Journal of Economic Behavior & Organization 142 (2017) 11–23

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

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

Tony So[a,b], Paul Brown[c], Ananish Chaudhuri[a,*], Dmitry Ryvkin[d], Linda Cameron[c]

[a] Department of Economics, University of Auckland, New Zealand
[b] Economics Division, International Business School Suzhou, Xi'an Jiaotong-Liverpool University, China
[c] School of Social Sciences, Humanities and the Arts, University of California, Merced, United States
[d] Department of Economics, Florida State University, United States

ARTICLE INFO

Article history:
Received 7 March 2017
Received in revised form 13 July 2017
Accepted 14 July 2017
Available online 21 July 2017

JEL codes:
C91
J24
J33
J39

Keywords:
Experiment
Payment scheme
Tournament
Piece-rate
Learning

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 noisy 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., the absolute distance between the actual and predicted values of X. We implement four treatments which are based on two different payment schemes: (1) piece rates, where subjects are paid linearly 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 a smaller absolute error earns a fixed payoff, while the other earns nothing. We find that it is only in the tournament payment scheme, and particularly in a more complex version of the task, that subjects show significant evidence of learning over time, in that their predictions get closer to the actual value of X. This learning process is driven by the all-or-nothing nature of the payoff structure in tournaments.

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 cardinal 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 (e.g., Lazear and Rosen, 1981; Green and Stokey, 1983; Nalebuff and Stiglitz, 1983) show that in many cases tournaments are effective in eliciting effort at a level analogous to piece rates. This insight is borne out by results in a classic laboratory experiment by Bull et al. (1987), where they show that, on average,

* Corresponding author at : Department of Economics, University of Auckland, 660 Owen G Glenn Building, 12 Grafton Road, Auckland 1142, New Zealand.

E-mail addresses: tony.so@xjtlu.edu.cn (T. So), pbrown3@ucmerced.edu (P. Brown), a.chaudhuri@auckland.ac.nz (A. Chaudhuri), dryvkin@fsu.edu (D. Ryvkin), lcameron@ucmerced.edu (L. Cameron).

http://dx.doi.org/10.1016/j.jebo.2017.07.016

numerical effort choices made in tournaments are statistically no different than those under piece rates, though the variance of effort choices in tournaments is 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 that 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, which always chooses the same effort number and subjects are informed of the computer's effort choice.[2] This has the effect of transforming the 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 subjects' choices in the final round 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 mostly due 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 in the field and certainly all so-called "white-collar" jobs require cognitive effort, it is 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 a variable (X), which is an unknown noisy function of those two cue values. The cue values shown to subjects change from one round to the next but the (deterministic part of the) underlying function does not. 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 absolute prediction error, i.e., |(Actual value of X) − (Predicted value of X)|, is our metric for performance. The smaller the absolute error, the better the productivity. By *learning* we will refer to decreasing absolute errors (increasing productivity) over time, which, in turn, implies increasing prediction accuracy. We implement four different treatments that are based on two different payment schemes: *piece-rate* refers to a linear payment scheme that relies only on the subject's own absolute error; in the *winner-take-all tournament* payment scheme, in each round subjects are paired and the winner earns a fixed amount, while the loser earns nothing. The remaining treatments manipulate the nature of the feedback provided to the subjects, allowing us to isolate the factors that impact learning. We also manipulate task difficulty, by employing two versions of the task described above. In the simpler, single cue version, one of the cues (cue A) is fixed for the duration of the experiment, whereas in the more complex, dual cue version, both cues are changing randomly from one round to the next.

We observe that while there are no differences in learning patterns, in terms of increasing prediction accuracy, across pay schemes for the simpler task, learning in the more complex task is facilitated most by a winner-take-all tournament. Evidence from an additional control treatment suggests that it is the winner-take-all nature of the payment scheme that fosters this effect of tournament incentives on learning, rather than the provision of relative rank information. The effect is particularly pronounced for those who were adept at the task to start with; but even those who were not, perform relatively better over time under a tournament payment scheme as compared to the others. 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 discuss the results and make some concluding comments.

[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 mechanical 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 and Levin, 2005; Erev and Roth, 1998; Rick and Weber, 2010; Roth and 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.

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

[3] Chaudhuri et al. (2006) extend Merlo and Schotter's (1999) study by adopting an inter-generational framework, where a group of subjects are recruited into the lab and play the same stage game for 10 rounds. Each player can then leave advice for his laboratory descendant, who then plays the game for another 10 rounds as a member of a fresh group of subjects. Chaudhuri et al. find that the presence of advice makes a difference in that the experimental groups who get advice perform better – their decisions are closer to the Nash equilibrium – compared to a control group of subjects that plays the game with no recourse to such advice. Iyengar and Schotter (2008) also rely on the Merlo and Schotter (1999) framework but use 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 advisees learn to make decisions that are closer to the Nash equilibrium.

Table 1
Actual Cue Values 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

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.[4] The variable X_t can be thought of as the underlying price of a stock; the cues as variables that affect the value of the stock; and the task at hand as one of forecasting stock prices. The stock value is determined by the 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 interval $[-5, 5]$ in each round t, i.i.d. across rounds. 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 and the noise term 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 in each round. This is designed to be less difficult than the *Dual Cue* task, where both cue values change from one round to the next. For both tasks, the sequence in which the cue values appear from one round to the next, is identical across treatments. Table 1 shows the cue values and corresponding stock prices for each round.

Our metric of performance is the absolute error, defined as the absolute distance, $|X_t{}^P - X_t{}^*|$, between the predicted value ($X_t{}^P$) and the actual value ($X_t{}^*$). 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.

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).[5]

[4] MCPL tasks are commonly used in psychology to study learning (see Balzer et al. (1989) for a review). In economics, besides Brown (1995, 1998), this task has been used by Vandegrift and Brown (2003) and Vandegrift et al. (2007) as well.

[5] We implemented two other treatments: (1) salary (S), where subjects are paid a flat $20 fee regardless of task performance and (2) salary-win-lose (SWL), where subjects are paired and paid the same flat fee of $20 regardless of performance but in addition, after each round, subjects are informed whether they did better or worse, in terms of forecasting errors, than the other member of the pair. Contrary to some prior studies (cf. Gill et al., 2015 and references therein), we do not find significant differences in average performance between S and SWL. However, we do not report detailed results here because directly comparing S or SWL with other treatments is not feasible, since treatments S and SWL did not have the initial 5 rounds of learning under piece-rate incentives. That said, the results from S and SWL are consistent with the main message of this paper in that we do not find evidence of learning

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

Fig. 1. Screenshot for PR treatment.

Under PR, the earnings for each subject in a particular round is given by NZ $1 minus the absolute forecast error (in cents). 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 earnings 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. Fig. 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.

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.

Our second treatment is piece-rate win-lose (PRWL). This treatment is designed to address the fact that in going from piece-rates to tournaments, the underlying incentives change in two ways. First, under piece rate one's payoff depends 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*.

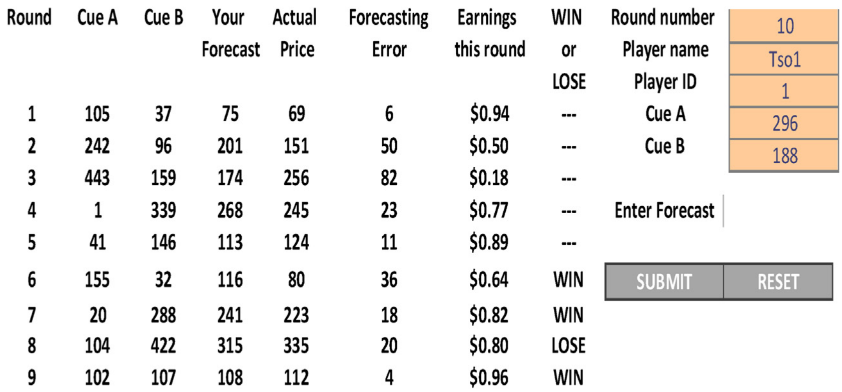
But, there is a second component to the change in incentives. In a tournament, agents must outperform their peers in order to attain a higher rank. While a higher rank may correspond to a higher tangible reward (such as promotion tournaments), agents may simply be motivated by the higher rank itself, 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).[6] 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.[7]

The first 5 rounds in PRWL are identical to PR but, from round 6 onwards, 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, respectively, smaller or larger than her paired subject's. However, whether a subject won or lost a particular round has no bearing on her earnings for that round since each subject continues

over time in the salary treatments, which stands in contrast to the learning that we observe in WTAT. As we show below, learning in WTAT is driven primarily by its all-or-nothing payoff structure. This is possibly why we do not see similar learning effects in the salary treatments, even with the provision of relative rank information. Given the design differences, we have omitted a discussion of S and SWL from the current study. We also find interpretable gender differences across the different treatments that, in the interests of parsimony, we have chosen not to pursue here.

6 The idea behind rank competition in our study is similar to what Kräkel (2008) describes as *"emotions"*, where positive and negative emotions are derived from winning and losing respectively. See also Gill et al. (2015).

7 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 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 treatments 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 to 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 et al. (2010) show that, by and large, it does not matter whether the rank information is provided publicly or privately for that information to have a positive impact on performance.

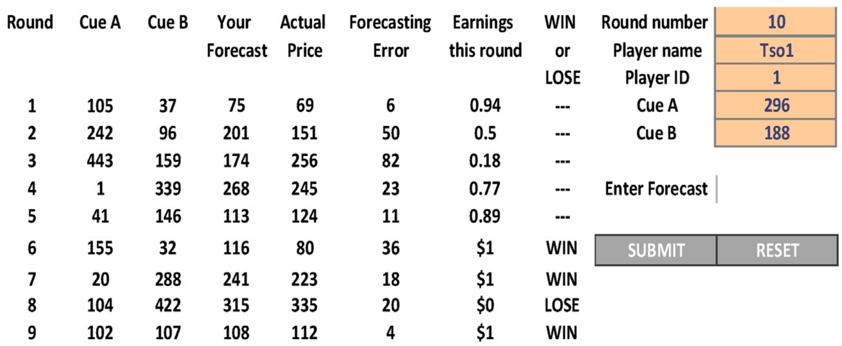


Round	Cue A	Cue B	Your Forecast	Actual Price	Forecasting Error	Earnings this round	WIN or LOSE
1	105	37	75	69	6	$0.94	---
2	242	96	201	151	50	$0.50	---
3	443	159	174	256	82	$0.18	---
4	1	339	268	245	23	$0.77	---
5	41	146	113	124	11	$0.89	---
6	155	32	116	80	36	$0.64	WIN
7	20	288	241	223	18	$0.82	WIN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

Fig. 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	0.5	---
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

Fig. 3. Screenshot for WTAT treatment.

to get paid on the basis of one's own absolute errors. The rank information is simply designed to capture the utility from winning and/or disutility from losing. Fig. 2 shows a screenshot of this treatment.

The WTAT treatment also starts in round 6, following 5 rounds of piece-rate payments. As in PRWL, from round 6 onwards, we divide 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.[8] Fig. 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.

The PRWL treatment is designed to help disentangle the two effects: *competing for higher payoff* versus *competing for higher rank*. 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 payoffs.

The fourth and final treatment is the winner-take-all-tournament with no information (WTAT-NI). This 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 divide 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 until the end of round 20*. Note that subjects still receive feedback on their absolute performance, i.e., the realized value of X and absolute error, similar to the PR treatments. Fig. 4 presents a screen-shot. We run this control treatment with the dual cue task only. The reason is that we only identified differences in learning patterns across the pay schemes for the dual cue task, and the WTAT-NI treatment is designed to isolate the effects of rank information and winner-take-all payoffs on learning. Table 2 provides a graphical overview of the similarities and differences between the four different treatments.

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

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

Fig. 4. Screenshot for WTAT-NI treatment.

Table 2
Graphical overview of the different treatments.

	Provision of feedback on wins/losses in each round	
Payment Scheme	No Feedback	Feedback
Piece-rate payment	Piece-rate (PR)	Piece-rate win-lose (PRWL)
Tournament payment	Winner-take-all- tournament no information (WTAT-NI)	Winner-take-all- tournament (WTAT)

Table 3
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

2.3. *Experimental procedure*

Experimental sessions were conducted at the DECIDE lab of the University of Auckland, using primarily first year students in business and economics. There are a total of 274 subjects (58% female) 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 et al., 1983). The questionnaire (provided in the Appendix) consists of 20 questions that are answered on a 1–4 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 explain below why we choose to control for trait anxiety.

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, for the PR, PRWL, WTAT and WTAT-NI treatments the first five rounds are identical, using a piece rate payment scheme. Subjects are told that they will be provided further information prior to the start of round 6. They are also provided with ten examples (shown in Table 3) for Cue A, Cue B and X and given some time to study these examples. These examples are generated by the same process as the numbers in the actual tasks and subjects are made aware of that fact.

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 get to learn whether they won or lost a round. They are also told 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.[9]

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.

2.4. Research hypotheses

In this section we formulate hypotheses that will guide our analyses later in the paper. We start by looking at the impact of rank competition on productivity by comparing the PR and PRWL treatments, both of which pay piece-rates, except PRWL provides additional information on wins/losses in each round. If people are motivated solely by earnings, then we would not expect rank competition to have any effect on productivity.

However, as noted above, there is now voluminous evidence that providing relative rank information leads to improved performance. This effect may consist of either an ex-ante anticipation effect (as in Blanes i Vidal and Nossol, 2011 or Kuhnen and Tymula, 2012) that may be associated with preferences for status and respect (see Ellingsen and Johannesson, 2007 for a review), or an ex-post revelation effect when people react to the feedback received (as in Azmat and Iriberri, 2010 or Tran and Zeckhauser, 2012). Either of these effects or both together, would lead to better performance in PRWL compared to PR, implying *(PRWL Errors) < (PR Errors).*

Turning to the issue of competing for payoffs, we can compare the PRWL and WTAT treatments. Both feature rank feedback, but differ in how subjects are paid: piece-rates and rank-dependent prizes respectively. Prior research, such as Bull et al. (1987) suggests that tournaments elicit similar effort levels as piece-rates. This would imply *(PR Errors) ≈ (WTAT Errors).* If, as argued above, competing for higher rank means *(PRWL Errors) < (PR Errors)*, then we would expect *(PRWL Errors) < (WTAT Errors).* Both Hannan et al. (2008) and Eriksson et al. (2009) provide evidence indicative of this; piece-rates with relative feedback perform better than tournaments featuring identical feedback. We combine the arguments above into our first main hypothesis.

Hypothesis 1. *When relative feedback is added to piece-rates, this will lead to better performance than either in tournaments or in piece-rates without rank information. This implies (PRWL Errors) < (WTAT Errors) ≈ (PR Errors).*

Our next research question deals with learning over time. How does rank competition affect learning over time? Research in psychology suggests that feedback may not improve task performance if it does not inform people about how to go about performing the task proficiently (see Kluger and DeNisi, 1996 for a review). In this regard, since relative performance feedback itself merely provides performance benchmarking and does not assist people to improve forecasts, we would not expect competition for rank to have any impact on learning. This leads to:

Hypothesis 2. *Relative performance feedback has no effect on learning. The rate of learning is identical in both PR and PRWL.*

When it comes to comparing PRWL and WTAT, we would expect competing for payoffs to promote faster learning in WTAT. While we may expect PRWL to perform better either early on or on average compared to WTAT, as posited by H1, the binary win/lose nature of the payoffs will likely encourage players to increase their effort, which in turn should improve their forecast accuracy over time. In PRWL, marginal increases in productivity improve pay only by a small amount. In contrast, in WTAT, those who win receive a large $1 prize, equivalent to what a perfect forecast yields under piece-rates, while losing implies getting nothing.[10] This suggests the following hypothesis.

Hypothesis 3. *Rank-dependent payoffs motivate learning more than piece-rates. We expect (WTAT Errors) to decline more quickly over time than (PRWL Errors).*

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

[9] 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 the two tournament payments schemes, 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.

[10] Dutcher et al. (2015) show that in their tournaments, the "avoid-being-last" objective has a greater effect than the "strive-to-be-first" objective in terms of eliciting effort. While our tournament, with a binary outcome, is not directly comparable to Dutcher et al. (2015), since their tournaments feature multiple prizes, nevertheless we believe that the desire to avoid a loss provides a stronger impetus to subjects to try and improve their forecasts, more so than in PRWL.

Table 4
Average errors across treatments.

Treatments	Single Cue		Dual Cue		Pooled	
	N	Average Error	N	Average Error	N	Average Error
Piece rate (PR)	42	10.2	39	26.6	81	18.1
Piece rate win lose (PRWL)	42	9.6	35	24.0	77	16.2
Winner take all tournament (WTAT)	40	10.0	38	30.7	78	20.1
Tournament no information (WTAT-NI)	*NA*	*NA*	38	27.74	*NA*	*NA*
Total	124		150		274	

Table 5
Random effects regression for absolute errors using data from both Single and Dual Cue Tasks.

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
Round	−0.136** (0.063)	−0.161** (0.068)	−0.142*** (0.024)	−0.158*** (0.039)	−0.158*** (0.039)
Dual Cue	16.4*** (3.503)	15.906*** (2.45)	15.906*** (2.451)	15.501*** (2.659)	9.748*** (1.003
PRWL	−0.557 (0.498)	−1.474*** (0.41)	−3.353** (1.559)	−3.039* (1.681)	−1.559 (1.309)
WTAT	−0.145 (0.958)	−0.563 (1.068)	2.115 (2.365)	−0.057 (2.838)	0.922 (2.184)
PRWL*Dual Cue	−1.977 (3.614)	−2.633 (2.411)	−2.633 (2.411)	−3.434 (4.157)	−2.737 (4.119)
WTAT*Dual Cue	4.318 (3.788)	3.636 (2.953)	3.636 (2.954)	7.902* (4.153)	6.732** (2.658)
Female	–	6.315*** (1.769)	6.315*** (1.769)	6.315*** (1.77)	4.126*** (1.511)
Trait Anxiety	–	0.041 (0.153)	0.041 (0.153)	0.041 (0.154)	0.042 (0.099)
PRWL*Round	–	–	0.145 (0.116)	0.12 (0.125)	0.120 (0.125)
WTAT*Round	–	–	−0.206* (0.118)	−0.039 (0.137)	−0.039 (0.137)
Dual Cue*Round	–	–	–	0.031 (0.04)	0.031 (0.040)
PRWL*Dual Cue*Round	–	–	–	0.062 (0.243)	0.062 (0.243)
WTAT*Dual Cue*Round	–	–	–	−0.328** (0.159)	−0.328** (0.159)
Round 1–5 Median Error	–	–	–	–	0.497*** (0.101)
Constant	11.93*** (0.877)	8.356 (6.811)	8.107 (6.696)	8.321 (6.746)	2.681 (5.033)
R^2	0.11	0.12	0.12	0.12	0.19
No. of observations	3540	3180[a]	3180	3180	3180
No. of participants	236	212	212	212	212

Note: Standard errors in parentheses; ***, ** and * denote significance at 1%, 5% and 10% respectively.
Dependent variable = Absolute errors = |Predicted value − Actual Value|.

[a] 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.).

subjects from the four treatments, for much of the analysis below, we will confine our attention to data from 236 subjects in the PR, PRWL and WTAT treatments, excluding the control treatment WTAT-NI for now. 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 4 provides a broad overview of the absolute errors in different treatments.[11] Not surprisingly the errors are much smaller in the single cue task than the dual cue one. It is also noticeable that, in both the single and dual cue tasks, average absolute errors are highest in WTAT, followed by PR, and smallest in the PRWL treatment. To explore these issues in more detail we turn to regression analysis next.

3.1. Productivity across treatments

3.1.1. Result 1: (a) on average, there are no significant differences in performance across treatments. (b) WTAT promotes faster learning over time in the dual cue task

As noted before, our metric for performance is absolute error (=|(Predicted value of X)– (Actual value of X)|). The smaller the absolute error, the higher the productivity. In Table 5 we present results for random effects regressions with the absolute forecasting error as the dependent variable. Errors are clustered by session to control for interactions due to session-wide random re-matching. (See Fréchette, 2012 for arguments as to why this is the appropriate approach in such cases.) In running these regressions, we pool the data from the single cue and dual cue tasks. 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, a dummy for the task difficulty (Dual Cue = 1 for the dual cue task; 0 otherwise), two treatment dummies (PRWL and WTAT with the PR treatment as the reference category) and finally treatment interactions with task difficulty. With 236 subjects

[11] We had 36 subjects in the dual cue 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.

making 15 decisions each we have 3540 observations here. None of the coefficients for the treatment dummies are significant at conventional levels. But it is clear that subjects perform worse in the dual cue task and this is especially pronounced for the WTAT treatment. This suggests that for tasks that are cognitively challenging and require real effort, we should expect tournament payment schemes to perform worse than the others, at least early on. Differences will likely be smaller with tasks that are simpler or more mechanical. The negative and significant coefficient on *Round* indicates that errors decrease over time in all treatments.

In Model 2 we include two additional regressors: female (=1 for women and 0 for men) and each subject's trait anxiety score.[12] Conditional on individual characteristics, PRWL produces lower errors than both PR and WTAT, particularly in the dual cue task. A large positive coefficient on the female dummy shows that, on average, women performed worse than men. This result is consistent with the large literature on the gender gap in mathematical performance (e.g., Guiso et al., 2008; Hyde et al., 1990; Hyde and Mertz, 2009).[13]

In Model 3, in order to pick up possible differences in learning across treatments we include two additional terms involving the treatment dummies interacted with round. The results are similar to Model 2 with the PRWL dummy negative and significant at 5%. Also noteworthy is the negative coefficient for the interaction term *WTAT*Round*, which suggests that in WTAT the errors diminish over time faster than in other treatments.

In Model 4, we add additional interaction terms to explore differences in learning between the single and dual cue tasks. Model 4 shows that the faster learning observed in the WTAT treatment in Model 3 is driven entirely by subjects' performance in the dual cue task. The regression results and associated Wald tests, which we have omitted for the ease of exposition, show that in the single cue task there are no differences between treatments at the beginning and no differences in time trends. In the dual cue task, WTAT errors are higher than both PR and PRWL at the beginning, but over time WTAT has a stronger downward trend as compared to either PR or PRWL.

Finally, in Model 5, we include an additional control for each subject's ability by including the subject's median error from Rounds 1 through 5, during which subjects in all treatments are paid piece-rates. We use the median error in those rounds as a benchmark for a subject's basic facility with the task. As expected, the coefficient on median error is highly significant. This model corroborates the finding that errors decline over time in WTAT in the dual cue task. Note that once we control for individual ability, the coefficient for PRWL is no longer significant.

3.2. Heterogeneous ability and learning in the dual cue task

3.2.1. Result 2: learning in the dual cue task in WTAT is particularly pronounced for those who are adept at the task to begin with; for those less adept, forecast performance stays about the same in WTAT but worsens over time in the other treatments

The results in Table 5 suggest that the differences in either forecast errors or learning are negligible in the single cue task, while being more pronounced in the dual cue task. Further, these results suggest that when we look at the dual cue task, the WTAT treatment leads to better learning and forecasts over time. Therefore, in what follows we will rely on data from the dual cue task only. This leaves us with data for 112 subjects in PR, PRWL and WTAT in the dual cue task.

Recall that the first 5 rounds in each treatment are identical, with subjects getting paid an identical piece-rate. As noted above, the median error in those 5 rounds can serve as a benchmark for a subject's ability in the task. If a subject's median error for the first 5 rounds is *larger (smaller)* than the overall median (across subjects) then we refer to this subject as a *low (high)* ability.[14] The results are presented in Table 6. The first column presents data for all 112 subjects in the dual cue task, while in the next two columns we split this up for the two separate groups: high ability (n = 58) and low ability (n = 54). Independent variables include a subject's median error during the first 5 rounds, round, dummies for PRWL and WTAT, with PR as the reference category, and two interaction terms involving the treatment dummies and round.

Looking at the data for all 112 subjects in the dual cue task, the median error for rounds 1 through 5 is positive and highly significant, similar to the last column in Table 5, which is expected as it works as a "fixed effect." This suggests that initial ability has significant impact on subsequent performance in the task. Round has a negative and marginally significant coefficient, showing that there is a small downward trend in PR. Errors are higher in WTAT at the beginning, given the

[12] We choose to control for trait anxiety because prior research suggests that females tend to exhibit higher levels of trait anxiety than males. Such anxiety may lead to a tendency to see competitive situations as being threatening and to approach them with a degree of apprehension or tension. Hishinuma et al. (2000, 2001), McKnight and McKnight (2012) and Segal and Weinberg (1984) provide evidence in favour of higher trait anxiety for women. Foot and Koszycki (2004), however, do not find significant gender differences in trait anxiety. Given the nature of the task and the treatments in our study, we felt that it made sense to control for trait anxiety.

[13] In recent studies, gender differences in math performance have been attributed to cultural and sociological factors, such as gender stereotypes and peer effects (Guiso et al., 2008; Hyde and Mertz 2009). These differences are context-specific and can be moderated via priming (e.g., Schmader 2002). In our case, there are interpretable gender differences across the different treatments. For instance, when we interact treatments and trends by gender, we find that there is significant learning for both men and women in the dual cue WTAT treatment, although men learn at a faster rate. We also find that gender differences in trait anxiety scores are able to explain why women perform worse than men in the WTAT treatment. In the interests of parsimony, we have refrained from elaborating on these results here since we feel that undertaking a meaningful analysis of gender differences is better left to another paper. There is also a voluminous literature in the area already. See, for instance, Croson and Gneezy (2009) and Eckel and Grossman (2008).

[14] The overall median for the first 5 rounds is 21 for the dual cue task. Therefore, subjects with a median error of more (less) than 21 for the first 5 rounds are labelled low (high) ability.

Table 6
Learning over time; random effects regression for absolute errors using Dual Cue task data only.

Independent variables	All Subjects (n = 112)	High Ability (n = 58)	Low Ability (n = 54)
Rds 1–5 Median Error	0.518*** (0.131)	0.522** (0.218)	0.516*** (0.192)
Round	−0.092* (0.051)	−0.523*** (0.016)	0.411*** (0.007)
PRWL	−2.956 (3.546)	−3.169 (4.57)	−2.82 (2.818)
WTAT	8.885*** (1.961)	5.042*** (1.827)	13.759** (5.801)
PRWL*Round	0.163 (0.152)	0.301 (0.207)	0.01 (0.065)
WTAT*Round	−0.389*** (0.086)	−0.174* (0.102)	−0.698*** (0.207)
Constant	14.686*** (2.588)	20.004*** (2.693)	8.513 (5.756)
R^2	0.09	0.02	0.07
No. of Observations	1680	870	810
No. of participants	112	58	54

Note: (1) Standard errors in parentheses; ***, ** and * denote significance at 1%, 5% and 10% level respectively; (2) We use data for Dual Cue task only, since we do not find significant differences among treatments in the Single Cue task in Table 5 above.
Dependent variable = Absolute errors = |Predicted value − Actual Value|.

Table 7
Average Actual Forecast Errors in Round 20.

	Actual Absolute Errors in Round 20	
	Single Cue	Dual Cue
PR	4.79 (3.18)	24.15 (33.22)
PRWL	4.43 (2.61)	27.66 (28.53)
WTAT	5.25 (4.94)	19.32 (18.14)
WTAT-NI		17.55 (23.03)

Note: Standard Deviations in parentheses.

positive coefficient on WTAT, but the interaction term between WTAT and round is negative and highly significant showing that the negative trend in WTAT is stronger than in other treatments.

Next, we look at the two types – high and low abilities – separately. The results for the high ability type are very similar to those for the pooled data, except that the high ability types improved over time in both PR and WTAT. The low ability types, on the other hand, actually performed worse over time in both PR and PRWL. While the coefficient for *WTAT*Round* is negative and significant for the low ability types, a Wald test shows that the overall negative trend in WTAT is not significant. This leads to the conclusion that there is divergence in learning patterns for high and low ability subjects. High ability types become better over time in PR and especially in WTAT, and stay the same in PRWL. Low ability subjects become worse in PR and in PRWL and stay the same in WTAT. Overall, WTAT facilitates better learning for all types in the dual cue task.

3.3. Average 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 early in the session, particularly vis-à-vis the PRWL treatment. If learning occurs gradually round-by-round, consistent with standard models, then we should expect cumulative improvements in performance to be the greatest in the final round. Table 7 shows average forecast errors for round 20. As expected, errors are substantially smaller for the single cue task; however, there are no statistical differences in errors across treatments within each task difficulty level. Thus, while the WTAT errors are much larger than those in either PR or PRWL in the early rounds, these differences narrow over time, so much so that by Round 20, the differences are no longer significant.

3.4. A dynamic learning model

3.4.1. Result 3: the finding of enhanced learning in WTAT in the dual cue task is robust to a dynamic specification incorporating reinforcement learning

Next, we look at dynamic reinforcement models of learning, where we control for two things: first, a subject's lagged error, i.e., her error in the previous round and second, the impact of winning or losing in the previous round. We present our results in Table 8, where once again we rely only on the dual cue task data, given our earlier finding of no significant treatment differences in the single cue task. We no longer control for ability levels in these regressions because a subject's lagged error controls for own ability and past performance. Looking at Model 1, as expected, errors are persistent (coefficient on lagged error is positive and significant). PRWL errors are lower and WTAT errors are higher than the baseline PR treatment at the beginning but, as before, WTAT errors go down over time. We also control for feedback (winning in the previous round), and for possible differences in the effects of feedback across treatments (through interactions of winning in the previous round with treatments PRWL and WTAT), but those do not appear to be significant.

Model 2 in Table 8 includes, in addition to all the variables of Model 1, the interactions of winning in the previous round with treatment dummies PRWL and WTAT also interacted with the subject's lagged error. The goal of including these

Table 8
Dynamic learning model for Dual Cue task only.

Independent variables	Model 1 OLS	Model 2 OLS
Lagged error	0.236*** (0.4)	0.223*** (0.045)
Round	−0.043 (0.07)	−0.046 (0.07)
PRWL	−5.028* (2.392)	−4.999* (2.404)
WTAT	8.825*** (2.272)	8.553*** (2.344)
PRWL X Round	0.064 (0.186)	0.068 (0.181)
WTAT X Round	−0.382*** (0.094)	−0.341** (0.113)
PRWL X Lagged Win	5.394 (2.854)	5.485** (1.621)
WTAT X Lagged Win	−1.495 (1.145)	−6.534* (3.007)
PRWL X Lagged Win X Lagged Error	–	−0.028 (0.159)
WTAT X Lagged Win X Lagged Error	–	0.292*(0.131)
Constant	20.691*** (1.166)	21.092*** (1.105)
R^2	0.07	0.07
No. of observations	1680	1680
No. of participants	112	112

Note: (1) Standard errors in parentheses; ***, ** and * denote significance at 1%, 5% and 10% level respectively; (2) We use data for Dual Cue task only, since we do not find significant differences among treatments in the Single Cue task in Table 5 above.
Dependent variable = Absolute errors = |Predicted value − Actual Value|.

Table 9
Comparing WTAT with WTAT-NI in the Dual Cue task only.

Independent variables	Model 1 OLS	Model 2 Dynamic OLS controlling for Lagged Errors	Model 3 Dynamic OLS controlling for Lagged Errors
Rounds 1–5 Median Error	0.494*(0.194)		
Lag error		0.209** (0.039)	0.200** (0.045)
Round	−0.481*** (0.073)	−0.448*** (0.062)	−0.417*** (0.062)
WTAT-NI	−2.613 (3.245)	−2.565 (2.567)	−2.972 (3.189)
WTAT-NI*Round	−0.015 (0.126)	0.013 (0.106)	−0.018 (0.103)
Lag Win			−2.603 (1.137)
Lag Win*WTAT-NI			1.728 (1.440)
Constant	24.191*** (3.804)	29.968*** (1.236)	31.106*** (1.747)
R^2	0.05	0.05	0.05
No. of observations	1140	1140	1140
No. of participants	76	76	76

Note: (1) The reference category is now the WTAT treatment. Standard errors in parentheses; ***, ** and * denote significance at 1%, 5% and 10% level respectively. (2) We use data for Dual Cue task only, since we ran the WTAT-NI treatment with this task only.
Dependent variable = Absolute errors = |Predicted value − Actual Value|

interactions is to assess whether, and how, the effect of winning in the previous round in PRWL and WTAT depends on the subject's past performance. The positive and significant coefficient on *PRWL*Lagged Win* in Model 2 shows that for subjects who made a zero or relatively small error in the previous round, current error is higher after winning in PRWL. In contrast, the coefficient on *WTAT*Lagged Win* is negative, suggesting that low errors in the previous round lead to a reduction in error in response to winning. Assuming better performance is associated, on average, with higher effort, these results imply that subjects who perform well and win in PRWL tend to "relax" in the following round, while subjects who perform well and win in WTAT tend to exert an even higher effort (or learn better) in the next round. Note, however, the positive (albeit marginally significant) coefficient on the interaction *WTAT*Lagged Win*Lagged Error*, which shows that in WTAT this improvement vanishes for those subjects who won in the previous round despite their error being relatively high.

3.5. What drives tournament learning?

3.5.1. Result 4: it is the all-or-nothing nature of the payment scheme in WTAT, rather than the provision of relative rank information, that promotes learning

In order to understand better how different components of WTAT incentives affect learning, we now turn to the WTAT-NI treatment, which features the exact same monetary incentives as the WTAT treatment but no feedback pertaining to winning/losing (and hence, earnings) between rounds. This means that in WTAT-NI subjects do not know whether they have won or lost the prior round of play. They, however, do receive feedback on their forecasting errors, just like in other treatments, which allows them to learn. 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 drives learning in WTAT, since there is no feedback in WTAT-NI.

We present the estimated time trends for the WTAT and WTAT-NI treatments in Table 9. As before, we use data from the dual cue task only (we ran the WTAT-NI control treatment only for the dual cue task). We also focus only on the WTAT and WTAT–NI treatments here, with the former serving as the reference category. Model 1 controls for ability level by including the median error in the first five round. Model 2 introduces dynamics by including lagged errors, which control for subjects' individual characteristics and past performance. Model 3 adds a variable indicating whether a subject won or lost in the previous round. As before, we continue to observe learning in the WTAT treatment (as indicated by the significant negative coefficient of the Round variable). The insignificant coefficient on *WTAT-NI*Round* shows that the rate of learning in the WTAT-NI treatment is no different from that of the reference WTAT treatment.[15] Except, that there is no reaction to winning in WTAT-NI, as shown by the insignificant coefficient of the *Lag Win*WTAT-NI* variable. This is to be expected, and serves as a falsification test, because subjects in this treatment are not privy to this information until the end of the session.

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. Discussion and concluding remarks

In this paper we have looked at the temporal dimension of learning across treatments using a multiple cue probabilistic learning task. We find that our experimental treatments do not result in significant differences in average productivity across treatments. However, the WTAT treatment stands out in terms of superior learning in the more complex dual cue task. It is noteworthy that initially the switch from piece-rates to tournament payoffs leads to a significant increase in errors (indicated by the positive and highly significant coefficient for the WTAT treatment dummy in Table 6). This is particularly true for the low ability subjects, for whom the average increase in the magnitude of errors is nearly three times as much as that for the high ability types.

Given that the payment mechanism does not change in going from PR to PRWL, it is likely that in both treatments subjects are quite focussed on minimizing their errors from one round to the next. The switch to WTAT, on the other hand, confronts subjects with a different payoff structure and introduces a strategic element, which may attenuate their focus on minimizing errors. This is because, under WTAT, one's own errors do not matter as much, as long as one outperforms the other member of the pair. This may have resulted in subjects expending relatively less cognitive effort in WTAT at the outset, leading to larger errors, and the effect is initially more pronounced for those who are less adept at the task. This also explains, in part, why WTAT does not perform better than other treatments *on average*, given the relatively larger errors in the early rounds of WTAT.

But while forecast accuracy in this treatment is worse at the beginning, it improves at a significantly faster rate than in any other treatment so much so that by Round 20 of the dual cue task, absolute errors in WTAT are not significantly different from those in either PR or PRWL. These differences in learning between WTAT and other treatments in the dual cue task are robust to various specifications and estimation methods. We find that subject ability matters, in the sense that there is divergence between those who were adept at the task at the outset as opposed to those who were not. The former got better over time in WTAT. Low ability subjects performed worse over time in PR and PRWL, while staying the same in WTAT.

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 suggests that feedback about relative performance is less important for learning; it is the rank dependent payoffs that drive learning.

A key difference between this study and other experiments on the effects of pay schemes on learning is the more cognitively challenging nature and complexity of the task. Prior studies mainly used more mechanical effort based tasks, such as number addition, multiplication or counting the number of zeros in a matrix. The MCPL task we use is more suitable to mimic work environments that require nontrivial cognitive effort and learning-by-doing. The differences we observe in learning patterns between the single cue and dual cue variations of the task further confirm that there may be important differences in learning depending on the nature and difficulty of the task.

The few papers that have looked at learning using a real-effort task, do so within a single tournament-type payment scheme, where the focus is primarily on the impact of exogenous interventions such as timing of payments (Merlo and Schotter, 1999) or the role of advice, either inter-generational (Chaudhuri et al., 2006) or peer-to-peer (Iyengar and Schotter, 2008). To the best of our knowledge, ours is the first paper to specifically study the impact of different payment schemes on learning in a cognitively challenging real-effort task. Our results then have interpretable implications for designing appropriate work-place compensation schemes depending on task difficulty, heterogeneity of worker ability and whether the goal is to have an immediate impact on productivity or to foster long-term learning.

Acknowledgements

We are grateful to the University of Auckland for funding this study. We are indebted to two anonymous referees of this journal for their detailed and thoughtful feedback. We thank Erwann Sbai, Ryan Greenaway-McGrevy, Choon Wang, Tibor

[15] Round 20 forecast errors in the WTAT-NI treatment, from Table 7, are similar to those in other treatments.

Neugebauer as well as seminar participants at the University of Auckland, Monash University, Queensland University of Technology, University of Zurich and participants at a number of conferences including the Economic Science Association North American Meetings in Dallas, TX in 2015 and the Economic Science Association Annual Meetings in Jerusalem in 2016 for many helpful suggestions. We are responsible for all errors.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2017.07.016.

References

Azmat, G., Iriberri, N., 2010. The importance of relative performance feedback information: evidence from a natural experiment using high school students. J. Public Econ. 94 (7–8), 435–452.
Balzer, W., Doherty, M., O'Connor, R., 1989. Effects of cognitive feedback on performance. Psychol. Bull. 106 (3), 410–433.
Blanes i Vidal, J., Nossol, M., 2011. Tournaments without prizes: evidence from personnel records. Manage. Sci. 57 (10), 1721–1736.
Brown, P., 1995. Learning from experience, reference points, and decision costs. J. Econ. Behav. Organ. 27 (3), 381–399.
Brown, P., 1998. Experimental evidence on the importance of competing for profits on forecasting accuracy. J. Econ. Behav. Organ. 33 (2), 259–269.
Bull, C., Schotter, A., Weigelt, K., 1987. Tournaments and piece rates: an experimental study. J. Polit. Econ. 95 (1), 1–33.
Cadsby, C., Engle-Warnick, J., Fang, T., Song, F., 2010. Psychological Incentives, Financial Incentives, and Risk Attitudes in Tournaments: An Artefactual Field Experiment Working Paper. Department of Economics, University of Guelph.
Cardella, E., 2012. Learning to make better strategic decisions. J. Econ. Behav. Organ. 84 (1), 382–392.
Charness, G., Levin, D., 2005. When optimal choices feel wrong: a laboratory study of bayesian updating, complexity, and affect. Am. Econ. Rev. 95 (4), 1300–1309.
Charness, G., Masclet, D., Villeval, M.-C., 2014. The dark side of competition for status. Manage. Sci. 60 (1), 38–55.
Chaudhuri, A., Schotter, A., Sopher, B., 2006. Learning in tournaments with inter-generational advice. Econ. Bull. 3 (September (26)), 1–16.
Croson, R., Gneezy, U., 2009. Gender differences in preferences. J. Econ. Lit. 47 (2), 448–474.
Dutcher, G., Balafoutas, L., Lindner, F., Ryvkin, D., Sutter, M., 2015. Strive to be first or avoid being last: an experiment on relative performance incentives. Games Econ. Behav. 94, 39–56.
Eckel, C., Grossman, P., 2008. Differences in the economic decisions of men and women: experimental evidence. In: Plott, Charles, Smith, Vernon (Eds.), Handbook of Experimental Economics Results, 1. North-Holland, Amsterdam, pp. 509–519 (Chapter 57).
Ellingsen, T., Johannesson, M., 2007. Paying respect. J. Econ. Perspect. 21 (4), 135–149.
Erev, I., Roth, A., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am. Econ. Rev. 88 (4), 848–881.
Eriksson, T., Poulsen, A., Villeval, M.-C., 2009. Feedback and incentives: experimental evidence. Lab. Econ. 16 (6), 679–688.
Foot, M., Koszycki, D., 2004. Gender differences in anxiety-related traits in patients with panic disorder. Depress. Anxiety 3, 123–130.
Fréchette, G., 2012. Session-effects in the laboratory. Exp. Econ. 15, 485–498.
Gill, D., Kissová, Z., Lee, J., Prowse, V., 2015. First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. Working paper (Available at https://papers.ssrn.com/sol3/papers2.cfm?abstract_id=2641875).
Green, J., Stokey, N., 1983. A comparison of tournaments and contracts. J. Polit. Econ. 91 (3), 349–364.
Guiso, L., Monte, F., Sapienza, P., Zingales, L., 2008. Culture, gender, and math. Science 320 (5880), 1164–1165.
Hannan, R., Krishnan, R., Newman, A., 2008. The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. Account. Rev. 83 (4), 893–913.
Hishinuma, E., Miyamoto, R., Nishimura, S., Nahulu, L., 2000. Differences in state-trait anxiety Inventory scores for ethnically diverse adolescents in Hawaii. Cult. Divers. Ethnic Minor Psychol. 6 (1), 73–83.
Hishinuma, E., Miyamoto, R., Nishimura, S., Goebert, D., Yuen, N., Makini, G., Andrade, N., Johnson, R., Carlton, B., 2001. Prediction of anxiety disorders using the state-trait anxiety inventory for multiethnic adolescents. J. Anxiety Disord. 15 (6), 511–533.
Hyde, J., Mertz, J., 2009. Gender, culture, and mathematics performance. Proc. Natl. Acad. Sci. 106 (22), 8801–8807.
Hyde, J., Fennema, E., Lamon Susan, J., 1990. Gender differences in mathematics performance: a meta-analysis. Psychol. Bull. 107 (2), 139–155.
Iyengar, R., Schotter, A., 2008. Learning under supervision: an experimental study. Exp. Econ. 11 (2), 154–173.
Kluger, A., DeNisi, A., 1996. The effects of feedback interventions on performance: a historical review, a meta-Analysis, and a preliminary feedback intervention theory. Psychol. Bull. 119 (2), 254–284.
Kräkel, M., 2008. Emotions in tournaments. J. Econ. Behav. Organ. 67 (1), 204–214.
Kuhnen, C., Tymula, A., 2012. Feedback, self-esteem, and performance in organizations. Manage. Sci. 58 (1), 94–113.
Lazear, E., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. J. Polit. Econ. 89 (5), 841–864.
McKnight, J., McKnight, M., 2012. Gender and anxiety: a comparison of student anxiety levels in face-to-Face and video conferencing courses. Creat. Educ. 3 (1), 92–95.
McNaughton, N., 2011. Trait anxiety: trait fear and emotionality: the perspective from non-human studies. Pers. Individ. Diff. 50, 898–906.
Merlo, A., Schotter, A., 1999. A surprise-quiz view of learning in economic experiments. Games Econ. Behav. 28, 25–54.
Merlo, A., Schotter, A., 2003. Learning by not doing: an experimental investigation of observational learning. Games Econ. Behav. 43 (1), 116–136.
Nalebuff, B., Stiglitz, J., 1983. Prizes and incentives: towards a general theory of compensation and competition. Bell J. Econ. 14 (1), 21–43.
Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? Q. J. Econ. 122 (3), 1067–1101.
Rick, S., Weber, R., 2010. Meaningful learning and transfer of learning in games played repeatedly without feedback. Games Econ. Behav. 2010, 716–730.
Roth, A., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games Econ. Behav. 8, 164–212.
Schmader, T., 2002. Gender identification moderates stereotype threat effects on women's math performance. J. Exp. Soc. Psychol. 38 (2), 194–201.
Segal, J., Weinberg, R., 1984. Sex, sex role orientation and competitive trait anxiety. J. Sport Behav. 7 (4), 153–159.
Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P., Jacobs, G., 1983. Manual for the State-Trait Anxiety Inventory. Consulting Psychologists Press, Palo Alto, CA.
Tran, A., Zeckhauser, R., 2012. Rank as an inherent incentive: evidence from a field experiment. J. Public Econ. 96 (9–10), 645–650.
Vandegrift, D., Brown, P., 2003. Task difficulty, incentive effects, and the selection of high-variance strategies: an experimental examination of tournament behavior. Lab. Econ. 10 (4), 481–497.
Vandegrift, D., Yavas, A., Brown, P., 2007. Incentive effects and overcrowding in tournaments: an experimental analysis. Exp. Econ. 10 (4), 345–368.