Are Smarter Groups More Cooperative?
Results for Corrected and Extended Datasets

Garett Jones
September 2013

Abstract

This short note summarizes corrections and extensions to my paper, "Are Smarter Groups More Cooperative? Evidence from Repeated Prisoner's Dilemma Experiments, 1959-2003," published in the *Journal of Economic Behavior and Organization* in 2008. That paper, a metastudy of published prisoner's dilemma experiments run at various U.S. universities, reported that students at schools with higher average test scores tended to cooperate more often. Results are reported here both for a corrected original dataset and for an extended dataset that includes all relevant observations from the meta-analysis of Balliet, Li, Macfarlan, Vugt (*Psych Bulletin,* 2011). In both datasets, experiments run at schools with higher average test scores tend to have statistically and economically significantly higher rates of cooperation, reinforcing the key result of the original paper.

I. Introduction

This note presents results from a corrected and extended dataset for my 2008 metastudy, "Are Smarter Groups More Cooperative?" As in the original paper, the database draws on previously published journal articles that report the results of repeated prisoner's dilemma experiments run at U.S. universities. Some of these journal articles include multiple studies run at the same school or different schools. The "study" or "experiment" as defined by a paper's author(s) is hence the unit of observation. The revised results document a positive relationship between cooperation rates in a given study and the average SAT or ACT score at the universities where the study was performed: On average, smarter groups appear to be more cooperative.

The original dataset recorded some rates of cooperation incorrectly, sometimes combining rates of individual cooperation (plays of "C") and rates of joint cooperation (plays of "C-C") along with errors of interpretation. However, the effects of these errors were mitigated by the positive correlation between correct and incorrect measures. This revision also corrects other, less common errors in control variables and test score measures where necessary, and omits studies that did not meet the original inclusion criteria. In addition, this revision includes data from all applicable studies in the most recent metastudy of social dilemma games, Balliet et al. (2011).

The test scores used in this update are the original 1966 and 1970 SAT scores as well as 2013 ACT and SAT test scores. The 2013 measures replace the earlier 2006 measures which came from online sources that no longer exist in many cases. The earlier test scores, though available for fewer schools, are more representative: Most of the underlying papers in the sample were published in the 60's and early 70's, and some schools had large

changes in their relative ranking during this period, as noted in the original paper. I provide a summary of the revised results below.

II. Corrected Original Dataset

Since rates of "joint" cooperation (plays of C-C) were reported for few of the original studies, all results below use rates of individual cooperation (i.e., rates of an individual pressing "Cooperate" rather than "Defect") as the dependent variable. In the eight cases where the two cooperation measures are both available (11 in the extended sample) the Pearson correlation is greater than 0.95. Table 1 reports summary statistics for the original dataset, the corrected original dataset, and the extended dataset described below.

In the 2008 paper, some studies were included in the sample that did not meet the inclusion criteria: The games were not repeated prisoner's dilemmas with two players knowingly playing each other. Instead, players were unknowingly paired with a computer program or a confederate or played multiple one-shot games (simultaneously or against new partners) rather than a true repeated game. One paper was a small metastudy that overlapped with an included study. Finally, in some studies the text of the paper did not explicitly identify the university where the students were enrolled (although it did note the affiliation of the authors). In regressions that attribute these unidentified students to the author's institution statistical significance is comparable.

I have dropped all of these studies from the corrected dataset. These results also omit observations used in the original paper where only median rates of cooperation were reported and one study (Lave (1962), run at MIT), where players were permitted to write out a supergame strategy rather than manually play the full 300 trials. Likewise, one inadvertently omitted study that was part of an included paper is now part of the corrected original dataset.

The original maximum sample size was 35 studies, as reported in tables and regressions, and drew on students from 24 schools. The original text instead stated, "The earliest of these 36 studies was in 1959….and only data on these 36 schools are used in the statistics reported below"(Jones, 492). In the corrected original dataset, there are a maximum of 23 studies with students from 19 schools (when a school's recent ACT is the test score).

As Tables 2 and 3 indicate, in this corrected original dataset, Pearson and Spearman correlations between cooperation and test score are higher than before for 7 out of 8 possible correlations, and p-values are less than 1% in 6 of 8. Therefore, the "main results" (p. 492) as stated in Section III of the original paper still hold for the corrected dataset.

III. Extended dataset

The only meta-analysis of social dilemmas more recent than my own is, to my knowledge, Balliet, Li, Macfarlan, Vugt (2011). Including all of their applicable papers returns the sample size closer to its original 2008 value, from a maximum of 30 studies (with students from 24 schools) for 2013 ACT to a minimum of 18 for 1966 SAT. The Balliet et al. meta-analysis includes many of the Jones (2008) papers.

In the extended dataset, results are similar to the original 2008 paper for Pearson and Spearman correlations, as noted in Tables 2 and 3: Correlations are higher than before in 6 of 8 cases, and significant at the 1% level in 5 of 8.

IV. Additional Controls

I turn now to what Section 4 of the 2008 paper denoted robustness tests. I run the same OLS regressions reported in Tables 4 and 5 of the original paper; the numbering of tables is retained here. Table 4 controls for number of trials, a dummy for whether players were able to interact during the game (either verbally or visually), and a dummy for whether there were monetary incentives for good game performance. For the corrected dataset, coefficient magnitudes are comparable to the original paper and have p-values under 5% in two of four cases, while in the extended dataset statistical significance is comparable to the 2008 paper though with slightly larger coefficient magnitudes.

Table 5 adds an additional control, a dummy for whether the school is private: Neither test scores nor the private school dummy are significant in the smaller corrected dataset, although SAT scores have magnitudes comparable to the 2008 paper. In the extended dataset the early SAT coefficients are as significant as in the 2008 paper and are of comparable magnitude. As in the 2008 paper, the private school dummy is never statistically significant.

V. Additional Information

While all test subjects here are university students, it is possible that some test subjects are non-degree or graduate students: While no paper noted those possibilities, studies were split in the extended dataset between thirteen denoting that students were drawn from undergraduate courses and fifteen denoting explicitly that students were all undergraduates. Two additional studies described test subjects only as university students.

Tables 4 and 5 exclude three observations (of which one is in the extended dataset) where it was unclear whether players were separated by a visual barrier; the observations were included in correlation results.

Finally, in all estimates reported here, dummy variables were set equal to 0.5 if a particular study included a combination of a public and a private school; included incentivized and unincentivized experimental subjects; or included students who did and did not interact with each other; 9% of dummies had this 0.5 value. In separate estimates where individually reported treatments were treated as individual observations and dummies were given 0-1 values, results were little changed.

VI. Conclusion

The new results indicate that all major claims of the original paper are supported: On average, students at high-scoring schools cooperated more often in all samples considered here. Particularly in the extended dataset, the results controlling for key observables are comparable to the 2008 paper.

Bibliography

Balliet, Daniel, Norman P. Li, Shane J. Macfarlan, and Mark Van Vugt. "Sex differences in cooperation: A meta-analytic review of social dilemmas." Psychological Bulletin 137, no. 6 (2011): 881.

Jones, Garett. "Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959–2003." Journal of Economic Behavior & Organization 68, no. 3 (2008): 489-497.

Lave, Lester B. "An Empirical Approach to the Prisoners' Dilemma Game." The Quarterly Journal of Economics (1962): 424-436.

Table 1a: Summary Statistics, Original Dataset

	Cooperation	ACT 2006	SAT 2006	SAT 1966	SAT 1970	Private
Mean	43%	27	1268	1151	1140	0.29
Median	39%	28	1273	1140	1145	0
Max	80%	32	1477	1428	1398	1
Min	19%	20	1074	991	955	0
S.D.	15%	3.3	118	143	138	0.44
n	35	35	34	20	20	35

Table 1b: Summary Statistics, Corrected Original Dataset

	Cooperation	ACT 2013	SAT 2013	SAT 1966	SAT 1970	Private
Mean	50%	28	1299	1173	1145	0.39
Median	46%	30	1330	1199	1170	0
Max	80%	33	1475	1345	1337	1
Min	29%	24	1085	991	980	0
S.D.	15.1%	2.7	117	124	118	0.48
n	23	23	22	13	16	23

Table 1c: Summary Statistics, Extended Dataset

	Cooperation	ACT 2013	SAT 2013	SAT 1966	SAT 1970	Private
Mean	49%	28	1281	1156	1128	0.37
Median	44%	29	1325	1164	1132	0
Max	80%	33	1475	1345	1337	1
Min	13%	23.5	1085	991	980	0
S.D.	17.5%	2.8	117	122	113	0.47
n	30	30	29	18	22	30

(Compare to Table 1 of 2008 paper)

Table 2:
Pearson Correlations with Rates of Cooperation

	Jones (2008)	Corrected Dataset	Extended Dataset
Recent ACT	0.47**	0.40	0.32
	[0.004]	[0.062]	[0.085]
n	35	23	30
Recent SAT	0.36*	0.59**	0.45*
	[0.039]	[0.004]	[0.013]
n	34	22	29
SAT 1966	0.67**	0.80**	0.79**
	[0.001]	[0.001]	[0.000]
n	20	13	18
SAT 1970	0.63**	0.73**	0.74**
	[0.003]	[0.001]	[0.000]
n	20	16	22
Private	0.36*	0.48*	0.28
	[0.036]	[0.022]	[0.136]
n	35	23	30

Note: Correlations, p-values, and sample sizes are reported, in that order.

*=5%, **=1% level of significance.

(Compare to Table 2 of 2008 paper)

Table 3:
Spearman Correlations with Rates of Cooperation

	Jones (2008)	Corrected Dataset	Extended Dataset
Recent ACT	0.42*	0.43*	0.39*
	[0.012]	[0.039]	[0.034]
n	35	23	30
Recent SAT	0.27	0.60**	0.51**
	[0.123]	[0.003]	[0.005]
n	34	22	29
SAT 1966	0.57**	0.80**	0.76**
	[0.009]	[0.001]	[0.000]
n	20	13	18
SAT 1970	0.61**	0.72**	0.75**
	[0.004]	[0.002]	[0.000]
n	20	16	22
Private	0.27	0.42**	0.30
	[0.121]	[0.046]	[0.112]
n	35	23	30

Note: Correlations, p-values, and sample sizes are reported, in that order. Jones (2008) values are re-estimates, with correlations here between 0.01 and 0.05 higher than published results: Published 2008 results were calculated with a Spearman formula that did not take account of the possibilities of ties in rank.

=5%, *=1% level of significance.

(Compare to Appendix of 2008 Paper)

Table 4:
Test scores as a predictor of cooperation, controlling for number of trials, along with monetary incentive and interaction dummies

Regression coefficients:

	Jones (2008)	Corrected Dataset	Extended Dataset
Recent ACT	2.5%**	2.1%	2.9%*
	[0.004]	[0.120]	[0.018]
n	31	20	26
Recent SAT	5.6%**	8.1%**	9.2%**
	[0.016]	[0.015]	[0.002]
n	30	19	25
SAT 1966	7.2%**	7.2%*	10.9%**
	[0.005]	[0.045]	[0.002]
n	18	11	15
SAT 1970	7.3%**	7.1%	11.4%**
	[0.005]	[0.061]	[0.001]
n	18	14	19

Note: Regression coefficients, p-values, and sample sizes are reported for each regression, in that order. ACT coefficient is measured per ACT point; SAT coefficients measured per 100 SAT points.

*=5%, **=1% level of significance.

(Compare to Table 4 of 2008 paper)

Table 5:
Test scores as a predictor of cooperation, controlling for number of trials, along with monetary incentive, interaction, and private school dummies

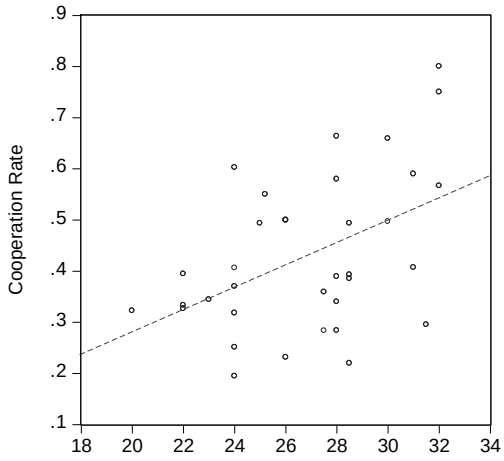
Regression coefficients:

	Jones (2008)	Corrected Dataset	Extended Dataset
Recent ACT	2.6%*	-0.3%	1.0%
	[0.022]	[0.864]	[0.544]
Private=1	*-0.5%*	*15.7%*	*13.8%*
	0.950	*0.109*	*0.139*
n	31	20	26
Recent SAT	5.7%	3.7%	5.4%
	[0.085]	[0.445]	[0.229]
Private=1	*-0.5%*	*11.3%*	*10.4%*
	[0.950]	*[0.242]*	*[0.293]*
n	30	19	25
SAT 1966	6.2%*	5.6%	9.4%*
	[0.020]	[0.101]	[0.012]
Private=1	*6.4%*	*8.8%*	*7.4%*
	[0.331]	*[0.207]*	*[0.335]*
n	18	11	15
SAT 1970	6.2%*	4.9%	8.9%*
	[0.019]	[0.176]	[0.015]
Private=1	*8.0%*	*10.8%*	*10.2%*
	[0.219]	*[0.133]*	*[0.165]*
n	18	14	19

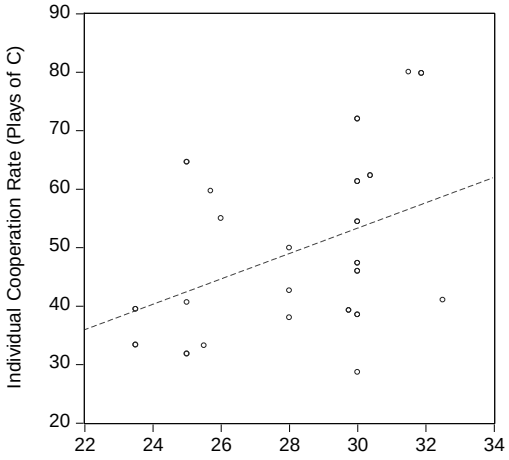
Note: Regression coefficients, p-values, and sample sizes are reported for each regression. ACT coefficient is measured in ACT points; SAT coefficients measured per 100 SAT points. Private school dummy for 2008 value in Recent SAT regression is a revised re-estimate, coincidentally identical to the 2008 value of Recent ACT dummy. Published 2008 value for the Recent SAT dummy was +5.6% (p=51.1%).
*=5%, **=1% level of significance.

(Compare to Table 5 of 2008 paper)

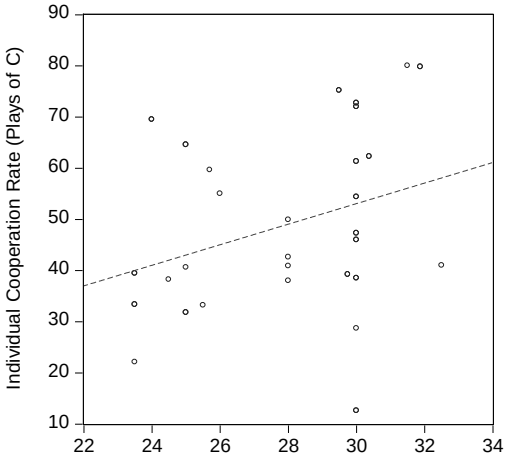
ACT 2013 and Rates of Cooperation



ACT 2006: Original Dataset

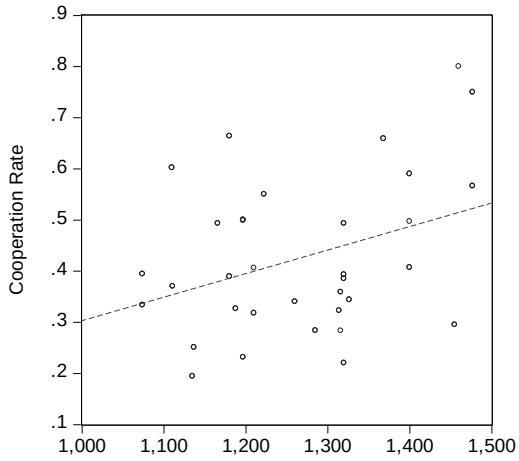


ACT 2013: Corrected Dataset

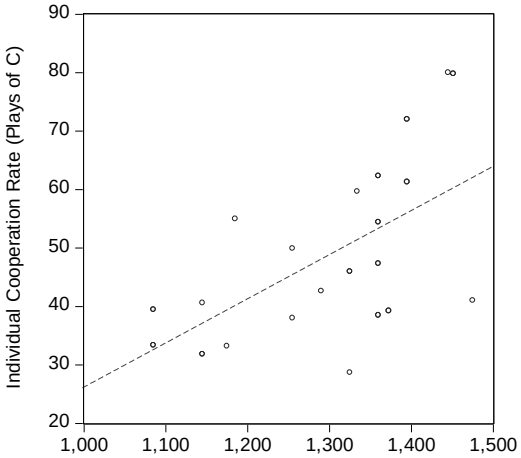


ACT 2013: Extended Dataset

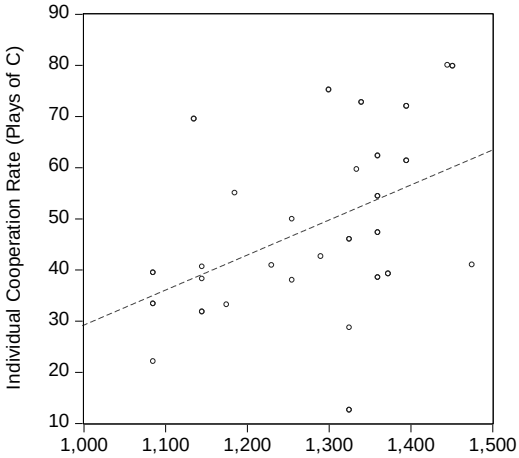
SAT 2013 and Rates of Cooperation



SAT 2006: Original Dataset

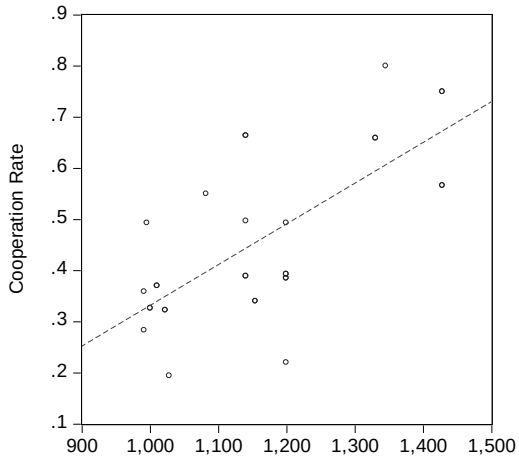


SAT 2013: Corrected Dataset

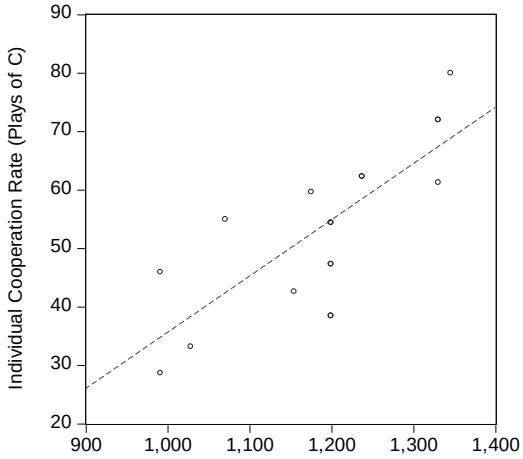


SAT 2013: Extended Dataset

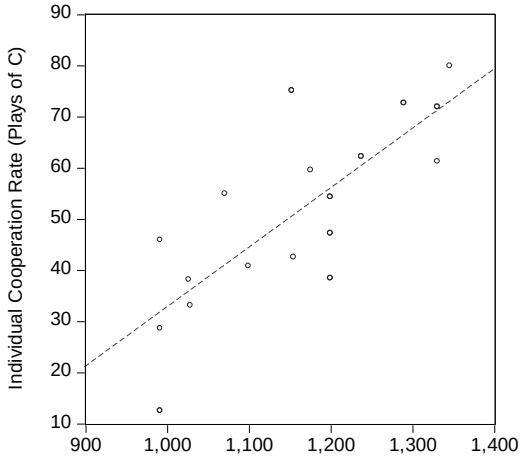
SAT 1966 and Rates of Cooperation



SAT 1966: Original Dataset

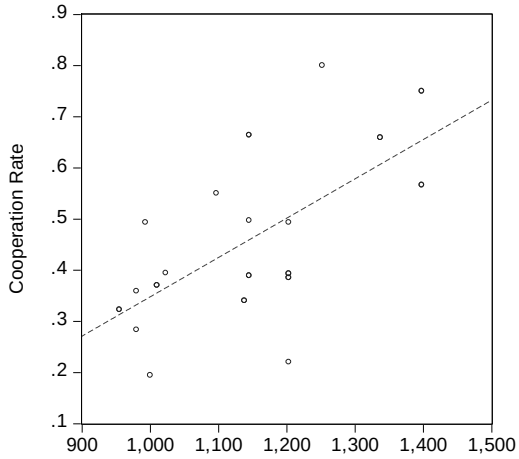


SAT 1966: Corrected Dataset

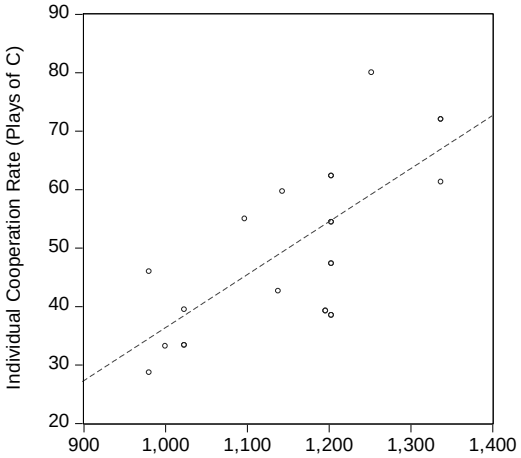


SAT 1966: Extended Dataset

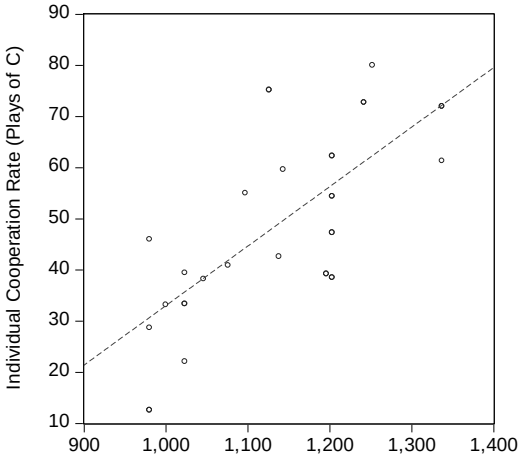
SAT 1970 and Rates of Cooperation



SAT 1970: Original Dataset



SAT 1970: Corrected Dataset



SAT 1970: Extended Dataset