

On Replication and Perturbation of the McKelvey and Palfrey Centipede Game Experiment

By James C. Cox and Duncan James

1. Introduction

McKelvey and Palfrey (1992) present one of the most thought-provoking results in experimental testing of game theory. The centipede game they test was purposely built as a paradox (Rosenthal, 1981), pitting geometric accumulation, naïve optimism, and noise (and maybe altruism) against the gravitational pull of backwards induction. Brought into the lab, it did not disappoint. Backward induction – or more formally subgame perfection – failed in the face of motley opposition, at least over 9 or 10 repetitions of the game with round-robin matching.

We experiment with the McKelvey and Palfrey (1992) 4-node, geometric accumulation, common information centipede game with non-zero payoffs for the player who does not choose to “Take” (the accumulating pot). Our results using this game, with stakes inflation-adjusted so as to be comparable to current-dollar value of the stakes in McKelvey and Palfrey’s (1992) “low treatment”, have more unravelling than most of the corresponding treatments in their paper (though it corresponds fairly closely to one of their Pasadena Community College sessions, PCC-2).

We also implement the game with the same nominal (unadjusted) payoffs as in McKelvey and Palfrey (hereafter MP); at these lower *current* dollar stakes, unravelling takes place more slowly. Our finding of faster unravelling at higher stakes is consistent with findings in MP, and with those in Rapoport, Stein, Parco, and Nicholas (2003).

We subsequently perturb the game from MP by setting equal to zero the payoffs to any player not choosing “Take”. We do this both as a robustness check on the results in MP, and as a way of connecting more closely their original results with those in recent centipede game experiments in Cox and James (2012). We find that this design change induces earlier unravelling, relative to otherwise comparable parameterizations of the original MP centipede game. Of note, the “marginal” effect appears to be not unlike that of moving to higher stakes. This also further clarifies results from Murphy, Rapoport, and Parco (2006) on changing the proportions according to which total payoffs at a node were allocated across the one “winner” and the $n-1$ “losers”.

The original centipede game still generates similar behaviour 25 years later and a continent away. Additionally, factors modulating such behaviour are better understood with ongoing research.

2. On Replication and Perturbation of the MP Experiment

2.1 Experimental Design and Protocol

The centipede game from MP, with doubled (CPI-inflation-adjusted) payoffs, appears in Figure 1.¹

¹ The payoffs at corresponding nodes in the MP “low” treatment are one-half of those in Figure 1. The MP paper was submitted to *Econometrica* in May 1990. We guess that their experiment was run in 1989 or earlier. Our experiment was run in May and June 2014. Using the CPI, one would calculate that \$1 in 1989 has the same purchasing power as \$1.92 in 2014. Doubling the MP nominal payoffs provides a reasonably good inflation adjustment that also provides nice numbers for use in the game tree.

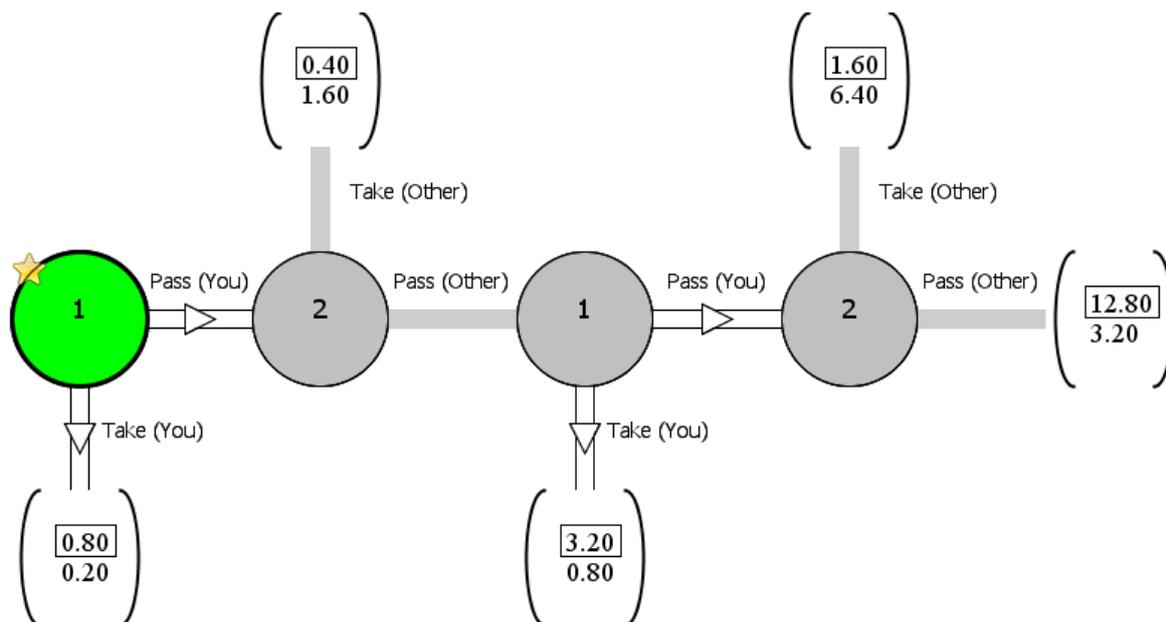


Figure 1. Game Tree for the T1 Inflation-Adjusted Treatment

It is a game of common information; each player can see all payoffs of both players. The summed payoffs (across the two players) double from one node to the next, left to right. The manner in which the split of this payoff “pot” alternates from one decision node to the next sets up a stark contrast at the final decision node – which is predicted to influence choice at the prior decision node, and so on, back to the first decision node.

In the original MP study, round-robin matching was used with 20 subjects, guaranteeing to the subjects that they would never face the same opponent twice, and thus that reputation formation was useless. We also use round-robin matching with 20 subjects in our replication treatment. In our perturbation treatments, we use larger pools of subjects so as to allow such matching for a greater number of rounds.

MP ran subjects through only 10 rounds. We use 10 rounds in our inflation-adjusted replication. In other treatments we run the subjects through an ABA or BAB sequence of three 10-round segments. In these other treatments the first 10-round segment should still be

comparable to the MP data because the existence of subsequent BA or AB rounds in an ABA or BAB treatment was not announced before completion of the first 10 rounds.

Note that our attempt at replication of the MP experiment takes place in the treatment we call the T1-Inflation-Adjusted treatment. We call the treatment with their nominal (lower real stakes) stakes T1. The payoffs in T1 are one-half the payoffs shown in Figure 1. The T2 treatment is the same as T1 except the payoffs to the non-Taker are set to zero. The T3 treatment is the same as T2 except there is a 10 second take opportunity which renders the “Pass” action passive (it occurs after 10 seconds, and not before, and not due to mouse clicks). Treatment T4 is like T3 except the tree institutional format is replaced by a digital price clock format familiar from the Dutch auction. The treatments with 10 second take opportunity and a price clock are used to produce data for comparison with the recent experiment on centipede games in Cox and James (2012) and for adding additional robustness checks on the MP experiment.

In all experiment sessions, subjects were paid the sum of their payoffs from all rounds, as in the MP experiment. 20 subjects participated in one session for the T1 Inflation Adjusted treatment, as in all MP treatments except MP PCC-2, where there were 18 subjects. In each of our three-treatment sessions there were 40 subjects.

Subjects read the instructions on their computer monitors and were invited to raise their hands if they had questions. When a hand was raised, an experimenter approached the subject’s privacy-screened computer station; the question was subsequently asked, and answered, in private. After all subjects had clicked on the “Finished Reading” button on their monitors, an experimenter projected and read aloud a summary of the subject instructions. Subjects were invited to raise their hands during this reading if they had questions; again, any

questions were asked, and answered, in private. Subject instructions and projection files for all treatments are available at <http://excen.gsu.edu/expinstructions/mp/>

2.2 Results from Replication Attempt for the MP Experiment

Figure 2 reports the average (across subject pairs of the) take node in the MP experiment and our replication. Our first finding is that the MP result that play does not completely unravel to a take at the first node can be replicated with the same five node game they used and the same real payoffs they used. Our T1-Inflation-Adjusted data are closest to data from MP's Session 2, run at Pasadena Community College (MP PCC-2 in Figure 2). But we find more unravelling than they found overall; our average take node is 1.55 in rounds 9 and 10 whereas their round 9 and 10 average take nodes were 2.5 or higher in all treatments except MP PCC-

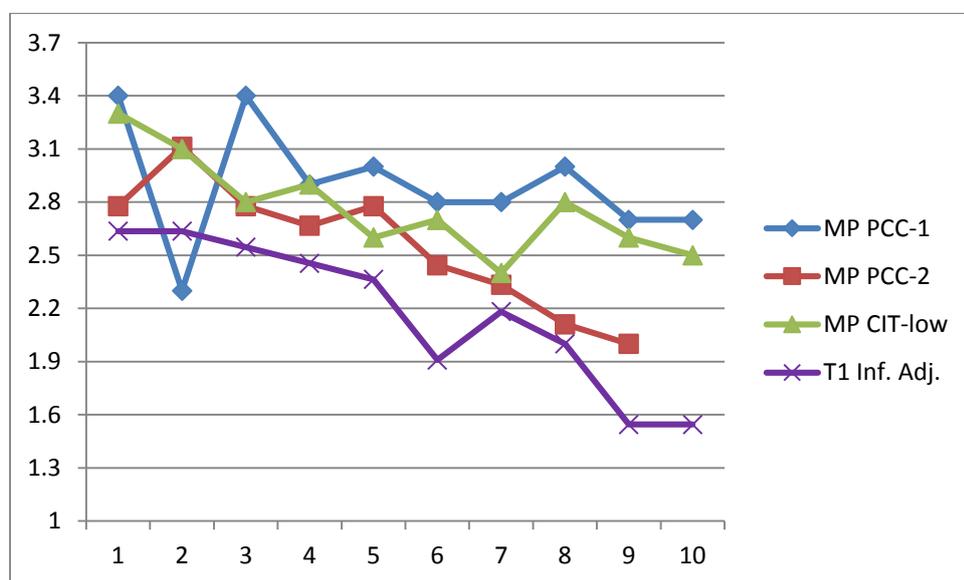


Figure 2. Direct Comparison to MP Data

3. Perturbation: Role of Payoff to the Non-Taker

Perturbation of the original MP design allows an assessment of the role of positive payoffs to non-Takers (versus zero, as would be natural in the centipede game's close relative, the Dutch auction). T1 employs the MP design with their nominal payoffs (there is no inflation adjustment). T2 is the same as T1 except payoffs to non-Takers are zero.

Figure 3 plots average take node data for the session with 10 rounds of T1, followed by 10 rounds of T2, followed by 10 rounds of T1: the T1-T2-T1 session. In this session we observe strong unravelling in treatment T2 beginning in round 2 for this treatment (round 12 in the session).

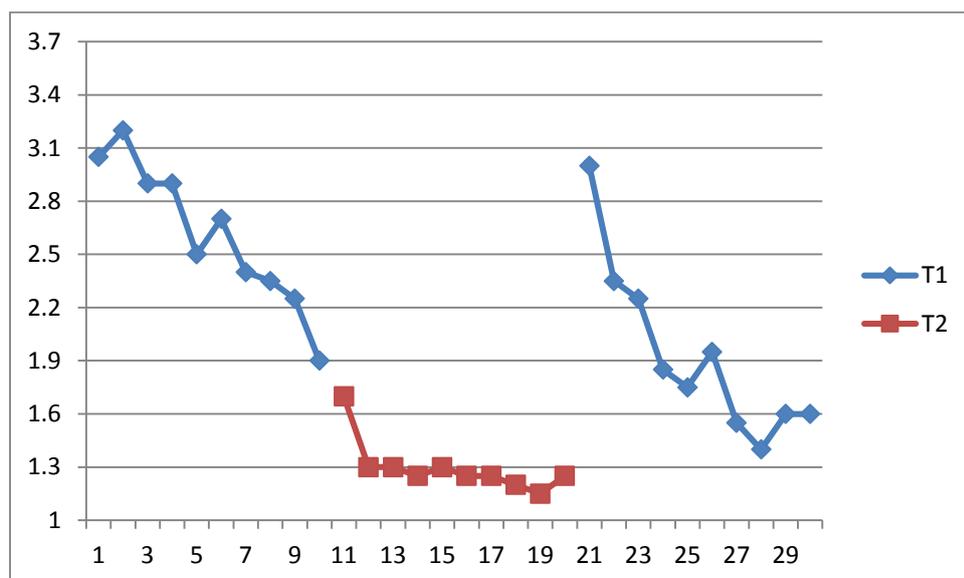


Figure 3. T1-T2-T1 Session Data

Figure 4 plots data for the session with 10 rounds of treatment T2, followed by 10 rounds of treatment T1, followed by 10 rounds of treatment T2: the T2-T1-T2 session. There is strong unravelling in the second implementation of T2 (see rounds 25 – 30 in Figure 4).

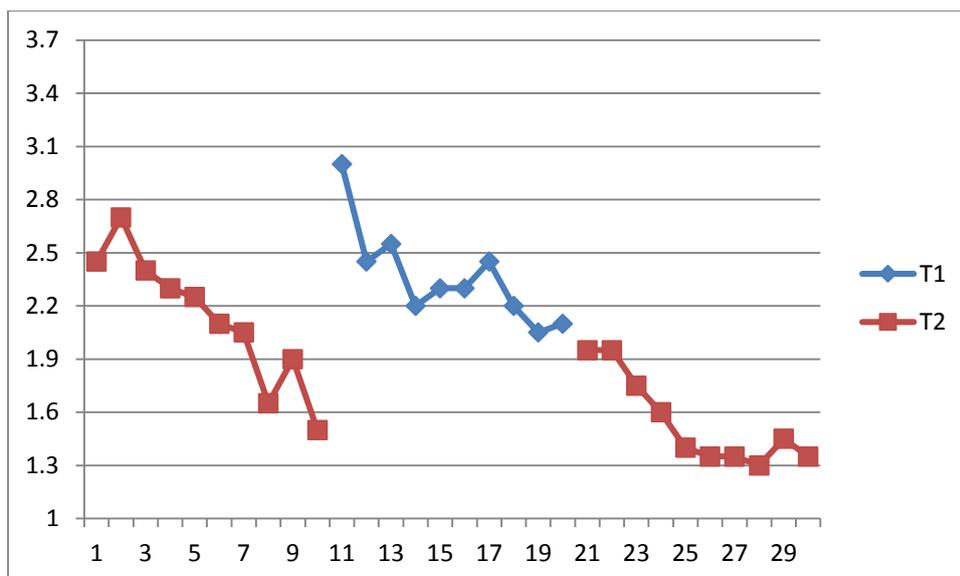


Figure 4. T2-T1-T2 Session Data

The data plotted in Figures 3 and 4 can be nested, and tested statistically, in the regression specification:

$$\text{Average take node} = \text{intercept} + b1*\text{time} + b2*T1 + b3*(\text{time}*T1) + b4*(2/1/2 \text{ session intercept dummy}) + b5*(2/1/2 \text{ session intercept dummy}*\text{time}*T1) + e \quad (1)$$

This yields the following coefficient estimates, and an R^2 of 0.81.

Table 1. Effects on Unravelling of Zero Non-Taker Payoffs

	<i>Coefficient</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	1.823499	0.115115	15.8407	5.97E-22
Time trend	-0.03413	0.005378	-6.3461	4.73E-08
Intercept dummy for T1	1.097136	0.150257	7.3017	1.33E-09
Slope dummy for T1	-0.00767	0.007612	-1.0081	0.31789
Intercept dummy for T2-T1-T2 session	0.543199	0.096483	5.6299	6.64E-07
Slope dummy for T1 in T2-T1-T2 ses.	-0.02945	0.008592	-3.4268	0.00117

Clearly, T1 generates later takes. Later takes in T1 are still observed when subjects have had 10 rounds of T2 in which to move towards taking at the first node, as in the T2-T1-T2

session; here we see the trend toward taking at the first node, as established in the first T2 segment, is broken by the introduction of T1, which postpones average taking by around 1.5 nodes, relative to the end of the immediately preceding T2 segment.

4. Perturbation: Role of Stakes

Data from our T1 Inflation-Adjusted treatment (presented along with comparable MP data in Figure 2) can be compared to the first segment T1 data from Figure 3, in order to get an idea of the “marginal” effect of stake-level in our study. Plotting those series together yields Figure 5.

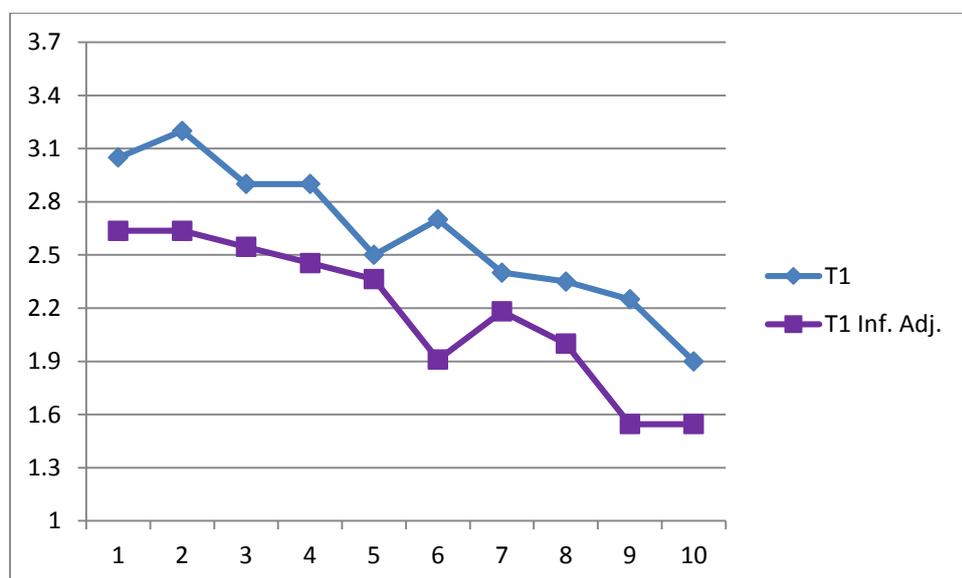


Figure 5. T1 and T1 Inflation-Adjusted Data

One interesting conclusion that might be drawn, from the results in this section and those in the section on payoffs to non-Takers, is that zero payoffs to non-Takers (as in Cox and James 2012) yields earlier unravelling – but so do higher stakes (as in MP, or Rapoport et al. 2003).

5. Perturbation: Role of Clock Format versus Tree Format

The MP design can be adapted to incorporate some of the features present in the recent centipede game experiment in Cox and James (2012). Maintaining the common information, stationary-across-rounds payoff structure of MP, rather than the independent private values (redrawn each round) environment of Cox and James (2012), but augmenting MP with a 10 second clock and passive “Pass” actions (forced by the 10 second timer) yields our treatment T3. Treatment T3 translated to a Dutch-auction-style clock format yields treatment T4. Note that these treatments introduce changes at the margin to the MP complete information game, not to the Cox and James (2012) private information games, and the results are thus not directly comparable to the results from Cox and James (2012).

Employing analogous graphs and regressions to those used in comparing T1 and T2, we find the following for T3 and T4. The second set of T3 rounds in Figure 6 exhibit almost complete unravelling, with 19 takes at the first node and 1 take at the second node in the repeat T-3 treatment rounds 7 and 9 (session rounds 27 and 29 in Figure 6).

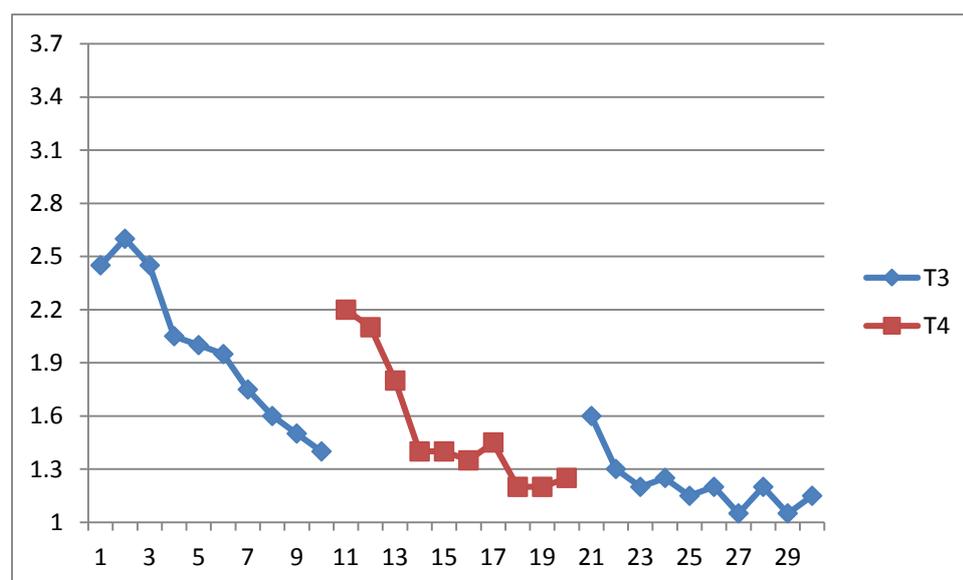


Figure 6. T3-T4-T3 Data

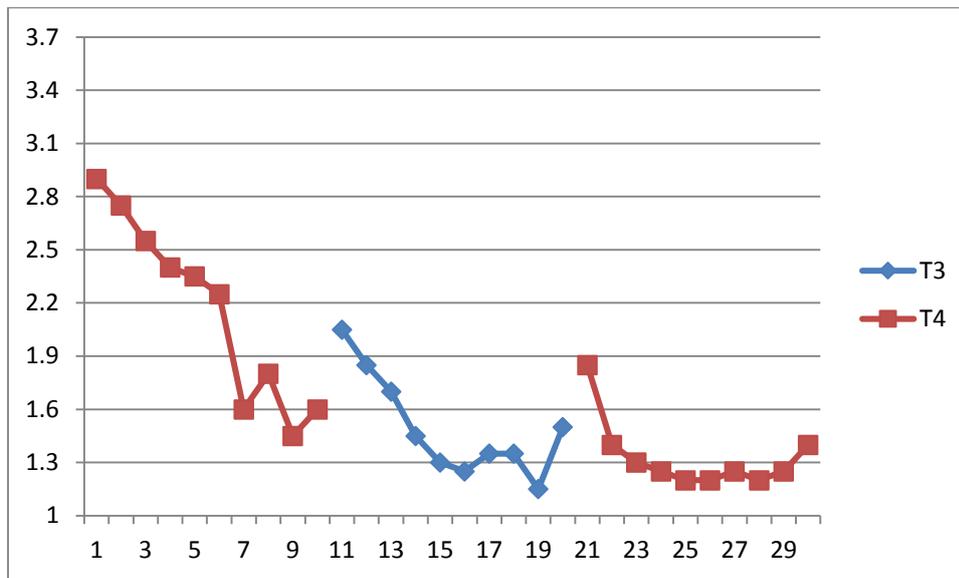


Figure 7. T4-T3-T4 Data

There does not appear to be much overall difference between treatment T3 and T4 data, barring a slight jump at the switch over of treatments (arguably akin to so-called “restart effects”). Statistical analysis bears this out, with Equation (1) (adapted for use with T3 and T4) yielding the following coefficient estimates, and an R^2 of 0.74.

Table 2. Effects of Clock and Tree Institutional Formats

	<i>Coefficient</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	2.279743	0.114164	19.9691	1.34E-26
Time trend	-0.04836	0.005333	-9.06875	1.93E-12
Intercept dummy for T3	-0.02999	0.149015	-0.20125	0.84126
Slope dummy for T3	0.006281	0.007549	0.83207	0.40903
Intercept dummy for T4-T3-T4 session	0.219868	0.095686	2.29780	0.02547
Slope dummy for T3 in T4-T3-T4 ses.	-0.02111	0.008521	-2.4775	0.01639

These results suggest why earlier researchers implementing the centipede game may not have considered institutional (tree or clock) format to be a treatment of interest: for repeated payoff structures with *common* information, format may not make a difference. But this also

illustrates the potentially important interaction between environment and institution, for in the independent private values environment institutional format most certainly does make a difference: see Dorsey and Razzolini (2003) and Turocy, Watson, and Battalio (2007) for auction results and see Cox and James (2012) for both centipede game and auction results.

6. Perturbation: Role of Timed Decision Opportunity

Some light on the role of the timer (or its lack) might be shed by comparing the first segments of the T3-T4-T3 and T2-T1-T2 sessions. Comparing the early T2 and T3 segments in cross section yields the following.

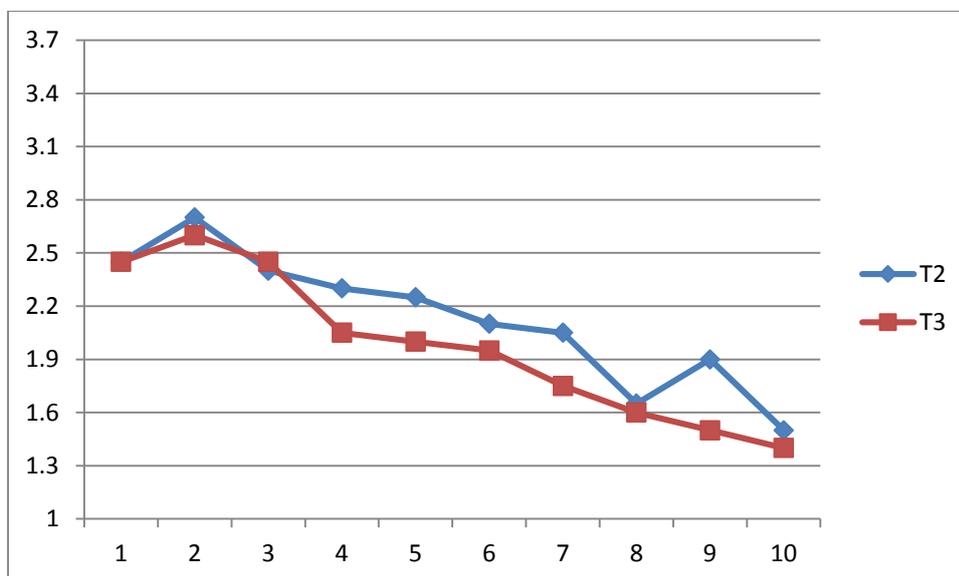


Figure 8. T2 and T3 First Segment Data

Clearly, the average take is indistinguishable across treatments (a t-test for difference in means yields a t-statistic of 0.55). In this setting, at least, having a 10-second timer versus untimed actions does not appear to change behaviour.

7. Conclusions

We observe more unravelling in our replication treatment than was reported by McKelvey and Palfrey (1992) for their “low” treatment. We do replicate earlier results for the effects of changing monetary stakes. New within-subjects comparisons of the MP game with an alternative game with zero payoffs to non-Takers shows that imposing zero payoffs to non-Takers has an effect on average Take node that is roughly similar to that of increasing stakes – at a much lower expenditure in subject payment money. Results on the role of introducing timed moves suggest that relative to the zero-payoffs-to-non-Takers version of MP, a timer does not make a difference. And in a common information, unchanging payoffs setting, clock versus tree format appears not to make a difference either. This contrasts with the significant effect of clock vs. tree format for both centipede games and Dutch auctions in the independent private values environment of the experiment reported by Cox and James (2012).

REFERENCES

- Cox, James C. and Duncan James, (2012), "Clocks and Trees: Isomorphic Dutch Auctions and Centipede Games," *Econometrica*, 80(2), 883-903.
- Dorsey, Robert and Laura Razzolini, (2003), "Explaining Overbidding in First Price Auctions Using Controlled Lotteries," *Experimental Economics*, 6(2), 123-140.
- McKelvey, Richard and Thomas R. Palfrey (1992), "An Experimental Study of the Centipede Game," *Econometrica*, 60, 803-836.
- Murphy, Ryan O., Amnon Rapoport, and James E. Parco (2006), "Breakdown of Cooperation in Iterative Real-Time Trust Dilemmas," *Experimental Economics*, 9, 147-166.
- Rapoport, Amnon, William E. Stein, James E. Parco, and Thomas E. Nicholas (2003), "Equilibrium Play and Adaptive Learning in a Three Person Centipede Game," *Games and Economic Behavior*, 43, 239-265.
- Rosenthal, R. (1981). "Games of Perfect Information, Predatory Pricing, and the Chain Store", *Journal of Economic Theory*, 25, 92–100.
- Turocy, Theodore L., Elizabeth Watson, and Raymond C. Battalio (2007), "Framing the First-Price Auction," *Experimental Economics*, 10(1), 37-51.