Risk and Temptation: A Meta-Study on Prisoner's Dilemma Games*

Friederike Mengel †
University of Essex
and Lund University

August 3, 2017

Abstract

This paper reports the results of a meta-study of 96 prisoner's dilemma studies comprising more than 3500 participants. We disentangle the role of "risk" (to cooperate unilaterally) and "temptation" (to defect against a cooperator) and find that (i) an index of risk best explains the variation in cooperation rates across one-shot games, while (ii) an index of temptation best explains the variation in finitely repeated games. Risk and temptation indices also affect gender comparisons. Women are more cooperative than the average man if risk is low and less cooperative if risk is high. There are no gender differences on average.

JEL Codes: C72, C90, D01, D70.
Keywords: *Prisoner's dilemma, Cooperation, Meta-study, Experiments, Game Theory;*

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. Social dilemma situations involve a tension between self and group interest that is at the heart of many interactions including effort provision in teams, tax compliance, public good provision or simply good citizenship behaviour. Achieving and maintaining high rates of cooperation in (many of) these situations seems central to create well functioning societies. In the literature different theories of human cooperation have been proposed (Dresher, 1961; Kreps et al., 1982; Axelrod, 1984; Hamerstein, 2003; Fischbacher et al.,

*I thank Jim Andreoni, Matt Embrey, Simon Gaechter, David Hugh Jones, Willemien Kets, Arno Riedl, Stephanie Rosenkranz, Sigrid Suetens, Till Weber, two anonymous reviewers and seminar participants at Fort Lauderdale (ESA North America 2014), Kiel (SBRCR workshop 2015), JRC Ispra, London (LEW 2016), Nottingham (NIBS workshop 2015) 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. Financial support from the European Union (grant PERG08-GA-2010-277026) is gratefully acknowledged.

†Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ (UK); Department of Economics, Lund University, SE-220 07 Lund (SE); *e-mail*: fr.mengel@gmail.com

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). Despite this long history of research using the prisoner's dilemma, there is still no agreement on why people cooperate and on how repetition (the number of stages and matches) affects cooperation (Normann and Wallace, 2012; Embrey et al., 2016). Another open question concerns gender differences in cooperation (Croson and Gneezy, 2009).

In this paper we try to organize existing evidence by disentangling the role of "risk" (to cooperate unilaterally) and "temptation" (to defect against a cooperator), the two defining properties of the prisoner's dilemma. We measure temptation by the percentage gain when unilaterally defecting against a cooperator (TEMPT) and risk by the percentage loss of unilaterally cooperating against a defector (RISK).[1] Understanding which of the two motives is more important for cooperation failure can help decide in which of two games higher cooperation rates can be expected, which can in turn inform the design of mechanisms and policy. It can also help discriminate between theories and inform new theory creation, as we discuss in more detail below. A second aim of the paper is to understand to which extent seemingly contradictory results in the literature can be explained along these two dimensions.

Our analysis is based on a meta-study of 96 prisoner's dilemma studies (combinations of payoff parameters, number of stages and matches) with more than 3500 participants across 6 countries. Studies include one-shot settings (including random rematching) and finitely repeated games with the 2×2 prisoner's dilemma as the stage game. Average cooperation rates across these 96 studies range from 0.04 to 0.84 with a mean of about 0.35.

We find that RISK best explains variation in cooperation rates across random matching and one-shot treatments, where people have no prior experience with their matches and hence face substantial uncertainty. Neither TEMPT, nor a measure of efficiency nor any of eleven other indices used in previous literature can explain this variation once RISK is controlled for. RISK also explains more of the

[1] On top of the requirements that RISK>0 and TEMPT>0, often a third condition is assumed in the prisoner's dilemma namely that mutual cooperation yields higher payoffs than mutual defection (efficiency gains). Without this condition there is no tension between individual and social rationality. We will control for possible efficiency gains using a variable we call EFF . If TEMPT$=0$, then the prisoner's dilemma becomes a stag hunt game and if RISK$=0$ it is a Hawk Dove game. If RISK , TEMPT and EFF are all zero then the game is a trivial game where all payoffs are equal.

variation in average cooperation rates by itself than any of the other indices.

Results are different in the repeated game condition. Here RISK plays a minor role and, if at all, it is our measure of temptation (TEMPT) that can explain variation in cooperation rates. However, a number of other indices from the literature play a role as well in the partner setting. When we compare RISK , TEMPT , efficiency and eleven other indices from the literature in terms of the share of the variation in cooperation rates they explain, we find that TEMPT explains variation well when the length of the game is accounted for. Otherwise other indices from the literature perform better.

We then ask whether seemingly contradictory results in the literature can be understood in terms of RISK and TEMPT. We first focus on the comparison between "partner" (repeated game) and "stranger" (one-shot) settings and find that overall there is no difference in cooperation rates between the two settings. There is more cooperation in "partner" than "stranger" if and only if RISK is high (above median) and TEMPT is low (below median). Interestingly, this differs from findings by Zelmer (2003) who finds that there is more cooperation (higher contributions) overall in "partner" than in "stranger" in the related public good game. In Section 3 we discuss possible explanations for this difference.

We also find that women are more cooperative than the average man if and only if RISK is low and less cooperative if and only if RISK 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 could be one explanation for why previous literature (usually relying on one set of payoff parameters) has found such mixed results (see Andreoni and Croson (2008) or Croson and Gneezy (2009) for surveys).[2]

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

[2]Andreoni and Croson (2008) summarize some of the research on the "partner" vs "stranger" question in a handbook article on the public good game. In the prisoner's dilemma Andreoni and Miller (1993) and Cooper et al. (1996) found more cooperation in the "partner" condition, Boone et al. (1999) found no difference and Andreoni (1988) found more cooperation in the stranger setting. Dal Bo (2005) find more cooperation in the "partner" condition with a long horizon in the repeated game and no difference with a short horizon of the repeated game. See also the more detailed discussion of differences between these papers in Section 4. Most articles on the prisoner's dilemma we found do *not* contain data on both the "partner" and the "stranger" case. The findings on gender are consistent with evidence in Simpson (2003), who argues precisely that the reason why existing literature on the prisoner's dilemma has, by and large, not found gender differences is because of the presence of both temptation and risk in these games. See also Kuwabara (2006).

Γ_1 has higher RISK and lower TEMPT compared to Γ_2, then our results suggest that Γ_2 may be more conducive to high cooperation rates than Γ_1 in one-shot (stranger) settings, while the reverse would be true in the repeated game. 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 one-shot games. Those theories assume that agents are intrinsically motivated to cooperate as long as others do so as well (Fischbacher et al., 2001). On the other hand, theories focused on creating incentives for cooperation (or avoiding temptation), should have better chances in the repeated game. 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 one-shot (random rematching) games, where risk explains most of the variation in cooperation rates.[3]

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 disentangle the role of 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 3 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) 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. Variation in these parameters is low, though, with RISK ranging from 0.66 to 0.83 and TEMPT ranging from 0.18 to 0.36. They find that cooperation rates correlate with all three indices,

[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). Other examples of interventions designed to reduce temptation could include taxation schemes, with very progressive schemes better suited to curb temptation compared to linear schemes.

but do not find systematic differences across their different designs. In a class of prisoner's dilemma games Capraro et al. (2014) find that the "benefit to cost ratio" increases cooperation rates. 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 trustor 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. Balliet et al. (2011) conduct a meta-study focused on sex-differences in cooperation. Consistent with our evidence, they find that there are no differences on average between cooperation rates of men and women. They do find, however, that men cooperate more in repeated interactions. We do not find evidence of the latter. They also find that male-male interactions are more cooperative than female-female interactions and that women cooperate more in mixed interactions. We cannot say anything about this question, since in all the studies we consider participants do *not* know the gender or sex of their opponent. By contrast, Balliet et al. (2011) do not analyze how gender differences are mediated by risk and temptation. In a recent meta study Embrey et al. (2016) ask whether behaviour in the finitely repeated prisoner's dilemma ("partner" setting) 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. We include this index in our analysis and find that it can also explain a substantial part of the variation in average cooperation rates in the repeated game. Rezaei Khavas (2016) studies the effect of culture on cooperation rates in the prisoner's dilemma. A meta-study on indefinitely repeated prisoner's dilemma games is provided by Dal Bó and Fréchette (2016).[4]

[4]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

The paper is organized as follows. In Section 1 we explain how we collected our data both from existing literature and from additional lab experiments we conducted. We also define and discuss our measures `RISK` and `TEMPT`. In Section 2 we present our main results. In Section 3 we discuss how they compare to the indices used by Murnighan and Roth (1983) and Embrey et al. (2016). Section 4 studies dynamics and Section 5 exploits questionnaire data using subsets of of our full data set for which this additional information is available. Section 6 concludes. Additional tables and figures can be found in an Online Appendix.

1 The meta study

1.1 Procedures

We study laboratory experiments on prisoner's dilemma games. Our data set comprises 96 studies involving more than 3500 participants in experiments conducted in Germany, Japan, the Netherlands, Spain, the UK and the USA. A "study" $\mathcal{S} = (\mathbf{\Pi}, T, M, X)$ is defined by a combination of payoff matrix $\mathbf{\Pi} \in \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} | a, b, c, d \in \mathbb{R} \right\}$, the length of the game (number of stages) $T \in \mathbb{N}$, the number of matches $M \in \mathbb{N}$ and an indicator X for the paper from which the data are taken. Hence two independent observations of the same paper (with the same $(\mathbf{\Pi}, T, M)$) would be part of the same "study", while two observations with the same $(\mathbf{\Pi}, T, M)$ that are part of different papers would constitute two different "studies".

Our data set contains 57 such studies stemming from 23 research papers in 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.) with a finite number of matches, (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) there is no deception of experimental participants, (viii) the

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. Habetinova and Suetens (2015) focus on the role of feedback in public goods and oligopoly games.

cooperation rate is either reported, data were made available to check it or it could be reasonably inferred from a graph and (ix) the study has been published before or in 2013.[5] We found studies via a keyword search on google scholar and via an e-mail to the ESA discussion group on January 6, 2014. Two papers (7 studies) were included after being suggested by a Reviewer. All the studies are listed in Table A3 in Online Appendix A. Note that often the main treatments of a paper don't satisfy our criteria listed above, but there is a control treatment that does. Such treatments are, of course, only included if they stem from between subject designs or if there is little risk that the data could be "contaminated" by subsequent treatments.

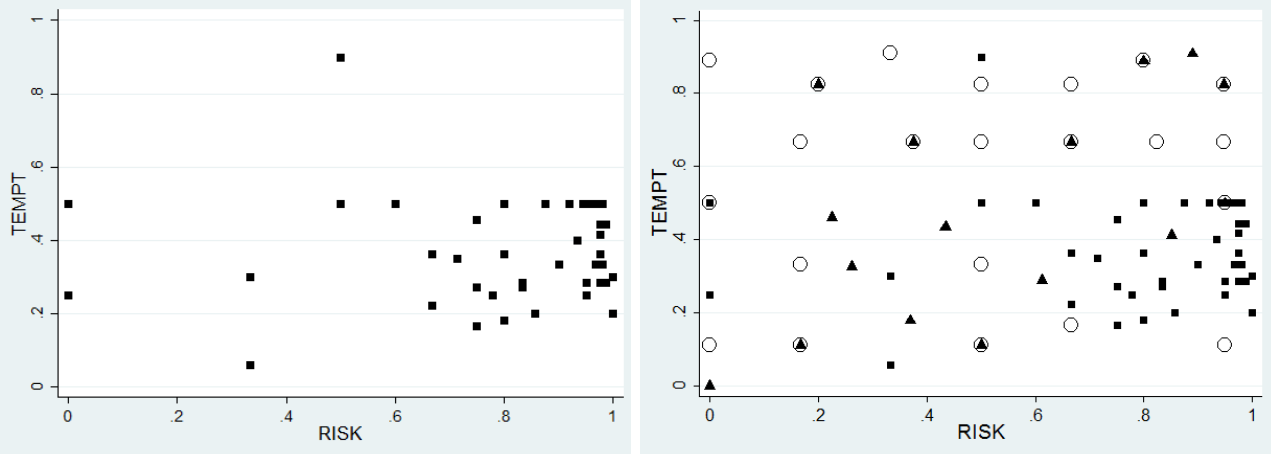
In addition, we conducted our own lab experiments to cover a larger parameter space (16 studies). In our own experiments we conducted 10 period prisoner's dilemma games in either a repeated game setting ($T = 10; M = 1$) or a random rematching setting ($T = 1; M = 10$). We also conducted some one-shot studies ($T = 1; M = 1$) on Amazon Mechanical Turk, AMT (23 studies). 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)) and after adding both our own lab studies and studies on AMT (Figure 1(b)). Figure 1(a) shows that the variation in the existing literature is not very high. With few exceptions all studies have values of `TEMPT` lower than 0.5 and values of `RISK` higher than 0.3. We added our own 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.[6] 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 (repeated and random rematching games) or AMT (one-shot). The parameter values used in our own studies are summarized in Tables A1 and A2 in Online Appendix A and data sets permitting replication are available online. Balancing tests, where we check whether the distributions of `RISK` and `TEMPT` are

[5] 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 B11 in Online Appendix B.2).

[6] We selected payoff parameters based on a slightly different definition of `RISK` and `TEMPT` used in an earlier version. With few exceptions all studies fall in the same square (though not the exact same coordinates) as they did previously.

balanced across the repeated and random rematching games can be found in Table B1 in Online Appendix B.1.



(a) Existing Literature (b) All Studies

Figure 1: Variation in `RISK` and `TEMPT` parameters in the existing literature and including our own studies. Existing literature is marked by squares, own lab studies by triangles and own AMT studies by circles. If several studies have the same parameters only one is shown.

1.2 Risk and Temptation Indices

We now proceed to describing our three key variables `RISK`, `TEMPT` and `EFF`. Consider a payoff matrix $\mathbf{\Pi}$ as the one illustrated in Figure 2 (left panel). Two fundamental inequalities on the entries of $\mathbf{\Pi}$ need to be satisfied in order for it to describe a Prisoner's dilemma: (i) $c > a$, which we will refer to as "temptation" and (ii) $b < d$, which we refer to as "risk". Usually, we also require $a > d$, i.e. that mutual cooperation is socially desirable and hence that there is a tension between efficiency and individual rationality. If any of these were not satisfied we would not refer to the game defined by $\mathbf{\Pi}$ as prisoner's dilemma game. In particular, assume that $a > d$. Then, if $c > a$ (temptation), but $b > d$ (no risk), $\mathbf{\Pi}$ describes an Anti-Coordination or Hawk Dove Game. If, by contrast, $c < a$ (no temptation) and $b < d$ (risk), then the game is a Coordination game. If $c < a$ and $b > d$ (neither temptation nor risk), then cooperation is a dominant strategy in the game. The relative amounts of risk and temptation also determine whether the game is supermodular, i.e. of strategic complements or substitutes. In particular, in order for the game to be supermodular $c - a \leq d - b$ is required, i.e. temptation cannot exceed risk. Our interest in this study is how these two defining payoff differences shape behaviour.

Now in a meta-study we will compare many different studies conducted in different countries at different times with different payoff scales and different exchange rates from experimental to local currency. In other words, a participant facing payoff matrix $\mathbf{\Pi}$ will be paid according to $\alpha\mathbf{\Pi}$ where $\alpha \in \mathbb{R}^+$ reflects the exchange rate from experimental currency (tokens) into purchasing power in the location and at the time when the experiment was conducted. The evidence on whether such linear transformations of payoffs affect behaviour in experimental games is somewhat mixed and depends on what type of linear transformation (changing stakes and/or frames) one has in mind.[7] Because we do not want to impose invariance to linear transformations, we focus on a percentage based definition rather than absolute differences in defining our indices of temptation and risk.[8] We define the percentage based measures `TEMPT` and `RISK` as follows.

	C	D
C	a	b
D	c	d

PD game.

	C	D
C	1	-$\texttt{RISK}^{Norm}$
D	$1+\texttt{TEMPT}^{Norm}$	0

Normalized game.

Figure 2: Left: Prisoner's dilemma (PD) game with payoff parameters $c > a > d > b$. The additional condition $c + b < 2a$ ensures that joint cooperation is more efficient than alternating between (C,D) and (D,C). Table B10 in Online Appendix B.2 splits the sample according to whether or not this condition is satisfied. Right: Normalized game with $a = 1$ and $d = 0$ (hence `EFF`$= 1$) represented using the normalization based indices $\texttt{RISK}^{Norm}$ and $\texttt{TEMPT}^{Norm}$.

Temptation (`TEMPT`) We measure the extent of temptation the game presents by the percentage gain when unilaterally defecting against a cooperator, more specifically by $\texttt{TEMPT}= \frac{c-a}{c}$, which ranges between 0 and 1 as long as $c > a > 0$, which is the case in all our studies. `TEMPT`$=0$ means there is no temptation ($a = c$) and `TEMPT`$=1$ means maximal temptation, where a is negligible compared to

[7] When it comes to linear transformations which do affect stakes the direction of effects is typically unchanged, while effect sizes can change (Camerer and Hogarth, 1999). Framing effects have been shown to affect behaviour despite not affecting stakes (Andreoni, 1995; Sonnemans et al., 1998; Iturbe et al., 2011). Linear transformations which do neither affect stakes nor involve substantial changes to the frame (other than payoffs being multiplied by a constant), such as exchange rates (between experimental currency and "real" currency) are thought not to affect behaviour, but to our knowledge this has never been directly tested.

[8] Note that cases where one of the entries in $\mathbf{\Pi}$ is zero (in our data those are always $b = 0$) are somewhat problematic, since those cases may imply that percentage based measures do not change even under some non-linear transformations such as e.g. $(\alpha_1, \alpha_2)\mathbf{\Pi}$, $\alpha_1 \neq \alpha_2 \in \mathbb{R}$, where we would expect behaviour to change. This happens because 0 is invariant to scaling. To deal with such cases we slightly perturb all zero entries in $\mathbf{\Pi}$ by adding a noise term drawn uniformly from the open unit interval.

c. In our data the variable TEMPT ranges between 0.03 and 0.83.

Risk (RISK) We measure the extent of risk the game presents by the percentage loss when unilaterally cooperating against a defector, i.e. by RISK$= \frac{d-b}{d}$, which ranges between 0 and 1 as long as $d > b > 0$ with the natural interpretation that RISK$= 0$ if $b = d$ and RISK$= 1$ if b is "much smaller" than d. In our data the variable RISK ranges between 0 and 1.

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 measure the extent of possible efficiency gains in the game by EFF$= \frac{a-d}{a}$, which ranges between 0 and 1 if $a > d > 0$. In our data the variable EFF ranges between 0 and 0.83.[9] We are mostly focused on RISK and TEMPT , because EFF involves payoff comparisons based on bilateral changes (and hence does not affect standard incentives as long as $c > a$ and $d > b$ are satisfied). We use EFF as a control, however, in almost all regressions.[10]

Normalization An alternative to percentage based measures is to normalize EFF$= 1$ by subtracting d from all payoffs and dividing the resulting entries by $(a-d)$. This yields the game shown in Figure 2 on the right hand side, where RISK$^{Norm} = \frac{d-b}{a-d}$ and TEMPT$^{Norm} = \frac{c-a}{a-d}$.[11] This normalization has been frequently used in the literature (Stahl, 1991; Embrey et al., 2016). The downside of this approach is that, as efficiency $(a-d)$ is normalized to one, it is not possible to control explicitly for efficiency differences across games and hence to net out the effect of risk and temptation. The fact that both RISK^{Norm} and TEMPT^{Norm} are obtained after dividing by $(a-d)$ under this approach also induces collinearity between the risk and temptation measures, which is undesirable given our aim to disentangle their role. This is why we prefer to use the percentage based measures defined above for our main analysis. Still, we will show in Online Appendix B.2 that our main results (reported in Section 3) are qualitatively robust when considering this approach as well as when considering an approach based on absolute payoff differences.

[9] There are 6 studies in our sample, however, for which $0 \geq d > b$. For those studies we define RISK$= \frac{|d-b|}{|b|}$ to ensure that RISK ranges between 0 and 1. For 6 studies where $a > 0 > d$ we define EFF$= 1$ to avoid non-monotonicity, as EFF$= 1$ if $d = 0$. Dropping these studies does not affect our results qualitatively (Table B11 in Online Appendix B.2).

[10] Note that a game where RISK=TEMPT=EFF$= 0$ is a trivial game where all payoffs are equal.

[11] The pairwise correlation coefficient between RISK and RISK^{Norm} is 0.4257^{***} and between TEMPT and TEMPT^{Norm} it is 0.4876^{***}.

Correlation and Balancing Tests Across all our studies the correlation between the three key variables RISK , TEMPT and EFF tends to be not statistically significant. Between RISK and TEMPT the Spearman correlation coefficient is -0.0217 (0.0749 if the AMT studies are excluded), neither of which is statistically significant. Between RISK and EFF the coefficient is -0.0162 (-0.0143), again both insignificant. Between TEMPT and EFF the correlation coefficients are -0.0631 (-0.0833), both statistically insignificant. Balancing tests, where we regress RISK and TEMPT on a partner dummy are reported in Table B1 in Online Appendix B.1. They show that the whole sample as well as the sub-samples obtained by splitting according to the median RISK and TEMPT are balanced.[12]

1.3 Theoretical Background

In this section we try to develop some intuition for how RISK and TEMPT could affect behaviour in one-shot and repeated games. The purpose of the paper is not to test for specific theories. Still, we find it useful to give some idea of how and when we would expect RISK and TEMPT to affect incentives.

Stranger/one-shot settings Incentives in the one-shot game are governed by the expected payoff difference between cooperating and defecting which we denote $\Delta\pi_i^e$. We further denote by p the probability with which a player believes that her opponent cooperates.

The relative importance of RISK and TEMPT for short run incentives $\Delta\pi_i^e$ depends on p in a straightforward manner. If people are pessimistic, i.e. $p = 0$, then short run incentives will be governed by RISK.[13] If, by contrast, players in such situations are optimistic, i.e. assume that $p = 1$, then short run incentives will be governed by TEMPT. If, on the other hand, players follow Laplace's principle of insufficient reason and attach probability $p = \frac{1}{2}$ to their opponent being cooperative, then both RISK and TEMPT should matter equally.

[12] One advantage of adding our own studies to the existing literature is that it does give us some control over collinearity of RISK and TEMPT and allows us to ensure that the distribution of our indices is balanced across the different scenarios (repeated, one-shot, stranger matching) that we are interested in

[13] Note that the stranger and one-shot settings aim to capture interactions where people have no prior experience nor knowledge about their opponent's type. Gilboa and Schmeidler (1989) maintain that people would resort to pessimistic priors in situations characterized by Knightian uncertainty. To the extent that this is a good model to think about beliefs in the one-shot/stranger setting, this theory would give some role to RISK. Risk is also crucial in determining short run incentives in Blonski et al. (2011), as reflected by their Axiom 3.

Partner setting In the repeated game the role of RISK and TEMPT is not as straightforward. In these games people have a larger strategy set available. In the literature usually some assumptions on strategy sets or cooperative types are made to derive predictions that can be contrasted with experimental data. Embrey et al. (2016), for example, assume that players in the finitely repeated game decide only between GT and the strategy "allD", where GT denotes grim-trigger, i.e. the strategy that starts out by cooperating and switches to defecting until the end of the game as soon as the opponent defects once (see also Dal Bó and Fréchette (2011)). We denote this strategy set by $S_1 = \{GT, allD\}$. Since we are interested in finitely repeated games, we can also consider a larger strategy set $S_2 = \{GT, 1GT, 2GT, 3GT, ..., allD\}$, where 1GT denotes the best response to grim trigger in the finitely repeated game, which is to cooperate in all rounds except for the last one. 2GT denotes the best response to 1GT (cooperate in all rounds but the last two) etc. and "allD" denotes the strategy that chooses always defection.[14]

Based on S_1, Embrey et al. (2016) derive an index that captures the probability that a player must assign to the other player playing GT so that they are indifferent between GT and "allD" themselves (see also Blonski et al. (2011) and Dal Bó and Fréchette (2011)). Following Embrey et al. (2016) we can normalize EFF by setting $a = 1$ and $d = 0$ (hence EFF $= 1$). The normalized game is illustrated on the right of Figure 2. In the normalized game this index can be expressed as follows: $\frac{\texttt{RISK}^{Norm}}{(T-1)+\texttt{RISK}^{Norm}-\texttt{TEMPT}^{Norm}}$. Hence this index includes aspects of both RISK and TEMPT (as well as the normalized EFF) . It takes the value zero if $\texttt{RISK}^{Norm} = 0$. Since the "allD" and GT strategies can potentially differ in all periods, the index depends also on the time horizon T. If $T = 1$, the value of the index exceeds 1 reflecting the fact that cooperation cannot be sustained in the one-shot game under these assumptions. We will include this index in our regressions in Section 3 to see how it compares to RISK, TEMPT and other measures from the literature.

Based on S_2, the probability that a standard type must assign to the other player playing GT so that they choose 1GT rather than any less cooperative strategy is $\frac{c-a}{c-d}$. In the normalized game this probability can be expressed as $\frac{\texttt{TEMPT}^{Norm}}{1+\texttt{TEMPT}^{Norm}}$. If $\texttt{TEMPT}^{Norm} = 0$, then this threshold is zero, i.e. a cooperative equilibrium always exists. ((C,C) is a Nash equilibrium in the one-shot game in this

[14]One advantage of using the smaller strategy set S_1 is that the resulting index can also be applied in indefinitely repeated games (Dal Bó and Fréchette, 2011). Note that the intermediate strategies $1GT, ..., (T-1)GT$ in strategy set S_2 are not well defined in indefinitely repeated games.

case.) This threshold does not depend on the time horizon T, since 1GT and the optimal deviation strategy 2GT differ by just one period. Under these assumptions, RISK should not play much of a role in explaining participant's incentives to cooperate in the finitely repeated game.

2 Results: Average Cooperation Rates

Our first set of results presented in this section uses information from all the 96 studies contained in our data set. Those are all the studies listed in Tables A1-A3 in Online Appendix A, in all of which an average cooperation rate is available that we will use as endogenous variable in this Section. For the one-shot and Stranger settings ($T = 1$) we use the average cooperation rate across all matches. For the repeated game setting ($T > 1$) we use the average cooperation rate in the last match in our baseline specification.[15] The average cooperation rate thus computed ranges between 0.04 and 0.84 in the one-shot and Stranger treatments with a mean of 0.37. In the repeated game it ranges between 0.17 and 0.58 with a mean of 0.39. The main reason to consider these matches is that this is the data available in all 96 studies. Hence this choice maximizes our sample size. To understand to which extent results depend on this choice, we will also consider specifications, though, that consider only the first stage game both in the $T = 1$ and $T > 1$ cases as well as specifications where we consider all matches and stages both in the $T = 1$ and $T > 1$ cases. In Section 5 we also use other measures than just these average cooperation rate relying on the smaller sample of studies for which the data available allow us to do so.

We start by asking how much of the variation in average cooperation rates across studies can be explained by RISK or TEMPT. Table 1 shows the results of simple OLS regressions explaining the average cooperation rates with our variables of interest: RISK, TEMPT and EFF.

One Shot Games (Stranger) Columns (1)-(4) focus on the Stranger setting that is meant to capture interactions where people have no prior experience with their opponent. Column (1) focuses on one shot games ($T = M = 1$), where experimental participants played the prisoner's dilemma for one period only. Column (2) focuses on games with more than one period in the Stranger setting

[15] 16 out of the 23 repeated game studies in our sample do only have one match, in which case "all matches" is the same as "the last match".

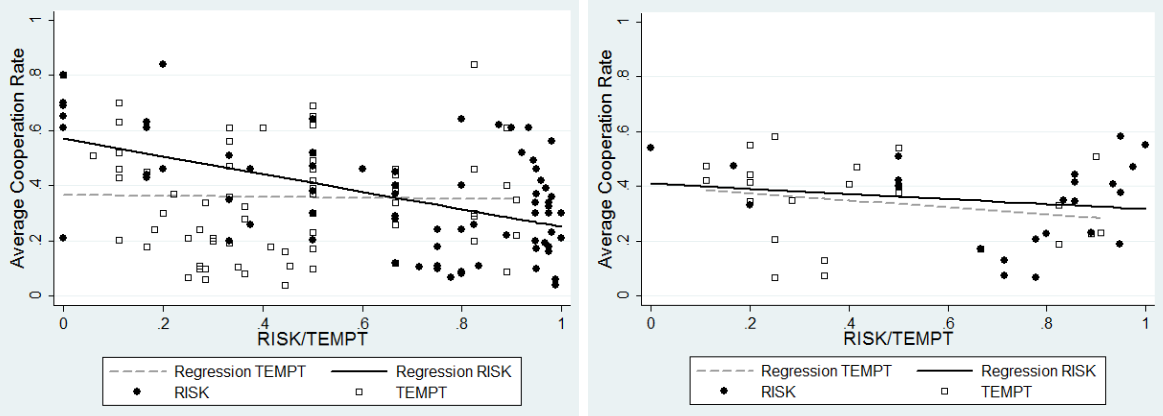
($T = 1; M > 1$) and column (3) pools these cases, i.e. contains all 73 studies conducted within the Stranger paradigm ($T = 1$). Column (4) focuses on the average cooperation rate in the first match, i.e. the first one-shot game played. All regressions consistently show a substantial negative effect of `RISK` on average cooperation rates. The coefficient on `RISK` ranges between -0.290*** and -0.197*** and is remarkably consistent across the 1-match and multi-match studies. The variable `TEMPT` , by contrast, does not have any statistically significant impact on cooperation rates in these studies and the coefficient size is smaller ranging between -0.110 and 0.002. Efficiency seems to have a substantial and positive impact on cooperation rates, which is not statistically significant in the multi-match games, though (column (2)). Figure 3(a) illustrates data points as well as a simple OLS regression of average cooperation rates on either `RISK` or `TEMPT`.

	Stranger ($T = 1$)				*Partner ($T > 1$)*	
	1 match	> 1 matches	Pooled	Pooled	Repeated	Repeated
				1st stage & match		1st stage & match
	(1)	(2)	(3)	(4)	(5)	(6)
RISK	-0.269***	-0.197**	-0.255***	-0.290***	0.008	-0.263
	(0.066)	(0.096)	(0.060)	(0.057)	(0.111)	(0.177)
TEMPT	-0.055	-0.110	0.002	-0.046	-0.208	-0.263*
	(0.096)	(0.118)	(0.079)	(0.073)	(0.107)	(0.134)
EFF	0.308***	0.192	0.291***	0.303***	0.320**	-0.012
	(0.100)	(0.133)	(0.089)	(0.086)	(0.114)	(0.348)
Constant	0.455***	0.332***	0.370***	0.451***	0.189	0.692**
	(0.098)	(0.118)	(0.083)	(0.078)	(0.141)	(0.289)
Observations	45	28	73	69	23	14
Sample	Lab/AMT	Lab	Lab/AMT	Lab/AMT	Lab	Lab
R-squared	0.484	0.333	0.377	0.458	0.353	0.478

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Average cooperation rate regressed on variables of interest.

Repeated Game Columns (5)-(6) in Table 1 show the results for the repeated game (partner) setting. Column (5) focuses on the average cooperation rate across all stages in the last match and Column (6) on the first stage of the first match. This regression has less power than the "Stranger" regressions, since the number of repeated game studies in our data set is smaller and since first-stage cooperation rates are not available in all studies. The variable `RISK` is not statistically significant and the coefficient size smaller than in the corresponding Stranger settings. `TEMPT`, by contrast, seems to affect cooperation rates in this setting. The variable `TEMPT` has a substantial coefficient size (-0.208 or -0.263^*, respectively) and is statistically significant at the 10% level if only the first game is considered. Figure 3(b) shows data points as well as a simple OLS regression of average



(a) Stranger (b) Partner

Figure 3: Average cooperation rate for different levels of RISK and TEMPT. Lines show fitted values from OLS regressions of average cooperation rate on either RISK (coefficients: -0.319*** in Stranger and -0.093 in Partner) or TEMPT (dashed lines; coefficients: -0.017 in Stranger and -0.128 in Partner), respectively

cooperation rates on either RISK or TEMPT (without controlling for EFF).

Robustness tests In Online Appendix B.2 we report a number of robustness checks. First, we show that, by and large, qualitatively the same results are obtained when an absolute rather than our percentage-based measure is used (Tables B2-B4.) Second, we show that results are robust when we normalize EFF= 1 and then use the normalization based measure as illustrated in Panel (b) of Figure 2 (Tables B5-B7). We then report another set of robustness checks. We show that the results shown in Table 1 are robust when we use weighted regressions, where we weigh our studies by the number of independent observations (Table B8). Results are also robust to dropping some studies with "special" details, such as studies were beliefs were elicited or studies that were paper-based (Table B9) and 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 B10). Table B12 includes interactions of our main variables of interest with a dummy variable "new" which indicates whether a study was conducted by us. This table shows that results are very similar across the set of existing studies and our new studies. Finally, Table B13 reproduces Table 2 below but focuses on average cooperation rates across all matches and stages for both settings. Results are qualitatively the same.

Partner vs Stranger In Table 2 we ask whether average cooperation rates are higher in "partner" ($T > 1$) or "stranger" ($T = 1$) settings. Existing research has delivered mixed results on this

	(1)	(2)	(3)	(4)	(5)
Partner	-0.143	0.149**	-0.060	-0.007	-0.014
	(0.098)	(0.069)	(0.096)	(0.080)	(0.045)
Constant	0.372***	0.296***	0.451***	0.307***	0.359***
	(0.048)	(0.040)	(0.045)	(0.035)	(0.022)
Observations	25	24	23	26	96
RISK	Small	High	Small	High	All
TEMPT	Small	Small	High	High	All
R-squared	0.084	0.174	0.018	0.000	0.001
	1st stage game only				
	(1b)	(2b)	(3b)	(4b)	(5b)
Partner	-0.016	0.101*	-0.059	-0.077	-0.012
	(0.105)	(0.061)	(0.130)	(0.084)	(0.053)
Constant	0.436***	0.318***	0.476***	0.350***	0.399***
	(0.052)	(0.043)	(0.041)	(0.034)	(0.022)
Observations	24	18	20	24	84
RISK	Small	High	Small	High	All
TEMPT	Small	Small	High	High	All
R-squared	0.001	0.071	0.012	0.037	0.001

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Partner vs Stranger. Average cooperation rates regressed on "Partner" dummy. Sample partitioned into four sub-samples according to median RISK and TEMPT. Column (5) shows entire sample.

question. 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). Dal Bo (2005) find more cooperation in the "partner" condition with a long horizon in the repeated game and no difference with a short horizon of the repeated game. Note also that, while Dal Bo (2005) compare matches of different lengths ($T = 1$ in the one shot case and $T \in \{2, 4\}$ in the "partner" case), the earlier literature has typically compared games with the same number of periods, where a period is a new match in the "stranger" case and a new stage within the same match in the "partner" case.

In Table 2 we regress average cooperation rates on a dummy indicating whether the study is a "partner" ($T > 1$) study. In the top panel of Table 2, the average cooperation rate is based on all matches in the "stranger setting" and on the last match in the "partner setting". In the bottom panel the average cooperation rate is the cooperation rate in the first (stage) game played in the first match. We partition our sample into four sub-samples according to the median RISK and median TEMPT. Columns (1) and (1b) show studies with below median RISK and below median TEMPT , columns (2) and (2b) with above median RISK and below median TEMPT , columns (3) and (3b)

with below median RISK and above median TEMPT and in columns (4) and (4b) both RISK and TEMPT are above the median. Columns (5) and (5b) shows the entire sample. Balancing tests, which show that the distributions of RISK and TEMPT are balanced across these sub-samples can be found in Table B1 in Online Appendix B.1.

Table 2 shows that there are no statistically significant differences between partner and stranger settings in the entire sample and in none of the sub-samples except for the case where RISK is high and TEMPT is small (columns (2) and (2b)). In these cases the partner setting is able to generate cooperation rates which are ≈ 15 percentage points higher than the stranger setting. What is also noticeable is that, while the partner dummy alone can explain around 18 percent of the variation if RISK is high and TEMPT low (7 percent for the first game), in all other cases the R^2 is lower and in Column (5) where all studies are aggregated it is very close to zero. This analysis shows that the effect of the matching technology (partner vs stranger) is mediated by the RISK and TEMPT indices. While for some values of RISK and TEMPT the partner matching can lead to substantially higher cooperation rates, this is not generally the case. It is interesting to compare these results to the public good game. In a meta-study for the public good game Zelmer (2003) has found that there are on average higher contributions in the partner compared to the stranger setting. There are a number of possible explanations for this difference. In the linear public good game studied by Zelmer (2003) it is not possible to create the same variation in RISK and TEMPT as both variables co-move with the same parameter, the marginal per capita rate of return. Hence it is possible that the RISK-TEMPT ratio in the public goods literature is simply favorable to higher contributions in the partner setting. It could also be that the larger group sizes in the public good game are a factor in this comparison.

Number of Stages/Matches In our analysis so far we have bundled together studies with a different number of matches and different lengths of the repeated game (number of stages). There are multiple ways, however, in which the number of stages and matches could affect average cooperation rates as well as the impact of RISK and TEMPT.

In the one-shot game the number of matches played in the past (m) can affect a participants' experience with the setting and hence possibly their behaviour. For the one-shot/stranger setting we

	One-Shot/Stranger			*Repeated Game*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RISK	-0.255***	-0.266***	-0.260***	0.008	0.007	-0.008	-0.015	0.164	0.003
	(0.060)	(0.053)	(0.071)	(0.111)	(0.103)	(0.123)	(0.127)	(0.655)	(0.116)
TEMPT	0.002	-0.092	0.116	-0.104	-0.299***	-0.087	-0.299**	0.057	-0.037
	(0.079)	(0.070)	(0.086)	(0.107)	(0.094)	(0.117)	(0.098)	(0.234)	(0.326)
EFF	0.291***	0.346***	0.370***	0.320**	-0.004	0.313**	-0.047	0.442	0.462***
	(0.089)	(0.081)	(0.105)	(0.114)	(0.164)	(0.123)	(0.211)	(0.447)	(0.151)
Matches/Stages			0.012					0.045	0.081
			(0.021)					(0.038)	(0.455)
RISK× Matches/Stages			-0.004					-0.018	-0.213
			(0.013)					(0.039)	(0.125)
TEMPT× Matches/Stages			-0.038***					-0.016	-0.022
			(0.010)					(0.027)	(0.512)
EFF× Matches/Stages			-0.015					-0.022	-0.081
			(0.020)					(0.000)	(0.002)
Constant	0.370***	0.451***	0.322***	0.189	0.496**	0.191	0.521**	-0.159	0.066
	(0.083)	(0.077)	(0.102)	(0.141)	(0.166)	(0.151)	(0.187)	(0.748)	(0.213)
Observations	73	73	73	23	23	23	23	23	23
NStages Fixed Effects	NO	NO	NO	NO	YES	NO	YES	NO	NO
NMatches Fixed Effects	NO	YES	NO	NO	NO	YES	YES	NO	NO
Match/Stage Interactions	NO	NO	Match	NO	NO	NO	NO	Stage	Match
R-squared	0.377	0.607	0.513	0.353	0.822	0.363	0.824	0.516	0.467

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Average cooperation rate regressed on variables of interest with number of stages/matches fixed effects and interactions.

would hence like to control for the average number of past matches played, i.e. $\frac{\sum_{m=1}^{M}(m-1)}{M} = \frac{M-1}{2}$. Remember that, since in this Section our unit of observation is a study $\mathcal{S}$, we cannot exploit variation in m *within* a study. We will do so, however, in Section 4.

In the repeated game setting the number of past matches played can also affect participants' experience. In addition, the number of stages T can play an important role in this setting. First, a longer time horizon (more stages) could lead to more cooperation, because agents have more to gain by establishing a reputation as cooperative types (Kreps et al., 1982; Mengel, 2014). A longer time horizon also decreases the Dal Bó and Fréchette (2011) and Embrey et al. (2016) threshold for cooperation discussed above. In fact, Embrey et al. (2016) find that, contrary to the backward induction prediction, T affects cooperation. In line with the unraveling logic of backward induction, Embrey et al. (2016) also find that participants cooperate on average more the higher the number of stages T and the first defection round moves earlier the more matches they have experienced in the past. For the repeated game we will hence want to additionally control for the number of stages.

The number of matches in our sample ranges between 1 and 100 in the Stranger condition and between 1 and 10 in the repeated game. The number of stages is, by definition, always 1 in the Stranger setting and ranges between 2 and 25 in the Partner setting. Table 3 controls for the number

of stages/matches in a variety of ways.

Columns (1)-(3) focus on the Stranger setting, where the number of stages is always one. Column (1) corresponds to our basic estimates from Table 1, Column (2) includes fixed effects for the number of matches (M) and Column (3) controls linearly for the number of matches and includes interaction terms.[16] The coefficient on `RISK` is very stable across these three specifications (between -0.255^{***} and -0.266^{***}). There is no statistically significant effect of the number of matches played on average cooperation rates. Note, though, that the R^2 is substantially higher if the number of matches is controlled for.

Columns (4)-(9) focus on the Partner setting. Column (4) corresponds to our estimation from Table 1 which does not control for the number of stages (T) or past matches ($M-1$). Columns (5)-(7) include either T- or M- fixed effects or both. The coefficient on `TEMPT` is substantially higher (and statistically significant) once T-fixed effects (for the number of stages) are included. Also the R^2 is substantially higher in this case. Closer inspection of the fixed effects reveals that cooperation rates are lower in shorter games (with fewer stages). This is consistent with our intuition: longer games make it easier to establish cooperation in early rounds, because participants can gain more by establishing a reputation as cooperative types. Including M-fixed effects, by contrast, does not change the coefficient on `TEMPT` much, nor does it increase the R^2 substantially. Note that this analysis merely asks how the number of stages or matches affects *average* cooperation rates. The dynamics of cooperation within matches will be analyzed in Section 4.

Result 1

- *Variation in average cooperation rates in one-shot games is best explained by `RISK`, while `TEMPT` best explains variation in cooperation rates in repeated games.*
- *There is more cooperation in the repeated (partner) compared to the one-shot (stranger) game if and only if `RISK` is high and `TEMPT` is low.*

3 Results: Other Indices

In this section we compare our measures `RISK` and `TEMPT` to a number of other indices based on payoff parameters that have been used in the literature, notably by Murnighan and Roth (1983). Most of

[16] Note that M is perfectly collinear to $\frac{M-1}{2}$. Controlling for the latter (or more precisely the smallest integer bigger than $\frac{M-1}{2}$) yields identical results in terms of the impact of `RISK`, `TEMPT` and `EFF`.

this literature has been motivated by creating an index of all payoff parameters that "summarizes" incentives in the Prisoner's dilemma. Hence typically these indices contain aspects of all three: RISK,TEMPT and EFF, but with different emphasis.

Murnighan and Roth (1983) discuss 10 different indices used previously in the literature.[17] 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), 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).[18] 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 and interpretation are adopted from Murnighan and Roth (1983) and will be preceded by MR- in the following. We also compare to the index used by Dal Bó and Fréchette (2011) and Embrey et al. (2016) discussed in Section 1.3, which is referred to as BAD for "Basin of attraction of defection".[19] While our measures RISK, TEMPT and EFF all range in $[0, 1]$ this is not generally the case for the indices discussed above. To be able to compare coefficient sizes we hence standardize all variables to mean zero and standard deviation one in this Section.

We first ask how our results from Section 2 change if these other indices are considered. We start with the one-shot/Stranger condition, where we have found above that the variable RISK can explain variation in cooperation rates well. Table B14 in Online Appendix B.3 shows the results of regressions, where in column (1) we regress average cooperation rates on our variables RISK , TEMPT and EFF (as in column (3) of Table 1) just that now we use standardized values of RISK , TEMPT and EFF. The results show that a standard deviation increase in RISK leads to an $\approx 8\%$ decrease in cooperation rates. The TEMPT coefficient is near zero (and statistically insignificant). Across columns (2)-(12) we add one each of the eleven indices described above.[20] RISK is statistically

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

[18] Index E1 corresponds to $\texttt{TEMPT}^{Norm}$ defined on the normalized game in Figure 2. It is also the inverse of the "benefit to cost ratio" employed by Capraro et al. (2014).

[19] As all indices used in the regressions, the BAD index used here is defined on the not normalized game, i.e. the left game in Figure 1. The index defined on the not normalized game is $BAD = \frac{-b+d}{(a-d)*T-b-c+2d}$.

[20] Adding all of them simultaneously would lead to considerable over-fitting in our sample of 73 observations as

significant in all columns (1)-(12) and the coefficient size is substantial ranging between -0.099^{***} to -0.046^{***} (compared to -0.081^{***} in column (1)). None of the additional indices included are statistically significant with the exception of MR-R3 (column (4)) which is interpreted by Murnighan Roth as a measure of risk and in fact shares the same numerator as our measure `RISK`. It is also interesting to note that in none of the regressions (1)-(12), the variable `TEMPT` plays a significant role.

Table B15 in Online Appendix B.3 shows the results for the repeated game. What emerges clearly is that, unlike in the Stranger condition, `RISK` here does not affect cooperation rates significantly. The `RISK` coefficient is 0.002 in the basic regression in column (1) and never statistically significant across columns (2)-(12) ranging in size between -0.014 to 0.037. Results are a bit less clear-cut when it comes to what does affect cooperation rates. `TEMPT` has a negative impact on cooperation rates in columns (4) and (12). `EFF` has a positive impact in most specifications and in particular BAD seems important, where a one SD increase leads to an $\approx 13\%$ increase in cooperation rates. Including the BAD index also substantially increases the R^2 compared to all other regressions, which is not too surprising as it is the only index that accounts for the number of stages T. None of the Murnigan-Roth measures has a statistically significant effect.

Comparing indices via R^2 We also conduct two "horseraces" between the 14 indices. In the first we simply run the following OLS regression $y_i = \alpha + \beta \mathbf{x}_i + \epsilon_i$, where y_i is the average cooperation rate in study i (defined as in Section 3) and $x_i \in \{\texttt{RISK}, \texttt{TEMPT}, \texttt{EFF}, MRR1, ..., MRK3, MRK4, BAD\}$. We then ask which index x produces the highest R^2, i.e. explains most of the variation in average cooperation rates. To obtain confidence intervals we bootstrap the R^2 using 3000 replications. Table 4 shows the results of this exercise. In the Stranger ($T = 1$) setting (leftmost column) `RISK` has the highest R^2 among all indices explaining around 28% of the observed variation in cooperation rates. MRR3 and `EFF` (whose R^2 are statistically not different from that of `RISK`) and to a lesser extent BAD and MRR1 can also explain a good part of the variation.[21] The R^2 of all other indices is not

well as possible collinearity. Tables B16 and B17 in Online Appendix B.3 show separate regressions for one-shot ($T = M = 1$) and multi-match ($T = 1; M > 1$) studies with qualitatively the same results.

[21]Note that BAD was designed with the repeated game in mind and assumes values above one if $T = 1$. It is hence harder to interpret in the one-shot game, but can nevertheless explain around 15% of the observed variation in cooperation rates.

	Comparison of Indices I		
	Stranger	Partner	Partner Nstages fe
1	RISK (0.280***) [0.093,0.467]	BAD (0.247*) [0.031, 0.526]	TEMPT (0.822) [0.719, 0.925]
2	MRR3 (0.267***) [0.089,0.445]	MRR1 (0.234**) [0.035, 0.434]	MRR2 (0.821***) [0.729, 0.913]
3	EFF (0.213**) [0.036, 0.371]	MRR2 (0.178) [0, 0.401]	MRR4 (0.821***) [0.733, 0.908]
4	BAD (0.170**) [0.015, 0.325]	MRR4 (0.178) [0, 0.400]	BAD (0.817***) [0.655, 0.840]
5	MRR1 (0.141*) [0.001, 0.277]	EFF (0.146*) [0.004, 0.321]	MRE1 (0.816***) [0.702, 0.930]
6	MRK3 (0.012) [0,0.066]	MRK3 (0.139) [0, 0.344]	MRR1 (0.792***) [0.689, 0.895]
7	MRE1 (0.010) [0, 0.039]	MRK2 (0.139) [0, 0.377]	MRE2 (0.780***) [0.652, 0.904]
8	MRK1 (0.007) [0, 0.054]	MRE2 (0.109) [0, 0.244]	MRK3 (0.778***) [0.652, 0.904]
9	MRE2 (0.003) [0, 0.041]	MRK4 (0.069) [0, 0.236]	MRK1 (0.705***) [0.521, 0.889]
10	MRK2 (0.001) [0, 0.041]	MRE1 (0.051) [0, 0.208]	EFF(0.670***) [0.492, 0.804]
11	TEMPT (0.000) [0, 0.031]	TEMPT (0.049) [0,0.233]	MRR3 (0.670***) [0.532, 0.812]
12	MRR2 (0.000) [0, 0.030]	RISK (0.028) [0, 0.175]	MRK4 (0.666***) [0.491, 0.862]
13	MRR4 (0.000) [0, 0.029]	MRR3 (0.018) [0, 0.150]	MRK2 (0.658***) [0.468, 0.808]
14	MRK4 (0.000) [0, 0.020]	MRK1 (0.009) [0, 0.128]	RISK (0.647***) [0.449, 0.804]

Table 4: Indices ranked by mean R^2 in simple OLS regression $y_i = \alpha + \beta \mathbf{x}_i + \epsilon_i$. Bootstrapped R^2 (3000 replications). Mean R^2 in brackets. Stars indicate whether R^2 is significantly different from zero (***p<0.01, ** p<0.05, * p<0.1). 95% confidence interval for R^2 (based on empirical distribution) in square brackets. With number of stages fixed effects not all bootstrap replicates allowed to estimate all coefficients. R^2 estimates are based on remaining replications.

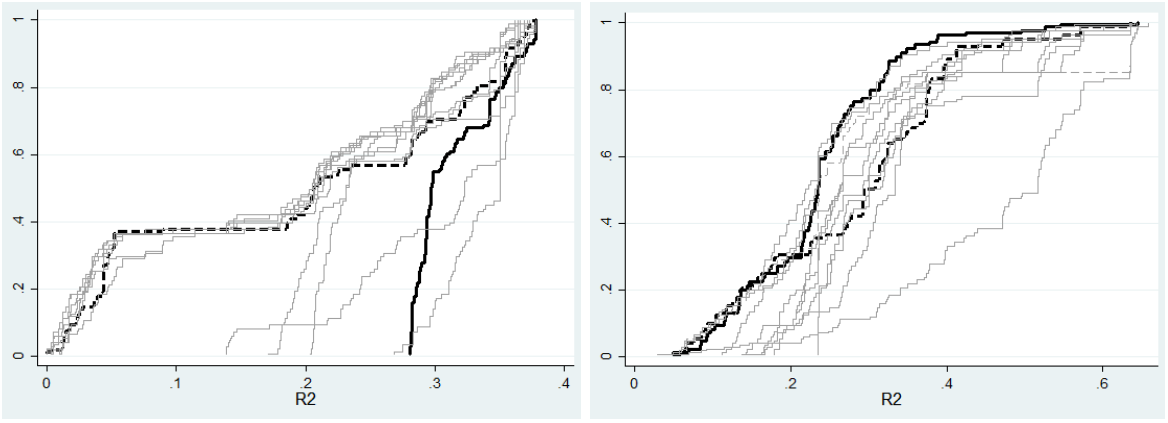
statistically different from zero.

In the repeated game (middle column) the BAD index has the highest R^2 explaining just short of 25% of the observed variation in average cooperation rates. The only other index with an R^2 statistically different from zero in this smaller sample is MRR1. TEMPT , by contrast, is not able to explain much of the variation in cooperation rates by itself. This changes as we include number of stages fixed effects in the regression (rightmost column). As already seen in Table 3 this substantially increases the explanatory power of TEMPT. In fact TEMPT now even has the highest R^2, even though the distributions of R^2 across the 3000 replications of the bootstrap are not significantly different for the first nine indices in Table 4. Since the number of stages is key to explaining average cooperation rates in the repeated game, the R^2 is substantially higher once number of stages fixed effects are included and, in fact, now all regressions have an R^2 clearly above zero.

In a second type of horse-race we allow for more than one index in the regression to account for the fact that certain "families of indices" might explain behaviour well. We hence run the following regressions

$$y_i = \alpha + \beta_1 \mathtt{x}_i + \beta_2 \mathtt{x}'_i + \beta_3 \mathtt{x}''_i + \epsilon_i, \tag{1}$$

where for each fixed index x we add two more indices $x', x'' \neq x$ to the regression from the set $\{\mathtt{RISK}, \mathtt{TEMPT}, \mathtt{EFF}, MRR1, ..., MRK3, MRK4, BAD\}$. This method yields a total of 169 regressions for each index x, some involving two (if $x' = x''$) and some involving three (if $x' \neq x''$) different indices.



(a) Stranger (b) Partner

Figure 4: Cumulative Density function of bootstrapped R^2 (100 reps) for 14 indices across all regressions based on equation (2) which involve the index in question. Bold solid line indicates `RISK` and bold dashed line indicates `TEMPT`.

Figure 4 shows the distribution of R^2 across these 169 regressions for each of the 14 indices. `RISK` is highlighted in bold and `TEMPT` by a bold dashed line. In the Stranger setting (shown in Panel (a)), the R^2 distribution of `RISK` first order stochastically dominates all other indices, except for two (MRR3 and `EFF`). `RISK` has the highest minimal R^2 and the highest R^2 overall. `TEMPT` , by contrast is among the worst performing indices in the one-shot (stranger) setting.

Results are different in the repeated game. `RISK` is now among the worst performing indices and `TEMPT` does relatively better compared to the one-index regressions when combined with other indices. Despite no stage-fixed effects being included in this analysis, there is only one index (MRR4) that first-order stochastically dominates `TEMPT`. Other indices that do well are BAD and MRR2.

The best family of indices in the Stranger combines `RISK` , MRR1 (a measure of efficiency)

and MRR3 (a measure of risk). Hence aspects of risk and efficiency are important, but not so much temptation. In the repeated game the best family includes BAD (which combines aspects of RISK ,TEMPT and accounts for the number of stages T), MRR4 (a measure of temptation) and MRK3, referred to by Murnighan and Roth (1983) as the "behavioural control" of one player over the other.

In the case of Stranger, the best regression has an R^2 of 0.3775 and in the case of the repeated game of 0.6587. Hence, while combining indices has a relatively small effect in Stranger (34% increase in R^2 compared to the regression with only RISK), it does increase the R^2 substantially in the repeated game (by 266% compared to the best one-index regression with only BAD).

To sum up, we have seen that RISK explains variation in cooperation rates in the one-shot game remarkably well also when compared to other indices. In the partner setting, several indices as well as the length of the game (T) are important in explaining variation in cooperation rates. Hence disentangling the role of RISK and TEMPT proves successful in the one-shot games highlighting RISK as a key influence on cooperation rates. This is in line with the idea of people using "pessimistic" beliefs in the face of large uncertainty (see Section 2.3). In the repeated game, by contrast, several indices matter. The analysis has highlighted, though, that TEMPT seems somewhat more important in these games than RISK. Also BAD seems particularly important in these games in line with the analysis presented in Section 2.3.

Result 2 *RISK explains variation in average cooperation rates in the Stranger setting irrespective of which of ten measures from the literature are included. In repeated settings RISK does not explain variation in cooperation rates, but TEMPT as well as several indices from the literature do.*

4 Results: Dynamics

We next ask whether RISK and TEMPT differentially affect the dynamics of cooperation. To answer this question we rely on data from all the studies on the 10-period prisoner's dilemma, either repeated ($T = 10$) or with random rematching ($T = 1; M = 10$) for which full data sets could be obtained. Focusing on 10 period games maximizes the number of available data. The included studies are

all our own lab studies (see Table A1) as well as the studies by Andreoni and Miller (1993), by Bereby-Meyer and Roth (2006), by Cooper et al. (1996), by Dal Bo et al. (2010) and by Normann and Wallace (2012). The resulting data set contains 7860 observations of 786 participants who participated in 23 different studies or treatments. Studying this subsample allows us to understand the effect of RISK and TEMPT on the dynamics of cooperation. It also allows us to exploit variation in the number of past matches m or the number of past stages t *within* studies.

	One-Shot games		*Repeated Games*			
	(1)	(2)	(3)	(4)		(5)
RISK	-0.239***	-0.208***	-0.018	-0.160*	RISK	-0.144
	(0.049)	(0.065)	(0.061)	(0.091)		(0.077)
TEMPT	-0.048	-0.105	-0.318***	-0.280***	TEMPT	-0.294***
	(0.036)	(0.148)	(0.072)	(0.098)		(0.091)
EFF	0.033	0.071	0.253**	0.118	EFF	0.294*
	(0.076)	(0.101)	(0.120)	(0.318)		(0.152)
m		-0.023**		-0.049***	t	-0.030**
		(0.009)		(0.015)		(0.014)
$m\times$ RISK		0.007		0.030***	$t\times$ RISK	0.022***
		(0.009)		(0.008)		(0.008)
$m\times$ TEMPT		0.012		-0.001	$t\times$ TEMPT	-0.004
		(0.007)		(0.011)		(0.010)
$m\times$ EFF		-0.008		0.021	$t\times$ EFF	-0.007
		(0.014)		(0.018)		(0.017)
Constant	0.374***	0.481***	0.358***	0.589***	Constant	0.523***
	(0.048)	(0.064)	(0.103)	(0.243)		(0.131)
Observations	5,180	5,180	2,680	2,680	Observations	2,680
Clusters	12	12	12	12	Clusters	11

Robust Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Cooperation regressed on variables of interest as well as number of past matches experienced m and interactions. Column (5) includes past stages t and interactions. Random effects panel regression with standard errors clustered at the study level.

Table 5 shows regression results. Columns (1)-(2) focus on the stranger setting and columns (3)-(5) on the repeated game. Column (1) replicates our main result for this subsample of studies. RISK is detrimental to cooperation in one-shot studies. Now neither TEMPT nor EFF do have a statistically significant impact on cooperation. Column (2) shows that the number of matches played m has a negative impact on cooperation. As participants gain experience, they cooperate less. This is a typical pattern observed in prisoner's dilemma studies with random rematching (Andreoni and Miller, 1993; Bohnet and Kuebler, 2005; Dal Bo, 2005; Grimm and Mengel, 2009). Experience m does, however, not interact with any of our indices in a statistically significant manner. In the repeated game TEMPT has the most negative effect impact in this sample. Both experience m and the number of past stages played t have a negative impact on cooperation rates. This is also a typical finding: cooperation decreases over time both within matches and across matches. Now

there is also an interesting interaction of m and t, respectively, with RISK. In line with Dal Bó and Fréchette (2011)'s basin of attraction effect, the detrimental effect of RISK is stronger in early stages of the repeated game. In terms of the relative importance of RISK and TEMPT columns (4) and (6) show that in early stages of a game and when players are inexperienced, RISK is comparatively more important, while in later stages of the game and when players are more experienced TEMPT becomes more important.

Partner vs Stranger As previous research has shown that dynamic considerations can also affect the partner-stranger comparison (Mengel and Peeters, 2011), we briefly revisit this comparison. Figure B3 in Online Appendix B4 shows the dynamics overall (main graph) and in four cases of interest: (i) below median RISK and TEMPT (subgraph (a)), (ii) below median RISK and above median TEMPT (b), (iii) above median RISK and below median TEMPT (c) and (iv) above median RISK and TEMPT (d). While overall there is more cooperation in Partner compared to Stranger in this subsample, this is not the case for all combinations of RISK and TEMPT. In particular, if TEMPT is high (subgraphs (b) and (d)), there seems to be no difference in cooperation rates between the partner and stranger settings.

These findings reinforce our finding obtained in Section 3. Because of the differential impact of RISK and TEMPT on the one-shot and repeated game, respectively, comparisons of cooperation rates across these settings will depend on the values of RISK and TEMPT. "High" RISK and "low" TEMPT will be conducive to there being more cooperation in the repeated game, while in other situations no difference or even the reverse ranking could be observed.

Result 3 *The number of past matches and stages played has a negative impact on cooperation. Furthermore, in the repeated game RISK is comparatively more important in early stages of a game and when players are inexperienced, while in later stages of the game and when players are more experienced TEMPT becomes more important*

5 Results: Questionnaire Data

This last section focuses on gender differences. There is a substantial literature on gender differences in cooperation behaviour, which 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 average individual cooperation rates. This last section exploits data from our own lab studies listed in Table A1 in the Online Appendix with 363 participants.[22]

	One-shot		*Repeated*	
	(1)	(2)	(3)	(4)
female	0.215***	0.014	0.291***	0.032
	(0.045)	(0.031)	(0.083)	(0.053)
female × RISK	-0.238***		-0.293**	
	(0.071)		(0.127)	
female × TEMPT	-0.094		-0.207*	
	(0.075)		(0.110)	
Constant	0.224***	0.224***	0.302***	0.302***
	(0.023)	(0.024)	(0.038)	(0.040)
Observations	238	238	125	125
R-squared	0.128	0.001	0.113	0.003

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Average individual cooperation rate regressed on gender dummy interacted with RISK and TEMPT . Simple OLS regressions.

Table 6 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 cooperate on average 23% of the time in the stranger setting (33% in the repeated game) and men 22% of the time in the stranger setting and 30% in the repeated game. Neither of these differences is statistically significant.

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,

[22]This excludes one participant who did not answer the questionnaire and one session where questionnaire data were not collected.

which are usually based on one set of parameters only.

This result adds to existing evidence in the literature. Simpson (2003) has argued precisely that the reason why existing literature on the prisoner's dilemma has, by and large, not found gender differences is because the presence of both temptation and risk in these games (see also Kuwabara (2006)). To test his hypothesis, he tested for gender differences in coordination games (with TEMPT= 0) and Anti-Coordination games (with RISK= 0) and, in line with our findings, observed that women are more cooperative than men when RISK= 0 and less cooperative when TEMPT= 0. We summarize as follows.

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.*

6 Conclusions

We conducted a meta-study of 96 prisoner's dilemma treatments with more than 3500 participants across 6 countries. We focused on two dimensions of the dilemma: RISK (the percentage loss of unilaterally cooperating against a defector) and TEMPT (the percentage gain of unilaterally defecting against a cooperator). While RISK explains variation in cooperation rates across random matching ("stranger") and one-shot treatments, TEMPT and other indices explain more of the variation in repeated interactions. These results are useful for discriminating between competing theories of why people cooperate. For policy making, it is useful to see which dimension of the dilemma 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 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 levels of RISK and TEMPT also affect the comparison between "partner" and "stranger" settings. 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. (1995). Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics 110(1)*, 1–21.

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.

Balliet, D., N. Li, S. Macfarlan, and M. V. Vugt (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin 12*, 1–30.

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.

Blonski, M., P. Ockenfels, and G. Spagnolo (2011). Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence. *American Economic Journal: Microeconomics 3(3)*, 164–192.

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.

Camerer, C. and R. Hogarth (1999). The effects of financial incentives in experiments: A review and capital-labor production framework. *Journal of Risk and Uncertainty 19*, 1–3.

Capraro, V., J. Jordan, and D. Rand (2014). Heuristics guide the implementation of social preferences in one-shot prisoner's dilemma games. *Scientific Reports 4:*, 6790.

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.

Dal Bo, P. (2005). Cooperation under the shadow of the future: experimental evidence from infinitely repeated games. *American Economic Review 95(5)*, 1591–1604.

Dal Bo, P., A. Foster, and L. Putterman (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *American Economic Review 100(5)*, 2205–2229.

Dal Bó, P. and G. Fréchette (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review 101(1)*, 411–429.

Dal Bó, P. and G. Fréchette (2016). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature in press.*

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 (2016). Backward induction in the finitely repeated prisoner's dilemma. SSRn working paper.

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

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.

Habetinova, L. and S. Suetens (2015). Transparency and cooperation in repeated dilemma games: A meta study. mimeo.

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

Iturbe, I., G. Ponti, J. Tomas, and L. Ubeda (2011). Framing effects in public goods: Prospect theory and experimental evidence. *Games and Economic Behavior 72*, 439–447.

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.

Kuwabara, K. (2006). Nothing to fear but fear itself: Fear of fear, fear of greed and gender effects in two-person asymmetric social dilemmas. *Social Forces 84(2)*, 1257.

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.

Mengel, F. (2014). Learning by (limited) forward looking players. *Journal of Economic Behavior and Organization 108*, 59–77.

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.

Rezaei Khavas, T. (2016). *Fairness concerns and cooperation in context*. Ph. D. thesis, University of Utrecht.

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.

Simpson, B. (2003). Sex, fear and greed: A social dilemma analysis of gender and cooperation. *Social Forces 82(1)*, 35–52.

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.

Sonnemans, J., A. Schram, and T. Offerman (1998). Public good provision and public bad prevention: The effect of framing. *Journal of Economic Behavior and Organization 34(1)*, 143–161.

Stahl, D. O. (1991). The graph of prisoners' dilemma supergame payoffs as a function of the discount factor. *Games and Economic Behavior 3*, 368–384.

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