Piece-rates and tournaments: Implications for learning in a cognitively challenging task

Tony So
Department of Economics
University of Auckland
t.so@auckland.ac.nz

Ananish Chaudhuri
Department of Economics
University of Auckland
a.chaudhuri@auckland.ac.nz

Paul Brown
School of Social Sciences, Humanities and the Arts
University of California, Merced
pbrown3@ucmerced.edu

Corresponding author: Ananish Chaudhuri
Department of Economics
University of Auckland
12 Grafton Road
Auckland 1142
New Zealand
Phone: +64-9-923-8307
Fax: +64-9-373-7427
Email: a.chaudhuri@auckland.ac.nz

Piece-rates and tournaments: Implications for learning in a cognitively challenging task

Abstract

We compare the impact of piece-rate and tournament payment schemes on learning in a cognitively challenging task. In each one of multiple rounds, subjects are shown two cue values, Cue A and Cue B, and asked to predict the value of a third variable X, which is a function of the two cue values. The subjects' aim is to predict the value of X as accurately as possible. Our metric of performance is the absolute error, i.e., distance between the actual value of X and the predicted value of X. We implement four treatments which are based on two different payment schemes: (1) piece rates, where subjects are paid on the basis of their own absolute errors and (2) a two-person winner-take-all-tournament, where subjects are paired and the one with the smaller absolute error earns a positive payoff, while the other earns nothing. We find that it is only in the tournament payment scheme that subjects show significant evidence of learning over time, in that their predictions get closer to the actual value of X. Learning in tournaments is particularly pronounced for those who are initially not adept at the task. The learning process is driven by the all-or-nothing nature of the payoff structure in tournaments.

JEL Codes: C91; J24, J33, J39

Keywords: Experiments; Payment schemes; Tournaments; Piece-rates; Productivity; Learning

1. Introduction

Piece-rates and tournaments are two oft-used mechanisms for paying workers. However, piece-rates, which pay individual workers on the basis of actual output, are hard to implement where output cannot be easily observed or measured. In such cases, employers often rely on tournament pay schemes that pay on the basis of relative rather than absolute output or performance. Theoretical analyses of tournaments (Lazear & Rosen, 1981; Green & Stokey, 1983; Nalebuff & Stiglitz, 1983) show that tournaments are effective in eliciting output at a level analogous to piece rates. This result is borne out by results in a laboratory experiment by Bull, Schotter, and Weigelt (1987), where they show that, on average, numerical effort choices made under tournaments are statistically no different than those under piece rates, though variance of effort choices under tournaments are larger.

However, prior studies have not really focused on which type of payment schemes foster better learning, especially in tasks that are complex and cognitively challenging. Part of this is due to the fact that most prior studies implement somewhat mechanistic tasks that do not provide scope for learning over time.[1] In fact, existing evidence suggests that in tasks, which require significant learning over time, the reward structure may play a crucial role in enhancing or impeding that learning. Merlo and Schotter (1999) study learning in the stylized two person tournament introduced in Bull et al. (1987) except in the former, one player is replaced by a computer, that always chooses the same number and subjects are informed of

[1] For instance, Kuhnen and Tymula (2012) and Cadsby et al. (2010) use an arithmetic task, where subjects are asked to add a sequence of five two-digit numbers without recourse to calculators, as in Niederle and Vesterlund (2007), while Charness et al. (2014) use a decoding task. These tasks mainly rely on effort in order to do well; there is nothing to 'learn' per se. Our task is different, in that, it is cognitively challenging. In order to improve forecasts, subjects need to uncover the underlying relationship between the cue values and the actual value of X, or at least, get as close to it, as possible. Our task relates more closely to those used to specifically study the processes and mechanics of learning. For example, in Merlo and Schotter (1999, 2003) players need to search for the equilibrium best response that maximises payoffs. In multi-player strategic games (Cardella, 2012; Charness & Levin, 2005; Erev and roth, 1998; Rick & Weber, 2010; Roth & Erev, 1995) the 'way to play' is often prescribed as a dominant strategy (or, at least, one that is not dominated), which players should learn to play over time.

the computer's effort choice.[2] This has the effect of transforming the underlying two-person tournament into an individual decision making exercise where subjects are essentially looking to find the maximum of the underlying payoff function. Merlo and Schotter (1999) report that subject choices are much closer to the Nash equilibrium in the *Learn-before-you-earn* (LBYE) treatment (where subjects play for 74 rounds without getting paid and then play a 75th round with substantial money at stake) than those in the *Learn-while-you-earn* (LWYE) treatment (where subjects play for 75 rounds with small payments in each round). This is due, to a large extent, to the fact that in the LWYE treatment subjects adopted a much more "myopic" view of the task by focusing on wins or losses in each round. Those in the LBYE treatment, on the other hand, engaged in greater "experimentation" in the non-payment rounds in an attempt to identify the optimum.[3]

Given that many, if not most, tasks and certainly all so-called white-collar jobs require cognitive effort, it is certainly of interest to understand which commonly used payment schemes, if any, lead to better facility at the task. Therefore, in this paper, we explore the impact of payment schemes on learning, using a multiple cue probabilistic learning (MCPL) task introduced by Brown (1995, 1998). We provide details of the task below in the section on experimental design. Here, we provide an overview. In each of multiple rounds subjects are shown two cue values (Cue A and Cue B) and asked to predict the value of an unknown variable (X), which is a function of those two cue values. The cue values shown to subjects change from one round to the next but the underlying function does not. Subjects do not know what the underlying functional form is but they do know that this function remains unchanged from one round to the next. The ***Appendix*** contains the

[2] This, in turn, implies that payoff is maximized by simply choosing what the computer is choosing in each round, i.e., 37.

[3] Iyengar and Schotter (2008) extend Merlo and Schotter (1999) by allowing for two-player teams where one player is allowed to pass advice to another, who can choose to ignore this advice. In one treatment ignoring advice is costly while in another it is costless. Iyengar and Schotter (2008) report that when advice is costly to ignore both advisors and advisee learn to make decisions that are closer to the Nash equilibrium.

instructions to the experiment. The goal for the subjects is to make accurate predictions on the basis of the cue values shown to them in each round, where accuracy is measured by the absolute distance of their predicted value from the actual value of the variable. This is our metric for performance: the absolute prediction error, i.e., |(Actual value of X) – (Predicted value of X)|. The smaller is the absolute error, the better is productivity. By *"learning"* we will refer to decreasing absolute errors (increasing productivity) over time, which, in turn, implies increasing prediction accuracy.

Before going on to explore the impact of piece rates and tournaments on learning, we need to address one issue that has not received enough attention in the prior literature. In moving from piece-rates to tournaments, the incentives change in two ways. First, under piece rate one's payoff depends on only on one's own performance while in a rank-order tournament it depends on one's rank. If the tournament happens to be of a winner-take-all type, then coming second implies zero monetary payoff. This can be thought of as *competing for higher payoff*.

But, there is a second component to this change, since in a tournament, agents must outperform their peers in order to attain a higher rank. While a higher rank may correspond to a higher tangible reward (such as promotion tournaments), agents may simply be motivated by the higher rank itself, in the sense that they derive pleasure or pain from the act of winning or losing respectively (as in a friendly game of tennis, squash or chess).[4] We will refer to this loosely as *competing for higher rank*. There is ample evidence that information about one's relative rank, vis-à-vis one's peers, has a positive impact on performance, even when that higher rank does not translate into higher monetary payoffs.[5]

[4] The idea behind rank competition in our study is similar to what Kräkel (2008, p. 206) describes as *"emotions"*, where positive and negative emotions are derived from winning and losing respectively.

[5] Blanes i Vidal and Nossol (2011) undertake a study of German warehouse workers, who were notified two months in advance that they would be receiving additional rank information in their payslips. The revelation of rank information was found to have a positive effect on productivity. In Kuhnen and Tymula (2012) participants solve multiplication problems over a number of timed rounds and are paid a fixed salary for their participation.

In order to disentangle these two effects and control for them, we will introduce a third treatment, which we will refer to as a piece-rate win-lose treatment. Here, subjects are paid of the basis of their own performance, exactly as in piece-rates, except in each round, subjects are put into pairs and each subject is told whether s/he did better or worse than the pair-member as in a tournament. However, this rank information does not have any payoff implications. Comparing piece-rate with piece-rate win-lose allows us to control for any additional impact of *competing for higher rank*; since the payoff mechanism is identical, except the latter provides additional rank information. Similarly, comparing piece-rate win-lose with tournament will allow us to understand the role of competing for higher payoff, since the rank information is the same in both, except in tournaments higher rank translates into higher payoff.

Finally, we have a fourth treatment, where subjects take part in a winner-take-all tournament, except they are not provided any information about whether they are winning or losing or their earnings till the very end of the session. This treatment is intended to bolster the findings of the winner-take-all tournament findings. We provide more details below.

We show that learning, in terms of increasing prediction accuracy, is facilitated most by a winner-take-all tournament. This learning effect is particularly pronounced for those subjects who were not particularly adept at the task to start with. Further, not only do subjects learn to make smaller errors in the tournament payment schemes, the variance of those errors

In one treatment participants receive relative performance feedback while in a second they receive such feedback with probability 0.5 while in a third treatment no feedback is provided. Players in both certain feedback and probabilistic feedback performed better than those who did not receive feedback while there are no differences in performance in the first two treatments. It appears that while feedback matters even the likelihood of receiving feedback can serve as a motivating force. Azmat and Iriberri (2010) use data from 1986-94 for Spanish high school students to understand whether providing relative rank information leads to improved student achievement. In the academic year 1990-91, due to exogenous changes, student report cards provided information about the average class grade alongside their own grade. This resulted in students attaining higher grades that year compared to previous and subsequent years where no such relative feedback was provided. Similarly, Tran and Zeckhauser (2012) found that Vietnamese English-language students who were notified of how they were ranked within their class, performed better than the control group who were not provided such information. Both Tran and Zeckhauser (2012) and Cadsby, Engle-Warnick, Fang and Song (2010) show that, by and large, it does not matter whether the rank information is provided publically or privately.

are also smaller. Finally, we show that it is the winner-take-all nature of the payment scheme that fosters this learning, rather than the provision of relative rank information. We proceed as follows. In Section 2 we explain our experimental design. In Section 3 we present our results and finally in Section 4 we make some concluding comments.

2. Experimental Design

2.1 Task

Our experiment is based on a multiple cue probabilistic learning (MCPL) task, where in each one of 20 rounds *(t)* subjects are required to predict the value of a variable X_t based on the observation of two numerical "cues" provided to them.[6] The variable X_t can be thought of as the underlying price of a stock; the cues as variables that affect the value of a stock; and the task at hand as one of forecasting stock prices. The actual stock value is determined by the underlying equation:

$$X_t = 10 + 0.3\times Cue\ A_t + 0.7\times Cue\ B_t + \varepsilon_t$$

where X_t is the actual stock value subjects are required to predict, $Cue\ A_t$ and $Cue\ B_t$ are the values of the two numerical cues provided to the subject, and ε_t is a random variable with uniform distribution drawn from the set [-5, 5] in each round t. Subjects do not know about the error term, the exact relationship between the cue values and the value of X, or even whether the relationship is linear or non-linear. They do know, however, that while the cue values change from one round to the next, the underlying relationship does not change.

We implement two variants of the task. In the *"Single Cue"* task, Cue A is fixed at the value of 150 for each of the 20 rounds, while Cue B changes each round. This is designed to be less difficult than the *"Dual Cue"* task, where both cue values change in each round. For both tasks, the sequence in which the cue values appear from one round to the next, is

[6] MCPL tasks are commonly used in psychology to study learning (see Balzer, Doherty & O'Connor, 1989 for a review). In economics, besides Brown (1995, 1998), this task has been used by Vandegrift and Brown (2003), Vandegrift, Yavas, and Brown (2007) as well.

identical across treatments. Table 1 shows the cue values and corresponding stock prices for each round.

<<Table 1 about here>>

Our metric of performance is the absolute error, the absolute distance between the predicted value (X_t^P) and the actual value X_t^*, i.e., ($|X_t^P - X_t^*|$). Forecast errors measure the accuracy of the predicted value, so smaller errors imply more accurate forecasts and therefore better performance (and higher productivity). Below, at times and where it makes for better exposition, we may refer to the absolute error simply as "error", but anytime we do so it implies absolute error.

<<Table 1 about here>>

2.2 Treatments

We report on data from four treatments for the purposes of this paper, which is part of a larger study involving other analyses. These four treatments are: (1) piece-rate (PR), (2) piece-rate win-lose (PRWL); (3) two-person winner-take-all tournament (WTAT) and (4) two-person winner take-all-tournament with no information (WTAT-NI). Under PR, the earnings for each subject in a particular round is given by NZ \$1 minus the absolute forecast error. For instance, if the absolute error is 20 then the payment for that round is NZ \$0.80. If the forecast error is greater than 100, then earning for that round is set to zero. Here, subjects are engaged in an individual decision making task where their aim is to minimize the absolute error in each round which in turn will lead to higher payoff. Figure 1 presents a screenshot to show what the subjects get to see at the end of a round. This is the information that a subject will be looking at prior to the beginning of round 10.

<<Figure 1 about here>>

Given the relative difficulty of the MCPL task and given the possibility of significant heterogeneity in ability levels, it is important to get a benchmark estimate of how adept or not a particular subject is at the task. Consequently, in each treatment, subjects are paid piece rates for the first 5 rounds. At the end of those 5 rounds, subjects are given further instruction as appropriate in the other treatments. In the PR treatment, there is no change in the payment mechanism following round 5 and subjects are instructed accordingly.

The PRWL treatment adds rank competition on top of the PR treatment. Here, starting with Round 6, subjects are paired, with random re-matching of pairs between rounds. They are paid according to their own absolute errors in each round, i.e. the payment scheme is the same piece-rate. But, from Round 6 onwards, in each round the subjects are also told whether they have "Won" or "Lost" depending on whether a particular subject's error was smaller than or larger than her pair member respectively. However, whether a subject won or lost a particular round has no bearing on her earnings for that round since each subject continues to get paid on the basis of one's own absolute errors. The rank information is simply designed to capture positive or negative emotions from winning or losing respectively. Figure 2 shows a screenshot of this treatment.

<<Figure 2 about here>>

The WTAT treatment also starts in Round 6, following 5 rounds of piece-rate payments. As in PRWL, from Round 6 onwards, we form subjects into pairs (with random re-matching from one round to the next), except here, we implement a winner-take all scheme where in each round, the subject with the smaller absolute error wins NZ $1 while the subject with the larger error gets zero.[7] Figure 3 presents a screen-shot. Note that compared to PRWL, from Round 6 onwards, WTAT not only provides the win/lose information but changes the payoffs as well. So WTAT adds payoff competition on top of the rank competition in PRWL.

<<Figure 3 about here>>

[7] If the forecast errors of a particular pair are equal in particular round, then the tie is broken by randomisation.

The no information tournament treatment (WTAT-NI) is very similar to the WTAT treatment and also starts in Round 6, following 5 rounds of piece-rate payments. As in WTAT, from Round 6 onwards, we form subjects into pairs (with random re-matching from one round to the next), and implement a winner-take all scheme where in each round, the subject with the smaller absolute error wins NZ $1 while the subject with the larger error gets zero. Except, while subjects are aware that they are in a winner-take-all environment and they are either winning of losing in every round with corresponding payoffs of $1 or nothing respectively, *they do not get to see any of this information till the end of round 20*. Figure 4 presents a screen-shot. We run this treatment with the dual cue task only. This is because performance is very similar in WTAT and WTAT-NI; furthermore, the latter treatment is primarily designed to reinforce insights gained from the WTAT treatment and to show that winner-take-all payment schemes foster learning over time. We provide more details below.

<<Figure 4 about here>>

2.3 Experimental Procedure

Sessions were conducted at the DECIDE lab at the University of Auckland, using primarily first year students in business and economics. There are a total of 274 subjects across the different treatments. Subjects are seated at computer cubicles with privacy partitions and are cautioned about not communicating with any other subject. To start with, subjects are asked to fill out a questionnaire which elicits subjects' trait anxiety level. See Spielberger, Gorsuch, Lushene, Vagg and Jacobs (1983). This is shown in the appendix. The questionnaire consists of 20 questions that are answered on a 1 to 4 scale. Questions 1, 6, 7, 10, 13, 16 and 19 are reverse scored. The questionnaire is designed to measure a subject's general tendency to feel anxious rather than their current level of anxiety (McNaughton, 2011). A higher score generated from the pre-task questionnaire indicates a higher level of trait anxiety associated with the individual. We use trait anxiety as a proxy of each subject's competitive preferences. (Segal and Weinberg, 1984).

Following this we hand out the instructions to the forecasting task. These instructions are also read out loud after subjects have had a chance to read them privately on their own. The ***Appendix*** contains a copy of the instructions. As noted above, subjects know that for the PR, PRWL, WTAT and WTAT-NI treatments the first five rounds are identical, using a piece rate payment scheme. They are told that they will be provided further information prior to the start of round 6. Subjects are also provided with ten examples for Cue A, Cue B and X and given some time to study these examples. This is shown in Table 2.

<<Table 2 about here>>

In the PR treatment, following round 5, subjects are told that there are no further instructions and they should continue as before. In the PRWL treatment, after round 5, they are told that in going forward they will be paired with another player in each round and told whether they won or lost a round but that this rank information has no bearing on their earnings, which still depend only on their absolute errors in any given round. In the WTAT treatment they are told both about the pairing and that from round 6 onwards they will earn either $1 or nothing in each round. In the WTAT-NI treatment, they are told that they will be paired from round 6 onwards and will either get $1 or nothing, except they will not learn about this till the end of round 20. In all relevant treatments, subjects are aware that they will be randomly re-matched from one round to the next.[8] See the ***Appendix*** for the details of individual treatments.

[8] We need to add a word here about expected earnings. In the dual cue task, the average errors per round were approximately 27. Under a piece-rate payment scheme, this implies earnings of NZ $0.73 per round or NZ $14.60 over 20 rounds. This along with the NZ $5 show-up fee meant a total payment of approximately NZ$19.60. In WTAT we assumed a 50:50 win-loss probability in each round. So because the first five rounds are paid on the basis of piece-rates we expected people to earn about NZ $3.65. If people won half the time over the next 15 rounds, then their expected payoff would be NZ $7.50. Prior to the start of round 6, we added NZ $4 to their earnings accounts. Including the NZ $5 show-up fee this also leads to an approximate earning of NZ $20.15.

At the conclusion of the session, subjects are asked to fill out a post-task questionnaire, which elicits information about subjects' intrinsic motivation, including self-reports of how competent they felt at the task, how motivated they were, how interesting they found the task, how much effort they exerted and how close they felt to other subjects in the room. We also collected basic demographic information including gender, age and ethnicity. We do not elaborate on the psychological questionnaires since we do not exploit data from them for the purposes of this study.

3. Results

We will start by providing a brief overview of differences in productivity, as measured by average absolute errors, across the different treatments. This is mostly to set the stage for the discussion on learning that follows. While we have data for 274 subjects, for much of the analysis below, we will confine our attention to data from 236 subjects in the PR, PRWL and WTAT treatments for reasons that will become clear as we go along. We will introduce data from the WTAT-NI treatment later, primarily to bolster insights gained via the WTAT treatment and corroborate findings from that treatment. Table 3 provides a broad overview of the absolute errors in different treatments along with the number of subjects and sessions in each treatment.[9] Not surprisingly the errors are much smaller in the single cue task than the dual cue one. What is noticeable is that, in both the single and dual cue tasks, average absolute errors are highest in WTAT, followed by PR. Errors are smallest in the PRWL treatment. To explore these issues more rigorously we turn to regression analysis next.

<<Table 3 about here>>

3.1 Productivity across treatments

[9] We had 36 subjects in the PRWL treatment but due to reasons beyond our control, one subject left early. Since we needed to put subjects in pairs from round 6 onwards, we discreetly replaced the departing subject with one of our under-graduate assistants (who had no prior experience with the game). We have excluded the choices made by this subject (and the replacement) from our analysis. However, given the random re-matching of subjects and the fact that subjects never get to see the ID numbers of their pair-members, we have retained the data for the remaining 35 subjects.

Result 1: Average performance is better in the PRWL treatment compared to PR or WTAT; there is no significant difference in performance between PR and WTAT.

As noted before our metric for good performance is absolute error (=|Predicted value – Actual value|). The smaller the absolute error, the higher the productivity. In the first three columns of Table 4 we present results for random effects regressions with the absolute forecasting error as the dependent variable with robust standard errors clustered on individual subjects. In running these regressions, we pool the data from the single cue and dual cue tasks. In a companion paper, Brown et al. (2016), we provide more disaggregated analysis broken up by single and dual cue tasks. We use a random effects specification because we have both time varying and time invariant variables among our regressors. In Column 4 of Table 4, we also present results of a quantile regression as a robustness check.

<<Table 4 about here>>

Recall that in all treatments subjects play the PR treatment for the first 5 rounds and the treatment (if any) is implemented only at the start of Round 6. Hence in running these regressions we use data for Rounds 6 through 20. We start with the simplest specification in Model 1 where we include the following independent variables: round and two dummies one for PRWL and another for WTAT with the PR treatment as the reference category. With 236 subjects making 15 decisions each we have 3540 observations here. It is noteworthy, and in line with the information in Table 3, that the coefficient of PRWL is negative while that for WTAT is positive. But none of the coefficients for the treatment dummies are significant at conventional levels.

In Model 2 we control for two additional regressors: female (= 1 for women and 0 for men) and each subject's trait anxiety score. This regression clearly indicates that the average errors are lower in PRWL compared to both PR and WTAT. What is also noticeable is the large positive coefficient for the gender dummy showing that, on average, women performed much worse than men across the board.

In Model 3, in order to pick up possible learning effects over time we include three additional terms involving the treatment dummies interacted with round. The results are similar to Model 2 except the coefficient for the PRWL dummy is now significant at 5%. Also noteworthy is the negative and significant coefficient for the interaction term *WTAT*Round*, which suggests that the errors diminish over time in WTAT. We will come back to this in greater detail shortly, as we study the issue of learning over time.

As a robustness check, in Model 4 we run a quantile regression which corroborates the previous findings including the negative and significant coefficient for PRWL. Further, the quantile regression suggests a role for trait anxiety; the coefficient is positive and significant at 5%, indicating that higher trait anxiety leads to higher errors. However, quantile regressions do not accommodate for the panel structure of our data and the dynamic nature of decision-making over time. Consequently, in what follows we rely primarily on random effects regressions. On the basis of these results, we conclude that average errors are lower in the PRWL treatment. We examine these productivity differences and the driving forces behind them in much greater detail in Brown, Cameron, Chaudhuri and So (2016). Therefore, we now move on to study learning over time in the different treatments.

3.2 Learning across treatments

Result 2: While WTAT does not perform well in terms of average errors, it is the only treatment where subjects demonstrated evidence of learning over time.

In this section we look at what happens to errors over time. In doing so, we look separately at rounds 6 through 20 and rounds 11 through 20. The rationale is this. In PRWL and WTAT, as opposed to PR, the treatment change comes into effect only after round 5. Therefore, it is arguable that it takes subjects some time to get used to the new treatment and therefore, looking at rounds 11 onwards, where subjects now have had experience in the new treatment for five additional rounds, produces more reliable estimates. Table 5 contains our main results.

The regressions are the exact same ones as that in Model 3 of Table 4; i.e., the regressors include two treatment dummies PRWL and WTAT with PR as the reference category as well as controls for gender (female = 1 for women, 0 for men) and trait anxiety and three interactions terms involving each of the three treatment dummies interacted with round. However, for ease of exposition, in Table 5, we have suppressed the coefficients for all other regressors except for the interaction terms. We present our results separately for rounds 6 through 20 (Model 1) and rounds 11 through 20 (Model 2). We note that the coefficient for *WTAT*Round* is negative and significant at for both Models 1 and 2.

<<Table 5 about here>>

Of course, an immediate riposte to this claim is that given higher average errors in WTAT this result is not surprising; errors decline more in WTAT because there is more scope for improvement in this treatment. In response, we note that average errors are at least as high in PR as WTAT, but learning in PR is less pronounced compared to WTAT. Further, as we show below, there are several other nuances to this finding that WTAT facilitates learning. We now turn to those factors.

3.3 Heterogeneous ability and learning

It turns out that learning in WTAT is driven to a large extent by improvements in performance by those who were not adept at the task to begin with. Given that the first 5 rounds in each treatment are identical, we can use performance in those rounds as a benchmark for a subject's basic facility with this task. We undertake a median split for absolute errors for the first five rounds. Then we calculate the median error for each subject.

If this is *larger (smaller)* than the overall median then we refer to this subject as a *low (high)* performer.[10]

We present our results broken up by ability level in Table 6. Models 1 and 2 refer to *high performers*, the former for rounds 6 through 20 and the latter for round 11 through 20. Models 3 and 4 do the same for *low performers*. We have already explained the rationale of looking at rounds 6 onwards and rounds 11 onwards separately. As before, for the ease of exposition, we only provide the coefficients for the interaction terms between the treatment dummies and round, while suppressing the other coefficients. It is clear that by and large the performance of high performers improved over time in all treatments, particularly so if we look at Model 2 for rounds 11 through 20. However, for low performer (Models 3 and 4) it is only the *WTAT*Round* that is negative and significant; again, this effect is particularly pronounced for rounds 11 through 20. So, the greater learning in WTAT is explained by the fact that while high performers learned to make better predictions in all treatments, it is only in WTAT that low performers also got better over time.

<<Table 6 about here>>

3.4 Predicted Errors in Round 20

A natural question to ask at this point is whether the learning in the WTAT treatment is enough to make up for the lower performance vis-à-vis the PRWL treatment. If learning occurs gradually round by round, as we model it, then we should expect the cumulative improvements in performance to be the greatest in the final round. Using parameters estimated from the regressions in Table 4, we can construct the predicted forecast errors in round 20 for each treatment *(k)*. We calculate the round 20 treatment differences as:

$$Predicted\ Round\ 20\ Forecast\ Error_k = \beta_0 + \beta_k + \beta_{k*Round} \times 20$$

[10] The overall median for the first 5 rounds is 21 for the dual cue task and 8 for the single cue task. So, for instance, in the dual cue task if a subject has a median error of more than 21 for the first 5 rounds then this subject will be labelled a low performer; a median error of less than 21 during the first 5 rounds means that this subject is a high performer.

Table 7 shows the predicted errors for round 20. Here, we generate the relevant numbers using the coefficients from the regressions run for (i) rounds 6 through 20 and (ii) rounds 11 through 20. (Not all of this data is provided in this paper, but are available from the authors). While the WTAT treatment performed poorly to start off with, the faster pace of learning makes it perform better than the PR treatment by the last round. In the single cue task, the predicted WTAT forecast errors at round 20 are lower than both PR and PRWL. However, in the dual cue task, WTAT errors in round 20 are smaller than those for PR, but not for PRWL. If we use the coefficients from the pooled data regressions for rounds 6 through 20 presented in Table 4, then our results suggest that the WTAT would out-perform PRWL, in terms of average errors, by round 27.

<<Table 7 about here>>

3.3. Dispersion of absolute errors

The above suggests that it is only in the WTAT treatment that absolute errors decline (and forecast accuracy increases) over time, demonstrating evidence of learning. If a subject displays significant learning over time, we would expect not only that their forecast accuracy improves, but also that their forecast errors become increasingly *consistent.* This is because a subject who learns to make better forecasts would be expected to settle on a forecast rule that converges to the underlying formula.

We use the *standard deviation of each subject's forecast errors* across time as our measure of consistency. Since we want to see how this changes over time, we break up the 15 post-intervention rounds (rounds 6 through 20) into 5 three-round blocks. We calculate the standard deviation for each subject for each of those three round-blocks. This yields 5 observations for each subject. We then estimate a random-effects regression where within-subject standard deviation is regressed against a time trend (denoted *"block"*, which takes the values 1, 2, 3, 4 or 5), two treatment dummies (with PR as the reference category), trait anxiety, gender and three additional terms interacting the treatment dummies with block to pick up differences in time trends across the different treatments. In Panel A of Table 8 we present the regression results. In Panel B, we present a selection of Wald tests pertaining to the interaction terms only.

<<Table 8 about here>>

Looking at Panel A one finds that compared to the reference category PR, standard deviations are initially higher in WTAT as shown by the positive value of the coefficient which is marginally significant. However, the coefficient for the interaction term involving WTAT is negative and significant at 5% indicating that standard deviations do decrease over time in this treatment. None of the other interaction terms are significant. Furthermore, Wald tests for time trends presented in Panel B suggest that the decrease in standard deviations in WTAT is significantly different from those for PR and PRWL. This result that the standard deviations decrease over time in WTAT, coupled with the previous results that absolute errors also decline over time in WTAT, provides evidence that it is only the WTAT treatment with its winner-take-all payoff structure that promotes learning over time.

3.4 What Drives Tournament Learning?

Why is learning more pronounced in the WTAT treatment than in other treatments? The answer lies in the rank-dependent reward structure, with the winner earning $1 while the loser earns nothing. Players improve their forecast errors in order to improve their chances of receiving the winning prize and/or improve their chances of avoiding the losing prize. Dutcher, Balafoutas, Lindner, Ryvkin, and Sutter (2015) show that the avoid-being-last objective has a greater effect than the strive-to-be-first objective in terms of eliciting effort. We believe that the prize structure of our tournament leads to a similar phenomenon.

In order to test this proposition more rigorously, we now turn to the WTAT-NI treatment, which features the exact same tournament incentives as the WTAT treatment. The only difference is that in the WTAT-NI treatment, there is no feedback pertaining to winning/losing or earnings, between rounds. This means that WTAT-NI subjects do not know whether they have won or lost the prior round of play. This, in turn implies, that if the WTAT-NI treatment leads to similar performance as WTAT, then it must be the case that the winner-take-all payment scheme, rather than relative feedback, is what is driving this result, since there is no feedback in the WTAT-NI scheme.

We present the estimated time trends for the WTAT and WTAT-NI treatments in Table 9. Here, we use data from the dual cue task only, since we ran the WTAT-NI treatment only for this task. Models 1 and 2 estimate a random effects specification with the WTAT-NI treatment dummy (with the WTAT treatment serving as the reference category), trait anxiety, gender and linear time trends for the T and WTAT-NI treatment. Regression models 3 and 4 run fixed effects regressions of the time trends, while getting rid of time-invariant variables. The regressions in models 1 and 3 are run over all post-intervention rounds 6 to 20, while models 2 and 4 repeat the same over rounds 11 to 20.

In model 1, across rounds 6 to 20, the linear trend for the WTAT treatment (*WTAT*Round*) has a slope of -0.494 and is significant with a p-value of 0.02. In the WTAT-NI treatment, we also see a significant downward trend of -0.447, with a p-value of 0.043. This is no different to that for the WTAT treatment. This implies that learning occurs even in the absence of relative performance feedback, lending more weight to the notion that the tournament incentives drive learning. We note that the coefficient for *WTAT*Round* is not significant in model 2, but a Wald test cannot reject the null hypothesis that the estimated trend lines for the WTAT and WTAT-NI treatment are the same. (p = 0.153). The fixed effects regressions in models 3 and 4 show similar results and the null hypotheses that the time trends for WTAT and WTAT-NI are equal are never rejected at conventional levels of significance.

Table 10 presents regressions of within-subject standard deviation of forecast errors for WTAT and WTAT-NI treatments. The regressors include the WTAT-NI treatment dummy, trait anxiety scores and gender of each subject, as well as a linear trend for T and WTAT-NI treatments, denoted as *WTAT*Block* and *WTAT-NI*Block* respectively. The first regression model is run across all round blocks, over rounds 6 to 20. We see that while the intercept is slightly lower in the WTAT-NI treatment compared to the WTAT treatment, although insignificantly so, consistency improves at a much faster rate in the WTAT treatment. The linear trend for the WTAT treatment has a slope of -1.838 and is significant at the 5% level. On the other hand, the WTAT-NI trend is negative but small and insignificant. A Wald test comparing the trends, however, does not reject the null that the trends are identical ($\chi^2(1) = 2.16$, $p = 0.142$).

Regression model 2 in Table 10 draws the trend lines over blocks 2 to 5. In other words, we are dropping 71 observations pertaining to the first block and estimating the trends starting from a different base. This allows us to test the sensitivity of these trends to their starting points. We note again that consistency improves in both WTAT and WTAT-NI treatments over time, but now at a faster rate in the WTAT-NI treatment. Despite these negative trends, neither are statistically significant at conventional levels. Wald tests again do not suggest any differences in these trends ($\chi^2(1) = 0.60$, $p = 0.437$). Finally, the third regression model in Table 10 replicates the previous regressions starting from the third block of rounds. Here, we observe an obvious downward sloping trend in the WTAT-NI treatment, with significance at better than 1%. Interestingly, we also see significant improvements in consistency for the WTAT treatment, also highly significant. As in the previous two regression models, hypothesis tests do not indicate any differences in the estimated trends for WTAT and WTAT-NI treatments ($\chi^2(1) = 0.73$, $p = 0.394$).

The above suggests that for both WTAT and WTAT-NI treatments, both the accuracy of forecasts and its consistency improves over time. More interestingly, the rates of learning are indistinguishable across these two treatments. Since the design differences between these treatments lie in the suppressed winning/losing feedback in the WTAT-NI treatment, the fact that both lead to similar patterns of learning, suggests that it is not the relative rank information that drives learning in the WTAT treatment. Rather, we can attribute learning to the fact that both of these treatments feature an all-or-nothing payoff structure. This rather extreme nature of the winner-take-all payment scheme seems to provide powerful incentives for players to improve their predictions over time, irrespective of whether subjects learn about winning/losing or their payoffs.

4. Concluding remarks

In this paper we have looked at the temporal dimension of learning across treatments and shown that the WTAT treatment stands out in terms of superior learning. Forecast accuracy improves at a significantly faster rate than any other treatment. Not only are forecasts becoming increasingly accurate in the WTAT treatment, we find evidence that these forecast errors are becoming increasingly consistent too. Both the improved accuracy and consistency of forecasts in the WTAT treatment constitute strong evidence for learning. The learning in the WTAT treatment is robust to various specifications and estimation methods and this learning is most pronounced for subjects who were initially not adept at the task.

It appears that the all or nothing payoff structure in WTAT is the driving force behind improved learning. Both WTAT and WTAT-NI provide rank dependent payoffs, but the WTAT-NI treatment withholds any feedback about winning or losing or the resulting payoffs. We see that the pattern of learning is similar in both treatments. This suggest that feedback about relative performance is less important for learning; it is the rank dependent payoffs that drive learning.

So where does it leave us in terms of the implications of this study? This depends on the extent to which our MCPL task mimics real-life work and which types of jobs are represented well by this task. It appears to us that most jobs in real life require some cognitive effort and require certain amount of learning by doing. To that extent a task like ours probably provides a better approximation of many work-places rather than the more mechanical number adding type tasks used in most prior studies. It also depends on the primary aim of an employer relying on a task of this nature. If the primary emphasis is on learning over time, then a tournament type scheme seems to provide better incentives for this.

References

Aoyagi, Masaki. (2010). Information Feedback in a Dynamic Tournament. *Games and Economic Behavior, 70*(2), 242-260.

Ariely, Dan, Gneezy, Uri, Loewenstein, George, & Mazar, Nina. (2009). Large Stakes and Big Mistakes. Review of Economic Studies, 76(2), 451-469.

Azmat, Ghazala, & Iriberri, Nagore. (2010). The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students. Journal of Public Economics, 94(7-8), 435-452.

Azmat, Ghazala, & Iriberri, Nagore. (2016). The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction. *Journal of Economics & Management Strategy, 25*(1), 77-110.

Balzer, William K., Doherty, Michael E., & O'Connor, Raymond. (1989). Effects of Cognitive Feedback on Performance. *Psychological Bulletin, 106*(3), 410-433.

Baumeister, Roy F. (1984). Choking under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance. Journal of Personality and Social Psychology, 46(3), 610-620.

Bellemare, Charles, Lepage, Patrick, & Shearer, Bruce. (2010). Peer Pressure, Incentives, and Gender: An Experimental Analysis of Motivation in the Workplace. Labour Economics, 17(1), 276-283.

Blanes i Vidal, Jordi, & Nossol, Mareike. (2011). Tournaments Without Prizes: Evidence from Personnel Records. Management Science, 57(10), 1721-1736.

Bowles, Samuel, & Polanía-Reyes, Sandra. (2012). Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature, 50*(2), 368-425.

Brown, P., Cameron, L., Chaudhuri, A. and So, T. (2016). Rank as incentive: Experimental Evidence from a Cognitively Challenging Task. Working Paper, Department of Economics, University of Auckland, September 2016.

Brown, Paul M. (1995). Learning from Experience, Reference Points, and Decision Costs. Journal of Economic Behavior & Organization, 27(3), 381-399.

Brown, Paul M. (1998). Experimental Evidence on the Importance of Competing for Profits on Forecasting Accuracy. Journal of Economic Behavior & Organization, 33(2), 259-269.

Bull, Clive, Schotter, Andrew, & Weigelt, Keith. (1987). Tournaments and Piece Rates: An Experimental Study. Journal of Political Economy, 95(1), 1-33.

Cadsby, C. Bram, Engle-Warnick, Jim, Fang, Tony, & Song, Fei. (2010). Psychological Incentives, Financial Incentives, and Risk Attitudes in Tournaments: An Artefactual Field Experiment

Charness, Gary, Masclet, David, & Villeval, Marie Claire. (2014). The Dark Side of Competition for Status. Management Science, 60(1), 38-55.

Deci, Edward L., Betley, Gregory, Kahle, James, Abrams, Linda, & Porac, Joseph. (1981). When Trying to Win: Competition and Intrinsic Motivation. Personality and Social Psychology Bulletin, 7(1), 79-83.

Deci, Edward L., Koestner, Richard, & Ryan, Richard M. (1999). A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. Psychological Bulletin, 125(6), 627-668.

Dutcher, E. Glenn, Balafoutas, Loukas, Lindner, Florian, Ryvkin, Dmitry, & Sutter, Matthias. (2015). Strive to be First or Avoid being Last: An Experiment on Relative Performance Incentives. *Games and Economic Behavior, 94*, 39-56.

Ehrenberg, Ronald G., & Bognanno, Michael L. (1990). Do Tournaments have Incentive Effects? Journal of Political Economy, 98(6), 1307-1324.

Ellingsen, Tore, & Johannesson, Magnus. (2007). Paying Respect. *Journal of Economic Perspectives, 21*(4), 135-149.

Eriksson, Tor. (1999). Executive Compensation and Tournament Theory: Empirical Tests on Danish Data. Journal of Labor Economics, 17(2), 262-280.

Eriksson, Tor, Poulsen, Anders, & Villeval, Marie-Claire. (2009). Feedback and Incentives: Experimental Evidence. Labour Economics, 16(6), 679-688.

Falk, Armin, & Ichino, Andrea. (2006). Clean Evidence on Peer Effects. Journal of Labor Economics, 24(1), 39-57.

Festré, Agnès, & Garrouste, Pierre. (2015). Theory and Evidence in Psychology and Economics about Motivation Crowding Out: A Possible Convergence? *Journal of Economic Surveys, 29*(2), 339-356.

Frey, Bruno S., & Jegen, Reto. (2001). Motivation Crowding Theory. Journal of Economic Surveys, 15(5), 589-611.

Gneezy, Uri, Meier, Stephan, & Rey-Biel, Pedro. (2011). When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives, 25*(4), 191-210.

Gneezy, Uri, & Rustichini, Aldo. (2000a). Pay Enough or Don't Pay at All. Quarterly Journal of Economics, 115, 791-810.

Gneezy, Uri, & Rustichini, Aldo. (2000b). A Fine is a Price. Journal of Legal Studies, 29(1), 1-17.

Green, Jerry R., & Stokey, Nancy L. (1983). A Comparison of Tournaments and Contracts. Journal of Political Economy, 91(3), 349-364.

Hannan, R. Lynn, Krishnan, Ranjani, & Newman, Andrew H. (2008). The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. *The Accounting Review, 83*(4), 893-913

Heyman, James, & Ariely, Dan. (2004). Effort for Payment: A Tale of Two Markets. Psychological Science, 15(11), 787-793.

Iyengar, Raghuram, & Schotter, Andrew. (2008). Learning under Supervision: An Experimental Study. Experimental Economics, 11(2), 154-173.

Kluger, Avraham N., & DeNisi, Angelo. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin, 119*(2), 254-284.

Kräkel, Matthias. (2008). Emotions in Tournaments. Journal of Economic Behavior & Organization, 67(1), 204-214.

Kristensen, Søren Rud, Meacock, Rachel, Turner, Alex J., Boaden, Ruth, McDonald, Ruth, Roland, Martin, & Sutton, Matthew. (2014). Long-Term Effect of Hospital Pay for Performance on Mortality in England. *New England Journal of Medicine, 371*(6), 540-548.

Kuhnen, Camelia M., & Tymula, Agnieszka. (2012). Feedback, Self-Esteem, and Performance in Organizations. Management Science, 58(1), 94-113.

Lazear, Edward P. (2000). Performance Pay and Productivity. American Economic Review, 90(5), 1346-1361.

Lazear, Edward P., & Rosen, Sherwin. (1981). Rank-Order Tournaments as Optimum Labor Contracts. Journal of Political Economy, 89(5), 841-864.

Mas, Alexandre, & Moretti, Enrico. (2009). Peers at Work. American Economic Review, 99(1), 112-145.

Masclet, David, Peterle, Emmanuel, & Larribeau, Sophie. (2015). Gender Differences in Tournament and Flat-Wage Schemes: An Experimental Study. Journal of Economic Psychology, 47, 103-115.

McNaughton, Neil. (2011). Trait Anxiety, Trait Fear and Emotionality: The Perspective from Non-Human Studies. Personality and Individual Differences, 50, 898-906.

Merlo, Antonio, & Schotter, Andrew. (1999). A Surprise-Quiz View of Learning in Economic Experiments. Games and Economic Behavior, 28(25-54).

Nalebuff, Barry J., & Stiglitz, Joseph E. (1983). Prizes and Incentives: Towards a General Theory of Compensation and Competition. Bell Journal of Economics, 14(1), 21-43.

Niederle, Muriel, & Vesterlund, Lise. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *Quarterly Journal of Economics, 122*(3), 1067-1101.

Prendergast, Canice. (1999). The Provision of Incentives in Firms. Journal of Economic Literature, 37(1), 7-63.

Rosenthal, Meredith B., Frank, Richard G., Li, Zhonghe, & Epstein, Arnold M. (2005). Early Experience with Pay-for-Performance: From Concept to Practice. *Journal of the American Medical Association, 294*(14), 1788-1793.

Segal, Jan D., & Weinberg, Robert S. (1984). Sex, Sex Role Orientation and Competitive Trait Anxiety. *Journal of Sport Behavior, 7*(4), 153-159.

Spielberger, Charles D., Gorsuch, Richard L., Lushene, Robert E., Vagg, Peter R., & Jacobs, Gerard A. (1983). Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Taylor, Beck A., & Trogdon, Justin G. (2002). Losing to Win: Tournament Incentives in the National Basketball Association. Journal of Labor Economics, 20(1), 23-41.

Tran, Anh, & Zeckhauser, Richard. (2012). Rank as an Inherent Incentive: Evidence from a Field Experiment. Journal of Public Economics, 96(9-10), 645-650.

Vandegrift, Donald, & Brown, Paul M. (2003). Task Difficulty, Incentive Effects, and the Selection of High-Variance Strategies: An Experimental Examination of Tournament Behavior. Labour Economics, 10(4), 481-497.

Vandegrift, Donald, Yavas, Abdullah, & Brown, Paul M. (2007). Incentive Effects and Overcrowding in Tournaments: An Experimental Analysis. Experimental Economics, 10(4), 345-368.

Yerkes, Robert M., & Dodson, John D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. Journal of Comparative Neurology of Psychology, 18(5), 459-482.

Table 1: Actual Cues and Stock Prices

Round	Single Cue Task			Dual Cue Task		
	Cue A	Cue B	Stock Price	Cue A	Cue B	Stock Price
1	150	201	192	105	37	69
2	150	263	243	242	96	151
3	150	88	117	443	159	256
4	150	248	232	1	339	245
5	150	201	200	41	146	124
6	150	196	194	155	32	80
7	150	353	305	20	288	223
8	150	173	173	104	422	335
9	150	270	248	102	107	112
10	150	243	222	296	188	231
11	150	60	102	413	266	321
12	150	320	274	165	412	353
13	150	340	289	172	167	174
14	150	361	311	359	262	298
15	150	321	285	271	418	385
16	150	361	309	227	31	98
17	150	148	155	381	435	426
18	150	309	275	262	339	323
19	150	135	145	316	92	164
20	150	142	156	196	285	269

Table 2: Cue Values given to subjects as practice examples

Single Cue Task			Dual Cue Task		
Cue A	Cue B	Actual Price	Cue A	Cue B	Actual Price
150	92	117	12	64	54
150	143	157	372	63	162
150	379	321	179	109	137
150	373	313	415	146	240
150	240	220	116	186	175
150	285	256	355	223	275
150	187	188	145	286	255
150	143	153	199	356	317
150	191	185	439	354	372
150	361	311	73	442	345

Table 3: Average errors across treatments

Treatments	Single Cue			Dual Cue			Pooled	
	Session	**N**	**Avg Error**	**Session**	**N**	**Avg Error**	**N**	**Avg Error**
Piece rate (PR)	1 2	16 26	10.2	1 2	20 19	26.6	81	18.1
Piece rate win lose (PRWL)	1 2	22 20	9.6	1 2	20 15	24.0	77	16.2
Winner take all tournament (WTAT)	1 2	16 24	10.0	1 2	18 20	30.7	78	20.1
Tournament no information (WTAT-NI)	NA	NA	NA	1 2	20 18	27.74	38	NA
Total		**124**			**150**		**274**	

Table 4: Pooled regressions for absolute errors; Columns (1) to (3) present results for random effects regressions while Column (4) presents results for a quantile regression

Dependent variable = Absolute errors = |Predicted value – Actual Value|

Independent variables	Model 1	Model 2	Model 3	Model 4
	Pooled	Pooled	Pooled	Pooled
	Random Effects	Random Effects	Random Effects	Quantile Regression
PRWL	-1.897 (2.186) [0.385]	*-4.388* *(2.425)* *[0.070]*	*-6.267* *(3.200)* *[0.050]*	*-1.318* *(0.743)* *[0.076]*
WTAT	2.052 (2.740) [0.454]	0.503 (2.958) [0.865]	3.181 (3.390) [0.358]	-0.442 (0.750) [0.556]
Female	---	*8.317* *(1.928)* *[0.000]*	*8.317* *(1.929)* *[0.000]*	*3.496* *(0.610)* *[0.000]*
Trait anxiety	---	0.162 (0.150) [0.281]	0.162 (0.150) [0.281]	*0.085* *(0.044)* *[0.052]*
Round	*-0.136* *(0.075)* *[0.070]*	*-0.161* *(0.082)* *[0.049]*	---	-0.093 (0.070) [0.182]
PR X Round	---	---	-0.142 (0.146) [0.331]	---
PRWL X Round	---	---	0.003 (0.153) [0.984]	---
WTAT X Round	---	---	*-0.348* *(0.119)* *[0.004]*	---
Constant	*19.83* *(1.878)* *[0.000]*	*10.90* *(6.164)* *[0.077]*	*10.65* *(6.630)* *[0.108]*	*5.915* *(2.052)* *[0.004]*
R^2	0.004	0.033	0.034	0.007 (Pseudo R^2)
Wald χ^2	6.27	29.93	38.92	---
$p > \chi^2$	0.099	0.00	0.00	---
No. of observations	3540	3180[11]	3180	3180
No. of participants	236	212	212	212
Wald test for PRWL = WTAT	$\chi^2 = 2.71$ p = 0.100	*$\chi^2 = 4.06$* *p = 0.044*	*$\chi^2 = 7.70$* *p = 0.006*	F = 1.43 p = 0.231

Standard errors in parentheses; p-values in square brackets.

[11] 24 subjects did not either fill in the trait anxiety questionnaire or provide gender information or both. This results in a loss of 360 observations (24 decisions per round for 15 rounds.)

Table 5: Learning over time; random effects regression for absolute errors for (i) rounds 6 through 20 (Column (1)) and (ii) for rounds 11 through 20 (Column (2)).

Dependent variable = Absolute errors = |Predicted value – Actual Value|

Independent variables include	**Model 1 Rounds 6 – 20**	**Model 2 Rounds 11 – 20**
PR*Round	-0.142 (0.146) [0.331]	-0.214 (0.226) [0.343]
PRWL*Round	0.003 (0.153) [0.984]	-0.349 (0.339) [0.303]
WTAT*Round	*-0.346* *(0.119)* *[0.004]*	*-0.776* *(0.187)* *[0.000]*
R^2	0.034	0.033
Wald chi-square	38.92	46.94
Probability > chi-square	0.000	0.000
Observations	3180	2120
Subjects	212	212

Standard errors in parentheses; p-values in square brackets.

Table 6: Learning over time; random effects regression for absolute errors for (i) high performers rounds 6 through 20 (Column (1)), (ii) high performers rounds 11 through 20 (Column (2)); (iii) low performers rounds 6 through 20 (Column (3)) and (iv) low performers for rounds 11 through 20 (Column (4)).

Dependent variable = Absolute errors = |Predicted value – Actual Value|

Panel A: Regression output				
	High Performers		**Low Performers**	
Independent variables include	**Model 1 Rounds 6 – 20**	**Model 2 Rounds 11 – 20**	**Model 3 Rounds 6 – 20**	**Model 4 Rounds 11 – 20**
PR*Round	*-0.346* *(0.124)* *[0.005]*	*-0.381* *(0.184)* *[0.039]*	0.088 (0.272) [0.746]	-0.027 (0.433) [0.951]
PRWL*Round	-0.158 (0.108) [0.141]	*-0.644* *(0.308)* *[0.036]*	0.210 (0.319) [0.511]	0.030 (0.662) [0.964]
WTAT*Round	*-0.352* *(0.138)* *[0.011]*	*-0.761* *(0.172)* *[0.000]*	*-0.343* *(0.200)* *[0.087]*	*-0.791* *(0.344)* *[0.021]*
R^2	0.025	0.026	0.032	0.033
Wald chi-square	23.58	35.59	16.35	21.98
Probability > chi-square	0.001	0.000	0.022	0.003
Observations	1710	1140	1470	980
Subjects	114	114	98	98

Standard errors in parentheses; p-values in square brackets.

Table 7: Predicted Forecast Errors in Round 20

Panel A: Rounds 6 to 20

	Pooled	Single Cue	Dual Cue
Piece Rate	11.201	11.373	17.274
Piece Rate Win Lose	7.977	11.089	14.025
Tournament	10.420	10.788	17.000

Panel B: Rounds 11 to 20

	Pooled	Single Cue	Dual Cue
Piece Rate	10.384	7.485	18.605
Piece Rate Win Lose	6.245	6.554	14.530
Tournament	8.281	5.142	17.414

Table 8: Learning over time; random effects regression for within-subject standard deviations

Panel A: Regression output	
	Pooled data for Rounds 6 - 20
PRWL	-2.621 (2.430) [0.281]
WTAT	*4.908* *(2.796)* *[0.079]*
PR*Block	0.535 (0.536) [0.318]
PRWL*Block	0.572 (0.578) [0.322]
WTAT*Block	*-1.211* *(0.491)* *[0.014]*
Trait Anxiety	0.084 (0.104) [0.420]
Female	*5.296* *(1.416)* *[0.000]*
Constant	7.447 (4.682) [0.112]
Observations	1060
Subjects	212
R^2	0.034
Wald χ^2	23.24
$p > \chi^2$	0.002
Panel B: Time trends (Wald test)	
PR*Block = PRWL*Block	$\chi^2(1) = 0.00$ p = 0.963
PRWL*Block = T*Block	***$\chi^2(1) = 5.53$*** ***p = 0.019***
PR*Block = T*Block	***$\chi^2(1) = 5.77$*** ***p = 0.016***

Standard errors in parentheses; p-values in square brackets.

Table 9: Time Trends in Dual Cue WTAT and WTAT-NI Treatments

	Random Effects		Fixed Effects	
	Rounds 6-20	Rounds 11-20	Rounds 6-20	Rounds 11-20
Dep Var: Forecast Errors	Model 1	Model 2	Model 3	Model 4
WTAT	(base)	(base)		
WTAT-NI	-4.659 (5.959) [0.434]	10.78 (12.22) [0.378]		
Trait Anxiety	0.412 (0.287) [0.151]	0.379 (0.333) [0.256]		
Female	*9.070* *(4.420)* *[0.040]*	*10.27* *(4.604)* *[0.026]*		
WTAT*Round	*-0.494* *(0.213)* *[0.020]*	-0.347 (0.295) [0.240]	*-0.481* *(0.201)* *[0.020]*	*-0.333* *(0.279)* *[0.236]*
WTAT-NI*Round	*-0.447* *(0.220)* *[0.043]*	*-1.226* *(0.540)* *[0.023]*	*-0.495* *(0.205)* *[0.018]*	*-1.311* *(0.506)* *[0.011]*
Constant	13.57 (13.38) [0.311]	11.85 (16.26) [0.466]		
Observations	1065	710	1140	760
Subjects	71	71	76	76
R^2	0.027	0.028	0.004	0.002
Wald χ^2	16.06	10.08		
$p > \chi^2$	0.007	0.073		
F			5.78	4.07
p > F			0.005	0.021
WTAT*Round = WTAT-NI*Round	$\chi^2(1) = 0.02$ p = 0.877	$\chi^2(1) = 2.04$ p = 0.153	F = 0.00 p = 0.959	F = 2.86 p = 0.095

Regressions are run with observations from the dual cue T and WTAT-NI treatments. Standard errors are clustered by subjects. Standard errors in parentheses; p-values in square brackets.

Table 10: Within-Standard Deviation Time Trends in WTAT and WTAT-NI Treatments

	Blocks 1-5 (Rounds 6-20)	Blocks 2-5 (Rounds 9-20)	Blocks 3-5 (Rounds 12-20)
Dep Var: Within-Subject Std Dev	Model 1	Model 2	Model 3
WTAT	(base)	(base)	(base)
WTAT-NI	-4.858 (4.608) [0.292]	6.014 (7.585) [0.428]	13.54 (16.11) [0.401]
Trait Anxiety	0.287 (0.221) [0.195]	0.255 (0.261) [0.328]	0.299 (0.290) [0.301]
Female	4.561 (3.524) [0.196]	5.019 (4.088) [0.220]	5.955 (4.371) [0.173]
WTAT*Block	*-1.838* *(0.896)* *[0.040]*	-0.199 (0.871) [0.819]	*-5.289* *(1.703)* *[0.002]*
WTAT-NI*Block	-0.179 (0.689) [0.795]	-1.291 (1.103) [0.242]	*-8.134* *(2.866)* *[0.005]*
Constant	11.72 (10.76) [0.276]	6.211 (12.01) [0.605]	*25.70* *(15.22)* *[0.091]*
Observations	355	284	213
Subjects	71	71	71
R^2	0.027	0.024	0.084
Wald χ^2	7.61	3.20	19.32
$p > \chi^2$	0.179	0.668	0.002
WTAT*Block = WTAT-NI*Block	$\chi^2(1) = 2.16$ $p = 0.142$	$\chi^2(1) = 0.60$ $p = 0.437$	$\chi^2(1) = 0.73$ $p = 0.394$

Figure 1: Screenshot for PR and FS treatments

Round	Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round
1	105	37	75	69	6	$0.94
2	242	96	201	151	50	$0.50
3	443	159	174	256	82	$0.18
4	1	339	268	245	23	$0.77
5	41	146	113	124	11	$0.89
6	155	32	116	80	36	$0.64
7	20	288	241	223	18	$0.82
8	104	422	315	335	20	$0.80
9	102	107	108	112	4	$0.96

Round number	10
Player name	Tso1
Player ID	1
Cue A	296
Cue B	188
Enter Forecast	
SUBMIT	RESET

Figure 2: Screenshot for PRWL treatment

Round	Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round	WIN or LOSE
1	105	37	75	69	6	$0.94	---
2	242	96	201	151	50	$0.50	---
3	443	159	174	256	82	$0.18	---
4	1	339	268	245	23	$0.77	---
5	41	146	113	124	11	$0.89	---
6	155	32	116	80	36	$0.64	WIN
7	20	288	241	223	18	$0.82	WIN
8	104	422	315	335	20	$0.80	LOSE
9	102	107	108	112	4	$0.96	WIN

Round number	10
Player name	Tso1
Player ID	1
Cue A	296
Cue B	188
Enter Forecast	
SUBMIT	RESET

Figure 3: Screenshot for WTAT treatment

Round	Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round	WIN or LOSE
1	105	37	75	69	6	0.94	---
2	242	96	201	151	50	0.5	---
3	443	159	174	256	82	0.18	---
4	1	339	268	245	23	0.77	---
5	41	146	113	124	11	0.89	---
6	155	32	116	80	36	$1	WIN
7	20	288	241	223	18	$1	WIN
8	104	422	315	335	20	$0	LOSE
9	102	107	108	112	4	$1	WIN

Round number	10
Player name	Tso1
Player ID	1
Cue A	296
Cue B	188
Enter Forecast	
SUBMIT	RESET

Figure 4: Screenshot for WTAT-NI treatment

Round	Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round	WIN or LOSE
1	105	37	75	69	6	0.94	
2	242	96	201	151	50	0.5	
3	443	159	174	256	82	0.18	
4	1	339	268	245	23	0.77	
5	41	146	113	124	11	0.89	
6	155	32	116	80	36		
7	20	288	241	223	18		
8	104	422	315	335	20		
9	102	107	108	112	4		

Round number	10
Player name	Tso1
Player ID	1
Cue A	296
CueB	188

Enter Forecast

SUBMIT RESET

Appendix: Instructions

The University of XXXX
Instructions for the Experiment

WELCOME.

PLEASE TURN YOUR CELL PHONES OFF NOW.

This is a study examining the manner in which people make decisions. The University of XXXX has provided the funds to conduct this research. If you follow the instructions and make good decisions you might earn a considerable amount of money.

At the beginning of the session each person will be given an Earnings Account with $5.00 in it. You will participate in a decision making task for each of 20 rounds. You will have the chance to earn money each round, with your earnings for each round being added to your Earnings Account. At the end of the experiment, the balance of your Earnings Account will be paid to you in cash.

DESCRIPTION OF THE TASK:

In each round you will be asked to predict the future value of a fictitious **'stock'**. The value of this stock is unknown to all subjects, but you will be able to observe two CUES that can help you form your forecast. These cues can be used to predict the stock's value much the same way that the amount of rainfall and the average temperature can be used to predict the quality of a corn crop, the number of unoccupied apartments and student enrolment this year can be used to predict next year's rent increases, or the demand for sports cars can be used to predict their future price.

In each round you will be shown the values for the two CUES.

NOTE: One of the CUE values will always be fixed at 150 for each of the 20 rounds. The other cue value will change each round. But the relation of the cue values to the stock's price will remain the same.

Example:

For example let the value of Cue A be fixed at 150. Suppose the values for the cues in a round were given as:

CUE A = 150
CUEB = 100

You will be asked to predict the price of the stock given these two cue values. The next round one of the cues will take on a different value, such as:

CUE A = 150
CUEB = 450

You will then predict that round's price using these new cue values. Remember that even though the values of the cues change, the underlying relation between the cue values and the stock's price remains the same. Thus, in order to make accurate forecasts you will need to determine the relation between the cues and the price of the stock.

YOUR FORECASTING ERROR

After making your forecast, the computer will calculate the distance between your forecast and that round's actual price (your absolute forecasting error). This amount will be your **forecasting error**.

Example:

Suppose your forecast was 230. If the actual price of the stock was 200 then your forecasting error would be 30:

Your forecasting error = 230 – 200 = 30

Suppose your forecast was 148. If the actual price of the stock was 200 then your forecasting error would be 52:

Your forecasting error = 200 – 148 = 52

YOUR EARNINGS IN EACH ROUND:

In each round, your earnings will depend on your forecasting error. Your earnings in each round will equal $1 **less** your forecasting error for that round.

That is, your earnings (E) in each round will be given by E = $1.00 – (forecast error).

Example:

Suppose your forecast error in a particular round is 30. Then you will earn $0.70 in that round. This is because:

$1.00 – $0.30 = $0.70

Suppose in another round your forecast error is 8. Then you will earn $0.92 in that round. This is because:

$1.00 - $0.08 = $0.92

Note that if your error is 100 or over, then you will earn nothing in that round. The minimum amount you can earn in a round is $0.00.

Suppose in another round your forecast error is 102. Then you will earn $0.00 because:

$1.00 - $1.00 = $0.00

SPECIFIC INSTRUCTIONS:

1. Before Round 1:

You will be shown 10 examples of cues and stock prices. You will have 5 minutes in which to examine these examples.

2. Round 1:

At the end of the 5-minute example round you will be shown the first two cue values and asked to forecast the price of the stock in Round 1. You will have **90 seconds** to make your forecast.

3. End of Round 1:

At the end of the **90 seconds** all subjects will have entered their forecasts. After all earnings have been calculated you will be shown your results for Round 1. The computer will then show you your earnings for the round, including:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earning this Round	Total Earnings

Please record this information on the RECORD SHEET provided to you.

4. Beginning of Round 2:

After examining and recording the earnings from round 1, you will be shown the values of CUE A and CUE B in Round 2. You will have **90 seconds** to form your forecast.

5. Subsequent Rounds:

Each subsequent round proceeds in the same way and will be repeated for each of the 20 rounds. In each round, you will make a forecast based on two new cue values. At the end of round 20, you will receive a cash payment in the amount indicated by the earnings account.

However, after you have finished the first five rounds of play, we will have a pause. It is possible that there will be a change in the way in which you earn money for the subsequent rounds 6 through 20. If there is no change then we will tell you so and ask you to simply continue playing the game in the same manner as in the first five rounds. However, if there is a change in payment, then we will provide you with further instructions at that point and explain these changes and also answer any questions you may have.

PRWL Specific Instructions

Rounds 6 to 20

Rounds 6 to 20 are played exactly as rounds 1 to 5 but with the following exceptions:

- Each period you will be paired with another subject in the session today. Your partner will change each round, so you will never be paired with the same partner for more than one consecutive rounds;

- After you have made your forecast, the computer will compare your forecasting error to your partner's forecasting error in that round;

- Your results will show whether your forecasting error was greater or less than your partner's for that round;

- If your error is less than your partner's, then you will be told you WIN that round. If your error is more than your partner's, you will be told you LOST that round. If your error is equal to your partner's, then the computer will randomly decide the winner and loser.

- Your payment will remain unchanged. That is, each round you will continue to be paid:

  Earnings = $1.00 – Forecasting Error

- You will also be shown your partner's forecast and forecasting error at the end of the round. That is, at the end of each round you will observe:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round	WIN or LOSE

Example: Suppose the actual price was 210, your forecast was 168, and your partner's forecast was 163. Your forecasting error would be 42 and your partner's forecasting error would be 47. You would see the following results for that round:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round	WIN or LOSE
		168	210	42	$0.58	WIN

Example: Suppose the actual price was 210, your forecast was 168, and your partner's forecast was 173. Your forecasting error would be 42 and your partner's forecasting error would be 37. You would see the following results for that round:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round	WIN or LOSE
		168	210	42	$0.58	LOSE

Do you have any questions?

WTAT Specific Instructions

Rounds 6 to 20

Rounds 6 to 20 are played exactly as rounds 1 to 5 but with the following exceptions:

- You will have an additional $4.00 added to your Earnings Account;

- Each period you will be paired with another subject in the session today. Your partner will change each round, so you will never be paired with the same partner for more than one consecutive rounds;

- After you have made your forecast, the computer will compare your forecasting error to your partner's forecasting error in that round;

- Your results will show whether your forecasting error was greater or less than your partner's for that round;

- If your error is less than your partner's, then you will be told you WIN that round. If your error is more than your partner's, you will be told you LOST that round. IF your error is equal to your partner's, then the computer will randomly decide the winner and loser;

- Your payment will depend upon whether your forecasting error is greater or less than your partners. That is, each round you will earn either $1.00 or $0.00. You will be paid either:

Earnings = $1.00 if you WIN

Or

Earnings = $0.00 if you LOSE

Example: Suppose your forecasting error was 42 and your partner's forecasting error was 47. You would see the following results for that round:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round	WIN or LOSE
				42	$1.00	Win

Example: Suppose your forecasting error was 42 and your partner's forecasting error was 37. You would see the following results for that round:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round	WIN or LOSE
				42	$0.00	LOSE

Do you have any questions?

WTAT-NI Specific Instructions

Rounds 6 to 20 are played exactly as rounds 1 to 5 but with the following exceptions:

- You will have an additional $4.00 added to your Earnings Account;
- Each period you will be paired with another participant in the session today. Your partner will change each round, so you will never be paired with the same partner more than once;
- After you have made your forecast, the computer will compare your forecasting error to your partner's forecasting error in that round;
- If your error is less than your partner's, then you will WIN that round. If your error is more than your partner's, you will LOSE that round. If your error is equal to your partner's, then the computer will randomly decide the winner and loser.
- Your payment will depend upon whether your forecasting error is greater or less than your partners. That is, each round you will earn either $1.00 or $0.00:

  Earnings = $1.00 if you WIN

  Or

  Earnings = $0 if you LOSE

- You will not know whether you won or lost until the end of the 20th round. That is, at the end of each round you will see the following information:

Cue A	Cue B	Your Forecast	Actual Price	Forecast ing Error

At the end of the 20th round, you will see the following information for each round:

Cue A	Cue B	Your Forecast	Actual Price	Forecast ing Error	Earning this round	WIN or LOSE
					$1.00	WIN

Whether you won or lost in each round will only be known at the end of the 20th round.

- ***Example***: Suppose your forecasting error was 42 and your partner's forecasting error was 47. At the end of that round you would observe:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earning this round	WIN or LOSE
				42		

At the end of the 20th round, you would observe:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earning this round	WIN or LOSE
				42	$1.00	WIN

- ***Example***: Suppose your forecasting error was 42 and your partner's forecasting error was 37. You would see the following results for that round:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earning this round	WIN or LOSE
				42		

At the end of the 20th round, you would observe:

Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earning this round	WIN or LOSE
				42	$0.00	LOSE

Do you have any questions?

Pre-Task Questionnaire

Player ID ____________

PLEASE ANSWER ALL OF THE FOLLOWING QUESTIONS

A number of statements which people have used to describe themselves are given below. Read each statement and, using the scale below, tick the appropriate number indicating **how you generally feel**. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe **how you generally feel**.

1	2	3	4
Almost never	Sometimes	Often	Almost always

	Almost Never	Sometimes	Often	Almost always
	1	2	3	4
1. I feel pleasant				
2. I tire quickly				
3. I feel like crying				
4. I wish I could be as happy as others seem to be				
5. I am losing out on things because I can't make up my mind soon enough				
6. I feel rested				
7. I am "calm, cool and collected"				
8. I feel that difficulties are piling up so that I cannot overcome them				
9. I worry too much over something that doesn't really matter				
10. I am happy				
11. I am inclined to take things hard				
12. I lack self-confidence				
13. I feel secure				
14. I try to avoid facing a crisis or difficulty				
15. I feel blue				
16. I am content				
17. Some unimportant thoughts run through my mind and bother me				
18. I take disappointments so keenly that I can't put them out of my mind				
19. I am a steady person				
20. I get in a state of tension or turmoil as I think over my recent concerns and interests				

Post-Task Questionnaire

Player ID ____________

PLEASE ANSWER ALL OF THE FOLLOWING QUESTIONS

A) For each of the following statements, please indicate how true the statement is for you using the following scale:

1	2	3	4	5	6	7
Not at all true			Somewhat true			Very true

	Not at all true			Somewhat true			Very true
	1	2	3	4	5	6	7
1. I enjoyed this activity very much							
2. I think I am pretty good at this activity							
3. I put a lot of effort into this activity							
4. I did not feel nervous at all which doing this activity							
5. This activity was fun to do							
6. I think I did pretty well at this activity, compared to other subjects							
7. I did not try very hard to do well at this activity							
8. I felt very tense while doing this activity							
9. I thought this activity was boring							
10. After working at this activity for a while, I felt pretty competent							
11. I tried very hard on this activity							
12. I was very relaxed doing this activity							
13. This activity did not hold my attention							
14. I am satisfied with my performance at this task							
15. It was important to me to do well at this task							
16. I was anxious while working on this task							
17. I would describe this activity as very interesting							
18. I was pretty skilled at this activity							
19. I did not put much energy into this							
20. I felt pressured while doing this activity.							
21. I thought this activity was quite enjoyable.							
22. This was an activity that I could not do very well.							
23. While I was doing this activity, I was thinking about how much I enjoyed it.							

B) The following items ask about how you felt about the other subjects during the session.

	Not at all true			Somewhat true			Very true
	1	2	3	4	5	6	7
1. I felt really distant to them							
2. I really doubt they and I would ever be friends							
3. I felt I could really trust them							
4. I'd really like the chance to interact with them more often							
5. I'd really prefer not to interact with them in the future							
6. I don't feel like I could really trust them							
7. It is likely that they and I could become friends if we interacted a lot							
8. I felt close to them							

C) How many of the people in this session did you know before the experiment? ____________

D) Basic information about you:

Your Gender (Male/ Female) ________

Age _________

Major: ____

Year in School (e.g., Stage 2) ______________________________

Ethnicity (Please circle one): Maori Pacific Island NZ European

Asian Other _______________

Country where you were born? __________________

If you were born outside of New Zealand, at what age did you move here? _________