A. Online Appendix

Table A.1: Pairwise robust rank-order test (average *contribution* per group is an independent observation)

	"No P" (14 obs.)	"Immediate P" (16 obs.)	"Delayed P1" (21 obs.)	"Delayed P2" (20 obs.)	"Delayed P3" (19 obs.)	**Average over periods**
"No P"		-6.62***	-7.71***	-4.23***	-2.49**	**1-15**
(14 obs.)		-6.28***	-6.51***	-3.97***	-2.50**	**10-15**
"Immediate P"			0.49	0.98	1.58*	**1-15**
(16 obs)			0.25	0.87	1.41*	**10-15**
"Delayed P1"				0.64	1.38*	**1-15**
(21 obs.)				0.71	1.16	**10-15**
"Delayed P2"					0.71	**1-15**
(20 obs.)					0.52	**10-15**

The values are the robust rank-order test statistic Ù. *, **, and *** denote $0.5 < p \leq 0.10$, $0.01 < p \leq 0.05$, and $p \leq 0.01$, respectively (using simulated small sample values - one sided tests - see Feltovich, 2005).

Table A.2: Pairwise robust rank-order test (average *punishment* per group is an independent observation)

	"Immediate P" (16 obs.)	"Delayed P1" (21 obs.)	"Delayed P2" (20 obs.)	"Delayed P3" (19 obs.)	**Average over periods**
"Immediate P" (16 obs.)		-0.72	1.12	-0.59	**1-15**
		-1.16	-1.20	-2.01**	**10-15**
"Delayed P1" (21 obs.)			1.89**	0.04	**1-15**
			0.61	-0.90	**10-15**
"Delayed P2" (20 obs.)				-1.56*	**1-15**
				-1.74**	**10-15**

The values are the robust rank-order test statistic Ù. *, **, and *** denote $0.5 < p \leq 0.10$, $0.01 < p \leq 0.05$, and $p \leq 0.01$, respectively (using simulated small sample values- one sided tests - see Feltovich, 2005).

.

Table A.3: Pairwise robust rank-order test (average *payoff* per group is an independent observation)

	"No P" (14 obs.)	"Immediate P" (16 obs.)	"Delayed P1" (21 obs.)	"Delayed P2" (20 obs.)	"Delayed P3" (19 obs.)	**Average over periods**
"No P"		-2.07**	-1.85**	-1.82**	-0.18	**1-20**
(14 obs.)		-3.06***	-2.95***	-2.14**	-0.60	**10-15**
"Immediate P"			0.50	0.25	1.63*	**1-20**
(16 obs.)			1.05	1.39*	2.15**	**10-15**
"Delayed P1"				-0.22	1.20	**1-20**
(21 obs.)				0.22	1.13	**10-15**
"Delayed P2"					1.26	**1-20**
(20 obs.)					1.02	**10-15**

The values are the robust rank-order test statistic Ù. *, **, and *** denote $0.5<p\leq0.10$, $0.01<p\leq0.05$, and $p\leq0.01$, respectively (using simulated small sample values - one sided tests - see Feltovich, 2005).

.

Table A.4: Tobit model estimations (prosocial and antisocial punishment)

Treatments	Coef. (Std. Err.) Prosocial punishment	Coef. (Std. Err.) Antisocial punishment
"Delayed P1"	0.33	0.18
	(0.29)	(0.64)
"Delayed P2"	-0.40	-2.02***
	(0.33)	(0.77)
"Delayed P3"	0.58*	1.18*
	(0.33)	(0.66)
Negative absolute deviation	0.49***	-
	(0.03)	
Negative absolute deviation x "Delayed P1"	-0.05	-
	(0.04)	
Negative absolute deviation x "Delayed P2"	-0.13***	-
	(0.05)	
Negative absolute deviation x "Delayed P3"	-0.04	-
	(0.06)	
Positive deviation	-	0.42***
		(0.08)
Positive deviation x "Delayed P1"	-	0.13
		(0.10)
Positive deviation x "Delayed P2"	-	-0.15
		(0.11)
Positive deviation x "Delayed P3"	-	0.02
		(0.10)
Constant	-5.04***	-5.61***
	(0.83)	(1.82)
Age, Man, Economist, Experience dummies	Yes	Yes
Periods	1-15	1-15
(Period dummies)	(Yes)	(Yes)
Observations	13680	13680
Uncensored Obs	1054	267
Pseudo-R²	0.24	0.21

Dep. variable "Prosocial punishment" ("antisocial punishment)" is defined as *the number of points received in period t when the individual contributed less (more) than his peer*. "Negative absolute deviation" ("Positive deviation") is defined as the *negative (positive) difference between an individual contribution and the other's contribution*. Note that for a prosocial (antisocial) punishment "Negative absolute deviation" ("Positive deviation") must be larger than 0. The estimation includes observations from Periods 1-15 only (3 observations per individual per period). Finally,*, **, and *** denote 0.5<p≤0.10, 0.01<p≤0.05, and p≤0.01. Standard errors are clustered across individuals.