

DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

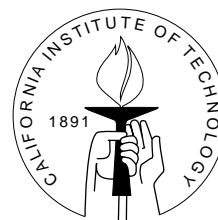
CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

THE ECONOMETRICS AND BEHAVIORAL ECONOMICS OF
ESCALATION OF COMMITMENT: A RE-EXAMINATION OF STAW &
HOANG'S NBA DATA

Colin F. Camerer

Roberto A. Weber



SOCIAL SCIENCE WORKING PAPER 1043

August 1998

The econometrics and behavioral economics of escalation of commitment: A re-examination of Staw & Hoang's NBA data

Colin F. Camerer

Roberto A. Weber

Abstract

We examine the phenomenon of escalation from an economist's perspective, emphasizing explanations which do not rule out rational behavior on the part of firms or agents. We argue that escalation cannot be established as a separate phenomenon unless these possible alternative explanations are properly accounted for. We present Staw and Hoang's (1995) study of NBA data as an instance of where evidence of escalation might be overturned upon more careful analysis. After performing several tests of our alternative explanations, we find that evidence of escalation persists, although it is weaker both in duration and magnitude.

JEL classification: D23

Keywords: Escalation; Sunk cost fallacy; Behavioral economics

The econometrics and behavioral economics of escalation of commitment: A re-examination of Staw & Hoang's NBA data*

Colin F. Camerer

Roberto A. Weber

1 Introduction

“Escalation of commitment” occurs when people or organizations who have committed resources to a project are inclined to “throw good money after bad” and maintain or increase their commitment to the project, even when its marginal costs (MC) exceed marginal benefits (MB). The escalation phenomenon, which is very similar to the “sunk cost fallacy”, is familiar to social psychologists and organizational theorists and is taken as well-established.¹

Compared to some other behavioral phenomena which question the rationality of firms and individuals, escalation has not captured the attention of economists. While escalation appears to be an important and large source of organizational (and individual) irrationality, economists have typically pointed to explanations that question the conclusion that escalation is common and suboptimal. In this paper we reopen the debate about suboptimal escalation in an effort to address concerns economists and econometricians are likely to have.

Our paper has two separate parts— one theoretical and one statistical. The theoretical section consists of an “economists’-eye” view of the large literature on escalation. We provide reasons why apparently irrational ongoing investments might actually be rational.

*We thank Amanda Rosenberg and Gail Nash for data entry, and participants at the Psychology-Economics Conference, Vancouver BC (June 6-7 1997), Linda Babcock, Max Bazerman, Richard Day, David Grether, Chip Heath, Keith Murnighan, Tom Ross, Barry Staw, Keith Weigelt and an anonymous referee for helpful comments. Corresponding author: Roberto A. Weber, Division of Humanities and Social Sciences 228-77, California Institute of Technology, Pasadena CA 91125

¹Escalation of commitment and the sunk cost fallacy are essentially the same phenomenon: both lead decision makers to exaggerate investments following previous commitment of resources. One distinction is that escalation may be associated with forms of commitment other than previous expenditures of economic resources, or sunk costs. For instance, it is possible for a decision maker to escalate following a verbal commitment. This implies that escalation is a more general phenomenon which includes escalating commitment to sunk costs.

We also argue that when escalation is irrational for firms, it is often explained by more basic forces which are familiar to behavioral economists (mutually-destructive rivalry, optimistic judgment biases, and agency costs).

Our statistical contribution is a closer look at existing field evidence of escalation. There are only four non-experimental multiple-observation field studies showing evidence of suboptimal escalation. One study shows that the tendency of banks to write off bad loans is correlated with managerial turnover (Staw, Barsade and Koput, 1997). Another study shows that employees' performance evaluations by supervisors were affected by whether the supervisors had hired the employees originally or not (Schoorman, 1988). A third study found that entrepreneurs who started their own businesses invested more than those who bought businesses from others (McCarthy, Schoorman and Cooper, 1993).

These three studies are consistent with economic explanations which suppose that people are rational but have different opinions (prior beliefs) about the quality of loans, prospective employees, and businesses. The findings in all three studies can be explained by prior beliefs which are not updated sufficiently quickly to converge to a common posterior belief. The studies are consistent with escalation as well, but they do not provide conclusive evidence because the differing-belief explanation has not been ruled out.

Our statistical analysis replicates and extends a fourth field study conducted by Staw and Hoang (1995), which demonstrates that NBA basketball players who were high-ranking draft picks played more minutes than was justified by their subsequent performance. Staw and Hoang interpret this result as evidence that teams escalated their commitment to high-ranking players, but their analysis does not entirely rule out some alternative rational explanations. We collected a new sample of data on NBA players and replicated their results. We then reexamined the data using variables and methods which could rule out alternative explanations. The results show a significant, temporary escalation effect which is about half as strong in magnitude and statistical strength as the original effect reported by Staw and Hoang, supporting their basic conclusion. Our results rule out rational alternative explanations more effectively than earlier studies did, and hence constitute the most conclusive available field evidence of irrational escalation.

2 Escalation in the NBA: A reexamination

Staw and Hoang tested whether players chosen earlier in the National Basketball Association (NBA) draft receive more playing time after controlling for differences in performance. The NBA draft is the annual process whereby players entering the professional league are allocated to teams. The draft consists of several rounds in each of which a team initially has one selection.² The use of a team's draft allocation on a particular player therefore reflects a commitment of resources to obtaining that player's talent –

²This can change, however, if teams make trades or deals with other teams for additional selections

the sooner a player is drafted, the more of a commitment it represents, both in terms of opportunity cost (foregone players) and financial resources (the NBA salary structure is such that rookie players' salaries are closely tied to their position in the draft).

As Staw and Hoang point out, however, draft order is not a perfect predictor of performance. It is often the case that players drafted early in the first round end up performing poorly and having short careers while players drafted later can turn out to be all-stars. Therefore, it is often the case that a team's initial evaluation of the value of a drafted player turns out to be wrong. The fact that resources have already been committed to such a player leads to a situation where escalation might arise: rather than decrease the player's court time based on poor performance, a team may over-use a player relative to his value if a large previous commitment has been made.

Staw and Hoang test the hypotheses that a lower draft number leads to more playing time, higher longevity in the league, and a lower likelihood of being traded. Draft order has a significant effect on all three dependent variables. They conclude that this is evidence of escalation of commitment in personnel decisions in the NBA.

The theoretical section below describes some of the conditions under which suboptimal escalation seems to occur. The findings of Staw and Hoang are surprising because they do not seem to fit these conditions. Escalation only occurs when there is ambiguity about future costs and benefits of an investment (like making loans, hiring employees, or drilling for oil); but there is less ambiguity about costs and benefits of an NBA draft pick because a player's salary is negotiated in advance and performance is easily measured and observed every night. Therefore, it seems that there is less room for optimistic or self-serving biases in judging costs and benefits which could cause escalation. Another condition for suboptimal escalation is an agency problem, in which managers continue investments which are bad for the firm because quitting would reveal their mistake. But NBA team management is usually very lean—draft choices are made by a general manager who consults closely with the team owner—so there is little room for an agency problem. Mutually-destructive rivalry is another condition for suboptimal escalation, but does not apply because teams don't play a mediocre draft choice to get back at other teams who are doing the same. Still another element of escalation is that marginal investments offer some hope of 'breaking even' or recouping the original investment. But teams spend draft picks and money initially, then make marginal investments in terms of playing time (minutes). They cannot directly recoup the sunk resource by spending further resources, because the resources are in different 'currencies' (draft picks and playing times). Taken together, the absence of the conditions under which suboptimal escalation appears to occur leads one to wonder why escalation occurs for NBA draft picks.

Staw and Hoang also acknowledge several rational explanations for escalation, but do not clearly rule them out. This leads to an important methodological question: How carefully must one rule out alternative explanations (even ones that seem implausible or are particularly difficult to test) before accepting the conclusion that irrational escalation occurred? Behavioral economists have learned that it is easier to win the hearts and minds

of economists with data that clearly rule out an alternative hypothesis than with mere reason and argument. We apply this principle to Staw and Hoang’s study and include some additional variables and methods to rule out alternative rational explanations.

Among the possible alternative explanations are:

1. Backup player ‘costs’. An explanation not considered by Staw and Hoang (mentioned to us by Chip Heath) is due to the nature of the draft itself. The draft is designed to give teams with worse records better chances of obtaining good players. Until recently, initial draft order was directly determined by a team’s prior season win-loss record.³ Therefore, it is reasonable to expect that teams which obtain the players with the lowest draft numbers tend to be teams which have less talent. Given that an important determinant of playing time is the expected performance of the next available player, players at worse teams, and therefore with lower draft numbers, might naturally be played more minutes than players taken later in the draft but at teams with more talent. Staw and Hoang include a measure of the team’s overall performance in their analysis, but a player-specific measure of the strength of the next-best player at their position is a better control. Including a measure of a team’s strength at a given player’s position controls for the marginal costs of not playing a draft choice, and might explain part of the draft-order effect.
2. Pre-draft expectations. An alternative which Staw and Hoang explore more carefully is the idea that draft order is an indication of the expected future performance or skill level of a given player. The same player’s performance during a season is then another signal of inherent quality. Teams could be gradually updating their beliefs about the player’s true ability, but if they begin with higher expectations for more highly-drafted players, then they should play those players more frequently until they become convinced that their initial expectations were too optimistic. It is hard to know whether this updating should take place over one year or five years, or ten. Draft order could also capture elements of player quality not encompassed by court statistics. Staw and Hoang recognize this possibility and measure the effect of draft order on performance by regressing a composite measure of performance on draft number and on player performance, position, trades and injuries in the prior year (the results of this are discussed but not presented in the paper). They find that draft order is a significant predictor of performance “surprise” in the second and third years, but not in the fourth and fifth years. They conclude that “while draft order does appear to contain some useful information on players’ early performance, it is not a significant predictor over longer periods of time” (p. 488).

The fact that draft order captures an element of performance not included in the previous year’s data for (possibly) the first three years is important enough to warrant further analysis. We try to capture information about expected performance

³Now the order is determined probabilistically, but teams with worse records have a higher chance of obtaining early draws.

in two ways. First, we include a measure of pre-draft player rankings by an outside expert. Second, we use a two-stage procedure in which performance is first predicted by lagged performance and draft choice (controlling for any predictive power of draft order), then expected performance is included in the playing-time regression. This procedure separates the informational component of draft order from the pure escalation effect.

3. Fan appeal. Highly drafted players may be especially popular with fans. If teams allocate playing time to maximize revenue, and fan popularity leads to increased revenue, it might be optimal to play highly-drafted players often to draw more ticket buyers to the games, even if the players are not performing well.⁴ Staw and Hoang note the possible effect but conclude that “the biggest problem with this alternative is that fan appeal is ephemeral. Though popularity among fans may be based on a player’s college reputation for the first year or two he is in the NBA, it is likely that popularity erodes quickly if it is not backed up by performance at the professional level.” (p 488). While this is probably true, it is still worthwhile to evaluate the possible impact of fan appeal on playing time rather than simply dismissing it.
4. Deescalating escalation: Draft x trade interaction. If escalation arises out of the draft as a commitment of resources, then when a player is traded to a new team, the new team should not inherit the escalation motives of the first team.⁵ So we can ask: When a player is traded to a new team, does the effect of draft number disappear? If so, then it is still possibly the case that escalation is present (since the second team should have no reason to commit resources based on the draft order). If not, however, then it is maybe evidence that draft order has a predictive value above prior statistics and that the new team recognizes and uses this information much as the first team. Notice that this test uses a prediction about when deescalation will occur as evidence that escalation occurs (cf. Heath, 1995).
5. First vs. second-round contract differences. There may be important economic differences between first and second round draftees. This is important because Staw and Hoang measured draft order by simply treating the first player in the second round as one position behind the last player in the first round. If there is a fundamental difference in first and second round costs or benefits, this should be controlled for by including a dummy variable for round.

Now (in 1997) it is typical for first-round draftees to get guaranteed “pay or play” contracts, which guarantee the players their salaries even if they are cut from the

⁴Some collateral evidence on this point comes from racial preferences. It is widely thought that in cities with larger white populations, teams keep white players, and perhaps play them more frequently than their performance warrants, because fans want them to. Nardinelli and Simon (1990) report evidence from baseball card prices which indicates that a player’s race (in addition to performance) affects the price of his card. This evidence suggests how it is possible that teams would not strictly maximize team performance and suggests a possible method for measuring fan appeal.

⁵One can argue that a trade represents a commitment which can also be escalated by excessive playing time, but it cannot be true that drafting players and trading for others both generate commitments which raise all players’ playing time.

team. Second round players usually do not have such contracts. This difference implies that first round players are “cheaper” because they must be paid even if another player is playing in their place. As result, teams have an economic incentive, *ceteris paribus*, to play first round players rather than second round players.

We are not certain when first-round guaranteed contracts became common (our sports-trivia sources think it might be around 1990). But if such contracts were common during much of our sample (and Staw and Hoang’s), this pay-or-play effect could be proxied for by draft order, producing a spurious finding of apparent escalation. To control for this possibility, we include a first-round dummy variable.

6. Use of aggregated lagged performance factors. To control for past performance, Staw and Hoang use lagged performance variables aggregated into three separate factors. There are two possible problems with this procedure. First, prior season statistics may not accurately capture a player’s performance on the court in a given season. This is important because it is current performance which should determine the amount of time a player spends on the court. Second, aggregation increases noise in these variables as measures of performance. If draft order contains information which predicts performance, the introduction of noise in measurement of performance means that any informational effect of draft order is picked up by the draft order coefficient in a regression on playing time, which inflates the coefficient on draft order, exaggerating the apparent escalation effect. We control for this problem by using ten separate performance variables, rather than aggregating them to three measures.

Our approach is to first replicate Staw and Hoang’s results as closely as possible, to see if we can replicate the effect of the draft-order variable which is consistent with escalation. Then we add control measures to see if draft order still predicts minutes when controlling for the above alternative explanations. We predicted that the draft order effect would disappear.

3 A Reexamination of Escalation in the Field: Analysis of Playing Time

Staw and Hoang’s sample includes all of the players taken in the first two rounds of the 1980-1986 NBA drafts. Our sample is similarly restricted to the first two rounds, but encompasses players selected in the 1986-1991 drafts. There are two reasons why we modified the years included in the analysis. First, we hoped to obtain data which would provide a measure of fan appeal for reasons given above. After considering endorsement contracts (for which data is unavailable) we decided upon pricing of player trading cards, which captures the extent to which a given player’s card is desired by fans and is usually accepted as a good measure of fan appeal. However, there is very little variance in card prices after 1990– only the cards of a few superstars exhibit any difference from the base price– and data were not available for some of the players in the sample. We also tried

to obtain records of the number of fans who voted for each player to be on the All-Star team, but could not obtain such data.⁶

Second, we wanted a ranking of expected player quality not directly related to draft order or expected order. A natural source proved to be pre-draft scouting reports. One such annual report is produced by NBA Analyst Don Leventhal. However, his reports are only available starting in 1985. Given that our goal is to first replicate Staw and Hoang's results, the use of different time periods should not be a problem if we can first establish their basic finding using data from a new time period.

Similarly to Staw and Hoang, we included in our sample only those players who were signed and played at least two years in the NBA. The reason for the two year minimum is that it is necessary to obtain prior performance data for the analysis and this data obviously only exists at the pre-professional level for all rookies. Therefore, while it would be beneficial to be able to conduct an analysis of first year playing time (particularly since this is when the draft order effect may be strongest under the draft-as-signal hypothesis), obtaining performance measures proves difficult.

Our sample, then, included a total of 229 players. To replicate Staw and Hoang's results, we obtained several season statistics for each player in our sample. Player statistics are carefully recorded during each game and are easily available from several sources. Our sources were primarily the following two books: *The Official NBA Basketball Encyclopedia* (Ed. A. Sachare, 1994) and *The Sports Encyclopedia: Pro Basketball* (Neft and Cohen, 1991). We also obtained information directly from the NBA through their website. As a replication of Staw and Hoang's main study, we used as our dependent variable the number of minutes each player was on the court during a season. In order to measure performance, we again used the same technique and obtained nine player statistics from the above sources. The following statistics were obtained for each player/season in our sample: total number of points scored, assists, steals, shots blocked, rebounds, personal fouls, free-throw percentage, field-goal percentage, and 3-point field-goal percentage. In addition, we controlled for the effect of playing time on total points, assists, steals, fouls, shots blocked and rebounds by dividing these measures by total minutes played.

In order to directly replicate Staw and Hoang's analysis, we conducted a factor analysis of the nine performance variables to determine broader performance measures, as they did. Using the same approach as Staw and Hoang, we divided players into two categories, guards and forwards/centers, and conducted separate principal components factor analyses with varimax rotation on each category. This procedure yielded results similar to Staw and Hoang's: the same three factors with eigenvalues greater than 1.0 surfaced for both samples. Together, these three factors explained 55 percent of the variance in the correlation matrix for forwards and centers and 57 percent for guards.⁷ Using the same structure as Staw and Hoang for labelling the components, we constructed the following three performance variables: 1) *scoring* (points per minute, field-goal per-

⁶The vote counts are not released by the NBA, apparently at the insistence of the players' union

⁷The three factors used by Staw and Hoang explained 58 percent of the variance for both samples.

centage, and free-throw percentage), 2) *toughness* (rebounds per minute, blocked shots per minute, and personal fouls per minute), and 3) *quickness* (assists per minute and steals per minute).⁸ Finally, we followed Staw and Hoang’s procedure and standardized the three measures so that all three variables had a mean of 0 and variance equal to 1, standardizing separately for guards and forwards.

In addition to the above performance measures, we included several other variables also present in Staw and Hoang’s study. First, in order to capture any effects of a player being traded, we included a variable *trade* which equaled 1 if a player was traded during or before a particular season and equaled 0 otherwise. Another variable, *win* was included to capture a team’s overall talent and consisted of the percentage of wins over the entire season. A dummy variable (*forward/center*) to capture any differences between forward/centers (1) and guards (0) was also included in the estimation. Finally, since injuries can obviously have an effect on playing time, a binary variable *injury* was included which was equal to 1 if a player was injured during a particular season and 0 otherwise.

The results obtained by Staw and Hoang are reported in Table 1 and the results of our replication are reported in Table 2. The model in both cases consists of regressing minutes in year t on performance in year $t - 1$, the above four control variables for year t , and on the player’s rank when drafted (draft number). Our results closely match those obtained by Staw and Hoang. There is no significant difference in either sample between players in the forward/center position and those in the guard position. Similarly, the variables win and toughness have no significant effect on playing time (in Staw and Hoang’s sample these variables are never significant at reasonable levels, and in our sample win is significant at $p < 0.05$ only in Year 2). In both samples, prior year scoring appears to be the performance measure which best predicts playing time. Injuries and trades also seem to have a consistently significant negative effect.

Draft order is a significant predictor of playing time in both samples. Our results indicated that a minimal increment in draft order decreased playing time from 22 minutes in year 2 to 11 minutes in year 5. These estimates are quite close to their estimates (23 to 14 minutes). In both samples, draft order is a significant predictor of playing time through the fifth year, and the effect shrinks over time.

In order to control for alternative explanations of the draft-order effect, we re-estimated the model with several new variables. The first set of additional variables were performance measures for the substitutes available for a team at the player’s position. That is, in order to measure the opportunity cost to a team of playing a particular player additional minutes, we obtained statistics for the (composite) back-up players in that position.⁹ The correlation between minutes played by drafted players in the sample and by their back-up players is $-.51$, which shows that drafted players and the composite

⁸These factors correspond to those used by Staw and Hoang with the exception that we included personal fouls in the toughness variable since this was consistent with the results of the component analysis for both samples.

⁹The back-up ‘player’ was actually a composite of all the likely players who would be on the court if

back-up are indeed substitutes. For each back-up, we collected data on the same ten performance variables as we did for each player in our sample. This data was used in the factor analysis described above, to construct the same three performance indices as we did for the other players. These indices are labelled *back-up scoring*, *back-up toughness*, and *back-up quickness*.

Among the alternative explanations outlined above is the idea that draft order captures beliefs about a player's skill which are updated but not perfectly determined by subsequent playing statistics. In order to test this hypothesis, we included a variable called *belief* to capture beliefs about player quality at the time of the draft. We use Don Leventhal's annual ranking of players available for the draft. Each year Leventhal, a draft analyst for the sports network ESPN, compiles a ranking of the top 180-200 players entering the draft. It is important to note that this ranking is intended to list players by quality and not by expected draft order. As Leventhal himself puts it, "The above list is not so much a prediction of the order of selection in the first-round, but rather a combination of my own and the NBA scouts' rankings of the top players . . ." (Leventhal, 1986 p. 6). In fact, in a separate section of the draft report, he includes a prediction of the order of the draft - taking into account teams' needs and other factors. This ranking includes all eligible players entering the draft and we include their rank number as a variable.¹⁰ Since we might expect a high degree of collinearity between belief and draft, we tested the correlation between the two variables and found it to be .67. While this does indicate that there is some collinearity, our hypothesis that the effect of draft will be reduced with the introduction of belief is still testable by estimating the marginal effect of belief on the predictive ability of draft.

In addition, in order to test the effect of trades on the escalation phenomenon (as we discussed above) we included a *draft x trade* interaction variable. Finally, we included a dummy variable, *first round*, which was set to 1 if the player was selected in the first round of the draft, and 0 otherwise.

the draftee was not playing, weighted by their (season total) minutes. If the player in our sample was a center, our variable included the two other centers with the most minutes. This procedure captures the fact that since there is only one center on the court at a given time, all the centers on the team are possible back-ups. If the player was a guard or forward, we first eliminated the player in that position with the most minutes, and then included the next three players in the same position with the most minutes. This procedure is appropriate because there are usually two guards and forwards on the court at a time; a guard or forward's back-ups should include only the remaining players after excluding the player with the most minutes (who is likely to be on the court). For draftees who are listed as playing two positions, or who play for more than one team during a season, we constructed a composite consisting of a simple average of backups for the two positions or their various teams. This procedure is very labor-intensive because it requires gathering information on at least three times as many players as there are draftees.

¹⁰For the first two years of our sample separate rankings are given for seniors and for underclassmen not among the top 24 players entering the draft. However, because almost all underclassmen entering the draft were among the top 24, we used the ranking of the top 24 along with the ranking of the top seniors to construct a ranking for all the players. There was only one non-senior who did not appear in the ranking of the top 24, and this player was given the highest possible ranking he could have received given that he was not on the list.

Table 3 presents the results of the regression including the additional control variables. Scoring and quickness are again the only two performance measures which are significant predictors of playing time. The coefficient on trade is again significant, but now only for years 2 and 3. However, the diminished effect of trade is not surprising given the introduction of the draft x trade interaction term.

Including the additional variables increases the adjusted- R^2 slightly, but Back-up scoring is the only additional control variable which is ever significant at reasonable levels. It is surprising that the coefficient on first round has the opposite sign from what we predicted for all four years.¹¹

The introduction of the control variables does not dampen the effect of draft order. However, the significance of the coefficient for the draft variable decreases more quickly across the years in our sample (for example, the coefficient has a t -statistic of -.67 in year 5).

In order to address the previously mentioned concerns about the use of lagged performance factors, we then re-estimated the model using contemporaneous performance variables. The dependent variable was again minutes in a season. As independent variables, however, we used all nine performance statistics for the same season, both for players and for back-ups. Because the number of times a player turns the ball over to the opposing team is an important measure of productivity which Staw and Hoang omitted, we included one additional performance variable: turnovers per minute, both for players in our sample and for their back-ups. In addition to draft order, we again included the control variables injury, trade, win, belief, first round, and draft x trade. Finally, since it was no longer necessary to use data from prior years, we conducted the estimation for all of the first five years of the players in our sample.

For this estimation, we restricted our sample in a given year to players for which we had data on all prior years. That is, we omitted any players with prior gaps in their playing careers.¹²

Table 4 contains the results for all five years. Notice that the corrected R-squared values are considerably higher in Table 4 than they are in Table 3 for all years in which the data overlap, indicating that current performance is a better predictor of current minutes than lagged performance (as expected). Among the performance variables, personal fouls per minute appears to be the strongest predictor of playing time. This is not surprising since it is usually the case that players are taken off the court once they

¹¹The estimate indicates that first round players actually receive less playing time than those selected in the second round. While our hypothesis called for a one-tailed test and, therefore, the coefficient is never significantly different from zero in the direction we predicted, it would only be significant in one out of the four years in a two-tailed test.

¹²This was done in order to enable us to make comparisons between the results of these regressions and those of future regressions where performance data on all prior years is necessary. We also conducted the same analysis using all of the players in our sample who played in the given year and found that our results were not altered significantly.

accumulate personal fouls in order to prevent them from exceeding the foul limit and being disqualified from further play. Finally, field goal percentage, points per minute and assists per minute also appear to have an effect on playing time, although none of these is significant throughout all five years.

As earlier results indicated, the effect of draft order on minutes played is strong in the first three years of a player’s career. For these years, a minimal decrease in draft number predicts an increase of up to 29 minutes in playing time. The magnitude of the coefficient is always greater than 25 and the coefficient is significantly different from zero at the $p < 0.001$ level in all three years. However, the effect of draft order in years 4 and 5 is dramatically reduced and insignificant.

While it is possible to interpret the decrease in the importance of draft order as a player’s tenure in the NBA increases as a diminishing effect of escalation, the decrease is also consistent with the idea that the draft captures prior information about the quality of players which is subsequently discounted. Given that these priors are subjective interpretations by teams, it is difficult to accurately measure them or estimate their effect. If these priors are overly optimistic, then it is possible that the observed “escalation” is simply a manifestation of this bias in judgment. However, even if the priors are unbiased, and represent accurate measures of overall quality, then it is possible that the rational use of prior beliefs is evident in the predictive ability of draft order.

One possible method for evaluating the existence of the rational use of unbiased priors is to determine the effect of draft order on performance, after controlling for prior performance.¹³ If information present in the draft is an unbiased predictor of future performance and accurately serves as a measure of a player’s overall quality, then the marginal effect of draft order on performance should capture such information not picked up by prior statistics. This provides us with a possible test of the draft order as escalation vs. information hypothesis. If draft order merely captures unbiased prior beliefs, then the effect of draft order on performance (after controlling for prior performance) should be similar in duration to the effect of draft order on playing time. However, if escalation is present, then the effect of draft order on minutes played should persist beyond the predictive value of draft order on performance.

In order to conduct this test, we regressed performance in year t on performance in prior years as well on draft order. We conducted separate regressions for Years 2 through 5. In order to aggregate the several performance variables we had for each player, we constructed an index, *performance*, by averaging the three performance variables: scoring, toughness, and quickness. This variable was then regressed on each of the three performance factors for all prior years and on draft order. In addition, we included the variable for injuries, since it is possible that a player’s performance may be lower if he is injured. The results are presented in Table 5.

The results indicate that draft order does have some significant power for predicting

¹³This is the method used by Staw and Hoang, but they reported no details.

performance (at $p < .05$), through the fourth year. Table 6 summarizes the effects of draft order in Tables 4-5, comparing the effect of draft order on playing time (minutes) and on performance. Staw and Hoang used this comparison to argue for escalation because the effect of draft order on performance became insignificant in years 4 and 5, while the effect on minutes was significant in all years. Our results are opposite. The draft order coefficient in the minutes regression is highly significant in years 2 and 3, but becomes insignificant in the fourth year, while the effect on performance is still marginally significant in the fourth year. This pattern suggests that the escalation effect of draft order on playing time ‘wears off’ before the information content of draft order for predicting performance does.

This finding goes against the escalation hypothesis. If the analysis were to stop at this point, one could conclude that there is no evidence of an irrational escalation effect. However, the effect of draft order on playing time is statistically stronger in years 1-2 than the effect on predicting performance (measured by t-statistics). This fact suggests that there might be an informational effect and an escalation effect in the first two years. The analyses summarized in Table 6 simply do not tell us whether there is a pure escalation effect of draft order on minutes, above that which can be explained by the information value of draft order, in years 1 and 2. There might be irrational escalation, but we simply cannot tell.

Fortunately, there is a simple two-stage procedure for separating the two effects in each year. First, we perform the following regression for each of the ten performance variables j and for each year T :

$$P_{jT} = \alpha_{jT} + \sum_j \left(\sum_{t=1}^{T-1} \beta_{jt} P_{jt} + \gamma_{jT} P_{jT}^{BAK} \right) + \sum_k \delta_{kT} x_{kT} + \rho D + \varepsilon_{jT} \quad (1)$$

The fitted portion of this regression, \hat{P}_{jT} , represents the expected performance in year T given prior performance and all the exogenous variables, including draft order. Hence, this measure contains any information present in draft number which may be relevant to the expected current output of a player in a given performance category.

We then used these expected performance measures in place of actual performance statistics in the minutes played regression:

$$M_T = \alpha_T^M + \sum_j \left(\beta_{jT}^M \hat{P}_{jT} + \gamma_{jT}^M P_{jT}^{BAK} \right) + \sum_k \delta_{kT}^M x_{kT} + \rho^M D + \varepsilon_{jT}^M \quad (2)$$

If draft order contains information about performance, but does not influence playing time after that information content is controlled for – i.e., there is no irrational escalation

effect – then the coefficient on draft order should be insignificant in the second-stage playing-time regression.

Table 7 shows the results of the second-stage regression.¹⁴ This regression keeps a constant sample of players and includes the expected performance measures (derived from the ten first-stage regressions of performance against lagged performance and all the other exogenous variables, including draft order). The effect of draft number is about half as large in magnitude as in the earlier results (e.g., Tables 2-3), declining from 14 minutes per increment to 2 minutes over years 2-5. The effect is also significant in years 2-3, at $p < .01$, and insignificant after that.

To recap our results: The new sample replicates Staw and Hoang’s finding of irrational escalation in NBA playing time using their methods (Table 2). The finding holds up when additional control variables are included (Table 3) and contemporaneous performance is used instead of lagged performance (Table 4). While a direct regression of lagged performance and draft order on performance shows that draft order does have some forecasting power for predicting future performance (Table 5), a simple comparison of the playing time and performance regressions cannot distinguish the (rational) information effect from the escalation effect (Table 6). However, a two-stage procedure disentangles the two (Table 7), and shows that when the information contained in draft order is controlled for, an apparent escalation effect persists in the second and third years.

The methodological moral of the story is that including more controls and using the two-stage procedure supports Staw and Hoang’s conclusion, establishing more firmly that there is an irrational escalation effect in field data. At the same time, the controls and procedure show that earlier results exaggerated the size and persistence of the effect by roughly a factor of two. The exact source of the effect is unclear from these analyses.¹⁵

¹⁴In the previous results (Table 4), the control variables for Belief, First round, and Draft*Trade were generally insignificant or did not help explain the apparent escalation effect. Therefore, we exclude these variables from the two-stage regression. In addition, to control for sample composition effects, we only included players who appear continuously in the sample for five years.

¹⁵The two-stage procedure essentially assumes that teams have rational expectations about the information contained in draft order for predicting performance (in the standard sense, that subjective beliefs match objective beliefs conditioned on available information). The results are still consistent with the hypothesis that teams are sluggishly rational, in the sense that they have prior beliefs but update them more slowly in the face of performance evidence than Bayesian updating assumes. The results are also consistent with the hypothesis that teams have optimistic priors (e.g., each team thinks its draft choice will perform better than average). A further possibility is that teams who are most optimistic about a player, and therefore have the strongest positive beliefs, are more likely to select that player in the draft and then overplay him until they gather enough information about his true ability. This argument is similar to the principle referred to as the winner’s curse, whereby the most optimistic bidder who highly overvalues an item is the most likely to win the item in an auction, and does not anticipate this selection bias when deciding how much to bid.

4 The behavioral economics of escalation

An economist looking at the escalation literature immediately wonders whether escalation is clearly established: Are all marginal costs and marginal benefits accounted for? If so, then one can proceed to ask whether causes of suboptimal escalation can be organized into categories which are familiar in behavioral economics. This section attempts to do so.

While previous sections have dealt specifically with a re-examination of the evidence of escalation in the NBA, the scope of this section is more general. Here we attempt to examine the escalation phenomenon more broadly and point to more general reasons why the phenomenon may not be as firmly established as is often believed. While not all of the categories are directly relevant to the NBA case, we draw a link where possible.

4.1 Is $E(MC) > E(MB)$?

We denote the “true” expectations of MC and MB , using all publicly available information, by $E(MC)$ and $E(MB)$. The first step to establishing suboptimal escalation is to check whether $E(MC)$ is indeed greater than $E(MB)$. If not, then by investing more a firm is not escalating suboptimally. Three observations should be made at this point. One is well-established in the escalation literature but bears repeating. The other two seem more novel and might not occur to non-economists.

1. First, accounting for MC and MB carefully and choosing optimally will sometimes produce decisions which will be suboptimal, after the fact, and which might look like escalation to an untrained eye. For example, the film “The Titanic” had an initial budget of \$150 million and was expected to gross, say \$200 million. Notationally, express this initial situation as $E_0(MC) = 150$, $E_0(MB) = 200$. Now partway through, the studio has actually spent \$200 million and thinks it will cost \$85 million to finish. Assuming a half-filmed movie is worthless, should the studio “escalate”? The answer is “Yes”. The marginal costs are $E_1(MC) = 85$ and, if revenue expectations have not changed, the marginal benefits are $E_1(MB) = 200$. Of course, if the revised estimates are accurate the firm will end up spending a total of \$285 million to earn \$200 million, losing money. But the blame should be pinned on optimistic initial cost forecasts, not on suboptimal escalation. This example reminds us that bad outcomes are not always evidence of suboptimal escalation.¹⁶

2. Inferring benefits from sunk costs: In some examples it is natural to assume that information about MC also affects estimates of MB . In these cases, when information arises which raises MC , and seemingly makes $E(MC) > E(MB)$, marginal benefits might sensibly rise too, which would justify escalation. For example, in the clever Arkes

¹⁶Indeed, the large literature on “hindsight bias” suggests that when bad outcomes do occur, the path which led predictably to those outcomes is psychologically available and leads to a tendency to overestimate how easy it should have been to avoid disaster.

and Blumer (1985) study, some students who approached a box office to buy tickets for a theater series were randomly given discounted tickets. They found that the students who got discounted tickets came to the theater less often (for the first half of the shows) than those who paid full-price. A picky critique of this study is that students who got discounts might have inferred something about the quality of the production from the fact that they got a discount (i.e., perceived MB of attending is affected by MC).¹⁷

Similarly, many of the most compelling examples of escalation come from studies in which subjects are presented with vignettes describing an investment project. One group of subjects are told that a large amount has already been invested and another group are told that a small amount has been invested. Often the large-amount subjects choose to invest more. Interpreting their behavior as suboptimal escalation is problematic unless one is very confident that all else is held equal (in the minds of the subjects) in the two cases. It may be natural for subjects to assume that firms or individuals investing a large amount are more optimistic and perceive higher marginal benefits, and hence should invest more. Notice that experimentally, it is hard to rule this possibility out in a between-subjects design: You can tell subjects to “assume nothing more than what you are told” but this is no guarantee that they will not make some inference about expected MB from previous investment. A better control is to conduct this experiment within-subjects, asking the same subject whether the large- and small-amount firms should invest more, and carefully instructing them to draw no inference about expected benefits from previous investment. Within-subject control seems better-suited to guaranteeing that subjects do not infer anything about benefits from past costs. To our knowledge, this precise experiment has not been done.

3. Information benefits of escalation: A proper accounting for MB should include the value of any information which is (only) gathered by investing further. Since information value is never negative, omitting this consideration always biases $E(MB)$ downward and makes suboptimal escalation appear more likely than it is. For all-or-nothing projects, the information value of escalating might easily justify continuing to invest, even when $E(MC) > E(MB)$. Consider a firm that is undertaking some modest investment which, if successful, will be repeated over and over (e.g., a restaurant franchise renovating the dining room in one of its restaurants). Suppose the renovation cost escalates, so that the cost of finishing one renovation, MC , is higher than the expected benefit for that restaurant. Then it may still pay to finish. The eventual benefit MB may turn out to be so low that, having experimented, the firm realizes it should not make the same investment at all its outlets. On the other hand, if the eventual benefit MB turns out to be high it can be realized at outlet after outlet. Since the information gained about whether to invest again is very valuable, it could be still worthwhile to escalate even when the current-project MC are much larger than MB .

In the case of NBA draft picks, it is possible that higher draft picks receive more

¹⁷A fussy control for this possibility is to tell all students that some students were awarded discounts, so that even students who paid full price would also have any information about production quality which is conveyed by the fact that discounts were available.

playing time because teams are trying to determine the true value of their investment. Since teams have only limited information about important factors such as a rookie player’s true NBA prospects or the ability of the player to fit into the team’s system, then they might be playing their early draft picks more minutes to gather information. While this would explain why the data reveal excess minutes given to early draft picks during their first season or two, it is unlikely that this effect would persist into later years.

4.2 Behavioral economics causes of suboptimal escalation

Assume $E(MC) > E(MB)$ and firms escalate anyway. What explanations might be offered for escalation? We organize them into four categories which are familiar in behavioral economics. Some of these are familiar in the organizational literature but others are not.

1. Gambling in the domain of losses. Many researchers have pointed out (e.g., Bazerman, 1984) that escalating commitment is utility-maximizing if people have a convex disutility function for losses, so that they prefer to gamble in the domain of losses in order to “break even” (as in prospect theory, Kahneman and Tversky, 1979). Formally, denote sunk costs by C_0 . Then this argument can be expressed formally by the condition $u(C_0) < u(MB - MC - C_0)$.¹⁸ Furthermore, Heath (1995) argues persuasively that escalation is often inhibited if marginal costs are paid in a different “currency” than sunk costs were paid in. Then MC and C_0 may not be naturally added together on a single scale and the escalation condition $u(C_0) < u(MB - MC - C_0)$ need not hold.

In corporate decisions, the importance of mental accounting implies that how decisions are combined organizationally may moderate escalation in important ways (as discussed by Kahneman and Lovallo, 1993). Consider biotech firms that undertake many R&D projects at a small scale, but produce very few marketable products. How do these firms avoid escalating their modest investments after initial failures (which are typical)? The answer is probably that divisional managers take a portfolio view of different projects, which makes it easier to “write off” sunk costs from any one project.

Heath (1995) makes the profound point that regarding escalation as a byproduct of gambling over losses, under suitable mental accounting conditions, implies that under predictable conditions firms will fail to escalate when they should (i.e., they suboptimally deescalate). He shows that when total costs will exceed a natural budget point, subjects often fail to invest further even though they should. In our view, this important study sharply points out that escalation, per se, is not a useful

¹⁸This argument depends on how costs and benefits are “mentally accounted” for. The condition above assumes that the person’s reference point for accounting for costs and benefits is set at zero (or the pre-sunk cost level). If subjects “reset” their reference point after costs are sunk, then the correct comparison is $u(0)$ versus $u(MB - MC)$, and suboptimal escalation cannot be easily explained this way.

category of phenomena. Instead, escalation is a byproduct of basic forces which are present in individual decisions, may sometimes be optimal, and may sometimes lead to the opposite of escalation.

2. Agency costs. In many organizational situations, $E(MC) > E(MB)$ for the firm but the private values to a decision-making agent favor escalation, denoted $\pi_{agent}(MC) < \pi_{agent}(MB)$. This argument has been made several times with various degrees of formality (Kanodia, Bushman, and Dickhaut, 1989; Prendergast and Stole, 1996). The basic theme is that escalation which is suboptimal for the firm, but privately rational for the agent, is just one example of how agents may take actions which are bad for the “principals” who hire them. These “agency costs” are frictions or losses which can be blamed on the fact that contracts and trust are not sufficient to completely eliminate conflicts between contracting parties with different interests.¹⁹

Here we give a brief sketch of how this “agency problem” could lead to escalation. Suppose agents are of two types, smart and dumb. Smart agents choose better projects (with higher benefits and lower costs, statistically speaking) than dumb agents do. Suppose that after a project is fully completed, the principals (e.g. a supervisor or board of trustees) can observe the project’s total costs and benefits, which gives them a clue – but an imperfect one – about whether the agent who chose the project was smart or dumb. Now consider the incentives of agents who, part of the way through the project, come to learn that the project is bad (so $E(MC) > E(MB)$). Should they continue investing? Under many conditions (though not all), the answer is “Yes”. If they abandon the project that is a clear probabilistic sign that the project was a bad one, and hence, that the agent was dumb. Agents may prefer to continue investing and take their chances of being fired, later, when the total costs and benefits are observed by their supervisors.

Applying the agency problem to the NBA, one could argue that coaches and general managers (who are usually primarily responsible for draft decisions) play earlier drafted players more minutes to validate their choices in the draft and not reveal any mistakes. However, as we argue above, NBA organizations tend to have small front offices where there is much interaction between owners, management, and coaches. An agency problem is therefore less likely to arise in this situation since owners are often actively involved in draft choices and subsequent team operations.

Staw (1996) comments, “My own view is that rational models such as agency theory have simply translated social and psychological concepts into cost-benefit analyses, providing few advantages in making a priori predictions about escalation” (p 195). We agree up to Staw’s comma, and disagree after that. It is true that many behaviors which have been well-documented in early organizational studies and given psychological interpretations can be “rationalized” as the outcomes of signaling games in which principals lack information that agents have, and agents have conflicting incentives about how to act on their information. Gibbons (in

¹⁹Another possibility, which is rarely mentioned in the agency theory literature, is that principals are making mistakes by not choosing better-designed contracts, by trusting untrustworthy agents, etc.

press) summarizes some of these ideas and argues cogently for their place in organizational research.²⁰

However, we disagree with Staw's conclusion that these models provide few predictive advantages. In most cases, the theories provide clear conditions under which agency problems will not occur. For example, escalation is less likely to occur as principals know more about the ability of agents or as agents have less career incentive to protect their reputations. Thus, agents who are older, or who are managing partners in partnerships or have job tenure are probably less likely to escalate than others.²¹ Zweibel's (1995) theory predicts that agents who have either very bad or very good reputations are more likely to take socially-risky projects (e.g., de-escalate when middle-reputation agents are escalating). Also, the agency theory view predicts less escalation in owner-managed firms, where the agency problem is small or nonexistent.

3. Overconfidence. In many examples of escalation, it seems likely that agents simply overestimate MB and underestimate MC , relative to estimates given by outsiders looking at the same situation. This distinction is important because, in principle, perceptions of MC and MB might be measured separately and related to escalation behavior. Digging deeper, these estimation biases might be caused by cognitive dissonance, self-justification, or other forces. In the case of the NBA, coaches might make overly optimistic predictions about a player's prospects and then update these beliefs too slowly. Furthermore, these biases could also cause deescalation— that is, if one person has refused to invest in a project, they may escalate their refusal to do so even when new information raises estimates of MB or lowers estimates of MC and makes investment optimal.
4. Mutually-destructive escalation of unfairness. The “all-pay” dollar auction (Teger, 1980) is often taken as a prototypical example of organizational escalation. In these auctions, players compete for a fixed prize, say a dollar, by bidding. The high bidder gets the prize but all bidders pay their bids. Bids begin at zero and can only be raised in small increments. A typical pattern is that bids start out timidly, and quickly escalate. Bidders often pause when bids get around a dollar, as they realize that they will end up paying more than a dollar if they continue. But players think that if they just bid a little more, the marginal cost is small (since the previous bid is already sunk) and the marginal benefit is a dollar. Winning bids often end up many times larger than a dollar. This example is instructive because it includes both escalation and a social or game-theoretic component. Part of the reason for escalation seems to be that players become angry at one another because when one

²⁰For example, several theoretical papers show how agents who are averse to revealing their skills will rationally “herd” and do what other agents do, even if the herd is likely to be acting inefficiently (Scharfstein and Stein, 1990; Zweibel, 1995). Prendergast (1993) shows how promotion processes which rely on supervisors rating their underlings creates economic incentives for the underlings to act in “yes-man” ways that please supervisors.

²¹For example, this idea predicts that lawyers settle cases more quickly if they are partners than if they are associates, and settle more quickly in small firms with fewer promotion rungs than in large firms.

person raises his or her bid, it harms others (by reducing their chance to get the dollar). The only way for a harmed bidder to inflict harm back is to raise his or her bid in turn.

In this example it seems reasonable to conclude that players are thinking about not just MC and MB , but about the social utility which arises from harming or helping others who have harmed or helped you. If one properly models the utility which arises from bidding, it may turn out that escalation is utility-maximizing even if it costs bidders money. An initial, productive attempt to include these forces into an extension of standard game theory has been made by Rabin (1993). In Rabin's model, a player A imagines that the other player knows her likely action, and then forms a judgment about whether the other player's response to A's action is "nice" (gives A a large economic payoff) or "mean" (gives A a low payoff). Rabin assumes that people get utility from economic payoffs and from the product of their own fairness and the other player's fairness. Thus, if the other player is nice to A (positive fairness), A maximizes utility by being nice as well. But if the other player is mean to A (negative fairness) then A prefers to be mean as well. We conjecture that mutually-destructive escalation in the dollar auction is a negative "fairness equilibrium" in a formal sense that Rabin makes clear.²²

5 Conclusion

There are two methodological messages implicit in how we approach the topic of escalation. The first is that carefully dissecting a phenomenon to understand its basic causes, and posing sharp questions about whether one cause or another can account for the phenomenon, is a useful way to do research. The escalation literature seems to have sometimes gone in the opposite direction, compiling long lists of various forces and considered escalation to be multiply determined— which it surely is— without trying to figure out which effects are more important or distill the long list into a short one.

²²Our argument is simply that mutually-destructive escalation in dollar auctions is distinctly different from other types of escalation because it involves more than one party. If players get utility from how others treat them and how they treat others, then escalation could be expensive but utility-maximizing in a way that is not possible in single-player decision problems. Furthermore, this conjecture suggests at least two types of interesting experiments which have not been done: (i) Conduct a single-player version of the dollar auction in which players bid against a nonhuman bidder who is programmed to stop at some unknown point. Unless people get social utility from how a program treats them (and vice versa), we expect to see much less escalation in this setting than when bidding against other people. (ii) Conduct a standard dollar auction, but halt the bidding at some previously-unannounced point, and give the high bidder a chance to "deescalate" the process by lowering his or her bid to just below the next-highest bid. If he or she does so, then the next bidder (who will now have become the high bidder) can deescalate, and so forth. While lowering one's bid is not money-maximizing for the ranking high bidder (it lowers costs, but means surrendering the dollar), in Rabin's terms this action is "nice" and could trigger a nice reciprocation by the next bidder, and so on. This also allows the possibility that players escalate niceness rather than simply escalating meanness (in the standard auction). Note that an upward escalation of bids, followed by a downward deescalation, would both be evidence of escalation— just in socially-opposite directions.

The second methodological message is that establishing an irrational behavior using field data is econometrically tricky. When an effect appears to exist, one can acknowledge alternative explanations and try to dismiss them with argument, anecdotal evidence, or reason. Or one can include control variables and see if the apparent effect then disappears. If the effect holds up, as the escalation effect in NBA playing time did, it becomes deserving of even more serious attention.

More generally, psychological and economic methods are productive complements, not substitutes. Economics ignores many aspects of behavior (details of thinking, social influences, limited rationality) in order to concentrate on two things: (i) simple formal assumptions about behavior (e.g. utility maximization) and (ii) predictions about naturally-occurring data. As a result, economic reasoning is excellent for checking whether behavior is rational and for drawing solid inferences from field data and testing alternative explanations tenaciously. We have argued that both of these features of the economic method could be brought more directly to bear on escalation. At the same time, psychological constructs and methods (particularly experiments) are useful for establishing sources of irrationality, which can then be plugged into economic theory with an eye to prediction of naturally occurring patterns. Thus, good work of one sort raises the marginal value of good work of the other sort.

5.1 Suggestions for further research

One way to demonstrate a firm grip on the phenomenon of escalation is to ask about the conditions under which its opposite, deescalation, occurs (see Heath, 1995) – that is, when do firms quit too soon? Asking this question could be a diagnostic tool to evaluate underlying explanations. For example, if escalation is due to gambling in the domain of losses, then one should see firms deescalating in the domain of gains.²³ Similarly, where escalation is due to mutually-destructive rivalry, we might also see deescalation due to mutually-beneficial cooperativeness. And if escalation is due to agency problems which make managers reluctant to pull the plug on losing projects, do those same agency problems lead to the failure to start promising projects in the first place? Surely this is the case.

It might be useful to look at instructive exceptions of firms or situations in which escalation is often avoided. For example, consumer products companies introduce many new products each year; and pharmaceutical companies work on hundreds of possible products (e.g., drugs), mostly with negative initial feedback. How do these organizations manage to avoid escalating?

Other naturally-occurring organizational examples do fit the recipe for escalation and could be studied in further research. For example, rich horse owners buy high-price

²³Data from individuals buying and selling stocks, for example, show that “losing” stocks are held substantially longer before selling them than “winning” stocks which are ‘cashed in’ more quickly. See Weber and Camerer (in press) and Odean (1997).

(unraced) yearlings at age one. These horses debut in “maiden special weight” races and often perform poorly. Horses that cannot win special weight races are usually dropped to “claiming races” in which every horse can be purchased for a prespecified price. An owner who spent a large sum for a yearling (\$500,000 and up are not unusual) may be reluctant to drop the yearling into the claiming category since the owner can prevent an irreversible loss by keeping the horse in the special weight races category. (Running in a claiming race risks “losing” the sunk cost.) The charm of this example is that betting odds established by the public give an objective measure of the horse’s true ability in each type of race, which can be compared with the owner’s (possibly biased) assessment. A clear escalation bias is revealed if the betting odds for high-priced yearlings in special weight races are higher than for low-priced yearlings.²⁴

We think our study illustrates two points: First, establishing systematic mistakes using naturally-occurring data is very difficult. Of course, this does not mean we should avoid such hard work and exploit the superior control of the lab; it just means that the standard of proof for mistakes outside the lab is high, and should be. It is also likely that important field anomalies will not be established by a single study, but by a series of studies which build on earlier results.²⁵ Behavioral economists have learned that the best way to win an argument about the existence of systematic mistakes is to take complicated rationalizations offered by critics seriously (no matter how cockamamie they are), and collect more data to test them. Our results on escalation in the NBA are an example of how an interesting result can be established more firmly when alternative explanations are properly accounted for.

²⁴Another example is contracts for new musical acts (bands). Most bands are signed to provide several records in several years and given a large financial advance. The record company has the right to refuse a record and drop the band. One can study whether record companies escalate their commitments to bands who they signed for the largest advances. (This example is similar to the NBA draft because establishing escalation clearly requires an outside measure of the record company’s pre-signing expectations of how well the band will do.)

²⁵Importantly, if journals are not receptive to publishing studies which build on earlier results, cumulating knowledge, then knowledge will not cumulate as swiftly as it could.

References

- [1] Arkes, H. R. and C. Blumer. 1985. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes* 35(1): 124-140.
- [2] Bazerman, M. H. 1984. The relevance of Kahneman and Tversky's concept of framing to organizational behavior. *Journal of Management* 10(3): 333-343.
- [3] Heath, C. 1995. Escalation and de-escalation of commitment in response to sunk costs : The role of budgeting in mental accounting. *Organizational Behavior and Human Decision Processes* 62(1): 38-54.
- [4] Kahneman, D. M. and D. Lovallo. 1993. Timid choices and bold forecasts : A cognitive perspective on risk taking. *Management Science* 39(1): 17-31.
- [5] Kanodia, C., R. Bushman and J. Dickhaut. 1989. Escalation errors and the sunk cost effect: An explanation based on reputation and information asymmetries. *Journal of Accounting Research* 27(1): 59-77.
- [6] Leventhal, D. 1986-1991. Don Leventhal NBA Draft Report. Annual reports.
- [7] McCarthy, A. M., F. D. Schoorman and A. C. Cooper. 1993. Reinvestment decisions by entrepreneurs : Rational decision-making or escalation of commitment? *Journal of Business Venturing* 8(1): 9-24.
- [8] Nardinelli, C. and C. Simon. 1990. Customer racial discrimination in the market for memorabilia: The case of baseball. *The Quarterly Journal of Economics* 105(3): 575-595.
- [9] Neft, D. S. and R. M. Cohen. 1991. *The Sports Encyclopedia 1891-1991, Pro Basketball* 4th Edition. New York: St. Martin's Press.
- [10] Odean, T. 1997. Are investors reluctant to realize their losses? Unpublished manuscript.
- [11] Prendergast, C. 1993. A theory of "yes men". *American Economic Review* 83(4): 757-770.
- [12] Prendergast, C. and L. Stole. 1996. Impetuous youngsters and jaded old-timers : Acquiring a reputation for learning. *The Journal of Political Economy* 104(6): 1105-1134.
- [13] Rabin, M. 1993. Incorporating fairness into game theory and economics. *American Economic Review* 83(5): 1281-1302.
- [14] Sachare, A. (Ed.). 1994. *The Official NBA Basketball Encyclopedia*, 2nd Edition. New York: Villard Books.
- [15] Scharfstein, D. S. and J. C. Stein. 1990. Herd behavior and investment. *American Economic Review* 80(3): 465-479.

- [16] Schoorman, F. D. 1988. Escalation bias in performance appraisals: An unintended consequence of supervisor participation in hiring decisions. *Journal of Applied Psychology* 73(1): 58-62.
- [17] Staw, B. M, S. G. Barsade, and K. W. Koput. 1997. Escalation at the credit window: A longitudinal study of bank executives' recognition and write-off of problem loans. *Journal of Applied Psychology* 82(1): 130-142.
- [18] Staw, B. M. and H. Hoang. 1995. Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. *Administrative Science Quarterly* 40(3): 474-494.
- [19] Teger. A. 1980. *Too much invested to quit*. New York: Wiley.
- [20] Weber, M. and C. Camerer. 1997. The disposition effect in securities trading: An experimental analysis. Forthcoming in *Journal of Economic Behavior and Organization*.
- [21] Zwiebel. J. 1995. Corporate conservatism and relative compensation. *Journal of Political Economy* 103(1): 1-25.