Dickinson, David L. (2009) "Experiment Timing and Preferences for Fairness." **Journal of Socio-Economics**, 38: 89-95. Published by Elsevier (ISSN: 1053-5357)
doi:10.1016/j.socec.2008.05.010

Experiment Timing and Preferences for Fairness

David L. Dickinson

ABSTRACT

Classroom experiments examining fairness preferences [Andreoni, J., Miller, J., 2002. Giving according to GARP: an experimental test of the consistency of preferences for Altruism. Econometrica 70 (2), 737–753] were conducted to examine two issues: first, are classroom points a salient reward medium (comparable to cash in research experiments)? Secondly, does experiment timing during the semester influence results. Subject choices are consistent with the existence of well-behaved utility functions, indicating that points experiments can be valid. Secondly, subjects are more likely to be "selfish" when the experiment is conducted early rather than late in the academic semester. This result has behavioral implications for environments where nonmonetary incentives prevail, as well as implications for the growing number of instructors using experiments and follow-up discussion in the classroom.

ARTICLE

1. Introduction

Experimental economics has proven quite useful as a pedagogical tool. Not only can classroom experiments be enjoyable for both students and instructors, but there is also evidence indicating that experiments can improve student comprehension of important economics concepts ([Frank, 1997], [Emerson and Taylor, 2004] and [Durham et al., 2007]). Many instructors who conduct classroom experiments use class points (or extra credit points) as the reward medium,1 and these instructors often discuss related results from the experimental economics literature. It is therefore important to know if predictable differences exist between comparable research (i.e., cash) and classroom (i.e., extra credit) experiments, and such differences are relatively understudied.

This paper's focus on reward medium effects in experimental outcomes has important behavioral implications. First, if the reward medium matters, then classroom experiments may provide valuable insights into naturally occurring decision environments where incentives are primarily nonmonetary. An example of a real-world allocation environment where fairness preferences matter, as with the experiments reported in this paper, might be an employee allocating credit or blame on a work project completed with a colleague. It may be quite important to explore whether any behavioral differences in credit/blame allocations arise if workplace rewards are primarily monetary (e.g., bonuses) or nonmonetary (e.g., recognition).[2] Secondly, for pedagogical purposes classroom experiments are often followed by discussion of related experimental research results. Significant reward medium effects merit some discussion for proper understanding of the classroom results.[3]

A second and unique variable of interest in this paper is the timing of the experiment over the course of the academic session. It would seem that there has been no discussion in the experimental economics or educational community as to whether it matters to conduct the same experiment early versus late in the semester, for example. To the extent that any sort of experiment timing effects have been considered, it has only been with respect to how much economics training a subject has had at the time of the experiment (i.e., timing within one's academic career). For example, Frank et al. (1993) find evidence that economics majors behave in a more self-interested fashion in the ultimatum bargaining game. They also do not trend towards more cooperative behavior, like other college students, as they progress towards graduation. This has led some to conclude that the study of economics “indoctrinates” subjects to behave more in accord with theories of purely self-interested behavior. However, Frey and Meier (2005) utilize a more natural experiment to identify a significant self-selection effect into economics studies, which casts doubt on the idea that the study of economics, per se, results in more self-interested experiment decisions.[4]

The hypothesized indoctrination effect is not a key concern of the present experiments, however, given that the experiment timing I examine is the timing within a semester – about 11 weeks apart – as opposed to across semesters or years of one's academic career. Nevertheless, an experiment's timing during the academic session is perhaps not an irrelevant detail. For example, if one conducts an experiment early in the semester, the potential value of a given number of class points may not be well known. Late in the semester, however, it becomes more certain how valuable a given amount of extra points may be. In effect, payoff uncertainty may be present if conducting a “points” experiment early in the session, and so one might expect risk-averse students to behave more selfishly over classroom points. Back to the behavioral implications of the workplace example, if the value of a “recognition” award is largely uncertain – perhaps your company does not have a history of using such awards – then risk-averse workers may behave more selfishly in their attempts to gain such awards. A manipulation of classroom experiment timing allows one to creatively explore such effects.

From a pedagogical standpoint, this article does not argue that one should conduct experiments only at one particular point in the semester. Rather, the secondary objective of this paper is to highlight how experiment timing during the semester can significantly alter experimental results. Students should be made aware of how experiment timing may alter results, and this can be exploited as an opportunity to compare/contrast experimental results.

The purpose of this paper is not to provide an exhaustive comparison of lab experiments and comparable classroom experiments. Rather, this paper presents the results from classroom experiments conducted using the well-known Andreoni and Miller (2002) design to examine preferences for altruism. Because their experiment examines a general issue of “fairness” preferences, it not only has significant behavioral appeal, but the topic can also be of significant interest in a typical economics principles class where students often think that economists assume only money matters. Of course, economists realize that this is not true, as evidenced by the often-mentioned trade-off between equity and efficiency in microeconomics principles. An experiment about fairness issues can, however, be quite useful for highlighting recent findings from economics research.5

Andreoni and Miller's (2002) cash reward experiments are replicated for comparison to classroom data generated using extra credit points as the reward medium. The classroom experiment is then conducted either early or late in the semester to examine experiment timing. Subject rationality (as determined by the data's consistency with the weak axiom of revealed preference, or WARP) is explored and found not to be significantly different in the classroom credit experiment compared to the cash experiments—this indicates that points constitute a salient reward medium. The cash experiment replication generates data showing similar proportions of selfish subjects, on average, to what Andreoni and Miller (2002) report—about 43% of our data are best described by a selfish-type utility function compared to 47% of their subjects (not statistically different).6 However, the classroom experiment results significantly differ based on experiment timing. Early in the semester the experiment generates 50% selfish subjects, compared to only 26% selfish subjects when the exact same experiment is conducted late in the semester (in different classes). Taken together, these data provide both important evidence in support of subject rationality in nonmonetary reward experiments and unique evidence that experiment timing can affect classroom experiment results.

2. The reward medium for classroom experiments

Most discussion of the reward medium in classroom experiments focuses on the proper or fair way to provide salient incentives. Some teachers offer a small cash incentive to students in their classroom experiments. For example, one might offer a small cash payoff to the high payoff outcome in the experiment, or one might offer some cash incentives by paying all subjects a small fraction of their experimental earnings (e.g., Holt, 1999). A natural alternative for those who desire a non-hypothetical decision environment is to offer some type of class points payoff. Considerable concern has been expressed over the "fairness" of affecting grades with experiments. Commentary in Fels (1993), and Williams and Walker (1993) leans towards continued use of extra credit points if their maximum effect on a student's grade remains relatively small (e.g., no more than one full letter grade). Some experiments contain a random element to outcomes, and there are those who find this inherently unfair (e.g., Stodder, 1998), while others find it quite parallel to how outcomes are sometime determined in the "real world" (e.g., Bell, 1993). Though I do not necessarily recommend it, offering regular class points based on experiment outcomes is another alternative. Depending on the instructor, the use of regular points versus extra credit points may have little practical effect on the final grading scale, but the idea of having regular points determined from outcomes is seen as less procedurally fair by students.7 It is important to note that my use of extra credit points for the experiments reported here was such that subject payoffs from the experiments could only improve one's grade. As such, even a zero payoff from the extra credit points experiment would not lower a student's grade below that which followed from performance on other course requirements.

The debate surrounding the use of class points will not abate, and it is not my purpose to convert anyone to the use of extra credit points as the reward medium for class experiments. But, because there is a large community of instructors who do use class points as a reward, the first question for this paper is whether or not the reward medium matters in terms of experimental outcomes. Presumably, the answer to this question is of interest both to those who actively practice the use of class points experiments as well as those who have a general interest in the potential significance of the chosen reward medium for an experiment. Isaac et al. (1994) conclude that results from their large-group public goods experiments are similar using class points or money incentives. However, control over this variable was not present in their study, given that the reward medium effect was not the objective of their research. Li (1991), in an unpublished study, conducts a controlled comparison of the reward medium effect of cash versus extra credit points in public goods experiments and finds no significant difference in the data on several outcome measures.

The limited amount of existing research on this issue should not be taken to imply that the reward medium used in an experiment does not matter. The present results are important in that they indicate that the timing of the experiment can confound the measurement of a reward medium effect. Indeed, the data show that, in the Andreoni and Miller (2002) experiments that I replicate, one set of classroom points data is statistically no different from the cash experiment replication, while the other generates significantly different levels of altruism (or selfishness) from the subjects.

3. The experiments

The design is motivated by the classical dictator game (Forsythe et al., 1988). Two subjects are paired, and one subject is allowed to divide a monetary pie between the two in any way he/she likes—the counterpart has no choice but to accept the allocation. Removing the opportunity for the counterpart to respond by rejecting the offer is meant to remove fear of rejection from the allocation decisions so that what remains is a more accurate measure of the subject's preference towards fairness. Andreoni and Miller, 2002 J. Andreoni and J. Miller, Giving according to GARP: an experimental test of the consistency of preferences for altruism, Econometrica 70 (2) (2002), pp. 737–753. Full Text via CrossRef | View Record in Scopus | Cited By in Scopus (168)Andreoni and Miller (2002) expand on the dictator game design by varying the “price” of giving away $1.00 from $.25 to $4.00. For example, if the price of giving away $1 is $2, this means that forgoing each $1 of payoff for yourself generates $.50 of payoff for your counterpart (in the prototypical dictator game, the price of giving away $1 is simply $1 less for yourself). Subjects make a series of 11 token allocation decisions designed to alter the price of giving (see Appendix A for the allocation decisions of the cash replication experiment. An additional appendix containing all instructions is available upon request). A standard dictator game is one in which the slope of the budget constraint in payoff-to-self/payoff-to-other space is negative one. The series of 11 token decisions, where each allocation decision is based on a distinct budget constraint in the own-payoff/other-payoff space, represent a mix of standard dictator games as well as allocation environments where the price of giving is greater or less than 1. Because the subject choice problem can be described in terms of budget-constrained choice bundles, revealed preference axioms can be used to examine consistency of preferences (i.e., rationality of choice). The game is simple to understand, yet foundational to many more complicated real-world decision environments. Fig. 1 gives an example of 3 of the 11 budget constraints from the Andreoni and Miller (2002) design used in the present experiments.

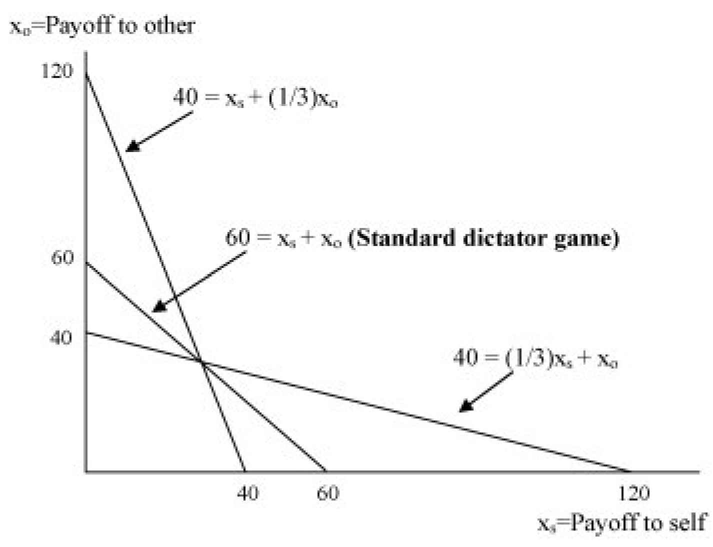

Fig. 1. Sample budget constraints for allocation experiment.

After making all 11 allocation decisions, each subject's decision sheet is randomly matched with those of two other anonymous subjects. During the first pairing of decision sheets, one subject has a randomly selected allocation decision chosen, and the subject earns the amount determined by that subject's own "hold" decision. The earnings from the first pairing are placed into the subject's payoff envelope by an experimenter. The decision sheets are then re-matched such that each subject is matched with a different subject, and this time those subjects who had earnings determined by their own randomly chosen "hold" decision received earnings based on the another subject's "pass" decision for another randomly chosen allocation decision (and vice versa). In this way, each subject had one payoff determined by one of their own "hold" decisions, and one payoff determined by someone else's "pass" decision. This was common knowledge, as was the fact that no subject would be matched twice with the same person, and that decisions would remain double-blind anonymous.8 Decision sheet pairings, random allocation choices, and envelope stuffing were always supervised by a compensated volunteer subject who did not take part in the decision-making portion of the experiment. The cash experiments did not involve students from my classes, and a given experiment group in the cash experiment was not a group of student classmates as in the points experiments.

The points experiment is identical to the procedures of the Cash experiment, with the exception that payoffs were in terms of class points that would be added to the student's Exam #1 (with no truncation at 100%). Whereas the average payoff in the cash experiments was about $15 (for a 30 min experiment), the average payoff in the points experiments was 9 class points. This amount of points was nearly a full grade on the 100-point Exam #1, and in total this average payoff amounted to 2% of the final class points amount (450). With the plus-minus grading system, this implies that the average amount of class points paid out for this experiment was small, but non-trivial.9 Over the course of a 16-week semester, this class points experiment was conducted in some classes in week 3 of the semester (the Early Points treatment), while in distinct classes it was conducted in week 14 (the Late Points treatment). The students had not yet completed Exam #1 in week 3, whereas all graded items except for the final exam had been completed in week 14 of the semester. All points experiments were conducted during regularly scheduled class sessions of my microeconomic principles classes – a required business class with varied student representation – and I personally conducted all experiments. To the extent that "experiment days" were preannounced, without revealing the topic or experiment details, participation is voluntary if students choose not to come. They were, however, aware that this experiment day would afford them the opportunity to earn some extra credit points.

4. Results

Two outcome measures are reported in order to compare the points experiments with the cash experiments, as well as for comparison with Andreoni and Miller (2002), who first focus on whether preferences for fairness are "rational". So, subject rationality is initially examined in Table 1. The data are analyzed in terms of their consistency with the WARP, consistency with which implies that the subjects' revealed preferences could be rationalized by some well-behaved utility function.10 The data in Andreoni and Miller (2002) are examined for their consistency with other revealed preference axioms, but the conclusions are not significantly altered if one focuses only on WARP violations. Importantly, there is no reason that data would abide by WARP if the classroom points incentives are not salient to the subjects, and so subject consistency with WARP will help indicate legitimacy of the nonmonetary incentives. The data in Table 1 report the number of violations of WARP at Afriat's (1972) Critical Cost Efficiency Index (CCEI) of .95. The CCEI is a measure of how much a budget constraint would have to shrink to avoid the preference axiom violation, and so the threshold CCEI of .95 allows for the small amount of decision error of the magnitude suggested by Varian (1991) and reported in the Andreoni and Miller (2002) study.

Table 1. Violations of WARP (violations reported for CCEI < .95)

	Experiment treatment		
	***Cash* (*N* = 30 subjects)**	***Class Points Early* (*N* = 80 subjects)**	***Class Points Late* (*N* = 42 subjects)**
No. of subjects with WARP violation(s)	5 (17%)	12 (15%)	5 (12%)
Average no. of violations (at CCEI = .95 level) per subject with a violation	1.20	1.42	1.40
% subjects for which the data are rationalizable with a quasi-concave utility function	83	85	88

The data in Table 1 indicate that subjects, in all treatments, are mostly rational in the sense that choices are consistent with revealed preference theory. Though somewhat more subject choices violate WARP in the present data (12–17%) than in Andreoni and Miller data (1.7%), it is still the case that the vast majority of subject choices are consistent with WARP. Rationality is not affected in the class points experiments by early versus late experiment timing ($p > .10$ for the binomial test).[11] There are no significant differences in subject rationality in comparing the *Cash* treatment with either of the class points treatments ($p > .10$). In sum, while the present data show relatively more frequent violations of rationality than in Andreoni and Miller (2002), choices are still over 80% rational across all treatments, with no significant differences across our *Cash* and *Points* treatments. In short, these data support the notion that classroom points can be a salient and valid reward medium for economics experiments.

For further comparison to Andreoni and Miller, preferences are categorized. Data consistent with WARP (i.e., "rational" subjects) merely imply that revealed preferences are justifiable by some well-behaved utility function. Andreoni and Miller examine three well-behaved preference types that are of particular interest: *Homo economicus* subjects (selfish), Rawlsian subjects (i.e., Leontief preferences), or Utilitarian subjects (i.e., perfect substitute preferences). Table 2 shows the results from a categorization of preferences for self-payoff, x_s, and other-payoff, x_o, into one of these three types: selfish type ($U = x_s$), Leontief preferences ($U = \min\{x_s, x_o\}$), or Utilitarian preferences ($U = x_s + x_o$). While the data can certainly be examined for consistency with other well-behaved utility functions, these three offer obvious benefits. Their variety includes preference types often considered in undergraduate economics class (at least in intermediate and advanced microeconomics courses). And, the distinct predictions of each of these utility functions makes it easy to examine differences across experiments.

Table 2. Categorizing subject preferences (percentage of subjects fitting weak and strong forms of preference type)

	Experiment Treatment (Dictator budget sets 6, 7, and 9 removed)		
	Cash Weak (Strong)	*Class Points Early* Weak (Strong)	*Class Points Late* Weak (Strong)
Selfish	30% (13%)	41% (9%)	14% (12%)
Leontief	47% (3%)	41% (3%)	43% (14%)
Perfect Substitutes	3% (3%)	3% (4%)	17% (0%)
TOTALS (weak or strong)			
Selfish	43%	50%	26%
Leontief	50%	44%	57%
Perfect Substitutes	67%	6%	17%
	Dictator Offers (only budget sets 6, 7, and 9)		
Average Dictator Offer	27%	28%	38%

Andreoni and Miller examine both a strong and weak preference classification for each category, depending on whether choices exactly match those required by their respective utility function. A weak preference classification simply means that subject choices are closer to those from one of the three utility functions than to either of the other two. For example, a subject facing the budget constraint 40 = xs + (1/3)xo who passes 30 tokens (and keeps 10) most closely matches the Utilitarian subject type for that decision. A strong Utilitarian type would pass all 40 tokens, a Leontief subject would pass 10 tokens, and a selfish subject would pass none. In what follows, I examine choices that do not exactly match these prototypical preferences by calculating the minimum sum of squared deviations of the subject's tokens “held” amount for all allocation decisions from the amounts that one would hold for either of these three preference types.12 The preference type that minimizes this sum of squared deviations is the subject's “weak” preference type. This comparison has the benefit of categorizing all subjects into one of these three categories, but it has the drawback of not differentiating between cases where a subject is relatively closer to a given preference type. The portion of subjects exactly fitting one of these three utility functions (i.e., the "strong" preference types, 16–26%) is somewhat less than reported in the Andreoni and Miller, 2002 J. Andreoni and J. Miller, Giving according to GARP: an experimental test of the consistency of preferences for altruism, Econometrica 70 (2) (2002), pp. 737–753. Full Text via CrossRef | View Record in Scopus | Cited By in Scopus (168)Andreoni and Miller (2002) data (43%).

The results from strong or weak preference type categorizations, which are shaded in Table 2, highlight a main result from of this paper. A large majority of the subjects are classified into selfish or Leontief preference types. Subjects are somewhat less selfish in the Cash experiment than what Andreoni and Miller (2002) report, but the difference is not statistically significant ($p > .10$ for the binomial test of whether our proportion of selfish subjects matches theirs (47%)).

In examining the class points experiments, the data show that subjects are somewhat more selfish, less Rawlsian (i.e., Leontief preferences), and less Utilitarian (i.e., substitute preference) for early-in-semester experiments relative to late-in-semester experiments. That is, the distribution of preference types (weak or strong) for the classroom experiment is more similar to the Cash experiment when conducted early in the semester, but quite different when the points experiment is conducted late in the semester. This result is consistent with what one would conclude from examining the subset of class points data from the more familiar dictator game budget sets (budget sets 6, 7, and 9). Dictator offers are significantly higher late in the semester versus early in the semester for the points experiments ($p = .00$ for the two-sample test of mean differences in dictator offers). Of course, there is no theoretical reason that subjects should be more selfish early versus late in the semester in these points experiments.

Thus, there are two main results: first, classroom points can be a valid motivator for subjects. Of course, individual subjects will view the reward as more or less salient depending on the value of those points to the subject. This is similar to the issue of wealth effects in cash research experiments, but the reward level is chosen in the present experiments so that virtually every subject can improve his/her final letter grade with the average level of points-payoff in the classroom experiment. The more surprising second result involved experiment timing. Preferences over classroom points change significantly depending on the timing of the experiments during the academic session. So, while points are a salient reward, preferences towards them are not independent of factors that change over the course of the session.

5. Concluding remarks

The main message from these experiments is simple. Though the results from a classroom experiment using extra credit points as the reward medium may replicate those from a cash experiment equivalent, they may also vary significantly given the timing of the class experiment. The reason results may vary can serve to highlight methodological issues and/or behavioral factors in decision-making. Because potentially important variables may differ over the course of a semester for a given student or classroom, there is a loss of control over certain classroom experiment details with respect to timing within the academic session—a methodological issue. However, I do find a regular pattern to outcome differences early versus late in the semester, at least for this particular experiment. One possible explanation for relatively more selfish behavior early in the semester is that the value of a given number of extra credit points is more uncertain early in the semester than late in the semester. So, risk-averse students will seek to maximize own-points in the uncertain reward environment, resulting in more selfish behavior.

An alternative hypothesis is that a given class of students may form some social bonds later in the semester. If this reduction in social distance generates less selfish behavior (see [Hoffman et al., 1996] and [Cox and Deck, 2005]), then one would also expect to see less selfish choices for experiments given late in the semester (and perhaps more Rawlsian or egalitarian preferences). However, some students might develop a feeling of greater isolation in the classroom as the session progresses, and so it is not immediately apparent that students will perceive a reduction in social distance among all classmates as the semester progresses. While it seems that one can argue for more or less social distance among subsets of students within a given classroom, the present design cannot properly evaluate the social distance hypothesis.13

In terms of real-world behavioral implications of these hypotheses, a workplace supervisor may exercise some control over the collegiality (i.e., social distance) among her workers, which may combat the uncertain payoffs of a new recognition award system (e.g., employee of the month). Or, an employer may go to extra lengths to clearly communicate the value of such awards to the company so that nonmonetary payoff value is more well-known. The present experiment results suggest that either practice may limit the extent of selfish behavior among workers. So, while additional research would be needed to determine whether the mechanism underlying this classroom timing result is one of uncertainty or social distance (or some other mechanism), this paper is the first to my knowledge to identify this intriguing result.

One potential confound in the classroom data is that students may feel that grades are based on a relative scale (though no grading curve was ever suggested to the students in these classes). If students perceive that their relative standing will affect their grades, then they may be more points-selfish when this perception is strongest, and this would also have behavioral implications in such areas as workplace rewards.14

One might criticize the present experimental design because Cash experiments were not conducted at two points (early versus late) in the semester. While early- and late-Cash treatments may appear to complete a 2 × 2 experimental design matrix, it is unclear what cash early versus late in the semester means. Perhaps a more analogous comparison would be cash early versus late in the month. Expense uncertainties may differ across a given month if one's paycheck of income flow is on a typical monthly cycle. This offers a potentially interesting avenue for future research, but one that was considered tangential to the present study given that the connection between semester points cycles and monthly expense cycles is somewhat speculative—it nevertheless suggests that future research could explore, for example, whether time together as a “group” engenders less selfish behavior within a group (even when decisions remain anonymous). The main objective of the Cash experiment in this paper was to replicate the Andreoni and Miller (2002) design and provide a baseline for comparison. Perhaps experimental economists should examine the potential effects of administering experiments at different points in one's paycheck cycle, but the present examination of experiment timing across a semester had immediate implications for the growing number of economists conducting classroom experiments.

Finally, it is also the case that no controls were implemented to prevent students who already took the Class Points Early experiment from talking to those who later took the experiment (though the same could be said of most experiments using a college student subject pool). If such communication occurs, the preferences may converge upon the early-player majority if subjects tend to follow advice from previous players (one sees such results in Schotter (2003), and Schotter and Sopher (2007)). In the present experiments, however, this concern should be minimal because different “generations” of players span several academic semesters. That is, early-points experiments were conducted in one semester, and late-points in a different semester, which decreases the likelihood that such communication occurred to any significant degree, if at all.

Overall, these results are important to those who question the viability of classroom points as a reward medium in experiments—subjects exhibit rational preferences over an important behavioral concept (fairness) when the payoff is in extra credit points. The results also highlight an additional pedagogical point: unanticipated confounds may exist over the course of an academic session, among other things. It would be interesting to extend this exploration of experiment timing effects, both in the classroom and in behavioral research, to a variety of experiments (e.g., individual decision-making, risky choice experiments, trust, etc.) in order to examine whether experiment timing affects outcomes in a variety of decision domains.

References

Afriat, 1972 S. Afriat, Efficiency estimates of production functions, International Economic Review 13 (1972), pp. 568–598.

Andreoni and Miller, 2002 J. Andreoni and J. Miller, Giving according to GARP: an experimental test of the consistency of preferences for altruism, Econometrica 70 (2) (2002), pp. 737–753.

Bell, 1993 C.R. Bell, A noncomputerized version of the Williams and Walker stock market experiment in a finance course, Journal of Economic Education 24 (4) (1993), pp. 317–323.

Cox and Deck, 2005 J.C. Cox and C.A. Deck, On the nature of reciprocal motives, Economic Inquiry 43 (3) (2005), pp. 623–635.

Dickinson, 2002 D.L. Dickinson, A bargaining experiment to motivate discussion on fairness, Journal of Economic Education 33 (2) (2002), pp. 136–151.

Durham et al., 2007 Durham, Y., McKinnon, T., Schulman, C., 2007. Classroom experiments: not just fun and games. Economic Inquiry 45 (1), 162–178.

Emerson and Taylor, 2004 T.L.N. Emerson and B.A. Taylor, Comparing student achievement across experimental and lecture-oriented section of a principles of microeconomics course, Southern Economic Journal 70 (3) (2004), pp. 672–693.

Fels, 1993 R. Fels, This is what I do, and I like it, Journal of Economic Education 24 (4) (1993), pp. 365–370.

Forsythe et al., 1988 R. Forsythe, J.L. Horowitz, N.E. Savin and M. Sefton, Fairness in simple bargaining games, Games and Economic Behavior 6 (1988), pp. 347–369.

Frank, 1997 B. Frank, The impact of classroom experiments on the learning of economics: an empirical investigation, Economic Inquiry 35 (4) (1997), pp. 763–769.

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

Frey and Meier, 2005 B.S. Frey and S. Meier, Selfish and indoctrinated economists?, European Journal of Law and Economics 19 (2005), pp. 165–171.

Hoffman et al., 1996 E. Hoffman, K. McCabe and V. Smith, Social distance and other-regarding behavior in dictator games, American Economic Review 86 (3) (1996), pp. 653–660.

Holt, 1999 C.A. Holt, Teaching economics with classroom experiments: a symposium, Southern Economic Journal 65 (3) (1999), pp. 603–610.

Isaac et al., 1994 R.M. Isaac, J.M. Walker and A.W. Williams, Group size and the voluntary provision of public goods: experimental evidence utilizing large groups, Journal of Public Economics 54 (1) (1994), pp. 1–36.

Li, 1991 Li, Y., 1991. Tests for a reward medium effect on experimental outcomes. University of Arizona Discussion Paper 91-11.

Schotter, 2003 A. Schotter, Decision making with naïve advice, American Economic Review 93 (2) (2003), pp. 196–201.

Schotter and Sopher, 2007 A. Schotter and B. Sopher, Advice and behavior in intergenerational ultimatum games: an experimental approach, Games and Economic Behavior 58 (2) (2007), pp. 364–393.

Stodder, 1998 J. Stodder, Experimental moralities: ethics in classroom experiments, Journal of Economic Education 29 (2) (1998), pp. 127–138.

Varian, 1991 Varian, H.R., 1991. Goodness of fit for revealed preference tests. University of Michigan CREST Working Paper Number 13.

Williams and Walker, 1993 A.W. Williams and J.M. Walker, Computerized laboratory exercises for microeconomics education: three applications motivated by experimental economics, Journal of Economic Education 24 (4) (1993), pp. 291–315.

NOTES

1 Other reward-medium options are to pay the subjects cash (or randomly select one subject to receive a cash payoff based on the experimental outcome), or to just conduct the experiment hypothetically.

2 One might further explore this in the lab by designing an experiment where two subjects work on a joint task, and subjects then anonymously allocate a level of credit/blame for the outcome. The experimenter might manipulate the payoff weights that map allocations to subject-specific final payoffs (cash or recognition payoffs), with subjects having full information on the payoff weights prior to making their allocation decisions.

3 This would be similar to the practice of discussing treatment ordering or learning effects which, while usually controlled by counterbalanced research designs, may confound interpretation of more loosely designed classroom experiments.

4 To go even further, Frey and Meier, 2005 B.S. Frey and S. Meier, Selfish and indoctrinated economists?, European Journal of Law and Economics 19 (2005), pp. 165–171. Full Text via CrossRef | View Record in Scopus | Cited By in Scopus (8)Frey and Meier's (2005) results indicate that certain subsets of economists (e.g., political economists) do not donate less money as a result of studying more economics.

5 Dickinson (2002) outlines a distinct classroom bargaining experiment to motivate the “fairness” discussion.

6 Using the binomial test on whether our subject's have a similar probability of being classified as selfish as in Andreoni and Miller, we fail to reject the null hypothesis of no significant difference ($p > .10$).

7 It was, in fact, during a class points experiment several years ago where I had the unpleasant and unique experience of a student wielding a knife during a class cartel experiment when he discovered that he was “cheated” upon. It is unclear whether the issue was the betrayal of his fellow cartel members, or the fact that such betrayal cost him valuable “regular” class points. Though he quickly claimed he was “just joking”, I (and presumably no one else in class) never again looked at that student the same.

8 A minor alteration was implemented to the double-blind procedure, in which neither subject decisions nor payoffs are known to the experimenter. Because the classroom points experiments could not be

conducted without knowing the final points payoff for each student, subjects in this cash replication were asked to show their total final payoff (but never their decisions) privately to the experimenter for record-keeping. In this manner, decisions were anonymous both to subjects and to the experimenter, but final payoff outcomes were known to the experiment just as in the class points experiments. This procedural difference did not result in significant differences in the proportion of “selfish” subject classifications, but I find relatively less “Utilitarian” subjects and more “Rawlsian” subjects (as discussed later) in the cash replication compared to Andreoni and Miller.

9 Given differences due to rounding of percentages and grades, it is somewhat complicated to determine exactly how many students’ grades were altered by the results of the experiment, but a conservative estimate is that at least 25% of those who took the experiment improved their grades (based on 2% improvement on average, with a range from 0% to 4.5%). More importantly, however, is the subject's belief that their grade might be improved.

10 WARP offers a way to examine revealed choices over goods (in this cases, the two “goods” the subjects chooses are self-payoff and other-payoff), and whether or not choice patterns can be generated from some well-behaved utility function. Basically, a violation of WARP implies that there is no set of (well-behaved) indifference curves that can be drawn for a consumer choice problem such that the actual chosen bundles are both utility maximizing.

11 For these binomial tests, the subject is the unit of observation – the subject either has a WARP violation(s) or not – and the treatment with the higher proportion of rational subjects in the pairwise comparison is used as the probability p for the test. For example, in comparing Class Points Late with Class Points Early, the binomial test null hypothesis is that the percentage of rational subjects in Class Points Early is equal to 88%, which is the percentage of rational subjects in Class Points Late.

12 Budget sets 6, 7, and 9 (the dictator games) are removed from the preference classifications in Table 2 because any allocation of tokens is consistent with substitute preferences.

13 It was noted earlier that any potential indoctrination effect, if it does exist, would likely not be that significant across just 11 additional weeks of a given course. Note that such an effect would be predicted to generate more selfish subjects late in the semester, not less selfish subjects, which only strengthens the present findings.

14 Depending on the course and instructor, students may think that points are more likely to influence a grading curve late in the semester if an instructor unpredictably softens his/her grading stance towards the end of the course. I gave no indication in any of his classes, however, that a grading curve was forthcoming or even a possibility, and so it is unlikely that such perceptions significantly affect these data.

Appendix A.

DECISION SHEET

Directions: Please fill in all the blanks be low. Make sure the number of tokens listed under *Hold* plus the number listed under *Pass* equals the total number of tokens available. Remember, all points are worth $0.10 (10 cents) to all subjects.

1) Divide 75 tokens: *Hold* _____@ 1 point each, and *Pass* _____@ 2 points each.

2) Divide 40 tokens: *Hold* _____@ 1 point each, and *Pass* _____@ 3 points each.

3) Divide 75 tokens: *Hold* _____@ 2 points each, and *Pass* _____@ 1 point each.

4) Divide 60 tokens: *Hold* _____@ 1 point each, and *Pass* _____@ 2 points each.

5) Divide 40 tokens: *Hold* _____@ 3 point each, and *Pass* _____@ 1 point each.

6) Divide 60 tokens: *Hold* _____@ 1 point each, and *Pass* _____@ 1 point each.

7) Divide 100 tokens: *Hold* _____@ 1 point each, and *Pass* _____@ 1 point each.

8) Divide 60 tokens: *Hold* _____@ 2 points each, and *Pass* _____@ 1 point each.

9) Divide 80 tokens: *Hold* _____@ 1 point each, and *Pass* _____@ 1 point each.

10) Divide 40 tokens: *Hold* _____@ 4 points each, and *Pass* _____@ 1 point each.

11) Divide 40 tokens: *Hold* _____@ 1 point each, and *Pass* _____@ 4 points each.

Appendix A - Supplementary Data

ADDITIONAL INSTRUCTIONS APPENDIX (NOT FOR PUBLICATION)

****

INSTRUCTIONS **(CASH TREATMENT)**

Welcome

This is an experiment about decision making. You will be paid in cash for your participation, and the amount of money your will earn depends on the decisions that you and the other participants make. The entire experiment should last less than one hour. At the end of the experiment you will be paid privately and in cash.

Your Identity

You will never be asked to reveal your identity to anyone during the course of the experiment, nor will any other participant know of your cash payoff in today's experiment. Your name will never be recorded by anyone. Neither the experimenter nor the other subjects will be able to link you to any of your decisions. In order to keep your decisions private, please *do not reveal your choices to any other participant.*

Claim Check

Attached to this page is a number on a colored piece of paper. This is your Claim Check. Each participant has a different number. You may want to verify that the number on your Claim Check is the same as the number on the top of page 4.

You will present your Claim Check to an assistant at the end of the experiment to receive your cash payment.

Please remove your Claim Check now and put it is a safe place.

EXPERIMENT SPECIFICS

This experiment will ask you to make a series of choices about how to divide a set of tokens between yourself and one other subject. You and the other subject will be paired randomly and you **will not** be told each other's identity.

As you divide the tokens, you and the other subject will each earn points. Every point that the subjects earn will be worth 10 cents. For example, if you earn 58 points you will make $5.80 in the experiment.

Each choice you make is similar to the following:

Example: Divide 50 tokens: *Hold* _____ @ 1 point each, and *Pass* _____ @ 2 points each.

In this choice you must divide 50 tokens. You can keep all the tokens, keep some and pass some, or pass all the tokens. In this example, you will receive 1 point for every token you hold, and the other player will receive 2 points for every token you pass. For example, if you hold 50 and pass 0 tokens, you will receive 50 points, or 50×$.10=$5.00, and the other player will receive no points and $0. If you hold 0 tokens and pass 50, you will receive $0 and the other player will receive 50×2=100 points, or 100×$.10=$10.00. However, you can choose to hold any number of tokens between 0 and 50. For example, you could choose to hold 29 tokens and pass 21. In this case you would earn 29 points, or 29×$.10=$2.90, and the other subject would receive 21×2=42 points, which is 42×$.10=$4.20.

Here is one other example:

Example: Divide 40 tokens: *Hold* _____@ 3 points each, and *Pass* _____@1 point each.

In this example every token you hold earns you 3 points, and every token you pass earns the other subject 1 point. Again, each point you earn is worth $.10 to you, and each point the other subject earns is worth $.10 to him/her.

Important Detail*:* In all cases you can choose any number to hold and any number to pass, but the number of tokens you hold plus the number of tokens you pass *must* equal the total number of tokens to divide. Please feel free to use a calculator or paper and pencil to calculate points and to assure that all of the tokens have been allocated.

EARNING MONEY IN THIS EXPERIMENT

You will be asked to make 11 allocation decisions like the examples we discussed on the previous page. We will calculate your payments as follows: These payments are independent of the initial $5 you received for showing up on time.

After your and everyone else's decision forms have been collected, we will shuffle the forms and randomly pair your form with that of another subject in this experiment. Using a table of random numbers, we will select <u>one</u> of your decisions to carry out. You will then get the points you allocated in the "hold" portion of your decision, and the other subject will get the points you allocated on the "pass" portion of your decision. These points will be worth 10 cents each, as was the case in the examples shown on the previous page. The earnings from your points will be placed in your earnings envelope.

Next you will be paired again with a different subject in the experiment. This time we will randomly choose <u>one</u> of the other subject's decisions. You will earn the points allocated in the "pass" portion from the other subject's decision sheet. Again these points will be worth 10 cents each. Your earnings from this pairing will also be placed in your earnings envelope.

> (Note: half of the subjects will receive their cash payoff from the counterpart's "pass" portion during the first shuffle, and the cash payoff from the "hold" portion of their own decision sheet during the second random matching of subjects and counterparts. In any event, each subject will earn a cash payoff that results from one of his/her own randomly chosen "pass" decisions, and also from one randomly chosen "hold" decision from being matched *with a different* counterpart).

The monitor chosen at the beginning of the experiment will verify that these procedures are followed.

After all the calculations have been made, another experimenter who was not involved in the experiment until this time will ask you to bring up your claim check and will hand you your earnings envelope. This will again help to guarantee your privacy.

On the following page are the choices we would like you to make for this experiment. Please fill out the form, taking the time you need to be accurate. When all subjects are done we will collect the forms.

Thank you very much for your participation.

DECISION SHEET (in preceding Appendix) FOLLOWS THIS PAGE FOR SUBJECTS

INSTRUCTIONS (Classroom POINTS TREATMENTS)

Note: Instructions are the same for the EARLY POINTS and LATE POINTS treatments. The experiments differ only in the timing of the experiment within the academic semester.

Welcome

This is an experiment about decision making. You will be paid in Econ 2030 [**micro-economics principles]** class points for your participation in this experiment, and the amount of points you will earn depends on the decisions that you and the other participants make. Each class point that you earn in today's experiment will be added to the points score you receive on Exam #1 in our class (with **no maximum** to the score one could have. Even if your Exam #1 score plus today's class points from the experiment sum to more than 100, I will still include all points in determining final grades). The entire experiment should last about one-half hour. At the end of the experiment you will be informed of your class points privately.

Your Identity

You will never be asked to reveal your identity to anyone during the course of the experiment, nor will any other participant know of your points payoff in today's experiment. Your name will never be recorded in connection with your decisions. Neither the experimenter nor the other subjects will be able to link you to any of your decisions. In order to keep your decisions private, please *do not reveal your choices to any other participant.*

Claim Check

Attached to this page is a number on a colored piece of paper. This is your Claim Check. Each participant has a different number. You may want to verify that the number on your Claim Check is the same as the number on the top of page 4.

You will present your Claim Check to an assistant at the end of the experiment to receive your payment of class points.

Please remove your Claim Check now and put it is a safe place.

EXPERIMENT SPECIFICS

This experiment will ask you to make a series of choices about how to divide a set of tokens between yourself and one other subject. You and the other subject will be paired randomly and you **will not** be told each other's identity.

As you divide the tokens, you and the other subject can each earn class points. Every 10 experiment-points (e-points) that the subjects earn will be worth 1 class point. Equivalently, this means that each e-point is worth .10 class points (I will keep track of fractional class points). For example, if you earn 50 e-points you will make 5 class points in the experiment (the equivalent of one-half letter grade on Exam #1).

Each choice you make is similar to the following:

Example: Divide 50 tokens: *Hold* _____ @ 1 e-point each, and *Pass* _____ @ 2 e-points each.

In this example choice you must divide 50 tokens. You can keep all the tokens, keep some and pass some, or pass all the tokens. In this example, you will receive 1 e-point for every token you hold, and the other player will receive 2 e-points for every token you pass. For example, if you hold 50 and pass 0 tokens, you will receive 50 e-points, or 50×.10=5 class points, and the other player will receive no e-points and therefore no class points. If you hold 0 tokens and pass 50, you will receive no class points and the other player will receive 50×2=100 e-points, or 100×.10=10 class points. However, you can choose to hold any number of tokens between 0 and 50. For example, you could choose to hold 29 tokens and pass 21. In this case you would earn 29 e-points, or 29×.10=2.9 class points, and the other subject would receive 21×2=42 e-points, which is 42×.10=4.2 class points.

Here is one other example:

Example: Divide 40 tokens: *Hold* _____@ 3 e-points each, and *Pass* _____@1 e-point each.

In this example every token you hold earns you 3 e-points, and every token you pass earns the other subject 1 e-point. Again, each e-point you earn is worth .10 class points to you, and each e-point the other subject earns is worth .10 class points to him/her.

Important Detail*:* In all cases you can choose any number to hold and any number to pass, but the number of tokens you hold plus the number of tokens you pass *must* equal the total number of tokens to divide. Please feel free to use a calculator or paper and pencil to calculate points and to assure that all of the tokens have been allocated. It is also important for you to realize that the class points you receive from this experiment ***will count*** in the actual determination of your grade (that is, each class point earned will be added to your Exam #1 score, with no maximum limit—your score can go above 100 if that is the case, and this will still benefit your final point and grade total).

EARNING CLASS POINTS IN THIS EXPERIMENT

You will be asked to make 11 allocation decisions like the examples we discussed on the previous page. We will calculate your payments as follows:

After your and everyone else's decision forms have been collected, we will shuffle the forms and randomly pair your form with that of another subject in this experiment. Using a table of random numbers, we will select <u>one</u> of your decisions to carry out. You will then get the tokens and resultant e-points you allocated in the "hold" portion of your decision, and the other subject will get the tokens and resultant e-points you allocated on the "pass" portion of your decision. These points will be worth .10 class points each, as was the case in the examples shown on the previous page. These class points earnings will then be written on a slip of paper and placed in your earnings envelope.

Next we will take the decision forms again and you will be paired *with a different* subject in the experiment. This time we will randomly choose one of the other subject's decisions. You will receive the tokens and resultant e-points allocated in the "pass" portion from the other subject's decision sheet (and the counterpart gets the tokens and resultant e-points from the "hold" portion of his/her decision this time). Again these points will be worth .10 class points each. Your class points earnings from this pairing will be written on a separate slip of paper and also placed in your earnings envelope.

> (Note: half of the subjects will receive the e-points from the counterpart's "pass" portion during the first shuffle, and the e-points from the "hold" portion of their own decision sheet during the second random matching of subjects and counterparts. In any event, each subject will earn e-points from one of their randomly chosen "pass" decisions, and also from one randomly chosen "hold" decision from a matching *with a different* counterpart).

The monitor(s) chosen at the beginning of the experiment will verify that these procedures are followed.

After all the calculations have been made, another experimenter who was not involved in the experiment until this time will ask you to bring up your claim check and will hand you your earnings envelope. I will then record your total class points earnings next to your name. Note that this process helps guarantee the privacy of your decisions. By using this process, I will know only what your final points outcome is, and I will not be able to link your points outcome to your individuals decisions, because decisions are only known by claim check number, which I will never see and which will never be recorded next to your name (and never be included in the data set from this experiment). **In other words, your individual decisions will remain anonymous not only to other subjects, but even to me (the experimenter).**

On the following page are the choices we would like you to make for this experiment. Please fill out the form, taking the time you need to be accurate. When all subjects are done we will collect the forms.

Thank you very much for your participation.

Claim Check Number _____

DECISION SHEET

Directions: Please fill in all the blanks below. Make sure the number of tokens listed under *Hold* plus the number listed under *Pass* equals the total number of tokens available. Remember, all e-points are worth 0.10 class points.

1) Divide 75 tokens: *Hold* _____@ 1 e-point each, and *Pass* _____@ 2 e-points each.

2) Divide 40 tokens: *Hold* _____@ 1 e-point each, and *Pass* _____@ 3 e-points each.

3) Divide 75 tokens: *Hold* _____@ 2 e-points each, and *Pass* _____@ 1 e-point each.

4) Divide 60 tokens: *Hold* _____@ 1 e-point each, and *Pass* _____@ 2 e-points each.

5) Divide 40 tokens: *Hold* _____@ 3 e-point each, and *Pass* _____@ 1 e-point each.

6) Divide 60 tokens: *Hold* _____@ 1 e-point each, and *Pass* _____@ 1 e-point each.

7) Divide 100 tokens: *Hold* _____@ 1 e-point each, and *Pass* _____@ 1 e-point each.

8) Divide 60 tokens: *Hold* _____@ 2 e-points each, and *Pass* _____@ 1 e-point each.

9) Divide 80 tokens: *Hold* _____@ 1 e-point each, and *Pass* _____@ 1 e-point each.

10) Divide 40 tokens: *Hold* _____@ 4 e-points each, and *Pass* _____@ 1 e-point each.

11) Divide 40 tokens: *Hold* _____@ 1 e-point each, and *Pass* _____@ 4 e-points each.