$= 0$), then the Prisoner's dilemma game is a Coordination game with a tradeoff between pareto dominance and risk dominance. If a is negligible compared to c, then there is maximal temptation (TEMPT$= 1$). In our data the variable TEMPT ranges between 0.03 and 0.83.

Risk (RISK) The second part of the dilemma is risk. Starting from a situation of mutual defection (with inefficient payoff d) players might want to reach a pareto improvement and move to cooperation. They face the risk, however, that the other player does not do so and that they end up with the "sucker payoff" of b. We measure the extent of risk the game presents by the percentage loss when unilaterally cooperating against a defector, i.e. by RISK$= \left|\frac{d-b}{d+b}\right|$, which ranges between 0 and 1. If there is no risk ($b = d$ or RISK$= 0$), then defection is only a weakly dominant strategy. If b is negligible compared to d, then there is maximal risk (RISK$= 1$). In our data the variable RISK ranges between 0 and 1. The cases where RISK$= 1$ in our data always have $b = 0$, which for different values of d may imply quite different levels of potential loss. We will treat those cases with special care below.

Efficiency (EFF) An important consideration for cooperation might also be efficiency, i.e. how much can be gained by mutual cooperation as opposed to mutual defection. We are mostly interested in RISK and TEMPT , because EFF cannot be controlled unilaterally. We use EFF as a control, however, in most regressions. We measure the extent of possible efficiency gains in the game by EFF$= \left|\frac{a-d}{a+d}\right|$, which ranges between 0 and 1. If there are no efficiency gains ($a = d$), then from a game theoretic viewpoint nothing changes. However, in that case, there seems to be no reason to cooperate as the lowest possible payoff obtainable upon defection is still (weakly) higher than the highest possible payoff obtainable upon cooperation. In our data the variable EFF ranges between 0 and 0.83.

Note that a game with neither risk, nor temptation, nor efficiency concerns, i.e. a game where RISK=TEMPT=EFF$= 0$ is a trivial game where all payoffs are equal. Hence these three factors (as well as their interactions) must capture everything that is going on in the Prisoner's dilemma. Note also that it is necessary to measure these three quantities using percentage gains or losses. The reason is that we are comparing many different experiments conducted in different countries at different times and in which different payoff parameters and exchange rates between experimental and local currency were used.

Correlation and Balancing Tests Across all our studies the correlation between the three key variables RISK , TEMPT and EFF tend to be not statistically significant. Between RISK and TEMPT the Spearman correlation coefficient is -0.0369 (0.0772 if the AMT studies are excluded), neither of which is statistically significant. Between RISK and EFF the coefficient is -0.0050 (-0.0067), again both insignificant. Between TEMPT and EFF the correlation coefficients are 0.1879^{*} (0.2312^{*}), both weakly significant at the 10 percent level. Figure 4 in Appendix A illustrates the correlation between the RISK , TEMPT and EFF variables across our studies. Balancing tests, where we regress RISK and TEMPT on a partner dummy are reported in Table 13 in Appendix B.2. They show that the whole sample as well as the sub-samples obtained by splitting according to the median RISK and TEMPT are balanced.

For most of the paper we will treat one-shot experiments and Stranger settings together. Both settings have in common that they represent situations where players have no or little knowledge about their opponent and in particular no prior observations on how the opponent might choose. In neither repeated game strategies should play a role and reputation models should not be effective. In fact, the stranger setting is often perceived as a sequence of one-shot games. In Table 19 in Appendix

B.3 we conduct a separate analysis for the two settings, which shows very similar results across both settings.

3 Theoretical Background

To give some intuition to the reader about how RISK and TEMPT affect incentives in Stranger/one-shot and Partner settings we consider the following simple model. Assume agent i gets utility $u_i(a_i)$ from taken action $a_i \in \{C, D\}$ given by

$$u_i(a_i) = \pi_i^e + g(\theta_i, a_i), \tag{2}$$

where $\theta_i \in \{0, \theta\}$ denotes agent i's type and $g(\theta_i, a_i)$ is a function satisfying $g(0, a) = 0$ and $\widetilde{g}(\theta) := g(\theta, C) - g(\theta, D) \geq 0$. The function g reflects the fact that θ-types may have additional motivations to cooperate such as altruism, warm glow or others. We denote agent i's prior on the probability that her opponent j is of type θ by $\Pr_i(\theta_j = \theta) = p_{ij}$.

Stranger/one-shot settings The stranger and one-shot settings aim to capture interactions where people have no prior experience nor knowledge about their opponent and hence face a large amount of uncertainty about his/her type. If we assume that people resolve this uncertainty by using the most pessimistic among multiple priors (Gilboa and Schmeidler (1989)), then we can write the difference in expected utility between cooperating and defecting as follows

$$\Delta EU_i = \underbrace{(d - b)}_{\text{Risk}} + \widetilde{g}(\theta). \tag{3}$$

Hence a θ-type will cooperate in this setting whenever

$$\theta > \widetilde{g}^{-1}(\text{Risk}) \tag{4}$$

A 0-type on the other hand will always choose defection in this setting.

Partner setting The partner (repeated) setting involves more trade-offs. Here, people can potentially learn about the type of their opponent θ_j. They also have a larger strategy set available. To give some intuition we focus on the following strategy set $S = \{GT, MGT, 2MGT, 3MGT, ..., allD\}$, where GT denotes grim-trigger, i.e. the strategy that starts out by cooperating and switches to forever defecting as soon as the opponent defects once. MGT ("mean grim trigger") denotes the (0-type's) best response to grim trigger in the finitely repeated game, which is to cooperate in all rounds, but the last one. 2MGT denotes the (0-type's) best response to MGT (cooperate in all rounds but the last two) etc. and "allD" denotes the strategy that chooses always defection. We will show conditions for an equilibrium with cooperation where the θ-type plays GT and the 0-type MGT. For each type equilibrium will require that she does not prefer deviating to another strategy in S and that she does not have incentives to deviate from her strategy to refine her prior p_{ij}. These requirements then imply that for the θ-type the following needs to hold

$$\theta > \widetilde{g}^{-1}(\text{Risk} + p_{ij}(\text{Temptation-Risk})). \tag{5}$$

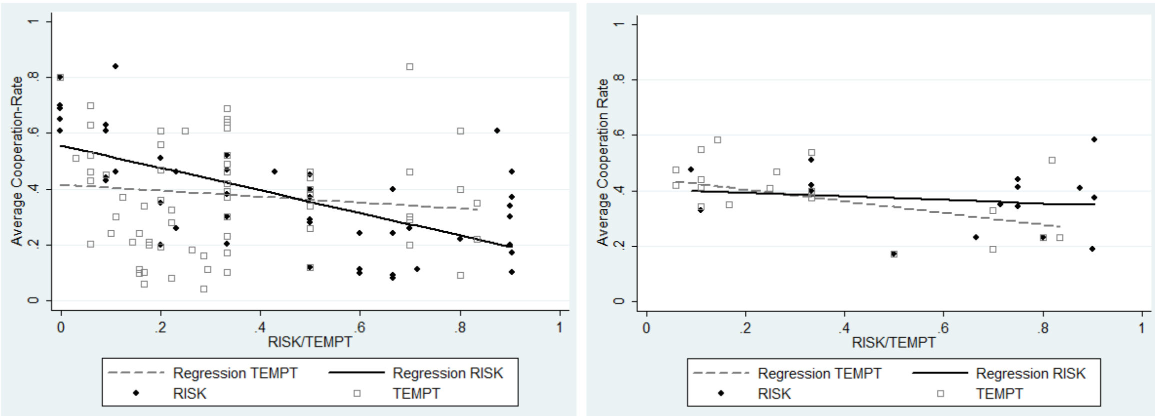
For the 0-type we need the following condition

$$p_{ij} > \frac{\text{Temptation}}{c - d} =: PA. \tag{6}$$

We denote the ratio $\frac{\text{Temptation}}{c-d}$ by PA and will include it in the regressions in Section 5. Proofs for the claims made can be found in Appendix B.1. The take-away message from this section is that, under the assumptions made, incentives are driven by `RISK` in the Stranger and one-shot settings, while both `RISK` and `TEMPT` play a role in Partner settings.

4 Results: Average Cooperation Rates

Our first set of results presented in this section uses information from all the 89 studies contained in our data set. Those are all the studies listed in Tables 10, 11 and 12 in Appendix A, in all of which the average cooperation rate is available. If the study consists of only one game (one-shot or finitely repeated), then the average cooperation rate we use is the average in that game. If the study consists of multiple repetitions of the same game with random rematching in between (as e.g. in Andreoni and Miller (1993)), then we use the average cooperation rate in the last match. We always average across all participants and all rounds in this section. The average cooperation rate thus computed ranges between 0.04 and 0.84 in the Stranger and one-shot treatments with a mean of 0.37. In the Partner experiments it ranges between 0.17 and 0.58 with a mean of 0.39.



(a) Stranger

(b) Partner

Figure 2: Average cooperation rate for different levels of `RISK` and `TEMPT`.

Stranger We start by analyzing the "Stranger" condition, that is meant to capture interactions where people have no experience with their opponent. Table 2 shows the results of simple OLS regressions of the average cooperation rates on our variables of interest: `RISK`, `TEMPT`, `EFF` and interaction terms. Columns (1) and (4) includes the whole sample of Stranger studies, but in columns (2), (3), (5) and (6) we omit the studies with `RISK`= 1. The reason is that, as discussed above, `RISK`= 1 are all studies where $b = 0$ (see Table 1). This implies that the `RISK` index will equal one but the associated loss can be very different.[9] In Appendix B.3, instead of omitting these studies we set b to a non-zero value and show that all results we present below are robust to this change (Table 16). Columns (2), (3), (5) and (6) further differ according to whether the AMT studies are included (columns (2) and (5)) or excluded (columns (3) and (6)) and to whether the interaction terms are included (columns (2) and (3)) or not (columns (5) and (6)).

[9] Defining `RISK` as absolute differences instead of percentage losses would avoid this problem. In this case, however, our indices would not be invariant to re-scaling or to multiplying all payoffs by a constant. In light of the fact that we consider experiments run in different countries with different currencies, exchange rates and payoff tables, invariance to such multiplication is, however, a necessary property of such an index.

Columns (2)-(6) consistently show that higher RISK implies lower cooperation rates. The variable TEMPT , by contrast, does not have any statistically significant impact on cooperation rates in these studies and the coefficient size is generally very small (columns (3)-(6)). If in column (5) we restrict only to studies where RISK is zero (those are 5 studies), then TEMPT is statistically significant with a coefficient of -0.1922^{*}. These results show that in anonymous or one-shot interactions participants seem to focus on the risk inherent in the dilemma when making their choice. When there is no risk, then temptation matters. The consideration of risk, however, seems to overpower any consideration of temptation or efficiency in these interactions. This is consistent with the theoretical reasoning given in Section 3.

Also the variable EFF and all interaction terms are statistically insignificant in all specifications (1)-(6), except for column (4), where RISK= 1 studies are included. This suggests that the impact of RISK is the same across the whole range of the variables TEMPT and EFF and that, conversely - apart from the extreme case just discussed (where RISK= 0) - the TEMPT or EFF variables have no significant impact on cooperation rates for the entire range of the variable RISK . To interpret these results it is useful to know whether these variables are correlated across the stranger treatments. It turns out that none of them are. The spearman correlation coefficient between RISK and TEMPT is -0.0440, between RISK and EFF it is 0.0538 and between TEMPT and EFF it is 0.0802. None of these is significant at the 10 percent level.

To sum up, RISK is the dominant factor explaining variation in average cooperation rates in "Stranger" settings. The results are illustrated in Figure 2(a), which shows data points as well as a simple OLS regression of average cooperation rates on RISK (coefficient: -0.405^{***}) or TEMPT (coefficient: -0.105).

Robustness tests In Appendix B.3 we show that the results are robust when we use weighted regressions, where we weigh our studies by the number of independent observations they contain (Tables 14 and 15). Results are also robust to different ways of dealing with the RISK= 1 cases (Table 16), are robust to dropping some studies with "special" details, such as studies were beliefs were elicited or studies that were paper-based (Table 17), the main effects appear both in games where mutual cooperation is efficient ($2a > c + b$) and cases where alternating between outcomes (C,D) and (D,C) is efficient (Table 18) and they appear when separating out the one-shot studies from the Stranger studies with > 1 periods (Table 19).

Partner Table 3 shows the results for the Partner treatments. These regressions have less power than those in Table 2, since the number of Partner studies in our data set is smaller. We include all studies in columns (1) and (3) and exclude those where RISK = 1 in columns (2) and (4). The variable RISK is not statistically significant in any of our specifications (1)-(4). Furthermore the coefficient size in columns (3) and (4) in Table 3 is much smaller than that of the corresponding columns (4)-(6) in Table 2. Temptation seems to be an issue, however, in Partner studies. The variable TEMPT is statistically significant with a substantial coefficient size.

As in the Stranger condition interaction terms are largely insignificant with the exception of the interaction EFF x TEMPT in column (2).[10] These two variables are pairwise strongly correlated in the partner treatment with a Spearman coefficient of 0.4328^{*}. This correlation can explain the drop in coefficient as we move from column (2) to columns (3)-(4). RISK is not significantly correlated with either TEMPT or EFF (Spearman coefficients: -0.0107 and -0.2358, respectively). Figure

[10]This interaction is also significant in both columns (1) and (2) of the weighted regression in Table 15 in Appendix B.3.

	STRANGER					
	(1)	(2)	(3)	(4)	(5)	(6)
RISK	−0.243	−0.669**	−0.910**	−0.204***	−0.374***	−0.345***
	(0.176)	(0.286)	(0.379)	(0.056)	(0.082)	(0.139)
TEMPT	0.437	0.159	0.040	−0.041	−0.044	−0.061
	(0.392)	(0.444)	(0.677)	(0.092)	(0.090)	(0.151)
EFF	0.020	−0.284	−0.532	−0.065***	−0.259	−0.340
	(0.619)	(0.668)	(0.921)	(0.187)	(0.288)	(0.366)
EFFx RISK	−0.142	0.787	0.637			
	(0.528)	(0.806)	(0.961)			
EFFx TEMPT	−0.186	−0.840	−1.365			
	(1.289)	(1.546)	(2.510)			
TEMPTx RISK	0.276	−0.157	0.637			
	(0.308)	(0.370)	(0.703)			
constant	0.537***	0.661***	0.701***	0.712***	0.642***	0.593***
	(0.184)	(0.189)	(0.244)	(0.074)	(0.091)	(0.112)
Lab Sample	All	RISK<1	RISK<1;	All	RISK<1	RISK<1;
AMT	YES	YES	NO	YES	YES	NO
Observations	70	46	24	70	46	24
R2	0.2491	0.4277	0.4800	0.2873	0.4098	0.3944

Table 2: Average cooperation rate regressed on variables of interest. Simple OLS regressions. (Pr>F)< 0.0009.***1%,**5%,*10% significance.

	PARTNER			
	(1)	(2)	(3)	(4)
RISK	0.306	0.127	0.026	−0.054
	(0.174)	(0.186)	(0.092)	(0.099)
TEMPT	−0.338	−0.455*	−0.257**	−0.258**
	(0.259)	(0.238)	(0.106)	(0.106)
EFF	−0.026	−0.270	0.064	0.135
	(0.612)	(0.610)	(0.179)	(0.179)
EFFx RISK	−0.553	−0.341		
	(0.558)	(0.647)		
EFFx TEMPT	0.817	1.057*		
	(0.604)	(0.573)		
TEMPTx RISK	−0.372	−0.278		
	(0.271)	(0.274)		
constant	0.395**	0.516**	0.444***	0.452***
	(0.190)	(0.186)	(0.092)	(0.092)
Lab Sample	All	RISK<1	All	RISK<1
Observations	19	16	19	16
R2	0.6823	0.7485	0.3171	0.3606

Table 3: Average cooperation rate regressed on variables of interest (Partner treatments). Simple OLS regressions. (Pr>F)< 0.0009.***1%,**5%,*10% significance.

2(b), which shows data points as well as a simple OLS regression of average cooperation rates on RISK (coefficient: −0.063) or TEMPT (coefficient: −0.2113**).

	(1)	(2)	(3)	(4)
Partner	0.046	0.1613**	−0.070	−0.024
	(0.154)	(0.069)	(0.095)	(0.103)
constant	0.401***	0.283***	0.398***	0.357***
	(0.049)	(0.045)	(0.043)	(0.043)
RISK	Small	High	Small	High
TEMPT	Small	Small	High	High
Observations	19	25	24	23
R^2	0.0053	0.1906	0.0243	0.026

Table 4: Partner vs Stranger. Average cooperation rates regressed on Partner dummy. Sample partitioned into four sub-samples according to median RISK and TEMPT. OLS regressions. ***1%,**5%,*10% significance.

Partner vs Stranger In Table 4 we ask whether contributions are higher in "partner" or "stranger" settings. Existing research on this question has delivered mixed results (see the survey by Andreoni and Croson (2008)).[11] In Table 4 we regress average cooperation rates on a dummy indicating whether the study is a "partner" study. We partition our sample into four sub-samples according to the median RISK and median TEMPT. Column (1) shows studies with below median RISK and below median TEMPT , column (2) with above median RISK and below median TEMPT , column (3) with below median RISK and above median TEMPT and in column (4) both RISK and TEMPT are above the median. Balancing tests, which show that the distributions of RISK and TEMPT are balanced across these sub-samples can be found in Table 13 in Appendix B.2. Table 4 shows that there are no significant differences between partner and stranger settings except for the case where RISK is high and TEMPT is small (column (2)). In these cases the partner setting is able to generate cooperation rates which are 16 percentage points higher than the stranger setting. The partner setting seems effective at disciplining behaviour in the presence of high RISK only as long as TEMPT is not too high. In all other cases (columns (1), (3) and (4)) the coefficient on the partner dummy is small and statistically insignificant.

Time Horizon We next ask how the impact of RISK and TEMPT is affected by the time horizon T. Reputational models, such as e.g. Kreps et al. (1982), predict that a longer time horizon is more effective in curbing temptation, simply because with a longer horizon agents have more to gain by establishing a reputation as cooperative types (see also Mengel (2014)). Table 5 shows the results of regressions where we regress average cooperation rates on our variables of interest RISK , TEMPT and EFF as well as interactions with T. The length of the horizon T ranges between 1 and 100 in the Stranger treatments (with a median of 10) and between 5 and 25 in the Partner treatments (with a median of 10). According to reputational theories we would expect a positive coefficient on the interaction TEMPT x T in the partner treatments (columns (3) and (4)), which is indeed what we

[11] Some authors have found that there is more cooperation in partner settings (Andreoni and Miller (1993), Cooper et al. (1996)), while others have found no difference (Boone et al. (1999)) or more cooperation in Stranger settings (Andreoni (1988)). A repeated game logic would suggest that, at least in indefinitely repeated games, there should be more cooperation in partner than stranger. However, this can be counteracted by path dependence and stronger negative reciprocity in the partner setting (Mengel and Peeters (2011)).

	STRANGER		PARTNER	
	(1)	(2)	(3)	(4)
RISK	-0.343^{***}	-0.340^{***}	0.174	-0.009
	(0.104)	(0.083)	(0.187)	(0.088)
TEMPT	0.034	-0.065	-0.099	-0.373^{***}
	(0.111)	(0.089)	(0.241)	(0.105)
EFF	-0.235	-0.306	0.085	-0.079
	(0.303)	(0.283)	(0.564)	(0.182)
RISK x T	0.002	-0.004^{*}	-0.016	
	(0.013)	(0.002)	(0.014)	
TEMPT x T	-0.024		0.011	0.016^{**}
	(0.015)		(0.023)	(0.007)
EFF x T	0.000		0.049	
	(0.022)		(0.034)	
constant	0.626^{***}	0.664^{***}	0.525^{***}	0.463^{***}
	(0.093)	(0.090)	(0.092)	(0.080)
Lab Sample	RISK<1	RISK<1	RISK<1	RISK<1
Observations	46	46	16	16
R2	0.4802	0.448	0.6455	0.5631

Table 5: Average cooperation rate regressed on variables of interest and interaction with time horizon T. Simple OLS regressions. (Pr>F)< 0.0009.***1%,**5%,*10% significance.

find (column (4)).[12] Existing theories do not make much of a prediction on the role of the horizon in Stranger treatments. We do find a small and negative coefficient on RISK x T (column (2)) which is marginally statistically significant ($p = 0.100$). We summarize the results discussed in this section as follows

Result 1
- *Variation in average cooperation rates in Stranger treatments is best explained by RISK , while TEMPT best explains variation in cooperation rates in Partner treatments.*
- *Consistently with reputational models, TEMPT is less detrimental in partner settings the longer the horizon of the game T.*
- *There is more cooperation in Partner settings if and only if RISK is high and TEMPT is low.*

[12] Based on only 16 observations the regression in column (3) is underpowered to detect significant effects. We show it, nevertheless, for the sake of completeness.

5 Results: Other Indices

In this section we compare our measures RISK and TEMPT to a number of different indices based on payoff parameters that have been used in the literature, notably by Murnighan and Roth (1983) who discuss 10 different indices used previously in the literature.[13] Most of these indices have been motivated not by disentangling risk and temptation, but by creating an index of all payoff parameters that "summarizes" incentives in the Prisoner's dilemma. Typically these indices contain aspects of all three: RISK,TEMPT and EFF, but with different emphasis.

Indices $R1 = \frac{a-d}{c-b}$, $R2 = \frac{a-b}{c-b}$, $R3 = \frac{d-b}{c-b}$ and $R4 = \frac{c-a}{c-b}$ measure efficiency (R1), value of mutual cooperation compared to unilateral cooperation (R2), inverse risk (R3) and temptation (R4) all relative to the difference between the "temptation" (c) and "sucker" (b) payoffs. Indices $E1 = \frac{c-a}{a-d}$ and $E2 = \frac{c-a}{a-b}$ both measure temptation relative to efficiency (E1) or relative to the difference between the value of mutual cooperation compared to unilateral cooperation (E2). Finally, indices $K1 = a + b - c - d$, $K2 = a - b + c - d$, $K3 = a - b - c + d$ and $K4 = a + b + c + d$ measure the "control one has over one's own outcomes" (K1), the "fate control" one player has over another (K2), the "behavioral control" of one over the other (K3) and the "overall level of outcomes in the game" (K4). The labels are adopted from Murnighan and Roth (1983).

We ask whether any of these measures provides a better explanation of the variation in cooperation rates observed. We start with the Stranger condition, where we have found above that the variable RISK can explain variation in cooperation rates well. Table 6 shows the results of regressions, where we regress average cooperation rates on our variables RISK , TEMPT and EFF (as in column (5) of Table 2) and add across columns (1)-(10) one each of the ten indices described above.[14] The table shows that the variable RISK is statistically significant at the 1% level in all columns (1)-(10) and the coefficient on RISK ranges between -0.275 to -0.396 irrespective of which other variables are included. Once RISK is included in the regression, not only are TEMPT and EFF not significant, but also none of the ten indices discussed by Murnighan and Roth (1983) can explain variation in cooperation rates in the Stranger setting.

Results are a bit less clear-cut in the Partner settings. Table 7 shows the results of regressions, where we regress average cooperation rates on our variables RISK , TEMPT and EFF (as in column (4) of Table 3) and add across columns (1)-(10) one each of the ten indices described above. As we have found in Section 4, TEMPT best explains variation in cooperation rates in these settings. The variable TEMPT is statistically significant in half of the regressions (in columns (2), (4), (7), (8) and (10)). Some indices used by Murnighan and Roth (1983), however, seems to explain the variation in cooperation rates better than the variable TEMPT. Those are in particular the indices R1, R2 and R4, that all involve the difference between temptation and sucker payoff in the denominator and measure the value of cooperation (R1 and R2) and temptation (R4) in the numerator. R1 and R2 affect cooperation rates positively and R4 negatively, as one would expect. Interestingly, also the index PA derived in Section 3 is significant. Hence both temptation and efficiency seem important in this setting. The reader should also note the lower number of observations in this case, which is potentially contributing to a less clear-cut picture than in the Stranger setting.

Result 2 *RISK explains variation in average cooperation rates in Stranger treatments over TEMPT , EFF and ten measures from the literature. In Partner settings RISK does not explain variation in cooperation rates, but TEMPT , PA and three indices from the literature (R1, R2 and R4) do.*

[13]Their eleventh index involves the discount rate in indefinitely repeated games and is hence not applicable.

[14]Adding all of them simultaneously would lead to considerable over-fitting in our sample of 46 observations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
RISK	−0.304***	−0.396***	−0.275**	−0.396***	−0.375***	−0.358***	−0.367***	−0.366***	−0.362***	−0.382***	−0.3518***
	(0.095)	(0.089)	(0.118)	(0.089)	(0.083)	(0.091)	(0.084)	(0.083)	(0.086)	(0.084)	(0.084)
TEMPT	0.186	0.212	−0.152	0.216	−0.059	−0.119	−0.027	−0.041	−0.076	−0.051	0.023
	(0.187)	(0.402)	(0.129)	(0.402)	(0.107)	(0.198)	(0.099)	(0.090)	(0.112)	(0.091)	(0.2683)
EFF	−0.079	−0.178	−0.186	−0.178	−0.281	−0.287	−0.279	−0.328	−0.280	−0.259	−0.099
	(0.313)	(0.314)	(0.294)	(0.314)	(0.302)	(0.298)	(0.295)	(0.306)	(0.294)	(0.290)	(0.321)
MR-R1	0.320										
	(0.229)										
MR-R2		0.247									
		(0.371)									
MR-R3			−0.321								
			(0.276)								
MR-R4				−0.247							
				(0.371)							
MR-E1					0.000						
					(0.001)						
MR-E2						0.006					
						(0.014)					
MR-K1							0.000				
							(0.000)				
MR-K2								−0.000			
								(0.000)			
MR-K3									−0.000		
									(0.000)		
MR-K4										−0.000	
										(0.000)	
PA											−0.2920
											(0.2633)
constant	0.366*	0.411	0.666***	0.658***	0.652***	0.654***	0.651***	0.682***	0.645***	0.668***	0.655***
	(0.216)	(0.359)	(0.093)	(0.095)	(0.099)	(0.097)	(0.094)	(0.108)	(0.092)	(0.101)	(0.092)
Observations	46	46	46	46	46	46	46	46	46	46	46
R2	0.4369	0.4163	0.4287	0.4163	0.4110	0.4125	0.4126	0.4170	0.4132	0.4153	0.4271

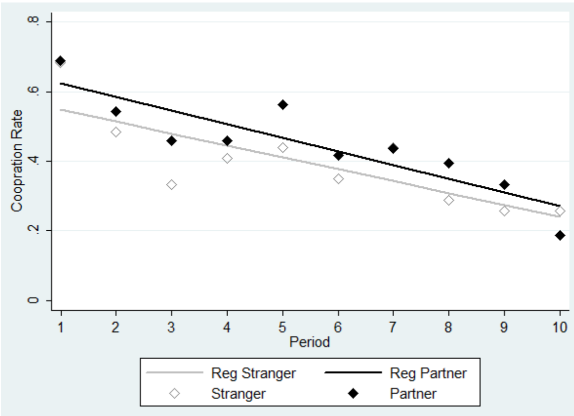
Table 6: Cooperation rates regressed on RISK , TEMPT , EFF as well as indices mentioned by Murnighan and Roth (1983). Stranger treatments.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
RISK	−0.102	−0.054	−0.167	−0.054	−0.040	−0.052	−0.048	−0.075	−0.140	−0.051	0.021
	(0.111)	(0.075)	(0.126)	(0.1016)	(0.106)	(0.096)	(0.106)	(0.104)	(0.097)	(0.100)	(0.086)
TEMPT	0.134	0.280*	−0.117	0.280*	−0.224	−0.037	−0.248**	−0.240**	−0.100	−0.303**	0.190
	(0.197)	(0.190)	(0.146)	(0.190)	(0.126)	(0.188)	(0.115)	(0.110)	(0.120)	(0.119)	(0.192)
EFF	0.063	−0.059	0.034	0.059	0.134	0.011	0.104	0.206	0.028	0.135	0.047
	(0.158)	(0.150)	(0.188)	(0.150)	(0.185)	(0.194)	(0.215)	(0.203)	(0.167)	(0.181)	(0.151)
MR-R1	0.569**										
	(0.253)										
MR-R2		0.518**									
		(0.165)									
MR-R3			0.405								
			(0.297)								
MR-R4				−0.518**							
				(0.165)							
MR-E1					−0.000						
					(0.001)						
MR-E2						−0.019					
						(0.013)					
MR-K1							0.000				
							(0.000)				
MR-K2								0.000			
								(0.000)			
MR-K3									0.000		
									(0.000)		
MR-K4										−0.000	
										(0.000)	
PA											−0.460**
											(0.175)
constant	0.040	0.016	0.405***	0.535***	0.437***	0.454***	0.466***	0.408***	0.497***	0.497***	0.509***
	(0.200)	(0.156)	(0.095)	(0.075)	(0.099)	(0.089)	(0.107)	(0.114)	(0.084)	(0.106)	(0.079)
Observations	16	16	16	16	16	16	16	16	16	16	16
R2	0.5613	0.6614	0.4527	0.6614	0.3773	0.4567	0.3654	0.3948	0.5403	0.4024	0.6059

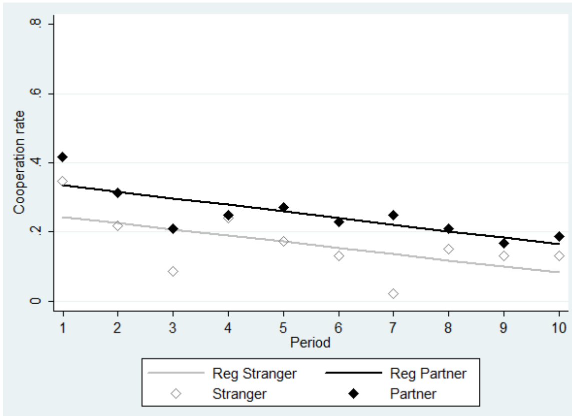
Table 7: Cooperation rates regressed on RISK , TEMPT , EFF as well as indices mentioned by Murnighan and Roth (1983). Partner treatments.

6 Results: Dynamics

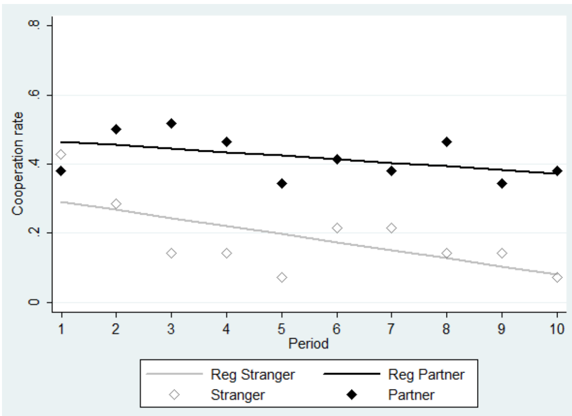
We ask whether RISK and TEMPT differentially affect the dynamics of cooperation. This section relies on data from all the studies on the 10-period repeated prisoner's dilemma (either partner or stranger) for which full data sets could be obtained. Those studies are all our own lab studies (see Table 10) as well as the studies by Andreoni and Miller (1993), by Bereby-Meyer and Roth (2006), by Cooper et al. (1996) and by Normann and Wallace (2012). The resulting data set contains 5100 observations of 510 participants who participated in 22 different studies or treatments.



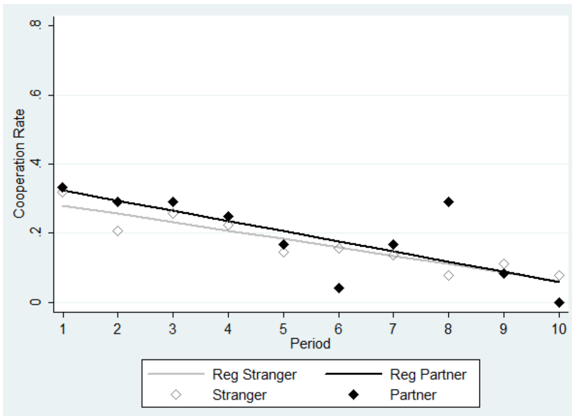
(a) Small RISK Small TEMPT



(b) Small RISK High TEMPT



(c) High RISK Small TEMPT



(d) High RISK High TEMPT

Figure 3: Dynamics of Contributions: Partner vs Stranger. The graphs show data averages as well a linear regression.

Figure 3 shows the dynamics in four cases of interest: (i) below median RISK and TEMPT (figure 3(a)), (ii) below median RISK and above median TEMPT (figure 3(b)), (iii) above median RISK and below median TEMPT (figure 3(c)) and (iv) above median RISK and TEMPT (figure 3(d)). Cooperation rates are decreasing over time in all cases (see also regression table 20 in Appendix B.4). In terms of the comparison between partner and stranger, the dynamic results obtained are consistent with the results obtained on average rates. The only significant difference between partner and stranger occurs in the cases where RISK is high and TEMPT is low (figure 3(c)).

Result 3 *Cooperation rates are decreasing over time in both the Stranger and Partner conditions and for all levels of RISK and TEMPT.*

7 Results: Questionnaire Data

This section exploits data from all the studies where the full questionnaire including our variables of interest are available. Those studies are our own 17 studies listed in Table 10 in the Appendix. This section contains two separate pieces of analysis. First, in subsection 7.1 we look at gender differences in cooperation. Second, in subsection 7.2 we analyze questionnaire data on how much risk and temptation participants perceived in the prisoner's dilemma they played.

7.1 Gender

There is a substantial literature on gender differences in cooperation behaviour, which, partly due to selective reporting issues, has remained inconclusive so far. Gender effects in the prisoner's dilemma have been studied by psychologists and some economists with about equally many studies showing that men cooperate more, women cooperate more or that there is no statistically significant difference (see the literature surveyed in Croson and Gneezy (2009)). We test for gender differences in the average cooperation rates among the 363 participants in our lab studies.[15]

	Stranger		Partner	
	(1)	(2)	(3)	(4)
FEMALE	0.196***	0.031	0.2614***	0.066
	(0.043)	(0.031)	(0.072)	(0.051)
FEMALE x `RISK`	−0.240***		−0.269**	
	(0.075)		(0.124)	
FEMALE x `TEMPT`	−0.079		−0.209*	
	(0.080)		(0.125)	
constant	0.206***	0.206***	0.267***	0.267***
	(0.023)	(0.024)	(0.037)	(0.039)
Observations	238	238	125	125
R2	0.1141	0.0044	0.1147	0.013

Table 8: Average individual cooperation rate regressed on gender dummy interacted with `RISK` and `TEMPT` . Simple OLS regressions. ***1%,**5%,*10% significance.

Table 8 regresses the average individual cooperation rate (across the ten periods played) on a gender dummy interacted with `RISK` and `TEMPT` . Columns (1)-(2) show stranger treatments and columns (3)-(4) partner treatments. Columns (1) and (3) include interactions and columns (2) and (4) don't. The table shows that on average across our games there is no gender difference in cooperation rates (columns (2) and (4)). Women are, however, more cooperative than the average man if `RISK` is low and less cooperative if `RISK` is high (above 0.7 approximately in the stranger and above 0.9 in the partner condition). This is consistent with women being more risk averse (Eckel and Grossmann (2008); Dohmen et al. (2011)) and can be one possible explanation for the contradictory findings in the existing literature, which are usually based on one set of parameters only.

Result 4 *Women cooperate more than the average man if `RISK` is low and less if `RISK` is high. There is no statistically significant gender difference on average.*

7.2 Perception of Risk and Temptation

In this subsection we exploit questionnaire answers on how much risk and temptation participants perceived in the game they played to see whether perceptions correlate with cooperation rates and

[15] This excludes one participant who did not answer the questionnaire.

our indicators of risk and temptation RISK and TEMPT. In particular we asked participants in a post-experimental questionnaire how strongly (on a scale from 0-8) they agree to each of the following statements:

R1 "The game we played is all about risk"

T1 "I was often tempted to choose B (defection)"

R2 "I was worried to choose A (cooperation) because the other participant might not"

T2 "The game we played is all about resisting temptation"

We then average their indicated agreement to the two risk (temptation) statements R1 and R2 (T1 and T2) to create an indicator of their risk (temptation) perception. Average agreement with the risk statements was 4.74 (standard deviation 2.08) in the Stranger treatments and 4.35 (2.10) in the Partner treatments. Average agreement with the temptation statements was 4.54 (1.96) in the Stranger treatments and 4.63 (1.94) in the Partner treatments.

	Stranger		Partner	
	(1)	(2)	(1)	(2)
RISK	−0.214***	−0.170**	0.095	−0.186**
	(0.061)	(0.055)	(0.196)	(0.088)
TEMPT	−0.054	−0.126	−0.223**	−0.188***
	(0.062)	(0.068)	(0.093)	(0.090)
RISK PERC	−0.005***	−0.009***	0.002	0.003
	(0.007)	(0.007)	(0.011)	(0.011)
TEMPT PERC	−0.015**	−0.015*	−0.039***	−0.040***
	(0.007)	(0.007)	(0.012)	(0.012)
EFF	0.518*		−1.372	
	(0.311)		(0.857)	
constant	0.296***	0.442***	0.963	0.624***
	(0.099)	(0.046)	(0.224)	(0.074)
Observations	238	238	125	125
R2	0.1752	0.1653	0.2239	0.2072

Table 9: Individual cooperation rates regressed on perception of RISK and TEMPT as well as actual RISK and TEMPT. Random Effects OLS regressions (3 observations dropped in partner condition, because of missing questionnaire answers).

Table 9 shows the results of regressions where we regress the average individual cooperation rate on the variables RISK, TEMPT, (and in columns (1) and (3) EFF) as well as the indicators for risk and temptation perception. The regressions for the stranger treatments (columns (1) and (2)) confirm results from before also in this sub-sample: RISK is strongly detrimental to cooperation rates, while there is no statistically significant impact of TEMPT. Risk perception also has a negative effect on cooperation rates on top of the actual value of RISK in the game. Interestingly, also temptation perception affects cooperation rates negatively. Hence those participants who perceive more temptation in the game do cooperate less.[16] If TEMPT is not able to explain the variation in cooperation rates in the Stranger condition, then this could be because most participants are focusing on the risk component in the game or because the TEMPT index does not correlate well with the concerns expressed in statements T1 and T2. For the partner condition (columns (3) and (4)) we

[16]It is important to note that we cannot and are not making any claims about the direction of the causality. People might claim they were tempted to defect because they defected or they might have defected because they were tempted.

find, as expected, that TEMPT matters, but the perception of temptation has an additional negative effect on cooperation rates.

We also ask whether RISK and TEMPT as measured via the payoff variables are related to perceptions of risk and temptation expressed in the post-experimental questionnaire. To answer this question we regress measured perception of risk and temptation on RISK or TEMPT , respectively. Table 21 in Appendix B.5 shows the results for the Stranger treatments. The table shows that if RISK is higher, then participants express a higher perception of risk in the post experimental questionnaire. This effect is particularly strong (and only statistically significant) if TEMPT is "small" (below the median). Perception of temptation in the stranger setting is generally lower and does not co-vary with the indicator TEMPT . Table 22 in Appendix B.5 shows the results for the partner treatments. In the partner condition risk perception is not significantly affected by the level of RISK in the game. Temptation perception, on the other hand, does co-vary with TEMPT. Taken together this evidence suggests, that participants focus on one of the two aspects of the game. TEMPT seems the dominant factor in the partner condition (where it also co-varies with temptation perception), possibly because participants feel that RISK can be controlled via the repeated interaction. In the stranger condition, on the other hand RISK seems the dominant factor. If participants perceive temptation it does lead to lower cooperation rates (Table 9), but most participants seem to be focused on the risk dimension in this game (Table 21). We summarize our results as follow.

Result 5 *Temptation perception is positively correlated with TEMPT only in the Partner condition and negatively correlated with cooperation rates in both the Stranger and Partner condition. Risk perception is positively correlated with RISK and negatively correlated with cooperation rates only in the Stranger condition.*

8 Conclusions

We conducted a meta-study of 89 prisoner's dilemma treatments with more than 3000 participants across 6 countries. We focused on two dimensions of the dilemma: "risk" (the percentage loss of unilaterally cooperating against a defector) and "temptation" (the percentage gain of unilaterally defecting against a cooperator). While risk explains variation in cooperation rates across random matching ("stranger") and one-shot treatments, temptation explains the variation in repeated ("partner") interactions. Consistently with reputational models (Kreps et al. (1982)) temptation is more effectively curbed in the Partner setting, the longer the time horizon of the game. These results are useful for discriminating between competing theories of why agents cooperate. For policy making, it is useful to see which dimension of the dilemma (risk or temptation) is best targeted to yield improved cooperation rates.

Our results can also contribute to understanding seemingly conflicting results in the existing literature. In terms of the "partner vs stranger" debate (Andreoni and Croson (2008)), we found that there is more cooperation in partner than stranger if and only if risk is high and temptation low. In terms of the debate on gender differences in altruism (Croson and Gneezy (2009)), we found that women are more cooperative than the average man if risk is low, but less cooperative if it is high. We also found that there are no gender differences on average. The fact that both these comparisons are mediated by the RISK and TEMPT measures can explain why previous literature (usually relying on one set of parameters) has found such mixed results.

References

Andreoni, J. (1988). Why free ride?: Strategies and learning in public goods experiments. *Journal of Public Economics 37(3)*, 291–304.

Andreoni, J. and R. Croson (2008). Partners versus strangers: random rematching in public goods experiments. In *Handbook of Experimental Economics Results*, Volume 1, pp. 776–783. New York: Elsevier.

Andreoni, J. and J. Miller (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *Economic Journal 103(418)*, 570–585.

Andreoni, J. and H. Varian (1999). Preplay contracting in the prisoner's dilemma. *Proceedings of the National Academy of Sciences of the United States of America 96*, 10933–10938.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution 54(1)*, 39–57.

Bereby-Meyer, Y. and A. Roth (2006). The speed of learning in noisy games: partial reinforcement and the sustainability of cooperation. *American Economic Review 96(4)*, 1029–1042.

Bohnet, I. and D. Kuebler (2005). Compensating the cooperators: is sorting in the prisoner's dilemma possible? *Journal of Economic Behavior and Organization 56*, 61–76.

Boone, C., B. D. Brabander, and A. van Witteloostuijn (1999). The impact of personality on behavior in five prisoner's dilemma games. *Journal of Economic Psychology 20*, 343–377.

Buhrmester, M., T. Kwang, and S. Gosling (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science 6*, 3–5.

Chaudhury, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics 14*, 47–83.

Cooper, D., D. DeJong, R. Forsythe, and T. Ross (1996). Cooperation without reputation: Experimental evidence from prisoner's dilemma games. *Games and Economic Behavior 12(2)*, 187–218.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47(2)*, 448–474.

Croson, R. and M. B. Marks (2000). Step returns in threshold public goods: A meta- and experimental analysis. *Experimental Economics 2*, 239–259.

Dawes, R. and R. Thaler (1988). Anomalies: Cooperation. *Journal of Economic Perspectives 2(3)*, 187–197.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2011). Individual risk attitudes: Measurement, determinants and bebehavior consequences. *Journal of the European Economic Association 9(3)*, 522–550.

Dresher, M. (1961). *The Mathematics of Games of Strategy: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall.

Eckel, C. and P. Grossmann (2008). Men, women and risk aversion: Experimental evidence. In *Handbook of Experimental Economics Results*, Volume 1, pp. 1061–1073. New York: Elsevier.

Embrey, M., G. Frechette, and S. Yuksel (2014). Backward induction in the finitely repeated prisoner's dilemma. manuscript in preparation.

Engel, C. and L. Zhurakhovska (2013). When is the risk of cooperation worth taking? the prisoner's dilemma as a game of multiple motives. mimeo.

Fischbacher, U., S. Gaechter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters 71(3)*, 397–404.

Frank, R., T. Gilovich, and D. Regan (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives 7(2)*, 159–171.

Friedman, D. and R. Oprea (2012). A continuous dilemma. *American Economic Review 102:1*, 337–363.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics 18*, 141–153.

Grimm, V. and F. Mengel (2009). Cooperation in viscous populations - experimental evidence. *Games and Economic Behavior 66(1)*, 202–220.

Hamerstein, P. (Ed.) (2003). *Cultural and Genetic Evolution of Cooperation*. MIT Press.

Hauk, E. and R. Nagel (2001). Choice of partners in multiple two-person prisoner's dilemma games: An experimental study. *Journal of Conflict Resolution 45*, 770.

Hayashi, N., E. Ostrom, and J. W. W. Yamagishi (1999). Reciprocity, trust and the sense of control: A cross-societal study. *Rationality and Society 11*, 27–46.

Horton, J., D. Rand, and R. Zeckhauser (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics 14*, 399–425.

Kagel, J. and P. McGee (2013). Team vs individual play in finitely repeated prisoner dilemma games. mimeo.

Khadjavi, M. and A. Lange (2013). Prisoners and their dilemma. *Journal of Economic Behavior and Organization 92*, 163–175.

Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoner's dilemma. *Journal of Economic Theory 27*, 245–252.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel and A. E. Roth (Eds.), *Handbook of Experimental Economics*. Princeton University Press.

Liberman, V., S. Samuels, and L. Ross (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin 30*, 1175.

Mengel, F. (2007). The evolution of function-valued traits for conditional cooperation. *Journal of Theoretical Biology 245*, 564–575.

Mengel, F. (2014). Learning by (limited) forward looking players. *Journal of Economic Behavior and Organization in press.*

Mengel, F. and R. Peeters (2011). Strategic behavior in repeated voluntary contribution experiments. *Journal of Public Economics 95*, 143–148.

Murnighan, J. and A. Roth (1983). Expecting continued play in prisoner's dilemma games. *Journal of Conflict Resolution 27(2)*, 279–300.

Normann, H. and B. Wallace (2012). The impact of the termination rule on cooperation in a prisoner's dilemma experiment. *International Journal of Game Theory 41(3)*, 707–718.

Pfeiffer, T., L. Tran, C. Krumme, and D. Rand (2012). The value of reputation. *Journal of the Royal Society - Interface*, 1–20.

Rand, D. (2012). The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology 299*, 172–179.

Rapoport, A. and P. Dale (1967). The "end" and "start" effects in iterated prisoners dilemma. *Journal of Conflict Resolution X(3)*, 363–366.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta- analysis of experiments 1958-1992. *Rationality and Society 7(1)*, 58–92.

Schmidt, D., R. Shupp, J. Walker, T.K.Ahn, and E. Ostrom (2001). Dilemma games: game parameters and matching protocols. *Journal of Economic Behavior and Organization 46*, 357–377.

Snijders, C. and G. Keren (1999). Determinants of trust. In I. E. D.V. Budescu and R. Zwick (Eds.), *Games and Behavior: Essays in Honor of Amnon Rapoport*, pp. 355–383. Lawrence Erlbaum Associates.

Suri, S. and D. Watts (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE 6*, e16836.

Swope, K., J. Cadigan, P. Schmitt, and R. Shupp (2008). Personality preferences in laboratory economics experiments. *The Journal of Socio-Economics 37*, 998–1009.

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics 6*, 299–303.

Online Appendix

A Details on Studies in Data Set

Table 10 shows the payoff parameters from our own lab studies. Table 11 shows the payoff parameters of our AMT studies and Table 12 shows the studies from the existing literature we use together with some key statistics. The column "details" in Table 12 mentions special details of some studies. Studies with such details are dropped in the robustness regressions shown in Table 17.

Partner	Stranger
$(400, 100, 450, 120)$	$(400, 100, 450, 120)$
$(400, 100, 450, 200)$	$(400, 100, 450, 200)$
$(10, 1, 90, 5)$	$(10, 1, 90, 5)$
$(150, 40, 850, 50)$	$(150, 40, 850, 50)$
$(150, 5, 850, 95)$	$(150, 5, 850, 95)$
$(250, 50, 750, 150)$	$(250, 50, 750, 150)$
	$(10, 1, 110, 9)$
	$(250, 100, 750, 160)$
	$(100, 50, 100, 50)$
	$(400, 100, 800, 200)$

Table 10: Payoff parameters (format (a, b, c, d), see the payoff matrix in Table 1) of additional studies run between December 2013 - January 2014 at EssexLab at the University of Essex. In all studies 10 periods were used with either a Partner matching or a Stranger matching in matching groups of size 6. 24 participants in each study.

$(400, 200, 450, 200)$	$(400, 10, 450, 200)$
$(400, 200, 800, 200)$	$(400, 10, 800, 200)$
$(400, 100, 450, 120)$	$(400, 100, 450, 200)$
$(10, 1, 90, 5)$	$(10, 5, 90, 5)$
$(150, 40, 850, 50)$	$(150, 5, 850, 95)$
$(250, 15, 750, 85)$	$(250, 5, 750, 95)$
$(250, 50, 750, 150)$	$(250, 100, 750, 160)$
$(10, 2, 110, 3)$	$(10, 1, 110, 9)$
$(150, 50, 850, 100)$	$(250, 50, 750, 150)$
$(400, 100, 600, 120)$	$(400, 100, 600, 200)$
$(400, 100, 1200, 120)$	$(400, 100, 1200, 200)$

Table 11: Payoff parameters (format (a, b, c, d)) of AMT studies, which are all one-shot games.

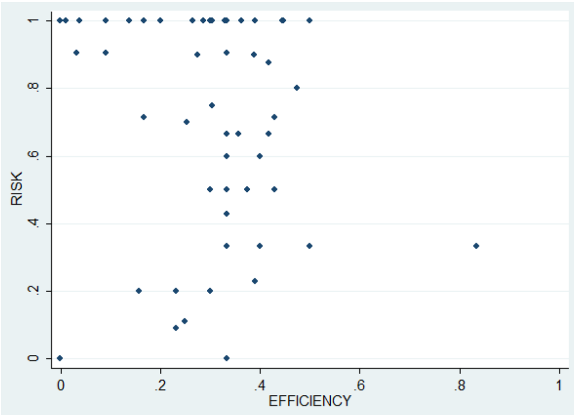
We provide some more information on the AMT platform. AMT is an online labor market in which employers can employ workers to complete short tasks (generally less than 10 minutes) for relatively small amounts of money (generally less than 1 US dollar). One major advantage of AMT is it allows experimenters to easily expand beyond the college student convenience samples typical of most economic game experiments. Among American subjects, AMT subjects have been shown to be significantly more nationally representative than college student samples (Buhrmester et al. (2011)). Running experiments online necessarily involves some loss of control, since the workers cannot be directly monitored as in the traditional lab; hence, experimenters cannot be certain that each observation is the result of a single person (as opposed to multiple people making joint decisions at the same computer). Moreover, although the sample of subjects in AMT experiments is more diverse

Reference	(a, b, c, d)	Partner	Periods	Other Details
Rapoport and Dale (1967)	$(1, -10, 10, -5)$	1	25	all male
Rapoport and Dale (1967)	$(5, -10, 10, -5)$	1	25	all female
Andreoni and Miller (1993)	$(7, 0, 12, 4)$	1	10	
Andreoni and Miller (1993)	$(7, 0, 12, 4)$	0	10	
Andreoni and Varian (1999)	$(6, 0, 8, 3)$	0	15	asymmetric
Andreoni and Varian (1999)	$(7, 0, 11, 4)$	0	15	asymmetric
Frank et al. (1993)	$(2, 0, 3, 1)$	0	1	paper based
Cooper et al. (1996)	$(800, 0, 1000, 350)$	0	10	(avg of 2 groups)
Cooper et al. (1996)	$(800, 0, 1000, 350)$	1	10	
Boone et al. (1999)	$(300, -600, 600, -30)$	0	12	Econ students
Boone et al. (1999)	$(300, -600, 600, -30)$	1	12	Econ students
Boone et al. (1999)	$(600, -400, 800, -20)$	1	12	Econ students
Hayashi et al. (1999)	$(10, 0, 15, 5)$	0	1	
Hayashi et al. (1999)	$(10, 0, 15, 5)$	0	1	
Hauk and Nagel (2001)	$(5, -6, 7, -1)$	1	10	
Schmidt et al. (2001)	$(40, 110, 10, 80)$	0	12	
Schmidt et al. (2001)	$(50, 110, 10, 70)$	0	12	
Schmidt et al. (2001)	$(50, 110, 10, 90)$	0	12	
Schmidt et al. (2001)	$(60, 110, 10, 80)$	0	12	
Schmidt et al. (2001)	$(30, 110, 10, 70)$	0	12	
Schmidt et al. (2001)	$(40, 110, 10, 60)$	0	12	
Liberman et al. (2004)	$(40, -20, 80, 0)$	1	7	(avg of 2 groups)
Bohnet and Kuebler (2005)	$(350, 0, 500, 150)$	0	5	
Bohnet and Kuebler (2005)	$(350, 100, 500, 150)$	0	5	
Bereby-Meyer and Roth (2006)	$(0.105, 0.005, 0.175, 0.075)$	1	10	
Swope et al. (2008)	$(10, 2, 20, 5)$	0	1	
Grimm and Mengel (2009)	$(800, 100, 850, 150)$	0	100	beliefs elicited
Grimm and Mengel (2009)	$(800, 100, 1100, 400)$	0	100	beliefs elicited
Friedman and Oprea (2012)	$(10, 0, 18, 4)$	0	1	
Friedman and Oprea (2012)	$(10, 0, 14, 4)$	0	1	
Friedman and Oprea (2012)	$(10, 0, 14, 8)$	0	1	
Friedman and Oprea (2012)	$(10, 0, 18, 8)$	0	1	
Friedman and Oprea (2012)	$(10, 0, 14, 2)$	0	1	
Normann and Wallace (2012)	$(800, 50, 1000, 350)$	1	22	
Normann and Wallace (2012)	$(800, 50, 1000, 350)$	1	10	
Normann and Wallace (2012)	$(800, 50, 1000, 350)$	1	5	
Pfeiffer et al. (2012)	$(30, -10, 40, 0)$	0	30	
Khadjavi and Lange (2013)	$(7, 1, 9, 3)$	0	1	
Engel and Zhurakhovska (2013)	$(5, 0, 10, [0, 5])$	0	1	
Kagel and McGee (2013)	$(105, 5, 175, 75)$	1	10	

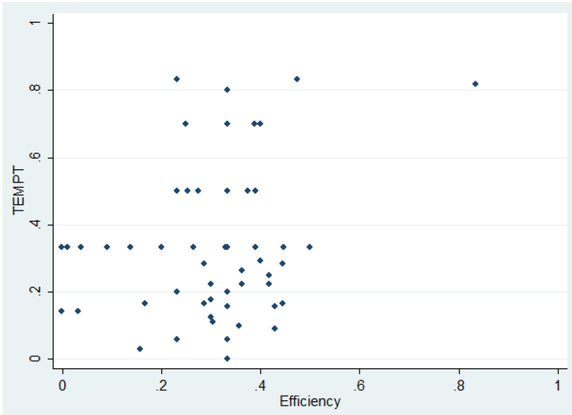
Table 12: Studies from existing literature used in the paper sorted by year of publication. Partner is zero if matching is one-shot or stranger (and one if partner). Studies with special "other details" as well as AMT studies are identified by a dummy.

than samples using college undergraduates, we are obviously restricted to people that participate in online labor markets. To address these potential concerns, recent studies have explored the validity of data gathered using AMT (for an overview, see Rand (2012)). Suri and Watts (2011), e.g., find quantitative agreement in contribution behavior in a repeated public goods game between experiments conducted in the physical lab and those conducted using AMT with approximately 10-fold lower stakes (see also Horton et al. (2011)). It has also been shown that AMT subjects display a level of test-retest reliability similar to what is seen in the traditional lab on measures of political beliefs, self-esteem, Social Dominance Orientation, and Big-Five personality traits (Buhrmester et al. (2011)) as well as belief in God, age, gender, education level and income (Rand (2012)). Hence we are confident that our measures obtained on AMT are at least somewhat representative. However, for our main results we always demonstrate that they are robust to excluding the AMT studies.

Figure 4 illustrates the correlation between `RISK` and `TEMPT` on the one hand and `EFF` on the other hand. As discussed above these variables are largely uncorrelated.



(a) `RISK` vs `EFF`



(b) `TEMPT` vs `EFF`

Figure 4: Correlation of `RISK` `TEMPT` variables with efficiency (`EFF`) across all studies considered.

B Additional Results

B.1 Additional theoretical background

We derive the conditions provided for the Partner setting in Section 3. For the 0-type we need to show (i) that she prefers to choose MGT over 2MGT and (ii) that there are no incentives to deviate in order to learn about p_{ij}. Condition (i) requires that

$$p_{ij}\frac{(T-1)a+c}{T}+(1-p_{ij})\frac{(T-1)a+d}{T}>p_{ij}\frac{(T-2)a+c+d}{T}+(1-p_{ij})\frac{(T-2)a+c+d}{T}, \quad (7)$$

where T denotes the length of the game. Note that we are not using a discount-factor here, which seems reasonable given the typical horizon in an experiment and the fact that participants are only paid after the experiment is finished. Including a discount-factor would not affect the conclusions we draw from this analysis. The condition above reduces to the condition given in Section 3

$$p_{ij}>\frac{\text{Temptation}}{c-d}=:PA. \quad (8)$$

There are no incentives to experiment to learn, because both types are indistinguishable until period T-1. For the θ-type the following condition needs to be satisfied in order not to switch to MGT:

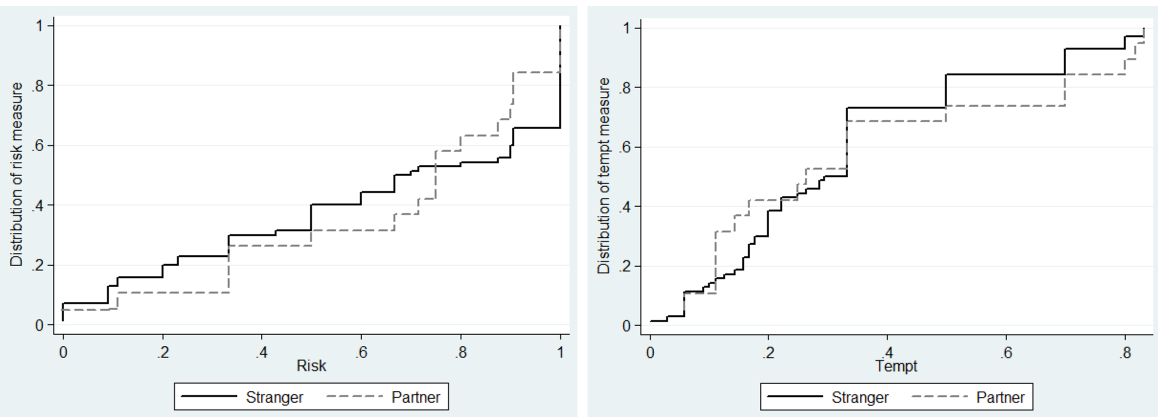
$$p_{ij}a+(1-p_{ij})\frac{(T-1)a+b}{T}+\widetilde{g}(\theta)>p_{ij}\frac{(T-1)a+c}{T}+(1-p_{ij})\frac{(T-1)a+d}{T}. \quad (9)$$

This equation reduces to the equation given in Section 3

$$\theta>\widetilde{g}^{-1}(\text{Risk}+p_{ij}(\text{Temptation-Risk})). \quad (10)$$

B.2 Balancing Tests

We regress our variables of interest `RISK` and `TEMPT` on a partner dummy to see whether the settings are balanced in terms of the mean of these variables. Results are reported in Table 13. We also illustrate the cumulative distributions in Figure 5.



(a) `RISK` (b) `TEMPT`

Figure 5: CDF of the `RISK` and `TEMPT` measures in the partner and stranger studies.

	RISK	RISK	RISK	RISK	RISK
Partner	0.031	0.027	0.081	0.048	−0.006
	(0.090)	(0.118)	(0.1023)	(0.058)	(0.063)
constant	0.637***	0.225***	0.237***	0.758***	0.824***
	(0.041)	(0.046)	(0.046)	(0.036)	(0.034)
Sample	All	LL	LH	HL	HH
Observations	89	19	19	18	14
R2	0.0014	0.0032	0.0359	0.0018	0.0001
	TEMPT	TEMPT	TEMPT	TEMPT	TEMPT
Partner	0.024	0.004	0.026	−0.042	0.083
	(0.061)	(0.075)	(0.098)	(0.045)	(0.125)
constant	0.330***	0.146***	0.562***	0.218***	0.583***
	(0.028)	(0.029)	(0.043)	(0.028)	(0.067)
Sample	All	LL	LH	HL	HH
Observations	89	19	19	18	14
R2	0.0018	0.0002	0.0039	0.0052	0.0353

Table 13: Balancing Tests. Simple OLS regressions. ***1%,**5%,*10% significance. LL denotes the sample where both RISK and TEMPT are below the median, LH where RISK is below and TEMPT above, HL where RISK is above and TEMPT below and HH where both are above.

B.3 Additional Tables: Robustness Checks Average Rates Regressions

The tables collected in this section all show different robustness checks for our regressions based on average rates (Tables 2 and 3 in the main text). Table 14 shows the results of a weighted regression where we weigh the different studies by the number of independent observations they contain. We were able to obtain this information for all studies except for six studies in the Stranger condition. For five of them we set the to the mean weight of all remaining studies (0.014). The sixth one was a study conducted with the strategy method and we set the weight proportional to $\frac{N}{cond}$ where N was the number of participants and *cond* the number of different conditions participants had to indicate their choices for. This resulted in a weight of 0.008 for this study. The table shows that results are robust. RISK is the predominant factor explaining the variation in cooperation rates in Stranger settings.

Table 15 shows the weighted regressions for the Partner treatment. Here the number of independent observations was known except for one study, where we set the weight to the average weight (0.052). The results are again qualitatively robust. TEMPT emerges as the dominant factor explaining variation in cooperation rates. Compared to Table 3 we see substantially higher values of R2 here and we see a significant interaction effect EFFx TEMPT suggesting that the extent of temptation is mediated by possible efficiency gains of joint cooperation.

Table 16 shows a different way of dealing with studies where RISK$= 1$ as a consequence of $b = 0$. Instead of dropping those studies, here b is set to 1 (if $d < 100$) or 10 (if $d > 100$) which generates values of RISK< 0.9. Columns (1)-(4) show results for the Stranger condition. Columns (1) and (3) contain all studies and columns (2) and (4) omit the AMT studies. Columns (5)-(6) show the results of the Partner condition. It can be seen that results are very robust. In the Stranger condition only RISK is statistically significant with a coefficient of around -0.3 as in Table 2. In the Partner condition RISK is not statistically significant, but instead TEMPT is with a coefficient size of about -0.25.

In Table 17 we drop studies with “special” details, e.g. where beliefs were elicited or were only women or men participated or studies that were conducted on AMT or were paper-based. These “special” details are listed in Table 12. Eliminating these studies reduces the sample size (and hence

	STRANGER					
	(1)	(2)	(3)	(4)	(5)	(6)
RISK	-0.289^{**}	-0.655^{***}	-0.686^{**}	-0.242^{***}	-0.363^{***}	-0.454^{***}
	(0.147)	(0.132)	(0.269)	(0.054)	(0.072)	(0.079)
TEMPT	0.125	0.028	−0.011	-0.144^{*}	−0.117	0.025
	(0.444)	(0.491)	(0.341)	(0.075)	(0.087)	(0.114)
EFF	−0.099	−0.307	−0.638	−0.200	0.009	−0.253
	(0.364)	(0.421)	(0.924)	(0.255)	(0.230)	(0.183)
EFFx RISK	0.234	1.005^{**}	0.914			
	(0.504)	(0.469)	(0.801)			
EFFx TEMPT	−0.790	−0.510	−0.467			
	(1.225)	(1.439)	(2.082)			
TEMPTx RISK	−0.046	0.035	0.287			
	(0.264)	(0.0.338)	(0.839)			
constant	0.628^{***}	0.716^{***}	0.703^{***}	0.673^{***}	0.628^{***}	0.609^{***}
	(0.104)	(0.108)	(0.115)	(0.080)	(0.077)	(0.064)
Lab Sample	All	RISK<1	RISK<1;	All	RISK<1	RISK<1;
AMT	YES	YES	NO	YES	YES	NO
Observations	70	46	24	70	46	24
R2	0.3090	0.5238	0.5987	0.2951	0.4880	0.5592
Sum of Weights	1	0.785	0.137	1	0.785	0.137

Table 14: Robustness check: Average cooperation rate regressed on variables of interest. OLS regressions weighted by number of independent observations. (Pr>F)< 0.0009.***1%,**5%,*10% significance.

	PARTNER			
	(1)	(2)	(3)	(4)
RISK	0.291^{*}	0.182	0.005	−0.017
	(0.141)	(0.110)	(0.042)	(0.049)
TEMPT	-0.398^{*}	-0.461^{**}	-0.262^{***}	-0.280^{***}
	(0.217)	(0.185)	(0.053)	(0.043)
EFF	0.095	−0.197	−0.118	−0.047
	(0.392)	(0.406)	(0.169)	(0.193)
EFFx RISK	-0.793^{*}	−0.439		
	(0.558)	(0.417)		
EFFx TEMPT	0.825^{*}	1.071^{**}		
	(0.387)	(0.408)		
TEMPTx RISK	−0.224	−0.293		
	(0.223)	(0.204)		
constant	0.392^{**}	0.490^{***}	0.517^{***}	0.508^{***}
	(0.154)	(0.135)	(0.055)	(0.062)
Lab Sample	All	RISK<1	All	RISK<1
Observations	19	16	19	16
R2	0.7936	0.8144	0.4953	0.5033
Sum of Weights	1	0.919	1	0.919

Table 15: Robustness check: Average cooperation rate regressed on variables of interest. OLS regressions weighted by number of independent observations. (Pr>F)< 0.0009.***1%,**5%,*10% significance.

	Stranger				Partner	
	(1)	(2)	(3)	(4)	(5)	(6)
RISK	-0.288^{**}	-0.362^{**}	-0.317^{***}	-0.300^{***}	0.248	-0.024
	(0.114)	(0.160)	(0.055)	(0.064)	(0.177)	(0.103)
TEMPT	0.177	-0.081	-0.019	-0.078	-0.408	-0.252^{**}
	(0.360)	(0.498)	(0.082)	(0.120)	(0.262)	(0.106)
EFF	0.428	0.135	-0.007	-0.074	0.151	0.033
	(0.426)	(0.520)	(0.206)	(0.225)	(0.612)	(0.174)
EFFx RISK	-0.519^{*}	-0.272			-0.820	
	(0.303)	(0.379)			(0.579)	
EFFx TEMPT	-0.837	-0.420			0.706	
	(1.210)	(1.714)			(0.600)	
TEMPTx RISK	0.205	0.332			-0.184	
	(0.307)	(0.513)			(0.294)	
constant	0.430^{***}	0.494^{***}	0.504^{***}	0.491^{***}	0.423^{***}	0.444^{***}
	(0.121)	(0.150)	(0.059)	(0.064)	(0.190)	(0.092)
Lab Sample	All	All	All	All	All	All
AMT	YES	NO	YES	NO	-	-
Observations	70	48	70	48	19	19
R2	0.4552	0.4977	0.4277	0.4847	0.6788	0.3171

Table 16: Robustness check: average cooperation rate regressed on variables of interest. Studies where $b = 0$ (and hence RISK$= 1$) are not dropped. Instead b is set to 1 (if $d < 100$) or 10 (if $d > 100$) to generate values of RISK< 0.9. Simple OLS regressions. (Pr>F)< 0.0009.***1%,**5%,*10% significance.

statistical power), especially for the Stranger studies but it can be seen that the main effects prevail and are, in fact, stronger and even more clear-cut.

	Stranger	Partner
RISK	-0.431^{**}	-0.076
	(0.180)	(0.080)
TEMPT	-0.002	-0.268^{***}
	(0.164)	(0.072)
EFF	-0.419	0.014
	(0.480)	(0.304)
constant	0.628^{***}	0.473^{***}
	(0.123)	(0.079)
Lab Sample	RISK< 1	RISK< 1
Observations	26	12
R2	0.5781	0.7168

Table 17: Robustness check: dropping studies with "special" details (beliefs elicited, only one gender, AMT or paper-based). Simple OLS regressions. (Pr>F)< 0.0139.***1%,**5%,*10% significance.

Table 18 shows the result of regressions, where we split the sample by whether joint cooperation or alternating between outcomes (C,D) and (D,C) is efficient i.e. whether $b + c \leq 2a$. Due to the smaller number of observations in the Partner condition, we conduct this analysis only for the Stranger condition. Columns (1) and (3) show the cases where mutual cooperation is efficient and columns (2) and (4) where alternating is efficient. The results, particularly across columns (3) and (4) look very similar and in line with the results seen before.

In Table 19 we separate the one-shot and Stranger treatments. The results across these conditions are very similar. RISK is the predominant variable explaining variation and in the model without

	Stranger			
	(1)	(2)	(3)	(4)
RISK	-1.423^{**}	-0.562	-0.341^{**}	-0.337^{***}
	(0.612)	(0.707)	(0.164)	(0.098)
TEMPT	-0.332	-0.697	-0.480	-0.058
	(1.083)	(1.350)	(0.481)	(0.173)
EFF	-0.364	-2.111	-0.363	-0.490
	(1.058)	(2.693)	(0.446)	(0.509)
EFFx RISK	2.657^{*}	0.842		
	(1.477)	(1.996)		
EFFx TEMPT	-5.270	1.971		
	(3.806)	(4.247)		
TEMPTx RISK	2.444	-0.128		
	(1.633)	(0.652)		
constant	0.825^{**}	1.251	0.715^{***}	0.719^{***}
	(0.325)	(0.896)	(0.152)	(0.187)
Lab Sample	RISK<1	RISK<1	RISK<1	RISK<1
AMT	YES	YES	YES	YES
$b+c \leq 2a$	YES	NO	YES	NO
Observations	23	23	23	23
R2	0.5246	0.4833	0.4051	0.4663

Table 18: Robustness check: average cooperation rate regressed on variables of interest. Sample split depending on whether joint cooperation or alternating between outcomes (C,D) and (D,C) is efficient i.e. whether $b+c \leq 2a$. Simple OLS regressions. (Pr>F)<0.0009.***1%,**5%,*10% significance.

interactions (columns (3) and (4)) the coefficient size for both the one-shot and Stranger settings is very similar.

	Stranger			
	(1)	(2)	(3)	(4)
RISK	-1.145^{*}	-0.548	-0.314^{**}	-0.308^{***}
	(0.614)	(0.413)	(0.146)	(0.074)
TEMPT	0.190	0.032	-0.090	-0.148^{*}
	(0.788)	(0.485)	(0.162)	(0.081)
EFF	-0.860	-0.225	-0.102	-0.249
	(1.236)	(0.668)	(0.461)	(0.267)
EFFx RISK	2.050	0.892		
	(1.757)	(1.210)		
EFFx TEMPT	-1.943	-0.445		
	(2.850)	(1.624)		
TEMPTx RISK	0.757	-0.104		
	(0.748)	(0.322)		
constant	0.789^{**}	0.696^{***}	0.491^{***}	0.726^{***}
	(0.347)	(0.195)	(0.157)	(0.085)
Lab Sample	RISK<1	RISK<1	RISK<1	RISK<1
AMT	-	YES	-	YES
>1 periods	YES	NO	YES	NO
Observations	21	25	21	25
R2	0.3995	0.5933	0.2993	0.5769

Table 19: Robustness check: separating one-shot and Stranger studies (with >1 periods). Simple OLS regressions. (Pr>F)<0.0009.***1%,**5%,*10% significance.

B.4 Additional Tables: Dynamics

In this subsection we present additional evidence regarding the dynamics of play across the ten periods (see Section 6). Table 20 shows the results of regressions where we regress the binary choice to cooperate on period (1,...,10) a "partner-dummy" indicating partner matching and the interaction of the two (columns (1), (3), (5) and (7)). The sample is split according to whether `RISK` and `TEMPT` are above or below median. The interaction between the partner dummy and period is rarely significant. Only when `RISK` is high and `TEMPT` small (column (5)) a marginally significant difference in dynamics between partner and stranger can be detected. Not only do partners cooperate significantly more in this case (columns (5) and (6)), cooperation rates also seem to decline less steeply in the partner compared to the stranger setting.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
period	-0.034^{***}	-0.036^{***}	-0.017^{***}	-0.018^{***}	-0.023^{***}	-0.013^{***}	-0.024^{***}	-0.025^{***}
	(0.005)	(0.003)	(0.005)	(0.004)	(0.006)	(0.003)	(0.003)	(0.003)
Partner	0.082	0.051	0.093	0.086^{**}	0.206^{***}	0.282^{***}	0.048	0.020
	(0.075)	(0.060)	(0.061)	(0.041)	(0.068)	(0.054)	(0.060)	(0.035)
period X part	-0.005		-0.001		0.013^{*}		-0.005	
	(0.007)		(0.008)		(0.007)		(0.008)	
constant	0.581^{***}	0.593^{***}	0.260^{***}	0.264^{***}	0.268^{***}	0.208^{***}	0.304^{***}	0.309^{***}
	(0.048)	(0.045)	(0.043)	(0.037)	(0.061)	(0.051)	(0.024)	(0.023)
`RISK`	Small	Small	Small	Small	High	High	High	High
`TEMPT`	Small	Small	High	High	Small	Small	High	High
Observations	1140	1140	940	940	1860	1860	1400	1400
Groups	114	114	94	94	186	186	140	140
ρ	0.380	0.380	0.162	0.162	0.353	0.353	0.089	0.089

Table 20: Dynamics of cooperation (Partner vs Stranger). Random Effects OLS regressions.***1%,**5%,*10% significance.

B.5 Additional Tables: Questionnaire Data

To answer this question we regress measured perception of risk and temptation on `RISK` or `TEMPT`, respectively. Table 21 shows the results for the Stranger treatments. Column (1) shows that if `RISK` is higher, then participants express a higher perception of risk in the post experimental questionnaire. This effect is particularly strong (and only statistically significant) if `TEMPT` is "small" (below the median), as columns (2) and (3) demonstrate. Columns (4)-(6) show that the perception of temptation is generally lower at a level of about 4.5 (out of 7). Furthermore it does not co-vary with the indicator `TEMPT`. There are at least two possible explanations for this. One is that participants are focused on risk (`RISK`) and not on temptation (`TEMPT`) in the Stranger condition. This is consistent with the evidence in the previous sections. Another explanation is that our measure of temptation `TEMPT` is not a good measure of temptation as expressed by questionnaire questions T1 and T2. Our evidence from the Partner treatment speaks against the latter explanation.

Table 22 shows the results for the partner treatments. In the partner condition risk perception is not significantly affected by the level of `RISK` in the game (columns (1)-(3)). Temptation perception, on the other hand, does co-vary with `TEMPT` and is higher than in the Stranger condition (as soon as `TEMPT` exceeds 0.2). Taken together this evidence suggests, that participants focus on one of the two aspects of the game. `TEMPT` seems the dominant factor in the partner condition (where it also co-varies with temptation perception), possibly because participants feel that `RISK` can be controlled via the repeated interaction. In the stranger condition, on the other hand `RISK` seems the dominant factor. If participants perceive temptation it does lead to lower cooperation rates (Table 9), but most participants seem to be focused on the risk dimension in this game (Table 21).

	STRANGER		
	(1)	(2)	(3)
RISK	0.345*	0.980**	0.304
	(0.369)	(0.358)	(0.520)
constant	4.778***	4.707***	4.780***
	(0.222)	(0.269)	(0.365)
Sample	All	Small TEMPT	High TEMPT
Observations	228	90	162
R2	0.0039	0.0363	0.0021
	(4)	(5)	(6)
TEMPT	0.497	0.879	0.357
	(0.375)	(0.788)	(0.805)
constant	4.514***	4.457***	4.580***
	(0.209)	(0.254)	(0.575)
Sample	All	Small RISK	High RISK
Observations	228	112	116
R2	0.0077	0.0112	0.0017

Table 21: Risk (Temptation) perception regressed on RISK (TEMPT). Stranger treatments. Simple OLS regressions. ***1%,**5%,*10% significance.

	PARTNER		
	(1)	(2)	(3)
RISK	−0.113	−1.890	−0.208
	(0.593)	(2.335)	(0.706)
constant	4.768***	5.005***	4.910***
	(0.289)	(0.570)	(0.422)
Sample	All	Small TEMPT	High TEMPT
Observations	120	48	72
R2	0.0003	0.0140	0.0012
	(4)	(5)	(6)
TEMPT	1.132**	1.075**	-
	(0.537)	(0.590)	-
constant	4.368***	4.374***	5.208***
	(0.267)	(0.255)	(0.425)
Sample	All	Small RISK	High RISK
Observations	120	96	24
R2	0.0362	0.0341	0.0000

Table 22: Risk (Temptation) perception regressed on RISK (TEMPT). Partner treatments. Simple OLS regressions. ***1%,**5%,*10% significance. No variation in TEMPT for the sample in column (6).

C Instructions Own Lab Experiments - Stranger treatments

Welcome and thanks for participating at this experiment. Please read these instructions carefully. They are identical for all the participants with whom you will interact during this experiment.

If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions. From now on communication with other participants is not allowed. If you do not conform to these rules we are sorry to have to exclude you from the experiment. Please do also switch off your mobile phone at this moment.

For your participation you will receive 2 pounds. During the experiment you can earn more. How much depends on your behavior and the behavior of the other participants. During the experiment we will use ECU (Experimental Currency Units) and at the end we will pay you in pounds according to the exchange rate 1 pound = 300 ECU. All your decisions will be treated confidentially.

THE EXPERIMENT

The experiment consists of 10 periods. At the start of each period you will be randomly matched with another participant. You will then in each period play the following game with the participant you were matched with in that period.

The other participant chooses		
	A	B
You choose A	a,a	b,c
You choose B	c,b	d,d

In the table your actions and payoffs are given in dark grey and your neighbours actions and payoffs in light grey. The table is read as follows (dark payoffs):

- If you choose A and the other participant chooses A, you receive a
- If you choose A and the other participant chooses B, you receive b
- If you choose B and the other participant chooses A, you receive c
- If you choose B and the other participant chooses B, you receive d

Note that the other participant (light payoffs) is in the same situation as you are. This means that for the other participant:

- If the other participant chooses A and you A, the other participant receives a
- If the other participant chooses A and you B, the other participant receives b
- If the other participant chooses B and you A, the other participant receives c
- If the other participant chooses B and you B, the other participant receives d

At the end of each period you will be informed of your choice, the choice of the other participant and your payoffs in that round.

Keep in mind that the other participant you are matched with changes in each period.

After the last period has been completed you will be asked to fill in a short questionnaire.

Your earnings in the experiment will be the sum of payoffs obtained in each round (exchanged into pounds according to the exchange rate above) plus the 2 pound show up fee.

Enjoy the Experiment!

D Instructions Own Lab Experiments - Partner treatments

Welcome and thanks for participating at this experiment. Please read these instructions carefully. They are identical for all the participants with whom you will interact during this experiment.

If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions. From now on communication with other participants is not allowed. If you do not conform to these rules we are sorry to have to exclude you from the experiment. Please do also switch off your mobile phone at this moment.

For your participation you will receive 2 pounds. During the experiment you can earn more. How much depends on your behavior and the behavior of the other participants. During the experiment we will use ECU (Experimental Currency Units) and at the end we will pay you in pounds according to the exchange rate 1 pound = 300 ECU. All your decisions will be treated confidentially.

THE EXPERIMENT

The experiment consists of 10 periods. At the beginning of the experiment you will be randomly matched with another participant. You will then in each period play the following game with the participant you were matched with.

The other participant chooses		
	A	B
You choose A	a,a	b,c
You choose B	c,b	d,d

In the table your actions and payoffs are given in dark grey and your neighbours actions and payoffs in light grey. The table is read as follows (dark payoffs):

- If you choose A and the other participant chooses A, you receive a
- If you choose A and the other participant chooses B, you receive b
- If you choose B and the other participant chooses A, you receive c
- If you choose B and the other participant chooses B, you receive d

Note that the other participant (light payoffs) is in the same situation as you are. This means that for the other participant:

- If the other participant chooses A and you A, the other participant receives a
- If the other participant chooses A and you B, the other participant receives b
- If the other participant chooses B and you A, the other participant receives c
- If the other participant chooses B and you B, the other participant receives d

At the end of each period you will be informed of your choice, the choice of the other participant and your payoffs in that round.

Keep in mind that the other participant you are matched with stays the same in each period.

After the last period has been completed you will be asked to fill in a short questionnaire.

Your earnings in the experiment will be the sum of payoffs obtained in each round (exchanged into pounds according to the exchange rate above) plus the 2 pound show up fee.

Enjoy the Experiment!