Is Regulation by Milestones Efficiency Enhancing?

– An Experimental Study of Environmental Protection –

JEL: C92, D78, H41, Q54.

Key Words: Cumulative Public Goods, Milestones, Climate Change, Experiment.

Andreas Freytag[a], Werner Güth[b], Hannes Koppel[b,c] and Leo Wangler[a,*]

2010

Abstract

Viewing individual contributions as investments in emission reduction we rely on the familiar linear public goods-game to set global reduction targets which, if missed, imply that all payoffs are destroyed with a certain probability. Regulation by milestones does not only impose a final reduction target but also intermediate ones. In our leading example the regulating agency is Mother Nature but our analysis can, of course, be applied to other regulating agencies as well. We are mainly testing for milestone effects by varying the size of milestones in addition to changing the marginal productivity of individual contributions and the probability to lose.

[a] Friedrich-Schiller-University, Jena, Germany

[b] Max Planck Institute of Economics, Jena, Germany

[c] Corresponding Author: Max Planck Institute of Economics, Kahlaische Str. 10, 07745 Jena, Germany. Tel.: +49 3641 686684. E-mail address: koppel@econ.mpg.de

*We are grateful to Christoph Engel, Gerhard Riener, M. Vittoria Levati, Michael Huettner, Oliver Kirchkamp and Sebastian Vergara, for valuable comments. We also would like to thank seminar audiences at the 2010 ESA world meeting in Copenhagen and the 2010 IAREP/SABE/ICABEEP conference in Cologne for their feedback. We are indebted to Christian Streubel for programming assistance.

1 Introduction

In real life, a number of long-term projects rely on intermediate targets or milestones. For individual choice problems imposing additional constraints may be detrimental to efficiency. External regulation does not make much sense if individuals are able to cope with problems on their own. However, this is different for collective choice problems. In the economic domain, we do indeed observe milestones mainly in social contexts. For instance, governments often announce official targets for budget reductions. Another prominent example is the GATT and its aim of continuous liberalization of international trade – a target which cannot be easily operationalized. Nevertheless, intermediate targets have been regularly set in trade rounds. At the end of each round, the negotiating parties agreed on an agenda to stepwise reduce trade barriers within a certain period.

A third example is environmental conservation. Here investments in climate protection could be imposed by an international environmental agreement (cf. Barrett 1994, 2003). Such an agreement often includes a long-term target which should be reached at a certain date, e.g., reduction of total emissions until 2050 by about 50 per-cent, based on the 1990 emissions (IPCC 2007).[1] Under such circumstances, milestones can proxy intermediate abatement targets to keep total emissions below a critical threshold (e.g., the emission reduction targets in the context of the Kyoto Protocol). If the international community fails to meet these intermediate targets, it will become more difficult to reach the final threshold, makingcatastrophic events more likely.

The rationale behind such milestones is to increase the intermediate credibility of policy announcements through a commitment to testable intermediate targets. Governments may thereby overcome pressure from vested interests and measure long-term goal achievements by short-term goal achievements. The above examples illustrate that milestones are often seen as disciplining factors for policy makers. In global economic policy making, the evidence seems clear – GATT has achieved considerable progress in the field of comprehensive international liberalization. Thus, applying a similar tool to climate policy may also be reasonable.

However, there is neither convincing empirical evidence nor a sound theoretical basis for this in the global governance literature or, more specifically, on how to promote efficiency by imposing additional restrictions like milestones. We do find evidence for markets, e.g., labor markets. Falk and Kosfeld (2006), for instance, have shown that an employer may suffer from imposing a minimum performance threshold for her employees. Similarly, Berninghaus et al. (2008) found that downward wage

[1] Without national commitments ("the business as usual" scenario) estimations predict a temperature increase with possibly catastrophic consequences (Stern 2007, Latif 2010).

flexibility, if exploited, may encourage shirking. In a more general sense, we will add to this literature, but not in a one-off interaction task? but a recursive interaction task.

We conduct a recursive game in which all players can gain by reaching a certain commonly known final target. The game can, however, be more strictly regulated by imposing intermediate targets to be reached earlier. In this way, regulation imposes additional risks of failure. Our main milestone hypothesis predicts that any additional regulation via milestones, i.e., intermediate performance targets, is efficiency enhancing.

Although our leading example is environmental protection to prevent global warming, we abstain from inventing a novel game and test the milestone hypothesis with a familiar experimental workhorse to compare our findings to those of other experiments. More specifically, we use the familiar linear public goods game (see Ledyard 1995, for an early survey of experimental studies) by interpreting contributions as investments protecting the environment, e.g., investments in emission reduction to limit or prevent global warming.

Thus, milestones and contribution targets set lower bounds for emission reduction our experiment. If one of the milestones or the final target is missed, the rather dramatic effect will be that all players sustain a total loss (all of their payoff) with a given probability. This implies additional (subgame perfect) equilibria for the usual free-riding equilibrium where the sums of the contributions accumulated at a certain point reach the targets exactly.

We test the milestone hypothesis as treatment effects with milestones as one treatment variable. Whereas for all treatments the final target is the same, we distinguish between high milestones (H) and considerably lower ones (L), the latter rendering the milestones rather inessential. We compare the H versus L effects in three different scenarios, leading to 2x3=6 different treatments. The three scenarios vary the individual marginal productivity of contributions and the probability of a total loss if one of the targets is not reached.

In spite of the impressive tradition of public goods experiments (Ledyard 1995), there are few studies which focus on environmental protection. Milinski et al. (2008) introduce and experimentally analyze a collective-risk social dilemma framed as a dangerous climate change. The players were endowed with €40 each and could repeatedly contribute €0, €2, or €4 to a "climate change account" over ten rounds. If they failed to reach the threshold after the last round, they sustained a total loss, left with a probability of 90%, 50%, or 10%, respectively. Results show that even with a losing probability of 90% half of the players failed to reach the threshold.

Fischbacher et al. (2010) rely on a linear public goods game, however, with only one trial contribution target with rather similar effects. But they do not address the question whether milestones would be efficiency enhancing. On the other hand, they made their final target stochastic by assuming that players would receive either private or common stochastic signals whose sum would determine the final target. We compare our findings with earlier related findings in the concluding section.

Section 2 describes our experimental design, including all treatments and the experimental protocol. In section 3 we present our results. Our conclusions in section 4 round off the paper.

2 Experimental Design

2.1 General Setting

To capture environmental protection problems, e.g., avoiding global warming, we rely on a linear public goods game (Isaac et al. 1985) as our experimental workhorse. Thus, monetary contributions mean investing in emission reduction for the sake of less global warming, whereas "free riding" stands for voluntarily abstaining from any individual attempt to protect the environment.

In all treatments, five players, respectively participants $i = 1, \ldots, 5$, are endowed with $e = 65$ tokens, which they can either keep or repeatedly contribute over six periods $t = 1, \ldots, 6$. Individual contributions $c_{i,t}$ must satisfy $0 \leq c_{i,t} \leq 10$, guaranteeing that after six periods each participant has some tokens left. In all treatments, furthermore, all players i sustain a total loss, i.e., what they have kept and what they could have gained from accumulated contributions $C^6 = \sum_{t=1}^{6} \sum_{i=1}^{5} c_{i,t}$ by all five players, with a certain probability $p \in (0, 1)$ if the contribution target of $C_6 = 150$ tokens is not reached ($C^6 < C_6$). Assuming constant individual marginal productivity ($\alpha \geq 0.2$) of individual contributions $c_{i,t}$, the payoffs for players $i = 1, \ldots, 5$ are thus

$$U_i = \begin{cases} e - \sum\limits_{t=1}^{6} c_{i,t} + \alpha C^6 & \text{for} \quad C^6 \geq 150 \\ (1-p)(e - \sum\limits_{t=1}^{6} c_{i,t} + \alpha C^6) & \text{if} \quad C^6 < 150. \end{cases}$$

Under the condition that $\alpha < 1 \leq 5\alpha$, opportunism in the sense of own monetary payoff concerns suggests to reduce own contributions in both ranges, $C^6 < C_6$ and $C^6 > C_6$, as long as this does not mean that C^6 becomes smaller than C_6, whereas $\alpha > 0.2$ renders efficient maximal individual contributions (in the sense of payoff maximization). Due to the discontinuity of the payoff function U_i at C_6, there exist many strict but only two symmetric and strict equilibria, leading to results

$$E^0 = \left[\sum_{t=1}^{6} c^0_{i,t} = 0 \quad \text{for} \quad i = 1, \ldots, 5\right] \text{and} \left[\sum_{t=1}^{6} c^*_{i,t} = 30 \quad \text{for} \quad i = 1, \ldots, 5\right] = E^*,$$

respectively.

Together with the efficiency outcome with $\sum_{t=1}^{6} c^+_{i,t} = 60$ for $i = 1, \ldots, 5$ serve as our benchmarks when we discuss actual behavior.[2] Since, in case of E^0, no individual player i can guarantee that the target of 150 is reached, it is obvious that E^0, based on 0-contributions throughout, is a (subgame perfect) equilibrium. This also holds for E^* since increasing $\sum_{t=1}^{6} c_{i,t}$ above 30 is clearly suboptimal, and contributing less than 30 would yield maximally 65 but only with probability p, whereas a player's payoff from E^* is $U_i = 150\alpha + 35$, which is at least 65 due to $\alpha \geq 0.2$.

Note that target C_6 could already be reached within three periods by all five players, contributing maximally ($c_{i,t} = 10$) in each of the three periods. Thus, viewing the first three periods as a base game with already two strict (symmetric) equilibria reveals that "finite-horizon Folk Theorems" (Benoit and Krishna 1987) can be applied, showing that there exist also nonstationary pure strategy (subgame perfect) equilibria.

In all treatments, subjects receive periodic feedback information, i.e., after each period $t = 1, \ldots, 6$ all five players $i = 1, \ldots, 5$ are informed of the individual contributions $c_{j,t}$ of all players $j = 1, \ldots, 5$ and can thus react accordingly when deciding on their next contribution $c_{i,t+1}$. Obviously, this allows for reciprocity and a variety of disciplining actions by future dealings on which the so-called Folk Theorems are based (Aumann 1981, Axelrod and Dion 1988, Benoit and Krishna 1985).

2.2 Milestones

Regulation is implemented by means of milestones (M), i.e., contribution targets on the way to reaching the final target of $C_6 = 150$, namely C_2 after period 2 and C_4 after period 4. Not reaching the intermediate targets has the same consequences as not reaching C_6. Although players $i = 1, \ldots, 5$ may sustain a total loss already after periods 2 and 4, they will, in the experiment, first decide successively for all six periods $t = 1, \ldots, 6$. Only then will it be decided randomly in view of C^2, C^4, and C^6 whether or not they will sustain a total loss already after period 2 if $C^2 < C_2$, after period 4 if $C^4 < C_4$, or finally if $C^6 < 150$.

[2]It is clear that the efficiency benchmark requires $\alpha > 0.2$.

Introducing these milestones, changes the payoff function to

$$U_i^M = \begin{cases} e - \sum\limits_{t=1}^{6} c_{i,t} + \alpha C^6 & \text{if} \quad C^2 \geq C_2 \,\&\, C^4 \geq C_4 \,\&\, C^6 \geq C_6, \\ (1-p)(e - \sum\limits_{t=1}^{6} c_{i,t} + \alpha C^6) & \text{if} \quad \begin{cases} C^2 < C_2 \,\&\, \geq \text{for the other} \\ \text{two restrictions} \quad \text{or} \\ C^4 < C_4 \,\&\, \geq \text{for the other} \\ \text{two restrictions} \quad \text{or} \\ C^6 < C_6 \,\&\, \geq \text{for the other} \\ \text{two restrictions,} \end{cases} \\ ((1-p)+(1-p)^2)(e - \sum\limits_{t=1}^{6} c_{i,t} + \alpha C^6) & \text{if} \quad \begin{cases} C^2 < C_2 \,\&\, C^4 < C_4 \,\&\, C^6 \geq C_6 \\ \text{or} \\ C^2 < C_2 \,\&\, C^6 < C_6 \,\&\, C^4 \geq C_4 \\ \text{or} \\ C^4 < C_4 \,\&\, C^6 < C_6 \,\&\, C^2 \geq C_2, \end{cases} \\ ((1-p)+(1-p)^2+(1-p)^3))(e - \sum\limits_{t=1}^{6} c_{i,t} + \alpha C^6) & \text{otherwise,} \end{cases}$$

where $C^2 = \sum\limits_{t=1}^{2} \sum\limits_{j=1}^{5} c_{j,t}$ and $C^4 = \sum\limits_{t=1}^{4} \sum\limits_{j=1}^{5} c_{j,t}$. Comparing U_i with U_i^M clearly reveals that implementing milestones on a sufficiently high level means implementing "regulation," where, in view of the environmental interpretation, the regulating agency is Mother Nature. We predict a milestone effect, i.e., a more efficient performance with stricter milestones. In order to test this effect, we distinguish two cases:

1. strict milestones (H): $C_2 = 50$ and $C_4 = 100$, and
2. less strict milestones (L): $C_2 = 5$ and $C_4 = 10$.

For the case of "strict milestones" (H) we set the intermediate targets such that contributions necessary to reach the final target of C_6 increase linearly. For the less strict case we do not omit the milestones but lower them by a factor of 10 which should render them inessential such that payoff U_i^M approximates U_i. The "philosophy" of such a manipulation is, of course, that the two cases, H and L, rely on the same verbal instructions and differ only in two numerical parameters, namely C_2 and C_4, which should not induce any difference in (sub)conscious demand effects between H and L, where "L" stands for –extremely– "low milestones."

2.3 Scenarios

We consider three different scenarios to test the potential milestone effect by comparing treatments with strict (H) and less strict milestones (L).

In the baseline scenario (B), we set $\alpha = 0.4$ and $p = 0.5$ in combination with a group size of five subjects.[3] Since the probability of sustaining a total loss seriously increases the free-riding "disincentives" the efficiency benchmark, $\sum_{t=1}^{6} c_{i,t}^{+} = 60$ for $i = 1, \ldots, 5$ may be expected more often than in usual public goods experiments.

Our experimental design might be criticized since linearly increasing total payoffs, even above the final target, may not adequately capture environmental protection. We therefore propose an alternative scenario (S), in which reaching the final target of C_6 merely preserves the status quo, i.e., a mean payoff of 65 tokens, and overshooting is not beneficial at all, i.e., removing the efficiency of $C^6 > C_6$. This is done by lowering the constant individual marginal productivity to $\alpha = 0.2$. Compared to scenario B, incentives to cooperate are, of course, smaller, namely below C_6.

This manipulation changes two aspects: it questions the efficiency benchmark and reduces the free-riding "disincentives," as measured by the expected payoff of a unilateral deviation from the E^*-equilibrium to constant 0-contributions (free riding). The difference in expected payoffs between the E^* -equilibrium and the payoff of a unilateral deviation to constant 0-contributions for scenario B is $95 - 56.5 = 38.5$ tokens, whereas it is 20.5 tokens for scenario S only. Since by comparing these scenarios, the two effects mentioned above cannot be disentangled, we consider a third scenario (P) and preserve the equilibrium and efficiency benchmarks of the baseline scenario by setting $\alpha = 0.4$ but keeping the free-riding "disincentive" equal to that of scenario (S) by lowering the probability of a total loss from $p = 1/2$ to $1/3$.[4] Altogether, this $2X3$ factorial design results in six treatments as listed in table 1.

2.4 Experimental Protocol

We ran 12 separate sessions for the six treatments. Three hundred sixty student participants were recruited from various disciplines of Friedrich Schiller University of Jena using the ORSEE software (Greiner 2004). The experiment was programmed and conducted with the software z-Tree (Fischbacher 2007). In each session, the 30 participants

[3] One might argue that setting $\alpha = 0.4$ is unrealistic in a climate change setting since investments in emission reduction are usually seen as preserving the status quo. This is because sustainability is the main argument for policy intervention.

[4] The probability is calculated by comparing the individual payoff that results when all players $i = 1, \ldots, 5$ play E^* to the individual payoff that results when a player deviates from E^* by free riding: $(95 - 20.5) = (1 - p)(65 + 0.4x120)$, implying $p = 38.5/113 \approx 1/3$.

Table 1: $2x3$ factorial treatment design

		Milestones	
		H	L
P	$\alpha = 0.4; p = 1/3$	PH	PL
B	$\alpha = 0.4; p = 1/2$	BH	BL
S	$\alpha = 0.2; p = 1/2$	SH	SL

per session were subdivided into two equally large matching groups and played the 6-period recursive games repeatedly. After each play of the 6-period recursive game the 15 participants of a matching group were randomly rematched to form three new groups with five players each who interacted in the next round of play. Since participants were only told that they were randomly rematched, they could expect that each of the 29 other participants might become an interaction partner. This should discourage reputation effects even more (participants can, of course, try to establish some reputation within the same rounds, i.e., across the six periods of a given round).

After entering the computer laboratory of the Max Planck Institute of Economics in Jena, participants received written instructions (see App. C for translated material), which were read aloud to establish their common knowledge. After answering questions privately participants had to answer a few control questions. The experiment only started when all participants had answered all control questions correctly. One session with altogether 12 rounds lasted, on average, 90 minutes, including reading instructions, answering control questions, and payment. Average payoffs were €17, with a minimum of €2.5 and a maximum of €29, including the €2.5 show-up fee.

3 Results

To begin with, we describe our findings on the group level, followed by a closer look at individual behavior. We first state the "Results and then proceed to validate them by descriptive and statistical data analysis.

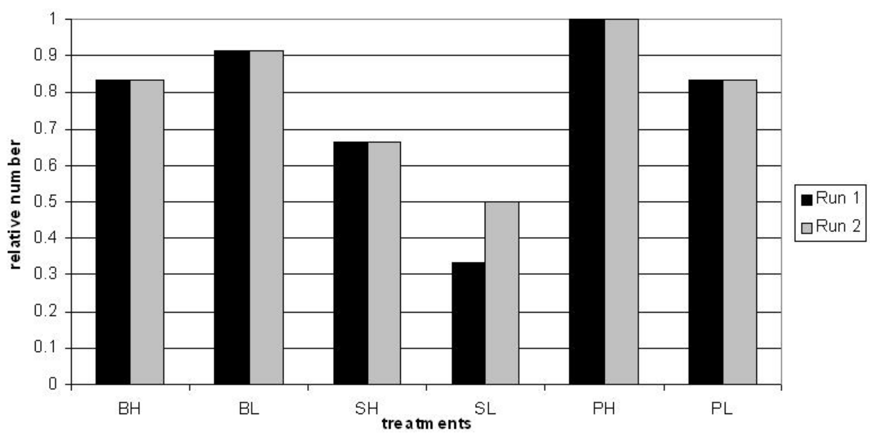
RESULT 1: Equilibrium play E^* and E^0 is negligible.

Only three out of 144 groups ended up in the E^* outcome of investing 150 tokens in total (two groups in treatment SH and one in treatment SL). One group in treatment SL was able to coordinate on the fair share equilibrium of contributing five tokens in each round. No groups totally free rode or contributed the maximum possible. However, we are not interested in testing equilibrium outcomes; rather, we want to study treatment effects and therefore turn to our main question, namely whether regulation by milestones is efficiency enhancing.

RESULT 2: Depending on the scenario, milestones increase the probability of reaching the final target.

Since expected payoffs are lower when the final target is not reached compared to when it is reached, it is more efficient in all scenarios to reach the final target. Figure 1 shows the probability of reaching the final target separately for scenario and treatment. In scenario B and P, almost all groups succeeded (10 of 12), and there is no significant treatment effect (H versus L). The picture slightly changes for scenario S with an almost significant milestone effect for the success probability in the first run, where 8 versus 4 out of 12 groups reached the final target (Fisher's exact, $p = 0.110$). However, the effect disappears since after the restart more groups, namely six, succeeded in SL, whereas for SH, there is no change (Fisher's exact, $p = 0.340$).

Figure 1: Final target reached



RESULT 3: On the group level, milestones increase average group contributions only in scenario S.

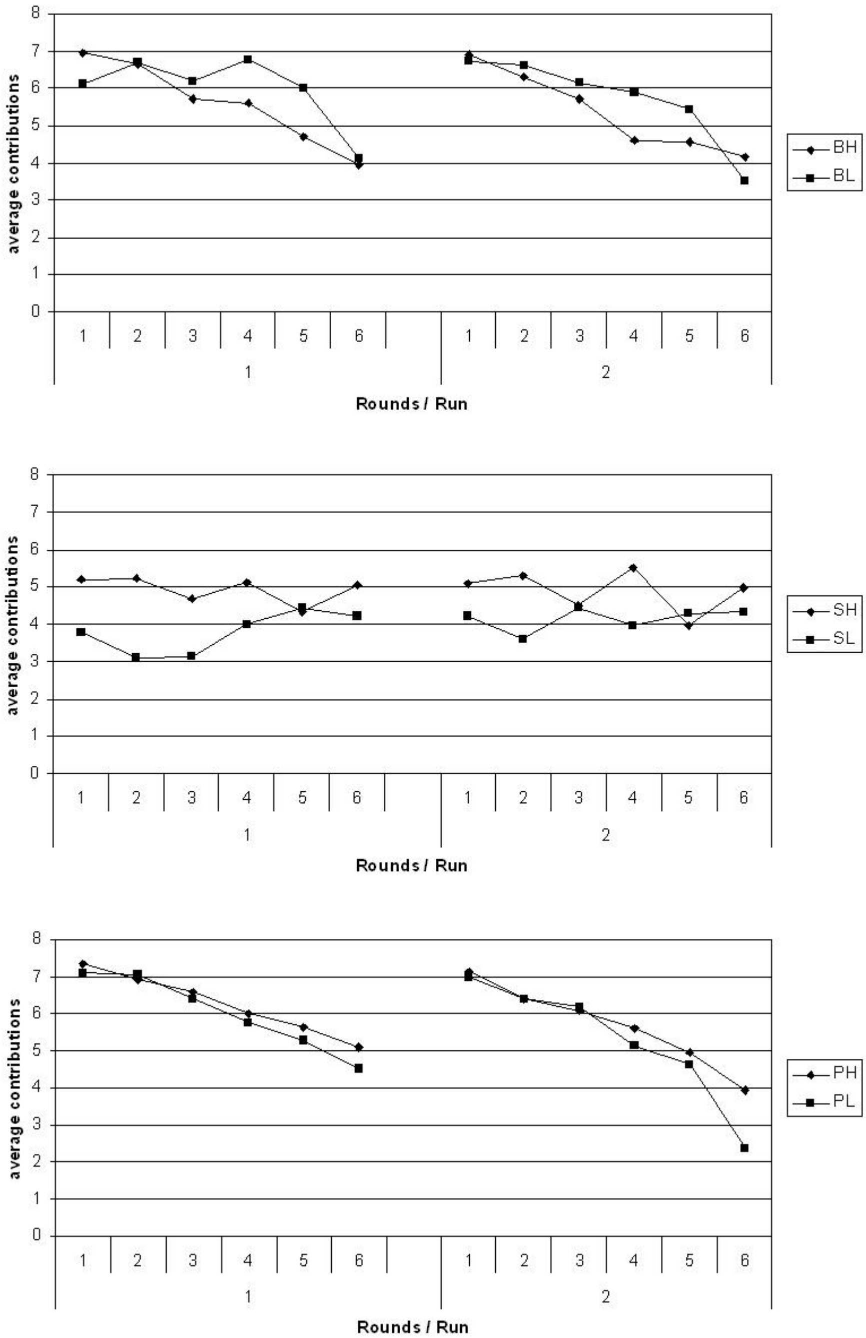
Figure 2 depicts, separately for the three scenarios (scenario B on top, scenario S in the middle, and scenario P at the bottom) and treatments, average contributions over the sequence of play, i.e., the six rounds of two runs. In the first run of scenario B, average contributions were lower in the treatment with strict milestones (5.6 tokens versus 5.98 tokens), and it seems that imposing additional risks by intermediate targets is detrimental to efficiency. The effect is not statistically significant (Mann-Whitney-U, $p = 0.2142$), however. For scenario S a significant milestone effect shows up in the first run. Imposing milestones increased contributions by approximately 30 percent from 3.78 to 4.93 tokens (Mann-Whitney-U, $p = 0.0831$) and is thus efficiency enhancing. No significant difference between treatments shows up in scenario P, which also goes for all three scenarios after the restart.

Thus, milestones increased the probability of success and contributions only in scenario S, which features investments in emission reduction as preserving the status quo by ruling out efficiency enhancement below and above C_6.

RESULT 4: An analysis of individual contributions shows that milestones are inspiring the former in scenario S and P.

Accordingly, on the level of individual behavior, the picture for scenario P changes when panel regressions are used. By design, the panel is strongly balanced and consists of 60 subjects per treatment cell (120 subjects per scenario). Taking group heterogeneity into account, we made use of a panel regression with adjusted standard errors on the group level (each group formed one cluster), i.e., in total, there were 24 groups per treatment (48 groups per scenario). Moreover, there were 24 groups for each run and 48 groups for both runs together. Contributions were explained by a dummy for the treatment with strict milestones (PH), dummies for one session of the respective treatments (S_{PH} and S_{PL}), lagged variables on own contributions, average contribution within the group, and accumulated contributions. Regression results are shown in table 2. There are no significant treatment effects in the first run (the first two columns). However, in columns 3 and 4, showing regression results for the sequence after the restart (second run), the treatment dummy is positive and significant. Controlling for sessions only (column 3), the effect is significant at the 5 percent level. Additionally controlling for various forms of information which subjects received (column 4) results in a better fit and a significant treatment effect on the 1 percent level, although it lower in magnitude. More precisely, subjects contribute, on average, 0.766 tokens more to the public good with strict rather than less strict milestones. Taking the two runs together in column 5 and controlling for the restart (including a dummy), the effect is weaker (on average 0.457 tokens more in SH) but still significant at the 5 percent level, whereas the restart dummy has no significant effect.

Figure 2: Average contribution per treatment



The effect is stronger in scenario S (see table 3). Although we do not find a significant treatment effect after the restart, there is a strong and high effect in the first run. Subjects in treatment SH contributed, on average, 1.080 tokens more with strict milestones, when controlling for the received information (column 2). In contrast to scenario P, the milestone effect disappears after the restart (columns 3 and 4) but is present when we consider both runs, controlling for information and the restart (column 5). Subjects in treatment SH contribute, on average, 0.873 tokens more than in SL, whereas the restart dummy is insignificant. Individual level analysis offers no further insights on scenario B (see appendix A).

Table 2: OLS Panel regression with clustered standard errors on the group level for scenario P

	Run 1 contribution	Run 1 contribution	Run 2 contribution	Run 2 contribution	Both runs contribution
PH	0.872	0.160	1.261*	0.766**	0.457*
	(1.46)	(0.65)	(2.53)	(2.93)	(2.46)
S_{PH}	-0.294	0.308	-0.372	-0.00931	-0.0242
	(-0.36)	(1.07)	(-1.59)	(-0.06)	(-0.13)
S_{PL}	0.956	0.403	1.350*	0.758**	0.483
	(1.18)	(1.02)	(2.29)	(2.83)	(1.68)
Lag contribution		0.479***		0.479***	0.426***
		(7.00)		(8.04)	(10.08)
Lag average contr.		0.320**		0.109	0.171*
		(3.26)		(1.01)	(2.17)
Lag accumulated		-0.00715*		-0.0164***	-0.000674
		(-2.56)		(-4.77)	(-0.34)
Restart					-0.0471
					(-0.29)
cons	5.533***	1.267**	4.606***	2.650**	1.824***
	(19.27)	(3.09)	(9.97)	(2.92)	(3.97)
N	720	600	720	600	1320
$N_{Indiv.}$	120	120	120	120	120
R^2_O	0.0102	0.333	0.0193	0.301	0.227

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Although the non-parametric group level analysis suggests no milestone effect in scenario P, we do find a significant milestone effect on the individual level, controlling for group, session, and information effects. Compared to scenario S, the effect is lower in magnitude and less significant. The milestone effect is therefore not only driven by the exclusion of efficiency above and below targets (scenario S) but is also due to higher free-riding "disincentives" (scenario P and S).

RESULT 5: The milestone effect in scenarios S and P is mainly driven by a higher share of individual contributions between 4 and 6 tokens.

To further scrutinize contributions on the individual level as well as the general sequence of play, we classified contributions as *low* (0-3 tokens), *medium* (4-6 tokens),

Table 3: OLS Panel regression with clustered standard errors on the group level for scenario S

	Run 1 contribution	Run 1 contribution	Run 2 contribution	Run 2 contribution	Both runs contribution
SH	2.156***	1.080*	1.000	0.473	0.873*
	(3.90)	(2.32)	(1.60)	(0.86)	(2.21)
S_{SH}	-0.100	-0.128	0.333	0.359	0.0898
	(-0.38)	(-0.93)	(1.07)	(1.52)	(0.58)
S_{SL}	1.917**	1.286*	0.844	0.538	0.850*
	(2.82)	(2.56)	(1.32)	(1.08)	(2.15)
Lag contribution		0.343***		0.387***	0.333***
		(3.68)		(4.74)	(5.75)
Lag average contr.		0.0874		-0.0340	-0.0196
		(0.51)		(-0.17)	(-0.16)
Lag accumulated		0.00925**		0.00530	0.00556*
		(2.73)		(1.44)	(2.55)
restart					0.0416
					(0.21)
cons	2.822***	1.052*	3.717***	2.079**	1.949***
	(5.36)	(2.41)	(6.87)	(2.78)	(4.85)
N	720	600	720	600	1320
$N_{Indiv.}$	120	120	120	120	120
r^2_O	0.100	0.232	0.0363	0.165	0.171

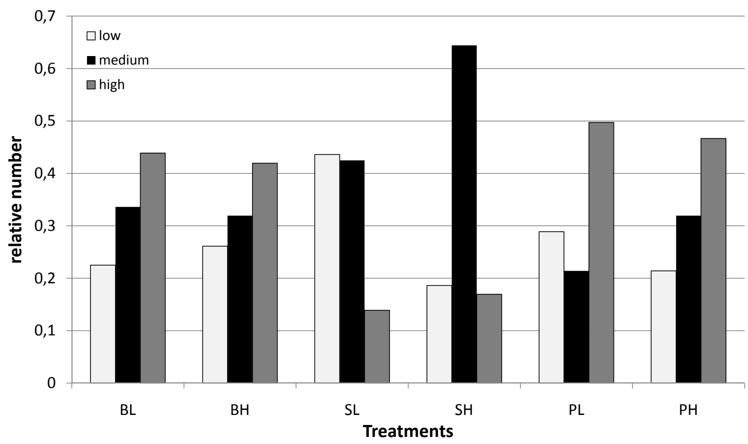
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and *high* (7-10 tokens). Figure 3 shows the resulting relative number of contributions for the respective classes in the six treatments in the first run.[5] It further shows that contributions are quite heterogeneous. However, in treatment SH most contributions (64.44 percent) fall into the *medium* class. Compared to treatment SL, milestones seem to discipline subjects to stay on track to reach the final target as the number of *low* contributions is significantly lower (Fisher's exact, $p = 0.000$) and *medium* contributions are significantly more frequent (Fisher's exact, $p = 0.000$). Though not particularly strong, a similar pattern is found for scenario P, where we observe a significantly lower frequency of *low* contributions and a significantly higher frequency of *medium* contributions in treatment PH (Fisher's exact, $p = 0.025$ for *low* and $p = 0.002$ for *medium*). No significant difference between contribution classes is found in scenario B (see also the graphical illustration in fig. 3). The share of contributions classified as *high* in scenario S is significantly lower than that in scenario B and P (Fischer's exact, $p = 0.000$, for BL vs. SL, BH vs. SH, PL vs. SL, and PH vs. SH), whereas we do not

[5]In the following, we show results for the first run only. Regarding the second run, the qualitative results for the classification are the same.

find any significant difference between *B* and *P*.

Figure 3: Contribution classes for *B*, *S*, and *P*



RESULT 6: Milestones stabilize individual behavior over the sequence of play in scenarios S and P.

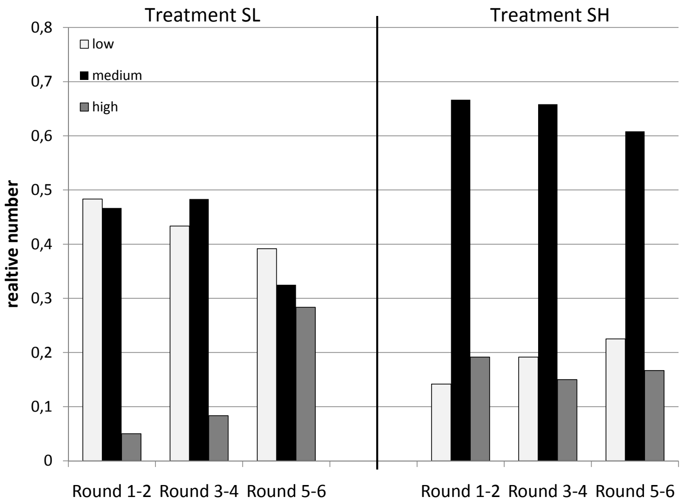
Is the classified behavior stable over the sequence of play ? To answer this question, we subclassified, additionally to the above classification, the relative number of contributions into three phases of rounds: *round 1-2* , *round 3-4*, and *round 5-6*.[6]

The results of the classification in scenario *S* are shown separately for the two treatments (*SL* and *SH*) in figure 4. There is a relatively stable share of *low* contributions in treatment *SL* over the three phases (Kruskall-Wallis, $p = 0.3588$), which is significantly higher than in *SH* (Fisher's exact, $p = 0.000$ for *round 1-2* as well as *round 3-4* and $p = 0.004$ for *round 5-6*). In contrast, a high and stable share of *medium* contributions is found in treatment *SH* (Kruskall-Wallis, $p = 0.5946$), which in all three phases is significantly higher than in *SL* (Fisher's exact, $p = 0.000$ for *round 1-2* , $p = 0.001$ for *round 3-4*, and $p = 0.004$ for *round 5-6*). Thus, the disciplining effect of the milestones operates through *medium* contributions, i.e., subjects seem to coordinate on medium contributions throughout. Subjects in treatment *SL* tried to make the best out of a bad job in *round 5-6*, with significantly more contributions in the *high*

[6] We chose this classification to capture differences in play between rounds including a target. Moreover, results do not change qualitatively if we take every single round into account.

class than in the previous rounds (Fisher's exact, $p = 0.000$). However, as shown in figure 1, they often failed to reach the long-term target. The comparison between SL and SH over the sequence of play shows that milestones stabilize average contributions and thereby offer a certain intermediate planning reliability.

Figure 4: Contribution classes over treatments and rounds for S

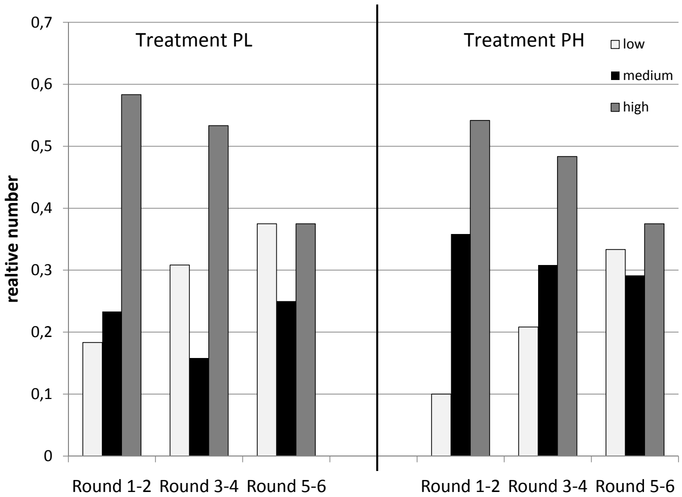


In Scenario P (see fig. 5), milestones have a significant disciplining effect especially in the first and second phase. The share of *low* contributions for PH is significantly lower than for PL (Fisher's exact, $p = 0.047$ for *round 1-2* and $p = 0.052$ for *round 3-4*). However, low contributions in both treatments increase steadily, indicating that participants anticipated that total contributions would exceed the critical thresholds. We also observe differences for the *medium* contribution class. In phases 1 and 2 *medium*, contributions in treatment PH are significantly higher than in treatment PL (Fisher's exact, $p = 0.024$ for *round 1-2* and $p = 0.005$ for *round 3-4*). This finding indicates that, as in scenario S, the milestone effect is driven by *low* and *medium* contributions. The overall contribution patterns for both treatments (besides the differences mentioned above) are rather similar.

4 Conclusions

To investigate if regulation by milestones is efficiency enhancing, we have imposed additional risks of failure. If, in a threshold public goods game featuring a final target

Figure 5: Contribution classes over treatments and rounds for P



after six rounds, the final target is not reached, a total loss is sustained with a given probability. The same consequences can be assumed if a milestone is not reached. Our treatments vary the magnitude of the milestones from less strict (approximately inessential) to strict (essential), the marginal productivity of contributions and thereby efficiency and free-riding incentives as well as the probability of a total loss in case of failure.

We find substantial differences between the three scenarios. Milestones do have a positive impact on efficiency when there is no efficiency benchmark and free-riding "disincentives" are low. The effect is strongest when higher contributions below and above targets are not efficiency enhancing and free-riding "disincentives" are low so that investments in emission reduction can only preserve the "status quo." A moderate effect is found when efficiency can be promoted but free-riding "disincentives" are still low. However, the result is mainly due to second run behavior. Since in the context of climate change, (there may not be) a restart or second chance may not be possible, learning might come too late. In the scenario where efficiency can be promoted and incentives to free ride are low, no milestone effect is found.

Our results are similar to Milinski et al. (2008) who find that half of the groups had difficulties reaching the final target. Note that Milinski et al. framed the game as a climate change, which might have increased contributions. Comparing our findings to Fischbacher et al. (2010), who do not implement any intermediate targets, we confirm their result that more serious losses promote cooperation when the threshold is missed. It is worth mentioning that commonly known targets (the common signal case), which we have implemented in a deterministic way, seem to provide a best-case scenario for environmental protection. Their and our observations imply that, depending on the specific scenario, regulation by milestones can be efficiency enhancing.

Caution should be exercised when generalizing our conclusions. Since we do not capture the advantages of early investments, our scenario is a kind of worst-case scenario for testing the milestone hypothesis. In the case of environmental protection, early investments can be considered superior to late investments. Without early investments, the costs of climate protection may increase because emissions accumulate, making it more difficult to reach a certain emission reduction target (cf. Kemfert 2005). Moreover, environmental returns might need some time to develop and accumulate. Capturing this, was not within the scope of this paper but would be a challenging topic for future research.

In our experiment, we have implemented and manipulated milestones exogenously. This seems unrealistic when thinking of environmental agreements where milestones are usually negotiated, as has been done in Kyoto. Implementing endogenous milestones in such a setting is problematic as one cannot invoke the punishment of Mother Nature. Before doing so, one has to think of the consequences if a milestone is missed.

In the actual debate on climate change, which discusses the investments in emission reduction needed to protect the climate in the long term, milestones may be essential to overcome the current coordination problem. Intermediate targets, as proposed by international environmental agreements such as the Kyoto Protocol, might help to solve this problem. However, our results reveal a high risk of failure. This has to be kept in mind when hoping for the milestone effect, especially when discussing coordination problems with possibly catastrophic consequences. In our scenarios, milestones provide a punishment mechanism imposed by nature, which in reality may is lacking so far and whose implementation through international agreements may be problematic but might prove effective nevertheless.

References

Aumann, R.: 1981, Survey of repeated games, Bibliographisches Institut, Mannheim/Wien/Zürich, pp. 11–42.

Axelrod, R. and Dion, D.: 1988, The further evolution of cooperation, *Science* **242**(4884), 1385–1385.

Barrett, S.: 1994, Strategic environmental policy and international trade, *Journal of Public Economics*, **Vol. 54**, pp. 325–338.

Barrett, S.: 2003, *Environment and statecraft*, Oxford University Press Oxford.

Benoit, J. and Krishna, V.: 1985, Finitely repeated games, *Econometrica: Journal of the Econometric Society* **53**(4), 905–922.

Benoit, J. and Krishna, V.: 1987, Nash equilibria of finitely repeated games, *International Journal of Game Theory* **Vol. 16(3)**, pp. 197–204.

Berninghaus, S., Bleich, S. and Güth, W.: 2008, Wage Flexibility in Ongoing Employment Relations - An Experiment with a Stochastic Labor Market, *in* W. Franz, H. Ramser and M.Stadler (eds), *Schriftenreihe des Wirtschaftswissenschaftlichen Seminars Ottobeuren*, Mohr Siebeck, Tübingen, pp. 171–192, Vol.37.

Falk, A. and Kosfeld, M.: 2006, The Hidden Costs of Control, *The American economic review* **Vol. 96(5)**, pp. 1611–1630.

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

Fischbacher, U., Güth, W. and Levati, V. M.: 2010, Crossing the point of no return: a public goods experiment (mimeo).

Greiner, B.: 2004, The online recruitment system orsee 2.0 - a guide for the organization of experiments in economics, *mimeo*, Department of Economics, University of Cologne.

IPCC: 2007, *International Panel on Climate Change, 30 April-4 May 2007: Mitigation of Climate Change – Summary for Policymakers*. <http://www.ipcc.ch/SPM040507.pdf>.

Isaac, R., McCue, K. and Plott, C.: 1985, Public goods provision in an experimental environment, *Journal of Public Economics* **Vol. 26(1)**, pp. 51–74.

Kemfert, C.: 2005, Weltweiter Klimaschutz - Sofortiges Handeln spart hohe Kosten, *DIW-Wochenbericht* **72**(12-13), 209–215.

Latif, M.: 2010, Die Herausforderung globaler Klimawandel, *Perspektiven der Wirtschaftspolitik* **Vol. 11 (s1)**, pp. 4–12.

Ledyard, J. O.: 1995, Public Goods: A Survey of Experimental Research, *in* J. H. Kagel and A. E. Roth (eds), *The Handbook of Experimental Economics*, Princeton University Press, Princeton, New York, pp. 111–194.

Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A. and Marotzke, J.: 2008, The collective-risk social dilemma and the prevention of simulated dangerous climate change, *Proceedings of the National Academy of Sciences* **Vol. 105(7)**, pp. 2291.

Stern, N.: 2007, *The Economics of Climate Change: The Stern Review*, Cambridge University Press, Cambridge, New York.

A Panel regressions

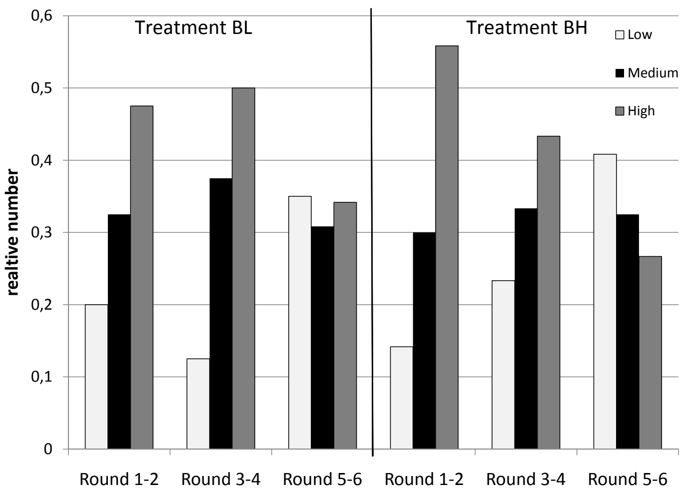
Table 4: OLS Panelregression with clustered standard errors on group level for scenario B

	Run 1 contribution	Run 1 contribution	Run 2 contribution	Run 2 contribution	Both runs contribution
BH	-0.367	-0.206	-0.400	-0.262	-0.315
	(-0.66)	(-0.46)	(-1.94)	(-1.52)	(-1.40)
S_{BH}	-0.111	-0.187	-0.0389	0.0564	-0.0731
	(-0.22)	(-0.48)	(-0.20)	(0.33)	(-0.35)
S_{BL}	-0.0889	0.186	-0.133	-0.0141	-0.0639
	(-0.19)	(0.56)	(-0.21)	(-0.03)	(-0.21)
Lag contribution		0.371^{***}		0.284^{***}	0.294^{***}
		(5.61)		(3.81)	(5.93)
Lag average contr.		0.213		0.263	0.0878
		(1.69)		(1.94)	(1.02)
Lag accumulated		-0.0142^{***}		-0.0137^{***}	-0.00495^{*}
		(-4.52)		(-3.70)	(-2.13)
Restart					0.181
					(0.98)
cons	6.022^{***}	3.509^{***}	5.794^{***}	3.466^{***}	3.956^{***}
	(22.54)	(3.89)	(31.86)	(3.99)	(7.03)
N	720	600	720	600	1320
$N_{Indiv.}$	120	120	120	120	120
r^2_O	0.00341	0.214	0.00324	0.189	0.115

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B Contribution classes over rounds

Figure 6: Contribution classes over treatments and rounds for *B*



C Instructions (English translation for treatment BH)

Welcome and thank you for participating in this experiment! Please read these instructions carefully, they are identical for all participants. For arriving on time you receive a show-up fee of €2.50. In the experiment, you will earn additional money, depending on your decisions and the decisions of other participants. During the course of the experiment, all amounts are stated in ECU (experimental currency units). At the end of the experiment, all earned ECU will be converted into cash and privately paid out according to the following exchange rate:

$$1ECU = 0.10€$$

From now on, please do not talk to your neighbors, switch of your cell phone, and remove unnecessary things from your desk. It is important that you follow these rules, otherwise we have to exclude you from the experiment and any payment. In case you have a question, please raise your hand, and we will answer your question privately.

The experiment will last for 6 rounds, and you will have to make a decision in each round. You are randomly assigned to groups of 5 participants, which remain unchanged for all rounds. At the beginning of the experiment, each participant of the group is endowed with 65 ECU just once. Your task in each of the 6 rounds is to make a decision on how you will use the 65 ECU.

The decision problem
As already described, you are a member of a group of five participants, in which each member is endowed with 65 ECU at the beginning. In each of the six rounds, you have the possibility to contribute any integral number between 0 and a maximum of

10 ECU to a joint account. The amount you have not contributed you may keep. After each member has made a decision about his or her contribution to the joint account, the next round starts, except for the sixth and last round.

The total income of each member of the group after the sixth round is calculated as follows:

Income from the joint account = Sum of all contributions over six rounds x 0.4

plus the ECU not contributed during the six rounds:

Total Income = Income from the joint account + ECU not contributed

For example, if, after six rounds, the sum of contributions of all group members to the joint account is 150 ECU, you and any other group member will receive an income of 150 x 0.4 = 60 ECU from the joint account. Additionally, you and all other group members will receive the respective ECU that have not been contributed to the joint account. If, after six rounds, the sum of contributions of all group members to the joint account is 150 ECU and you have not contributed 35 ECU, you will receive 60 + 35 = 95 ECU.

Thresholds

The total income at the end of round 6 also depends on whether the sum of contributions to the joint account has reached certain thresholds after the critical rounds 2, 4, and 6. The threshold for the sum of contributions after the second round is 50 ECU, after the fourth round 100 ECU, and after the sixth round 150 ECU. If the sum of contributions after a critical round does not reach the respective threshold, you will lose your total income with a probability of 50%.

All necessary random draws are made successively after round 6 (for rounds 2, 4, and six). This means you make a decision about your contribution to the joint account six times but will be informed whether you lost your total income if a threshold has not been reached after one of the critical rounds after the end of round 6. The result of the random draws will then be displayed on your computer screen.

The probability of a total loss

If your group contributes less to the joint account than is required for the respective thresholds after each of the three critical rounds (2, 4, and 6), you will lose your total income with a probability of $1/2 + 1/2 * 1/2 + 1/2 * 1/2 * 1/2$ =$875/1000 (= 87.5\%)$.

If your group contributes less to the joint account than is required for the respective thresholds after two of the three critical rounds (rounds 2 and 4, 4 and 6, 2 and 6), you will lose your total income with a probability of $1/2 + 1/2 * 1/2$ =$75/100 (= 75\%)$.

If your group contributes less to the joint account than is required for the respective threshold after one of the three critical rounds (round 2, 4, or 6), you will lose your total income with a probability of $1/2 (= 50\%)$.

In case your group has reached the respective thresholds after each of the three critical rounds (round 2, 4, and 6), you may keep your total income for good.

If the threshold has been reached after each of the critical rounds, you and your group members earn the income from the joint account (sum of contributions over six rounds x 0.4) plus the ECU that you have not contributed.

If the sum of contributions to the joint account is less than 150 ECU after round 6, even though the thresholds have been reached after the other two critical rounds before (i.e., one of the three thresholds), you and your group members will lose your total income with a probability of 1/2 (50%). With a probability of 1/2 (=50%), you will receive the income from the joint account (sum of all contributions over six rounds x 0.4) plus the ECU that you have not contributed. The probability of not losing the

	None of the three thresholds reached	One of the three thresholds reached	Two of the three thresholds reached	All three thresholds reached
Probability of losing the total income	87,5%	75%	50%	0%

total income is reduced analogously if more than one threshold has not been reached. After each round you are told how much each member of the group has contributed to the joint account.

Randomized Events
If thresholds have not been reached, it will be randomly decided whether you lose your total income after round 6. One number out of 1 to 1000 is randomly drawn. A number between 1 and 500 translates into a negative result (you lose your total income), while a number between 501 and 1000 translates into a positive result (you do not lose your total income). The number of random draws depends on the number of thresholds that have not been reached. If necessary, we start with the threshold after round 2, followed by, if necessary, the threshold after round 4, and finally, if necessary, the threshold after round 6. After the six rounds, your total income, the results of potential random draws, and your payoff/ earnings (in €) will be displayed on the screen. After have finished reading the instructions, please click **Continue**. You will then be asked to answer some control questions.

Please answer the following control questions. The experiment will only start after all participants have answered all questions correctly.

1. Each group member is endowed with 65 ECU. Assume that all five group members (including yourself) contributed 3 ECU to the joint account in each of the six rounds.

   (a) In which critical rounds is the threshold reached (please mark the correct answer)?
   Round 2 ☐ and or
   Round 4 ☐ and or
   Round 6 ☐ or
   None of the three rounds ☐

   (b) With which probability will you lose your total income?

   ......

2. Each group member is endowed with 65 ECU. After the second round, a total 4 ECU have been contributed to the joint account. In the third round, a total of 20 ECU and in the fourth round a total of 26 ECU are contributed to the joint account. After round 6, 165 ECU have been contributed to the joint account.

   (a) After which rounds is the threshold reached (please mark the correct answer)?
   Round 2 ☐ and or
   Round 4 ☐ and or
   Round 6 ☐ or
   None of the three rounds ☐

(b) With which probability will you lose your total income?

......

(c) Assume all random draws are to your advantage. Which income do you receive from the joint account?

......

3. Each group member is endowed with 65 ECU. You contribute a constant amount to the joint account in each of the six rounds. The other four group members contribute the same amount to the joint account in each of the six rounds.

   (a) What is the total income you will receive after round 6 if you and your group members contribute 10 ECU to the joint account in every round?

   ......

   (b) With which probability will you lose your total income if you and your group members contribute 0 ECU to the joint account in every round?

   ......

4. A total of 155 ECU have been contributed to the joint account. After round 6, you have 10 ECU left.

   (a) With which probability will you lose your total income if only the threshold after round 6 has been reached?

   ......

   (b) With which probability will you lose your total income if only thresholds after rounds 2 and 6 have been reached?

   ......

   (c) What is your total income (in ECU), if all thresholds have been reached?

   ......

Surprise restart (Instructions):
We are repeating this experiment once. You are once more assigned to a group of five participants, which will remain unchanged for the six rounds. Because of the high number of participants, it is unlikely that you will be assigned to the same group of five participants with the same group members.