ACQUISITION REFORM:
THEORY AND EXPERIMENTAL EVIDENCE FOR TOURNAMENT SPONSORS

Lt Col Richard Fullerton, USAF, Bruce G. Linster, Michael McKee, and Lt Col Stephen Slate, USAF

Billions of dollars worth of contracts are awarded to the winners of research and development competitions annually. In the past decade a number of theoretical papers have been published on the use of research tournaments to induce optimal research efforts. This paper gives a general overview of decision-making experiments conducted to test theoretical predictions in a controlled setting. Our results indicate that carefully designed research tournaments can be highly effective at promoting research efforts as predicted by theory.

Recent General Accounting Office (GAO) studies have identified acquisition reform as one of the Pentagon's highest priorities (GAO, 1997). With dwindling defense budgets, the Department of Defense (DoD) plans to defray the cost of force modernization with the savings from acquisition reform. One area where savings might accrue is DoD-sponsored research tournaments. A recent paper by Curtis Taylor (1995) demonstrates potentially huge implications for how the federal government can save money. By sponsoring a research tournament, the government can induce the efficient amount of research and development effort from industry without the need for costly regulatory oversight— oversight which the Carnegie Commission (1992) estimates costs the DoD about 40 percent of its acquisition budget.

Research tournaments have played an important role in the economic growth of nations since the Industrial Revolution. For example, the golden age of steam locomotion was spawned by a research tournament spon-sored by the Liverpool and Manchester Railway in 1829.[1] More

recently, research tournaments have been used to create a variety of products ranging from fuel-efficient refrigerators (Langreth, 1994) and digital televisions ("HDTV," 1993), to high-tech fighter aircraft for the military (Schwartz, 1991). Today, scientists and lawmakers are even considering the use of a research contest to propel the development of the first manned space mission to Mars.[2]

Most recently, Jacques S. Gansler, the Under Secretary of Defense for Acquisition and Technology, discussed the need to "continue and greatly expand our efforts to implement a 'revolution in business affairs' within DoD…" (1998, p. 8). Gansler's keynote address on "Realizing Acquisition Reform" discusses five areas that require specific attention. Many of these specific recommendations are inherent qualities of sponsored research tournaments. In the acquisition arena, research tournaments allow DoD to be "another buyer of high-quality, high-performance, differentiated items" by allowing for greater competition, less oversight, and more flexibility in the acquisition of new products for mission requirements (Gansler, p.11).

> **"In fact, the most prolific sponsor of research competitions is the federal government, and in particular the Department of Defense."**

In fact, the most prolific sponsor of research competitions is the federal government, and in particular the Department of Defense. Each year the federal government awards billions of dollars worth of contracts to winners of competitive research and development competitions.[3]

Taylor (1995) provided a theoretical model for evaluating the effect of the parameters (competitors and duration) of these research tournaments. Fullerton and McAfee (1996) extended this framework to tournaments with heterogeneous contestants. Taylor proved that, by limiting the number of competitors in a research tournament and charging each competitor an entry fee, the tournament sponsor could inspire an efficient amount of research effort. Fullerton and McAfee show that with heterogeneous competitors, sponsors can induce the best-qualified competitors to enter the competition by conducting specialized all-pay entry auctions. Thus there is a small but growing collection of literature on the theory of research tournaments, as well as some empirical evidence.

The focus of this investigation is Taylor's original model. In it, M risk-neutral firms and T periods are available for each firm to conduct research. Each period, firms pay research cost C to obtain a single independent draw, x, from the innovation distribution, $F(x)$. Research is performed with recall, so that at the end of T periods each firm submits its best innovation to the tournament sponsor. Taylor proved that the optimal strategy of firms competing in a research tournament is conducting research until drawing an innovation worth at least some value "z" and then stopping. This stopping rule strategy dominates all other strategies, and is what we tested in our laboratory setting.

Although the economic intuition behind the effectiveness of tournaments is straightforward, the empirical calculations required to compute equilibrium strategies are complex. The empirical question is whether individuals or firms are able to

compute these strategies. We conducted a series of laboratory experiments in which the subjects chose a search strategy in a research and development (R&D) setting to answer this question. The experiments we conducted tested Taylor's research tournament theory by examining whether subjects in a controlled economic laboratory setting can be induced to expend the predicted amount of R&D effort in an essentially unregulated environment.[4]

Despite the complexity of computing the equilibrium research strategy, we find the overall level of effort expended and the winning innovations are remarkably close to the predictions of Taylor's model. Although some subjects overinvest in R&D and others underinvest, the majority of subjects adopt a stopping rule strategy when conducting research, as predicted by Taylor. This stopping rule strategy is simply that a competitor will cease conducting R&D once a certain level of innovation is reached, as opposed to other strategies, to include continuous research throughout the tournament regardless of the level of innovation attained. Indeed, we find few instances of internal inconsistencies. We also find that the average behavior of subjects is close to that predicted by Taylor's theory. As a consequence, the R&D tournaments achieve very high levels of efficiency in the laboratory. The implication of our study is that the government needs to carefully monitor the length of the tournament and the number of competitors in any procurement action. The oversight should be aimed at achieving the optimal level of competition between firms, and not at the effort level of the individual firms. This shift away from a micro-oriented regulatory strategy to a more macro- or industry-

oriented strategy should result in substantial savings in the overall procurement budget, since less detailed oversight will be required.

EXPERIMENTAL DESIGN

To test Taylor's model, we designed a series of experiments to determine if subjects individually, or as a market, provide results similar to those predicted by the theory. The experiments were conducted at the University of New Mexico's computerized experimental economics laboratory with subjects recruited from undergraduate social science classes. Each subject was assigned a computer terminal, and the laboratory was

"...the R&D tournaments achieve very high levels of efficiency in the laboratory."

designed to limit the subject's view to his or her own terminal. This helped to ensure each subject's response was independently determined. Computerization of the experiments allowed for immediate feedback for the subjects and this feedback enhanced subject understanding of the actions required.

As the experiment began, subjects received a set of written instructions explaining that they would be participating in a market where the task was to decide whether to pay for a draw of a random number in an effort to win a prize. At the start of each round, subjects were given an endowment of francs (the laboratory currency) sufficiently large to ensure they could take a draw every period of the round without exhausting their endowment. Each draw generated a value

between 0 and 999, with each number equally likely. Subjects were told the maximum number of draws in each round that could be taken, the cost of taking a draw, and the number of competitors in their group. At the end of each round, the player in each group with the highest draw was awarded the specified prize. A subject's total payoff at the end of the experiment was equal to the sum of the prizes won in each round plus all unspent francs remaining from the endowment.

> **"In the context of a research tournament, choosing to make a draw corresponds to conducting research at a constant cost-per-unit."**

The subjects did not know how many rounds would be conducted during the session. Finally, they were told they would be assigned to a different group each round, and at the end of the session their francs would be converted to dollars at a stated exchange rate.

In the context of a research tournament, choosing to make a draw corresponds to conducting research at a constant cost-per-unit. Beforehand, the outcome of the research process is unknown, but the distribution from which the research results will be drawn is common knowledge in Taylor's model. Each draw corresponds to the realized level of research for that period, and the group high draw is the level of the winning innovation for that round. Again, a round comprises several periods in which research can be conducted, but each round is a separate, independent research tournament. While we realize that the assumptions of Taylor's model are simplistic, our purpose is to give

the predictions of his model a chance at succeeding before violating those assumptions and creating a much more complicated and difficult experiment. Certainly varying the cost of research across firms, across time, and allowing the distribution of winning innovations to remain unknown would add a greater element of realism and deserves greater study, both theoretically and in the laboratory.

Experimental sessions covering five treatments were conducted. A session refers to subjects interacting in the laboratory, whereas a treatment refers to the specific parameters subjects face in a given session. A total of 103 subjects participated in these experiments and no subject participated in more than one session. Across sessions, we varied the number of competitors in each group, the maximum number of draws, and the prize (in francs) awarded to the competitor—with the largest draw in each group for each round.

EMPIRICAL RESULTS

In this section we subject our data to various tests at the market and individual level. We find the subjects do not individually employ the symmetric equilibrium stopping strategies predicted by Taylor. Some individuals behave as if their z-stop is below the predicted level while others behave as if it is above. However, we find that the aggregate behavior in each tournament treatment is generally consistent with the predictions of the theory, and that the majority of subjects do appear to be following a stopping rule strategy.

In *every* cross-comparison of treatments, the mean winning innovation and

the mean research-to-prize ratio moved in the direction predicted by Taylor's theoretical model. This is not to say that each individual subject accurately computed and employed the correct *z*-stop strategy, but rather that, on average, the winning innovation in each tournament and each group's cumulative research effort moved in the proper direction. The theory predicts virtually identical levels of winning innovations, and means that research-to-prize ratios increase across certain treatments but decrease across others. This is precisely what we have observed. At the market level, the data are qualitatively consistent with the predictions of Taylor's model. Thus, statistically, it appears that Taylor's model is internally consistent. By this we imply that *relative to other treatments,* when the mean level of the winning innovation was predicted to rise as a function of changing one of the parameters, our experimental data are consistent with the prediction. Moreover, changes in the distribution of winning innovations across experimental treatments are not only in the proper direction, but they are also statistically significant.

Taylor's theoretical predictions arise from the argument that the competitors adopt the *z*-stop strategies constituting the symmetric equilibrium. Using our data, we can estimate the implicit stopping rule that is generated by a subject's observed behavior. For example, if a subject uses a stopping strategy, the imputed *z*-stop may be estimated. Therefore, by calculating the average number of draws in our experimental sessions, we can estimate the *z*-stop that would generate the same number of experimental draws. The imputed z-stop strategies are close to the predicted levels for all treatments except one.[5] This particular treatment has both a large number of competitors as well as a long time horizon. As we shall see repeatedly, this combination of conditions exhibits more violations of the theory than do settings in which the tournament is short-lived or in which there are fewer competitors.

In addition to predicting a stopping rule, the theory also predicts a level of draw (research) activity. In all treatments except one, the subjects made slightly fewer draws than the level predicted by Taylor's theory. This result supports the conjecture that a large tournament (with many players invited) that continues for several periods may lead to excessive expenditures on R&D.

The variance in subject behavior is more likely to create problems for the tournament sponsor when the tournament is permitted to continue for several periods. With longer tournaments the potential exists for the subjects who overshoot the theoretical *z*-stop to overwhelm those who undershoot. This overshooting phenomenon has some serious implications if it bears out in real-world research tournaments. In particular, sponsors may risk driving some of their potential R&D firms to bankruptcy if they sponsor tournaments with too long a time horizon or too many competitors.

> **"In *every* cross-comparison of treatments, the mean winning innovation and the mean research-to-prize ratio moved in the direction predicted by Taylor's theoretical model."**

CONCLUSIONS

The focus of our experiments was to evaluate the fixed-prize mechanisms as a means to obtain a given quality of research at as low a cost as possible under various market conditions. Overall, the results appear to support the theory. At the market level, the winning research product and level of research effort tended to be close to the theoretical prediction. In addition, the majority of our subjects appear to employ a stopping rule strategy rather than a rule of thumb. However, instead of observing a uniform level of effort across all competitors, as the symmetric equilibrium would predict, the research strategies varied significantly across subjects.[6]

This variance tended to affect the aggregate results of our experiments. When there were only two periods for research, there tended to be less total research than predicted because there was not enough time for those who did the most research to make up for those who did very little. In the longer research tournaments with several competitors, we tended to see levels of research at or above the predicted amounts. Here, the high-effort competitors had ample research time to make up for the low-effort players, and the result was higher levels of winning innovations and in some instances "excessive" levels of aggregate research, which reduced the tournament efficiency.

The effect of additional participants and more research periods can potentially be substantial. The evidence supports the intuitive notion that if these parameters are increased arbitrarily, participants in long research tournaments may lose money because of excessive competition. In the long run, this would be self-correcting because either the competitors would adjust their behavior and collectively engage in less research, or some competitors would be driven out of the market and aggregate research would naturally decline due to fewer number of competitors. It may not be in the sponsor's long-term best interest for this to happen; judicious selection of the prize, the time horizon, and the number of competitors seems to be indicated.

There is still much to be learned about how individuals and firms respond to research tournaments. We hope to continue our investigation into research tournaments by varying parameters and the assumptions of Taylor's model and encourage others to do likewise. However, the preliminary results provide some evidence that fixed-prize tournaments are highly effective and efficient.

Lt Col Richard L. Fullerton is the deputy division chief of the Airborne Reconnaissance Division, Directorate of Information Dominance. Fullerton, an F–15 pilot, earned his Ph.D. degree in economics from the University of Texas, where he began his academic investigation of research tournaments. After teaching economics at the U.S. Air Force Academy, he was assigned to the Pentagon as a program element monitor.
(E-mail address: fullertr@af.pentagon.mil)

Lt Col Bruce G. Linster, USAF (Ret.), is an associate professor of economics at the U.S. Air Force Academy. A former C–130 pilot, he earned his A.M. and Ph.D. degrees in economics from the University of Michigan. Dr. Linster has published extensively in the areas of game theory, experimental economic methods, and evolutionary economics.
(E-mail address: Linsterbg.dfeg@usafa.af.mil)

Michael McKee is a professor of economics at the University of New Mexico and is currently on sabbatical as a visiting professor at the Department of Economics, School of Policy Studies, Georgia State University. McKee, a motorcycle enthusiast, earned his Ph.D. degree in economics from Carleton University, Ottawa, Canada. He has published numerous articles in refereed journals on a variety of topics, including experimental economics, public sector bureau behavior, and environmental economics.
(E-mail address: Mckee@unm.edu)

Lt Col Stephen Slate is an associate professor of economics at the U.S. Air Force Academy. While stationed at Whiteman Air Force Base as a missileer, Slate earned an M.B.A. degree from the University of Missouri, then earned his Ph.D. degree in economics at the University of Colorado at Boulder. His research and teaching interests include experimental economics, public finance, and acquisition reform.
(E-mail address: slatest.dfeg@usafa.af.mil)

REFERENCES

Camerer, C. (1995). Individual decision making. In J. Kagel, & A. Roth (Eds.), *The handbook of experimental economics* (pp. 587–703). Princeton, NJ: Princeton University Press.

Carnegie Commission on Science, Technology, and Government. (1992, December 1). *A radical reform of the defense acquisition system.*

Day, J. R. (1971). *Trains.* New York: Bantam Books.

Fullerton, R. L., & McAfee, R. P. (1996). Auctioning entry into tournaments (working paper). University of Texas.

Gansler, J. S. (1998, January-February). Realizing acquisition reform. *Program Manager,* 6–11.

General Accounting Office. (1997, January). *Reports and testimony* (GAO/OPA-97-4). Washington: GPO.

HDTV all together now. (1993, May 29). *The Economist,* p. 74.

Langreth, R. (1994, January). The $30 million refrigerator. Popular Science, pp. 65–67, 87.

Schwartz, J.. et al. (1991, May 6). The $93 billion dogfight. *Newsweek,* pp. 46–47.

Taylor, C. R. (1995, September). Digging for golden carrots: An analysis of research tournaments. *American Economic Review,* pp. 872–890.

Zubrin, R. (1996, November). Mars on a shoestring. *Technology Review,* pp. 20–31.

ENDNOTES

1. The contest, known as the Rainhill Trials, was used to select an engine for the first-ever passenger railroad in Britain. The £500 first prize was won by George and Robert Stephenson, who built the Rocket, which attained a top speed of 46 km/h. See Day (1971) for details about the evolution of steam locomotives.

2. The mission to Mars contest was worked up for a member of Congress by the executive chairman of the National Space Society, Robert Zubrin. The proposal is a series of contests with prizes in the $1 billion range, culminating in a $20 billion first prize (Zubrin, 1996).

3. By law, federal agencies are required to conduct competitive procurements whenever practicable. For example, in 1991 the Air Force held a "fly-off" competition to select the new Advanced Tactical Fighter. Lockheed won that competition with its F–22, and won the production contract, which was estimated at the time to be worth more than $90 billion (Schwartz, 1991).

4. A far more rigorous treatment of the theory, statistical results, and testing, as well as a more complete description of our experimental design, is available upon request and will be published in an upcoming edition of *Economic Inquiry*.

5. For these tests we eliminated the rounds in which individuals did not make any draws. The justification for this is that these are simultaneous move games. That is, when an individual chooses a strategy, it is based on the expectation that the group is of the announced size. Thus, the strategy choice is unaffected by whether one or more competitors has decided to drop out.

6. Such variance in behavior has been observed in many other individual decision settings. Camerer (1995) reports several examples of experiments in which subjects systematically overstated risk while others understated risk. It is an interesting question whether markets correct such behavior or whether aggregating market observations merely masks it.