THE MATCHING RELATION AND SITUATION-SPECIFIC BIAS MODULATION IN PROFESSIONAL FOOTBALL PLAY SELECTION

STEPHANIE T. STILLING AND THOMAS S. CRITCHFIELD

ILLINOIS STATE UNIVERSITY

The utility of a quantitative model depends on the extent to which its fitted parameters vary systematically with environmental events of interest. Professional football statistics were analyzed to determine whether play selection (passing versus rushing plays) could be accounted for with the generalized matching equation, and in particular whether variations in play selection across game situations would manifest as changes in the equation's fitted parameters. Statistically significant changes in bias were found for each of five types of game situations; no systematic changes in sensitivity were observed. Further analyses suggested relationships between play selection bias and both turnover probability (which can be described in terms of punishment) and yards-gained variance (which can be described in terms of variable-magnitude reinforcement schedules). The present investigation provides a useful demonstration of association between face-valid, situation-specific effects in a domain of everyday interest, and a theoretically important term of a quantitative model of behavior. Such associations, we argue, are an essential focus in translational extensions of quantitative models.
Key words: choice, concurrent schedules, generalized matching equation, football, play selection

---

The present report concerns the generality of a relation described by the generalized matching equation (GME; Baum, 1974) as applied to situations outside the laboratory. The GME may be expressed as

$$\log\left(\frac{B_1}{B_2}\right) = a \log\left(\frac{r_1}{r_2}\right) + \log b \qquad (1)$$

in which B terms signify competing behaviors and the r terms signify reinforcement that is contingent on those behaviors. With logarithmic transformation the relationship between behavior and reinforcement ratios is a linear function in which a = slope (a measure of sensitivity to differential reinforcement) and $\log b$ = y-intercept (a measure of bias, or pervasive preference for one behavior beyond what the r terms predict). As an account of operant choice, the GME is neither conceptually complete nor universally applicable (e.g., Davison & Nevin, 1999), but it has advanced the analysis of behavior in a remarkable array of laboratory and nonlaboratory situations. Thus,

existing studies on the GME demonstrate broad generality, two aspects of which may be noted separately.

One form of generality is shown when a model accounts for substantial portions of the variance in behavior across many domains of investigation, or across many instances within a particular domain. For example, Baum's (1974, 1979) seminal papers on the GME showed that the GME described choice in many different laboratory investigations that used a variety of procedures and were designed to evaluate a variety of choice-influencing variables. In applied extensions, the GME has been found to account for a substantial amount of variance in the allocation across response options of behaviors as diverse as conversation (Borrero, et al., 2007; McDowell & Caron, in press-*a*), teen pregnancy (Bulow & Meller, 1998), classroom conduct (Billington & DiTommaso, 2003), and sport performance (Reed, Critchfield, & Martens, 2006; Vollmer & Bourret, 2000). The same consistent good fit also has been shown for numerous instances *within* selected domains of application (for example, over 300 college basketball teams; Alferink, Critchfield, Hitt, & Higgins, 2009). In these cases, the critical point is that the GME's defining variables (in Equation 1, B_1/B_2 and r_1/r_2 ratios) covary dependably as the model predicts. This type of generality can be termed *reliability of fit* (Stilling & Critchfield, in press).

A different type of generality is shown when a model such as the GME sheds light on situation-specific variations in behavior within a domain of application. This type of generality, which scales fitted parameter estimates to specific kinds of environmental events, may be termed *explanatory flexibility* (Stilling & Critchfield, in press), and is the focus of considerable basic research (for a detailed example of parameter scaling in laboratory experiments, see Davison & Nevin, 1999). For example, readers of concurrent-schedules studies in which changeover delay (COD) is manipulated know that strength of preference is an asymptotic function of COD duration (Mazur, 1991). The GME puts this effect into theoretical context by showing that it manifests as changes in the sensitivity parameter (Baum, 1974).

As Critchfield and Reed (2009) have noted, explanatory flexibility should be a primary focus in translational research because

> A model is of limited interest if its fitted parameters only show effects that are peculiar to some laboratory procedure. The working assumption, therefore, should be that these parameters apply in meaningful ways to the world outside of the laboratory…. Translational research can determine whether this is the case by evaluating the relationship between a model's fitted parameters and face-valid effects in an everyday domain. (Critchfield & Reed, p. 354)

For instance, in applications of the GME to basketball shot selection (in Equation 1, B terms were the number of two-point and three-point shots taken, and r terms were the number of those shots made), bias varied when rule changes affected the difficulty of making three-point shots (Romanowich, Bourret, & Vollmer, 2007), and sensitivity was higher for players on successful versus unsuccessful teams and for regular players versus substitutes (Alferink, et al., 2009). Unfortunately, the explanatory flexibility of the GME has been evaluated only rarely outside of the laboratory (for other examples, see McDowell & Caron, in press-*b*; Reed & Martens, 2008).

Of interest to the present discussion is the extent to which the GME's fitted parameters describe face-valid effects that make American-rules football (hereafter, simply *football*) interesting to its followers, and was prompted by a preliminary analysis reported by Reed et al.

(2006). No thorough explanation of football rules (Goodell, 2008) is possible here, but underpinning the offensive portion of the game is a team's imperative to move the ball toward a goal line to score points. Progress toward scoring may be accomplished through either passing plays (in which one player throws the ball to another) or rushing plays (in which one player runs with the ball). Across many opportunities within each game, someone, usually a coach, decides what kind of play to execute. In this sense the offensive side of football bears similarity to two-alternative operant choice. The parallel is accentuated by the fact that in choosing plays coaches routinely consider the success of previously selected plays, which is measured in terms of yards gained toward the goal (Edwards, 2002). Consistent with these observations, in applying the GME Reed et al. used the number of passing plays executed and rushing plays executed as the B terms, and the yards gained from those plays as the r terms, hence:

$$\log\left(\frac{Plays_{Pass}}{Plays_{Rush}}\right) = a\log\left(\frac{Yards_{Pass}}{Yards_{Rush}}\right) + \log b. \quad (2)$$

Reed et al. (2006) found that, as Equation 2 predicts, the plays-selected and yards-gained ratios were positively correlated in a variety of cases (i.e., good reliability of fit). With regard to explanatory flexibility, Reed et al. also compared play selection across three offensive situations. Each time a football team receives possession of the ball, it has four opportunities, or downs, to either score or advance the ball 10 yards, in which case another set of four downs is earned. Most often, if a new set of downs has not been earned by the completion of third down, then fourth down is reserved for a kicking play that transfers possession of the ball to the other team, leaving three downs on which passing and rushing plays tend to occur. According to football sources, rushing plays are especially attractive on first down, and passing plays are preferred for many third down situations (Allen, 2002; Westering, 2002). Consistent with this conventional wisdom, Reed et al. found a rushing bias on first-down plays and a passing bias on third-down plays, with an intermediate log b estimate for second down (although note that these effects were evaluated strictly through visual inspec-

tion of graphed parameter estimates, leaving unclear whether bias changes across down were statistically reliable). Taken at face value, these effects appear to show how a theoretically important term of the GME maps onto a situation-specific phenomenon of practical importance to football.

The present investigation sought to extend Reed et al.'s (2006) application of the GME to situation-specific play selection in football. The general strategy was to identify several types of game situations that football experts believe are relevant to play selection, and within each to identify several levels or categories across which play selection is thought to vary. The GME was used to evaluate play selection for each level of each of these situational variables so that, consistent with a consideration of explanatory flexibility, the resulting sensitivity and bias parameters could be compared across levels.

Descriptive Analysis

Consistent with the approach of Reed et al. (2006), the GME (Equation 2) was applied to play selection in a descriptive analysis of archival game statistics from the National Football League. It is axiomatic that the study of complex everyday behavior often precludes the use of experimental methods. Behavior analysts have, at times, been accused of preferring research questions that map conveniently onto preferred research designs (Baer, Wolf, & Risley, 1987), an approach that yields principles of debatable generality (e.g., Critchfield, Haley, Sabo, Colbert, & Macropoulis, 2003; Critchfield & Kollins, 2001; McDowell & Caron, 2010a). When experiments cannot be conducted, descriptive methods can shed light on behavior that, by virtue of the importance placed on it by laypersons, demands attention by any science claiming to offer a general-purpose explanation of behavior.

Not surprisingly, many translational extensions of the GME have employed descriptive designs in which neither the behavior of interest nor the putative reinforcers was under investigator control (e.g., Alferink et al., 2009; Borrero et al., 2007; McDowell & Caron, 2010a, b; Reed et al., 2006; Romanowich et al., 2007; Vollmer & Bourret, 2000). The assumption underlying such studies, of course, is that operant choice manifests similarly in everyday and laboratory environments. Be-

cause correlation does not support causal inferences, descriptive analyses cannot verify that this assumption is true (see Alferink et al., 2009; Critchfield & Reed, 2009; Reed et al., 2006; Vollmer & Bourret, 2000), but they *can* provide disconfirming evidence. In the present case, for instance, the GME could fail to adequately describe football play selection. Such an outcome would be unsurprising given that, in the everyday world, contingencies do not exactly parallel laboratory reinforcement schedules, and many factors operate in addition to those specified by Equation 2 (Reed et al., 2006).

This highlights a further difference between laboratory investigations and field extensions. Laboratory procedures minimize extraneous variance to give effects of interest every possible opportunity to emerge (Sidman, 1960). Uncontrolled natural environments confer no such advantage. As Reed et al. (2006) noted with respect to football, ''Few everyday environments are as complex and multiply determined as those in which elite sport competition occurs.... Many variables are believed to influence sport performance.... Any lawful principle or functional relation found to cut through all of these variables to reliably predict sport performance would be noteworthy indeed'' (pp. 281–282). In this limited sense, descriptive, translational investigations speak more directly to the possible robustness of functional relations than do highly controlled laboratory experiments (e.g., see McDowell & Caron, in press-*a*).

Overall, while descriptive methods cannot show unambiguously that operant choice manifests similarly in dissimilar environments, they can provide intriguing circumstantial evidence to this effect. From a scientific perspective, circumstantial evidence is better than no empirical evidence. Historically in behavior analysis, a common approach to examining complex everyday behavior has been the narrative essay that Skinner (e.g., 1953, 1957, 1991) popularized. Such treatises can be colorful and conceptually expansive, but they are not empirical and therefore easily undermined, as critics may dispute even the basic premises and observations that underpin narrative accounts (e.g., see Chomsky's, 1959, review of Skinner's *Verbal Behavior*). By contrast, descriptive analyses reveal patterns in everyday behavior that any theoretical inter-

pretation (whether inspired by basic behavioral research or not) must be able to explain (Alferink et al., 2009). In this way they serve as a valuable tool in the effort to analyze the many complex everyday situations that have received limited empirical attention in behavior analysis (e.g., Mace, Lalli, Shea, & Nevin, 1992).

Evaluating the Situational Modulation of Fitted Parameters

The strategy of the present study was to fit the GME to naturally occurring football data to evaluate whether fitted parameters change systematically across game situations. A quantitative model's fitted parameters are informative only if the model accounts for substantial variance in behavior (Lunneborg, 1994), so in the present investigation sensitivity and bias estimates could be evaluated only if the GME accounted for a nontrivial percentage of variance (R^2) in play selection. For present purposes we define ''nontrivial'' in the context of previous field applications in which the GME typically has accounted for ≥40% of the variance in a behavior of interest (Billington & DiTommaso, 2003; Borerro et al, 2007; Bulow & Meller, 1998; Reed et al, 2006). If this is the case across many football game situations then reliability of fit will have been demonstrated and comparisons of parameter estimates facilitated.[1]

Assuming that the GME accounts for a nontrivial amount of variance in play selection in all of the game situations considered here, play selection tendencies still might not be associated with systematic changes in bias or sensitivity (e.g., perhaps the effects described by Reed et al., 2006, were visually suggestive but not statistically reliable). Such a finding could arise if situation-specific play-selection preferences simply reflect points along a single

matching function (e.g., see Critchfield & Reed, 2009, Figure 5 and associated text). Relative frequency of passing and rushing plays might vary across game situations only as a function of relative success in earning yards. If the GME's fitted parameters do not vary systematically across game situations, then the GME, as applied to football play selection, could have good reliability of fit (consistently good R^2) but poor explanatory flexibility.

Alternatively, the GME's fitted parameters might detect situation-specific variations in play selection that football observers regard as interesting. We expected that any such effects would manifest in terms of bias (log b) rather than sensitivity (a). Sensitivity can be said to reflect ''knowledge'' (i.e., discrimination) of contingencies (e.g., Baum, 1974; Davison & Nevin, 1999), which increases with both accumulated experience in adjusting to contingencies (e.g., Todorov, Olivera Castro, Hanna, de Sa, & Barreto, 1983) and the quality of discriminative stimuli signaling behavior–consequence relations (e.g., Davison & Nevin, 1999). It may be relevant, therefore, that the coaches who select most NFL plays have extensive experience in football and, thus, extensive direct exposure to the game's contingencies. They also benefit from the supplemental stimulus control exerted by detailed statistics and other information (e.g., video records of past performances) about what kinds of plays tend to succeed in what situations. Because of these factors, a ceiling effect may exist in which sensitivity, while not optimal, may be as high as it can be under the naturalistic conditions of NFL play selection. Thus, in the present investigation sensitivity was not expected to vary systematically as a function of game situations.

Bias (see Baum, 1974) is thought to result from systematic changes in aspects of the behavior–consequence relations other than those subsumed by the r terms of Equation 1. In the present study, r terms reflected yardage gained from passing and rushing. Although football experts sometimes allude to situation-specific factors other than mean yardage gains that may influence play selection (e.g., Westering, 2002), it is not always clear how these factors map onto the reinforcement-based conceptual framework of the matching relation. For this reason, the present investigation focused primarily on *identifying* bias effects in

---

[1] A related issue is whether the GME accounts for different amounts of variance in play selection across game situations. Such an outcome would raise interesting questions about whether matching is differentially relevant to different game situations. Unfortunately, it appears that no objective means exists to determine whether R^2 values differ significantly when the same model is fitted to different data sets (instead, theorists have focused on comparing the fits to the same data of models with different numbers of fitted parameters; see Lunneborg, 1994; Motulsky & Christopoulis, 2006). For this reason we report R^2 values but offer no prediction or comment about the possibility of systematic R^2 effects.

play selection. We will return to the problem of a conceptual analysis of these effects in the Discussion.

METHOD

Data Transcription

Data were retrieved, and organized into spreadsheets for purposes of analysis, between October 13, 2007 and June 20, 2008 from http://www.espn.com. Recorded for each play of 192 targeted games (see below) was whether a pass or a rush was executed; the yardage gained; and the situational variables described below. Prior to data collection six transcribers read printed instructions (available on request) and collectively recorded and discussed a small sample of plays before individually transcribing a sample game. An investigator then checked for errors by comparing transcriber records to the data source, and provided feedback and answered questions. This process required approximately 1 hr. Thereafter transcribers created the data set, with each individually transcribing a different subset of the targeted games. For five randomly-selected games per transcriber, the experimenters compared transcriber records to the data source in order to check for transcriber drift. None was detected. Agreement (defined as exact match between the source and the transcribed data) occurred on 96.8% to 99.8% of several hundred data entries (five variables for at least 60 plays) per game. Errors, when they occurred, consisted almost exclusively of manual mistakes (e.g., typing ''332'' instead of ''32'' or accidentally replacing a number with a letter located beneath it on a QUERTY keyboard) rather than transcribing a value from the wrong location in the data source. When such errors were detected in records other than those on which accuracy was systematically evaluated, they were corrected by consulting the data source.

Limitations of the Data Source

The archival statistics on which the analyses were based have two limitations that could affect the precision of the present analyses. First, football statistics do not specify exactly who selects each play. On each team, a single individual (the offensive coordinator) is nominally charged with play selection (McCorduck,

1998); in this sense, play selection is individual behavior. Yet in at least some circumstances for some teams, multiple individuals may influence play selection, although this is not reflected in public data sources. For purposes of the present investigation, each team's offensive staff was considered as a single, collective ''organism'' (i.e., a group whose behavior, by virtue of exposure to shared contingencies, presumably was under common control). This approach is consistent with the findings of investigations in which several individuals working under a shared contingency produced collective behavior that was patterned like that exhibited by laboratory subjects working individually under similar contingencies (e.g., Buskist & DeGrandpre, 1995; Critchfield, Haley, Sabo, Colbert, & Macropoulis, 2003; Graft, Lea, & Whitworth, 1977; Grott & Neuringer, 1974; Mace et al., 1992; Sokolowski, Tonneau, & Friexi I Baque, 1999; Wolff, Burnstein, & Cannon, 1964). Nevertheless, the probable intermingling of play-selection behavior of multiple individuals was expected to adversely affect the percentage of variance for which the GME accounted.

Second, the data source categorized plays as passing or rushing based on what actually happened, not necessarily what was intended by the team's play selector(s). For instance, imagine that a pass play is planned but after the play begins the quarterback attempts to run instead. Such a play is identified in the record as a rushing play, even though a choice initially was made to select a passing play. Such eventualities probably impose unexplained variance on a matching analysis beyond what typically is encountered in the laboratory and in field settings where no analogous coding ambiguities arise (e.g., analysis of basketball shot selection; Vollmer & Bourret, 2000).

Levels of Analysis

Season-aggregate data. Archival sources traditionally include NFL offensive statistics pooled across an entire season. Recorded for each team in the 2006–2007 season were the total number of rushing and passing plays that were executed and the total number of yards gained from each type of play in each game of a 16-game season.

Play-by-play data. For each team, six games from the 2006–2007 season were randomly chosen from which to extract play-by-play data.

For purposes of this analysis a game was defined as the offensive performance of one team excluding any overtime (because special rules apply to offense during overtime and because overtime data were not always available from our source). For each offensive opportunity, the type of play (passing or rushing) and the number of yards gained were recorded (other types of plays are possible in which the ball is kicked but these were not considered relevant to the present investigation). Overall, 192 games were evaluated (6 games for each of 32 teams) in each of which a team's offense conducted approximately 60 rushing or passing plays, for a corpus of more than 12,000 total plays. This sample was expected to support analyses in which play selection was examined as a function of several game situations, an assumption that was largely but not universally borne out. Occasionally, a team had limited offensive opportunities in one of the categories. If fewer than 15 plays were available for analysis, the team was dropped from all categories of the relevant situational variable; specific instances are indicated below.

Each play was categorized according to the following types of game situations: down, yards needed to earn a new set of downs, time remaining, score, and field position. For each variable three categories were developed to reflect conventional wisdom about football as represented in professional publications on football, primarily authored by successful coaches and others with long-term involvement with the game at high levels of competition (e.g., Allen, 2002; Bryant, 1999; Kehres, 2006; Levy, 1999; McCorduck, 1998; Teaff, 1999; Westering, 2002). Hereafter, for economy of expression, these individuals will be referred to as football ''experts.''

Situational Variable Categories

Down. This variable, defined above, was included in the present study to determine whether the results of Reed et al. (2006) could be replicated for a different season of play. The levels were first down, second down, and third down.

Yards needed. The distance that a team must advance the ball in order to earn a new set of downs varies from play to play. The nominal range is 1 to 10 yards, but after losing ground through penalties or unsuccessful plays a team may need more than 10 yards to earn a new set of downs. Football experts suggest that play selection usually is rushing-oriented when ≤4 yards are needed (e.g., Allen, 2002; McCorduck, 1998). This may be the case in part because the average gain from an NFL rushing play is about 4 yards, with relatively little variance and few plays that yield no yards or a loss of yards (Rockerbie, 2008). By contrast, NFL passing plays yield about 7 yards on average, but the variance is high, meaning that some pass plays yield considerably bigger gains (Rockerbie, 2008). Perhaps for this reason, plays on which many yards are needed to earn new downs are regarded as passing-oriented (e.g., Allen, 2002). For purposes of the present analysis of yards needed the levels were 1–4, 5–10, and >10.

Time remaining to play. NFL games are divided into four 15-min quarters. Play proceeds without major interruption between the first through second quarters (collectively called the first half) and during the third through fourth quarters (collectively called the second half). Between the halves is a suspension of play (called halftime) lasting at least 12 min (sometimes longer to accommodate factors such as the broadcasting of television commercials). Football experts regard the last 2 min of each half as unusual given that opportunity to score is waning (e.g., Fulmer, 2002; Levy, 1999; Tranquil, 2006; Westering, 2002). Passing plays are said to be preferred during this interval for two reasons (McCorduck, 1998). First, pass plays have the potential to gain many yards quickly. Second, when a pass is not caught (incomplete) the game clock stops briefly, allowing the offensive team to regroup for the next play without expending game time. By contrast, at the end of rushing plays, the game clock continues to operate. The present analysis of time remaining thus focused on the final 2 min of each half. Because relatively few plays can occur during these brief intervals, the final 2 min of the two halves were combined to increase the relevant sample. For consistency, data from the remainder of the 2nd and 4th quarters were pooled prior to analysis, as were data from the entire 1st and 3rd quarters. Thus, the present analysis focused on time remaining in a half, and the levels were >15:00 (1st and 3rd quarters combined), 2:01–15:00 (2nd and 4th quarters combined, minus the final 2 min),

and ≤2:00 (final 2 min of the 2nd and 4th quarters combined). Two teams (Arizona and Atlanta) were excluded from this analysis because of insufficient data (as defined in the preceding section), leaving $N = 30$ teams.

Score. According to conventional football wisdom teams that are winning tend to select plays that will minimize the chance of losing the ball through turnover and consume as much game time as expediently as possible (e.g., Kehres, 2006; Levy, 1999; McCorduck, 1998). Both factors suggest rushing plays because they tend to consume more time from the game clock than passing plays and, as Reed et al. (2006) reported, interceptions (passes caught by the other team) are more common than rushing-related lost fumbles (when an opponent picks up a ball that was dropped). Teams that are losing are under pressure to use game time efficiently and to use each play to gain as many yards as possible toward scoring. Both factors suggest passing plays because they may allow the game clock to be stopped briefly, and the average gain is larger for passing plays than for rushing plays (Rockerbie, 2008). For the present analysis of score, the levels were winning, tied, and losing. One team (Tampa Bay) was excluded from this analysis because of insufficient data, leaving $N = 31$ teams.

Field position. While on offense, a team must attempt to move from wherever it receives possession of the ball to the goal line. Field position specifies the location on the field from which a given play is initiated. A football field is 100 yards long, ranging from a target team's own goal line (which the opponent must cross to score) to the opponent's goal line (which the target team must cross to score). For present purposes field position will be described in terms of yards separating a team from the opponent's goal line, i.e., a scale of 1 to 99 (the ball cannot be positioned on a goal line, and in football records field position is rounded to the nearest yard).

Football experts do not agree about the number of functional play selection zones that exist on the field or the strategies that are preferred for these zones. The present analysis focused on two zones that are discussed with some consistency across experts, who generally agree that plays executed at the extreme ends of the field should minimize turnovers and avoid zero- or negative-yardage outcomes (Bryant, 1999; Westering, 2002; Tressel &

Bollman, 2000; Tressel, 2000). When a team is near its own goal line, turnovers create scoring opportunities for the other team, while yardage gains increase the space available behind the line of scrimmage (the location from which a play begins) for the offensive team to execute plays. Additionally, the closer a team is to its own goal line, the greater the risk of being tackled behind it, creating a safety that scores two points for the other team. When a team is near the opponent's goal line, gaining yards means getting closer to scoring, and turnovers forfeit scoring opportunities. Both cases suggest advantages of selecting rushing plays.

For present purposes, ''near the opponent's goal line'' was defined as 1–8 yards from the opponent's goal line, and ''near one's own goal line'' was defined as 83–99 yards from the opponent's goal. The rest of the field (9–82 yards from the goal) was treated as a single zone, even though many experts recommend play-selection strategies for specific portions of this zone (e.g., Bryant, 1999). We conducted numerous exploratory analyses that divided the 9–82 zone into subzones but found no consistent differences in play selection among them. Note that, because the 1–8 category encompasses only a small portion of the field, relatively few plays per game occur there, and consequently six teams (Baltimore, Green Bay, Jacksonville, New York Giants, Philadelphia, and Washington) were excluded from this analysis because of insufficient data in this category, leaving $N = 26$ teams.

RESULTS

Season-Aggregate Data

Figure 1 (top) summarizes NFL play selection during the 2006–2007 regular season, Consistent with an approach employed by Reed et al. (2006), the GME was fitted to a function involving one data point for each NFL team ($N = 32$), with each data point representing the season-aggregate statistics of one team. When Reed et al. applied the GME to data from the 2003–2004 season, the line of best fit, $y = .72x − .13$, accounted for 75.7% of the variance in play selection. To expand this historical frame of reference, we repeated this analysis for other years in the decade of 1999–2008, and found that 2006–2007 outcomes fell within the ranges for sensitivity (.50 to .73), bias (−.14 to −.06),
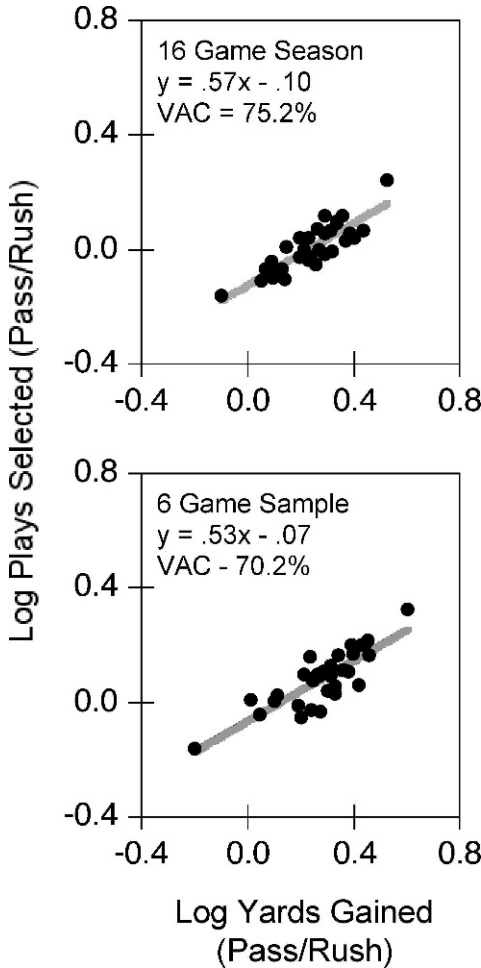
Fig. 1. Relationship between relative play frequency and relative yards gained through those plays, with both expressed as passing/rushing, for the 2006–2007 National Football League season. Each data point shows the season-aggregate data from one team. Shown in gray are lines of best fit as determined by applying the generalized matching equation (Equation 2) through least-squares linear regression. Top: Based on data from all 16 regular-season games. Bottom: Based on six randomly-selected games per team.

and variance accounted for (54.9% to 80.5%). As in other recent years, three features were evident in the 2006–2007 data: undermatching, a bias for selecting rushing plays (negative log b estimate), and a majority of play-selection variance accounted for by Equation 2. Overall, the 2006–2007 season may be considered a representative sample of contemporary NFL competition.

Because the analyses involved fitting a single matching function to data from multiple teams (Figure 1 and below), it is reasonable to ask how well individual cases are represented by such an aggregate function. Figure 1 (top) provides a partial answer. To the extent that a single function economically subsumes all 32 NFL teams, these teams may be said to exhibit a common form of global play-selection matching (in which case aggregating them does not intermingle incompatible functions). This assumption is consistent with the finding of Reed et al. (2006) that matching functions of individual teams of the 2003–2004 season usually were similar to a function that aggregated all of the teams. Reed et al. created individual-team functions by treating each of a team's regular-season games as a separate observation. We replicated this approach for the 2006–2007 season; Figure 2 summarizes the results. Central tendencies for sensitivity and bias were similar to the estimates based on the season-aggregate function (Figure 1, top), although Equation 2 tended to account for less variance in individual-team functions (median = 56%; not shown in Figure 2) than in the season-aggregate function. This outcome we attribute in part to the relatively small number of plays available for analysis in each game. Overall, Figure 2 is consistent with the view that interteam similarities in matching allow data to be aggregated from different teams (for a sophisticated empirical and conceptual evaluation of the underlying issues, see McDowell and Caron, 2010a, who concluded that aggregation of the sort employed here is, in at least some cases, defensible).

Figure 1 (bottom) summarizes NFL play selection during the six games per team that were randomly selected for situational play-selection analysis. As in the top panel, the GME was fitted to a function involving one data point for each NFL team ($N = 32$), with each data point representing the season-aggregate statistics of one team. Slope and bias estimates were similar to those derived from the full 16-game season, and Equation 2 accounted for a similar amount of play-selection variance. By these broad metrics, the six-game sample was representative of the full season from which it was drawn.

Game Situations

For each of the situational variables, each team's data were obtained by pooling plays from the six targeted games. Consistent with
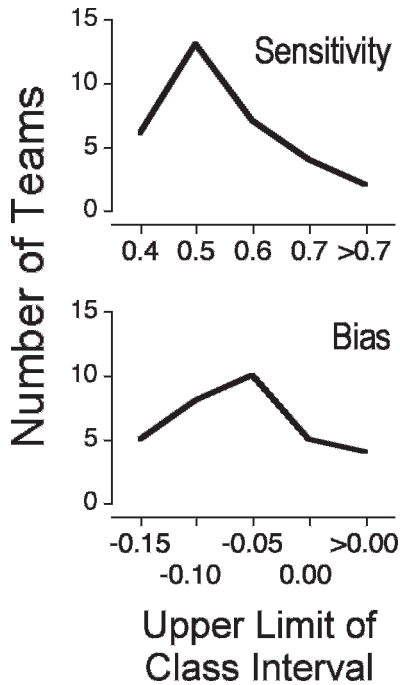
Fig. 2. Summary of fitted parameter estimates obtained from fitting the generalized matching equation (Equation 2) to the data from each of 32 National Football League teams, with each of a team's sixteen 2006–2007 regular-season games as an observation.

an approach employed by Reed et al. (2006), for each category of each situational variable (see below), the GME was fitted to a function involving one data point for each NFL team, excluding those (noted above) for which insufficient data were available.[2] For instance,

[2] For situational analyses, functions were fitted to the data of multiple teams, rather than individual teams, because of a limited supply of plays available to analyze at the team level. Taking Figure 2 as a frame of reference, we might have attempted to fit Equation 2 to data for each team, with different opponents counting as observations. This would yield a pool of roughly 60 plays per observation, which in turn would be divided across three situational categories. Plays are not distributed evenly across the categories for any of our five situational variables, so we expected to be unable to complete an analysis for most teams for most variables. To illustrate, in the present corpus based on six games, teams attempted an average of fewer than four eligible plays per game (excluding kicking plays) from within 1 to 8 yards of the opponent's goal, far too few for the ratio-based analysis of Equation 2. To address this problem, data might be combined from different seasons, although professional football rosters and coaching staffs are notoriously fluid from season to season, in which case play selection and

for the variable *down*, there were three GME analyses, one each for first, second, and third down plays. For each level of each variable, the ratio of passing and rushing plays was considered as a function of the ratio of yards gained from passing and rushing as per Equation 2. In each case, least squares linear regression was used to determine the line of best fit and to estimate the fitted parameters.

Goodness of fit. Figure 3 summarizes the success of the GME in describing play selection across levels of several types of game situations. The figure shows the percentage of variance for which the GME accounted in each analysis; the leftmost portion of the figure provides a frame of reference by showing the same outcome for all plays (both 16-game and 6-game totals). The remaining columns show outcomes for levels of the situational variables. The GME accounted for a majority of variance in most game situations (and >40% in all cases), but typically less than for all plays combined (Figure 1). The latter outcome may reflect, in part, the relatively small sample of plays involved in these subordinate analyses. Note that, across categories, the number of plays available for analysis (pooled for all teams) was positively correlated with the amount of variance for which the GME accounted ($r = +.50$). Also shown in Figure 3 are results of an analysis by down for the 2003–2004 season by Reed et al. (2006; open data points). All outcomes of the present situational analyses fell within the range of that previous analysis.

Statistical evaluation of situational variance in sensitivity and bias. Comparisons of sensitivity (slope $= a$) or bias (intercept $= \log b$) across levels of each situational variable employed an inferential statistical test based on analysis of covariance (ANCOVA; Motulsky & Christopoulis, 2006; Zar, 1999; for computational details and an example of application to behavioral research, see Magoon & Critchfield, 2008). For each type of game situation, the test began with an omnibus ANCOVA (alpha = .05)

play success for different seasons would represent the behavior of different personnel (the same drawback of the present corpus). Overall, we chose the present analytical strategy because alternatives appeared to be both more effortful and less likely to shed light on the research question, but we acknowledge that our approach precludes the examination of potentially interesting between-team differences in situational play-selection.
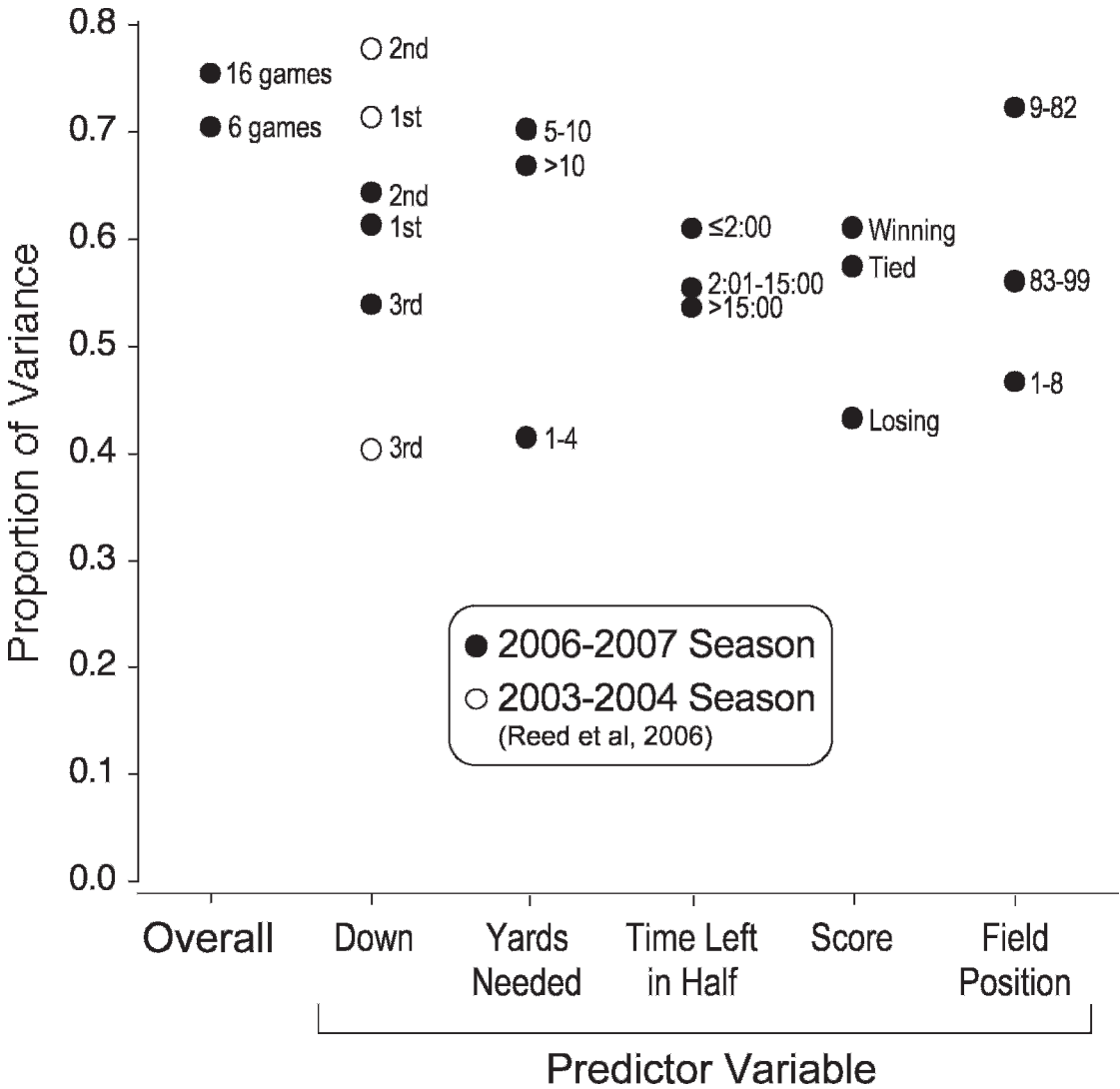
Fig. 3.   Percentage of variance accounted for by fitting the generalized matching equation (Equation 2) to play selection data for each level of five types of game situations. See text for details.

comparing sensitivity estimates across all categories. With this test, a statistically significant sensitivity (slope) effect has two implications. First, the same test can be used in paired comparisons of sensitivity among the levels of the same predictor variable (in the present case, with the Bonferroni adjustment of alpha = .05 divided by the number of comparisons as a control for Type 1 error risk). Second, meaningful tests of bias (intercept) are precluded because slope and intercept are confounded in linear regression (Zar, 1999; for approaches that more readily accommodate

intercept effects, see Milliken & Johnson, 2002). No omnibus sensitivity effects were found in the present investigation, which allowed the ANCOVA analysis to be used to evaluate bias effects.

As with sensitivity tests, for bias estimates a significant ANCOVA (alpha = .05) led to paired comparisons among the levels of a given situational variable. Each paired comparison began with a sensitivity test comparing two levels of a given variable. If a significant difference in sensitivity estimates was identified (alpha = .05), no bias comparison was
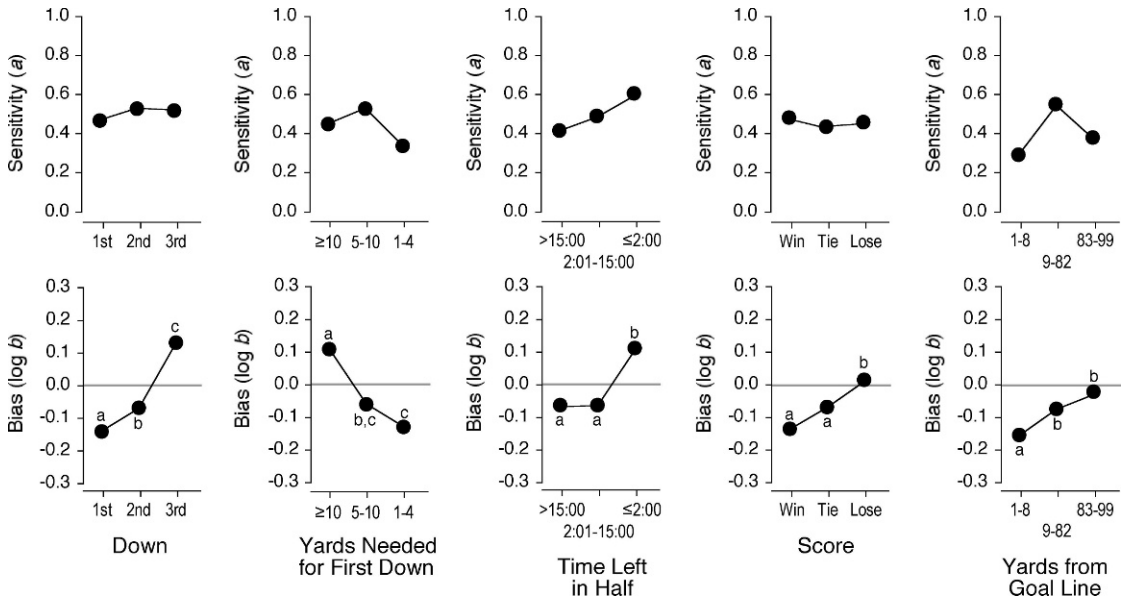
Fig. 4. Play-selection sensitivity and bias estimates for each level of five types of game situations. Within each panel, pairs of data points that do not share a letter code are significantly different according to paired comparisons; absence of letter codes indicates no statistically significant effects. See text, Table 1, and the Appendix for details of the statistical analyses.

conducted and no bias effect was assumed because of the slope–intercept confound noted above. Note that the decision criterion for these paired sensitivity comparisons (.05) was not adjusted for Type 1 error risk because this created a conservative criterion for determining pairwise bias effects (which could be evaluated only if associated slope effects were not significant). Because omnibus ANCOVAs yielded no significant results for sensitivity, in paired comparisons a low *p* value for sensitivity was not taken as evidence of an effect. For each paired comparison, if a significant sensitivity effect was absent, the associated bias comparison was conducted with alpha adjusted as described above to reduce Type 1 error risk.

Figure 4 shows the sensitivity and bias estimates for each level of the five types of game situations. Table 1 summarizes the outcomes of omnibus ANCOVA analyses for each of these variables. For each type of game situation, the omnibus ANCOVA revealed no significant slope effect (top row of panels in Figure 4) and a significant bias effect. For this reason the present discussion will focus on bias effects as revealed in paired comparisons among levels of each type of game situations (for statistical details of these comparisons, see

the Appendix). Results are summarized in the bottom row of Figure 4 through letter codes. For each type of game situation, data points that do not share a common letter code are significantly different.

Figure 4 shows that (1) play selection was biased toward rushing on first down and biased toward passing on third down, with an intermediate log *b* estimate on second down. This replicates the pattern described by Reed et al. (2006) based on visual inspection of graphed data, and improves upon Reed et al. by showing that all differences among log *b* estimates for the three downs were statistically reliable. Other findings include that play selection was (2) biased toward passing when 10 or more yards were needed to obtain a new set of downs, and biased toward rushing when ≤4 yards were needed; (3) biased toward passing when 2 min or less remained to play in a half, and otherwise biased toward rushing; (4) essentially unbiased when a team was losing, and otherwise biased toward rushing; and (5) strongly biased toward rushing when a team had possession within 8 yards of the opponent's goal line, with a less pronounced rushing bias at other field locations.

Table 1

Results of omnibus ANCOVA tests of sensitivity and bias estimates for each of five types of game situations. Significant ANCOVAs were followed by the paired comparisons that are summarized in the Appendix.

Situational Variable	Test	F	df	p
Down	Sensitivity	0.17	2,90	.845
	Bias	36.20	2,92	<.0001*
Yards Needed to Earn New Downs	Sensitivity	1.09	2,90	.340
	Bias	29.97	2,92	<.0001*
Time Remaining in Half	Sensitivity	0.56	2,84	0.57
	Bias	27.41	2,86	<.0001*
Score	Sensitivity	0.07	2,87	0.93
	Bias	16.69	2,89	<.0001*
Field Position	Sensitivity	1.48	2,72	0.23
	Bias	9.44	2,74	.0002*

* Statistically significant (alpha = .05)

DISCUSSION

Generality of the Matching Relation

Although football experts often describe situation-specific patterns of play selection, it was not clear at the outset of this investigation how (or whether) these patterns might translate into the outcomes that are examined in a matching analysis. One possibility is that the matching relation, described previously for season-aggregate data by Reed et al. (2006), would not hold for all specific game situations. Consistent with the findings in several other applied domains (e.g., Billington & DiTommaso, 2003; Bulow & Meller, 1998), however, the GME accounted for about 40% to 70% of the variance in play selection across the various game-situation categories. The percentage of play-selection variance for which the GME accounted did not appear to vary systematically across the levels of these situational variables, suggesting reliability of fit.

Another possible outcome was that play selection might vary across game situations, but strictly in accordance with a single covariant relationship between relative ratios of yards gained and plays selected. That is, perhaps what distinguishes the various game situations in which play selection occurs is the relative yards-gained ratio in Equation 2, with play selection shifting upwards or downwards along a single matching function. Contrary to this view, however, significant changes in the GME's bias parameter were associated with all five types of game situations. This suggests that some NFL game situations are best described with a unique matching function rather than as part of a single general function. Such

mapping of theoretical parameter to face-valid game situations establishes a degree of explanatory flexibility for the GME as applied to football play selection.

The effects just mentioned shed light on conventional football wisdom by providing an alternative to conventional ways of characterizing play selection. Alamar (2006) illustrated the traditional perspective by speaking of a general "passing premium puzzle" in the NFL, in the form of "a balance between the number of passing and running plays, even though there is a greater expected return in passing plays" (unpaginated abstract). From this perspective, NFL teams pass too little and rush too much. A matching analysis precisely defines "rushing too much" in terms of the bias parameter of Equation 2 (we take up the question of why this particular outcome may arise in a later section). Although football experts often speak of "rushing situations" or "passing situations," in the present findings these situations are distinguished behaviorally, not in terms of raw preference, but rather in terms of deviations from the level of preference that is predicted based on relative "reinforcement" (i.e., bias). Illustrating this distinction are five cases, among those shown in Figure 4, in which passing occurred on more than 50% of total plays although, in GME terms, play selection actually was biased toward rushing. These cases were: second down, 5-1 yards needed for a first down; 2:01–15:00 remaining in the half, score tied, and ball positioned 9–82 yards from the goal.

No situation-specific effects in play-selection sensitivity were identified. We suggested previously that this may reflect a ceiling effect in

which sensitivity of NFL play selection has reached a practical maximum through extensive experience and access to detailed discriminative stimuli (e.g., statistics and game films). This assumption does not preclude that sensitivity effects might be associated with NFL game situations other than those considered here; it merely predicts that such effects should be uncommon. Our perspective also anticipates situation-specific sensitivity effects for less experienced football play selectors (e.g., novice coaches); this is an interesting direction for future research.

Complexity of Play Selection

Although the present analyses were more detailed than those of the Reed et al. (2006) study that inspired them, a football expert might be dissatisfied with our approach of evaluating selected types of game situations separately, because play selection is assumed to reflect the joint influence of many types of game situations (McCorduck, 1998). For example, preference for passing versus rushing plays is thought to be especially volatile on third down, depending on whether the number of yards needed to earn new downs is large or small, respectively (Allen, 2002; Reed et al., 2006). To the preceding a statistician might add that our analyses all drew upon the same sample of NFL plays, so to the extent that various game-situation variables are intercorrelated their effects on play selection would not be independent. A brief example may illustrate the problem. In our analysis of score we examined play selection when a game is tied. All games begin with a score of 0-0, however, so plays chosen early in a game would be represented disproportionately in this category. Because we found that rushing bias was prevalent early in NFL games, our "tied" category might have underestimated actual preference for passing.

A logical resolution of this problem lies in multivariate methods (Lunneborg, 1994) in which the type of play selected is the binary predicted variable (making logistic regression suitable) and two or more game situations are the predictor variables. Such methods could evaluate the relative strength of association between various game situations and play selection—but they would not necessarily address the matching relationship and its conceptually-important fitted parameters.

Within the matching literature, multivariate issues have been addressed by proposing concatenated models that subsume several choice-influencing variables (e.g., Baum, 1974; Hamblin & Miller, 1977; Herrnstein, 1961). Concatenated models can be imagined that simultaneously consider many football game-situations in an analysis of play-selection matching, although two challenges confront the development of such models. First, it is not always clear how the various factors that distinguish football game situations translate into the behavioral concepts that matching equations are intended to represent. Second, due to limited theory development and empirical testing, much remains to be resolved about how to construct concatenated matching models (e.g., Critchfield, Paletz, MacAleese, & Newland, 2003; Davison, 1988; Davison & Hogsden, 1984; Davison & Nevin, 1999; Grace, 1999; Shahan, Podelsnik, & Jiminez-Gomez, 2006).

Plausibility of an Operant-Choice Interpretation

In developing and reporting the present study we took at face value Reed et al.'s (2006) operant-choice interpretation of football play selection, but doing so required a number of conceptual leaps. The most general issue is that a descriptive analysis cannot support strong cause–effect inferences like those that derive from experiments. An operant interpretation implies that play selection tracks yards-gained reinforcement (or "expectations" thereof that are derived from statistics and game films), but the matching relationship could be spurious if the converse is true. Imagine that, for each NFL team, passing and rushing plays produce different average yardage gains, although the numbers of pass and rush plays selected are controlled by something *other* than these gains (e.g., coach superstitions, instructions from a micromanaging team owner, etc.). Because total yards gained is the product of the number of plays selected and the average yards gained from those plays, relative ratios of behavior and reinforcement would covary as Equations 1 and 2 stipulate, but without the influence of behavior–consequence relations that are the GME's conceptual scaffolding. A spurious-correlations account cannot be ruled out in descriptive studies, but is directly testable through simulation methods (e.g., Rubenstein

& Kroese, 2008). The question of interest is whether it is possible to ''manipulate'' the play selection of hypothetical play selectors—with yardage gains modeled after those of real football teams but no reinforcement effects assumed—to create matching outcomes like those observed for real teams. If so, then an operant interpretation is undermined.

Another useful approach is to compare the patterns revealed in a descriptive analysis to benchmark effects in operant-choice experiments. Here the situational modulation of fitted parameter estimates is of special interest. To illustrate, consider that, in matching analyses of basketball shot selection, preference has been defined in terms of the relative frequency of two-point versus three-point shot attempts. In laboratory matching experiments, unequal reinforcement magnitudes create bias (e.g., Landon, Davison, & Elliffe, 2003). By analogy, a three-point shooting bias is expected in basketball, and has in fact been widely observed (e.g., Alferink et al., 2009; Hitt, Alferink, Critchfield, & Wagman, 2007; Romanowich et al., 2007; Vollmer & Bourret, 2000). Such empirical parallels lend a degree of confidence to an operant interpretation.

Evaluating the plausibility of an operant interpretation of play selection thus requires close attention to specifics of the operant choice literature, which does not always provide clear guidance. For example, choice is studied most often in concurrent schedules of constant-magnitude, variable-interval reinforcement (Davison & McCarthy, 1988; Mazur, 1991), whereas the schedules governing the yardage ''reinforcers'' that were considered here and by Reed et al. (2006) probably are ratio based and involve variations in relative magnitude (based on mean yards per play for passing versus rushing). If, as is widely believed, concurrent ratio schedules ''yield nearly exclusive responding to the schedule that yields richer reinforcement'' (Vollmer & Bourret, 2000, p. 144), then the orderly matching functions of the present study may be at odds with laboratory principles. Yet the extent to which the matching relation emerges in concurrent schedules with ratio-like properties remains a matter of some debate (Green, Rachlin, & Hanson, 1983; Herrnstein & Heyman, 1979; Herrnstein & Loveland, 1975; LaBounty & Reynolds, 1973; MacDonall, 1988; Rider, 1979; Savastano & Fantino, 1994;

Shimp, 1966; Shurtleff & Silberberg, 1990). For discussions of how some of the relevant issues apply to sport behavior, we refer the reader to Reed et al. (2006) and Vollmer and Bourret (2000).

Questions also may be raised about whether behavior allocation matches the relative ratio of reinforcement magnitudes. A small body of reports indicates that for nonhumans it does (e.g., Davison & Baum, 2003; Elliffe, Davison, & Landon, 2008; Grace, 1995, 1999; Kyonka & Grace, 2008; Landon et al., 2003; Lau & Glimcher, 2005). For human subjects, the few available studies show limited matching to reinforcer magnitude (Dube & McIlvane, 2002; Sanders, 1968; Schmitt, 1974; Wurster & Griffiths, 1979). Whether this reflects idiosyncrasies of the relevant experiments (e.g., Baron & Derenne, 2002) or of human beings per se remains to be determined. If humans do not match reliably to reinforcer magnitudes then, once again, the present matching functions may be at odds with laboratory principles. For now, the basic operant literature provides insufficient guidance for a reinforcement-based interpretation of the present findings.

Determinants of Bias: The Role of Risk

Football experts cite almost as many situation-specific reasons for play selection as they do situations in which plays may be selected (e.g., American Football Coaches Association, 1995), making a search for general principles in the present bias effects difficult. Across many types of situations, however, two factors are mentioned with some regularity. The first factor is turnover risk. In the NFL, turnovers occur more frequently on passing plays than on rushing plays (on a per-play basis, interceptions are more common than rushing fumbles, and fumbles also can result from passing plays). According to football experts, play selection favors rushing in cases where turnovers are regarded as especially costly, such as at selected field positions (near one's own or the opponent's goal line), when relatively few yards are needed to earn a new set of downs, and when a team is winning (e.g., Allen, 2002; Bryant, 1999; Levy, 1999; McCorduck, 1998).

The second factor is variance in yards gained through passing versus rushing plays. In the NFL this variance tends to be higher for

passing plays (Rockerbie, 2008) due to exceptionally large yardage gains from some passing plays and to the fact that roughly 40% of NFL passes are not completed (resulting in a gain of zero yards). According to football experts, play selection especially favors rushing in cases where uncertainty in gains should be avoided, such as at selected field positions (near the other team's goal line), when relatively few yards are needed to earn a new set of downs, and when a team is winning (e.g., Allen, 2002; Bryant, 1999; Levy, 1999; McCorduck, 1998). By contrast, turnover risk and yardage variance combine to define the situations in which passing is regarded as especially useful, namely when success can only be achieved through the big yardage gains attainable through passing, or when the adverse effects of turnovers are most easily tolerated (McCorduck, 1998).

Reed et al. (2006) conceptualized turnovers as punishment for play selection. Variance in yards gained may be conceptualized in terms of variable schedules of "reinforcer" magnitude. Given the obvious relevance to operant choice of punishment (e.g., Critchfield, et al., 2003; Farley & Fantino, 1978) and variable-magnitude reinforcement (e.g., Davison & Hogsden, 1984), we wondered whether turnover risk and yardage variance might shed light on effects observed across all levels of our five situational variables, even though football experts do not speak of all game situations in terms of risk. To obtain a global estimate of situation-specific turnover risk and yards-gained variance, plays were pooled from all 32 NFL teams for the six games per team that comprised the present data set, and two ratios were determined for each level of each of our five situational variables. The first was the ratio of turnover rate for passing (interceptions plus fumbles) versus rushing (fumbles). The second was the ratio of standard deviations of yards gained from passing versus rushing. Figure 5 shows the relationship between these measures (logarithmically transformed for consistency with GME analyses) and the bias estimates shown in Figure 4. With ratios calculated as passing/rushing, the strong negative correlations shown in Figure 5 ($r = -.74$ for turnover risk and $r = -.76$ for yards-gained variance) indicate that preference indeed shifts toward rushing plays as passing becomes relatively more risky.
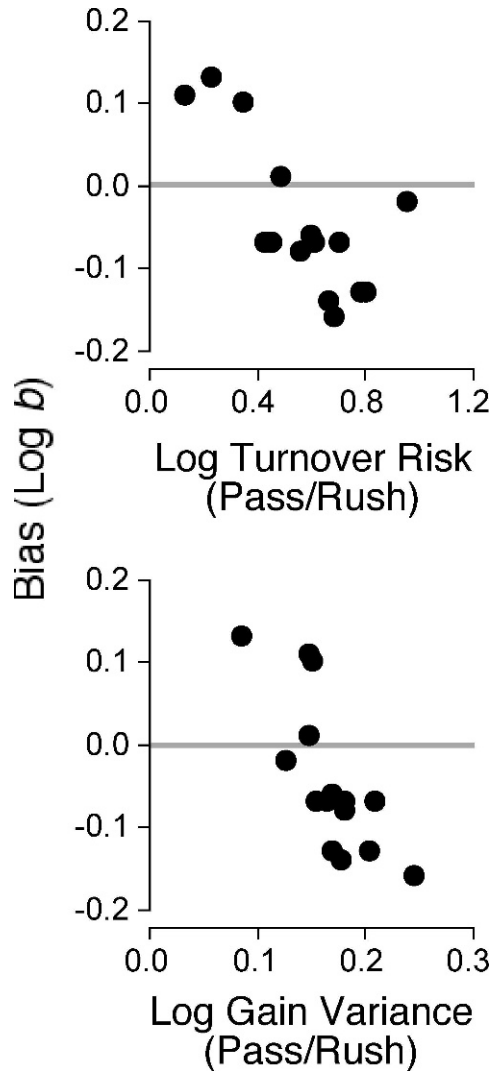

Fig. 5. Relationship between play-selection bias and two measures of play-selection risk. See text for details.

Although suggestive, the analysis of Figure 5 is flawed because it mixes two levels of analysis (the matching relations describe between-team differences in play selection, while risk ratios were determined by aggregating data from many teams). A more appropriate strategy might be to replicate the analyses for individual teams. Reed et al. (2006) showed matching for individual teams when each game in a season was treated as one observation. In theory, team-specific turnover rates and variance in yardage gains can be calculated and regressed against bias estimates at the team level, but this approach would require a

much larger sample of plays than the present study provided.

The relationship between risk and bias also could be examined by repeating the analyses of Figure 4 with a concatenated version of the GME that directly represents both punishment and reinforcer variability. Unfortunately, it is unclear whether punishment should be incorporated according to the dictates of one-factor theory, two-factor theory, or neither (Critchfield, et al., 2003), and reinforcer-magnitude variance has been the focus of extremely limited model-building. In the latter case, a large literature on risk aversion (e.g., Kahneman & Tversky, 1979) suggests that that bad outcomes (yardage losses or gains of zero yards) should affect play selection more than good outcomes (large gains), in which case preference may be a function of mean magnitude discounted according to the variance in some fashion that remains to be specified. Overall, the operant literature appears to offer no widely-endorsed model for incorporating punishment and reinforcer-magnitude variance into the GME.

The preceding concerns notwithstanding, the relationships shown in Figure 5 merit further consideration. From an operant-principles perspective, these results highlight the importance of developing better elaborated models of operant choice. From a football perspective, Figure 5 lends empirical support to the attention that football experts have placed on risk in certain play-selection situations, and simultaneously suggests that football experts may have underestimated the commonalities that exist across play-selection situations. That is, even game situations for which experts have not emphasized the role of risk appear to fall along the risk functions of Figure 5.

Concluding Observations

When a theoretical model is extended to a new everyday domain, the first question of interest is whether the model provides a good description of behavior in that domain (i.e., accounts for substantial variance). If so, then detailed mapping of model concepts, as defined by its fitted parameters, to domain-specific phenomena can proceed. For applications of the GME, the initial question is whether the relevant behavior follows some variation on the matching relation. If so, then the specifics of the matching function, and the conditions that influence these specifics, become of interest. Because the matching relation has been extended to many applied domains only recently (e.g., football by Reed et al., 2006), reliability of fit has been the focus of most investigations. The present findings demonstrate further reliability of fit (Figures 1 and 3), but also extend the generality of the GME by showing how a theoretically important fitted parameter (log b) is relevant to situation-specific play selection in football.

Such explanatory flexibility, though rarely explored in applications of the GME to date, is critical in two ways. First, to be taken seriously outside a small circle of behavior theorists, a theoretical account of any applied domain must address situation-specific differences in behavior that are well known to domain experts. As newspaper columns, radio talk shows, and web pages (e.g., http://www.twominutewarning.com) illustrate, football aficionados dissect their sport in great detail. An operant-choice account of play selection is unlikely to interest them unless it speaks to the rich play-selection variance that is part of the sport's appeal. Second, behavior theorists should be gratified when studies like the present one help to place the variability of an applied domain into a parsimonious conceptual framework. Although football fans tend to emphasize the uniqueness of various game situations, consistency across situations is shown both when play selection follows the linear pattern described by the GME and when game situations differ along a common dimension (e.g., the GME's bias parameter).

Exercises like the present one also serve basic behavior science by highlighting questions that have not received adequate attention in the laboratory. As noted above, although hundreds of concurrent-schedules studies have been conducted across several decades, relatively little is known about how factors such as punishment and moment-to-moment variability in reinforcer magnitude affect operant choice. Given the considerable challenges of simply understanding control of behavior by concurrent frequencies of positive reinforcement (e.g., Davison & Nevin, 1999), these omissions are understandable, but given the prevalence of aversive events (e.g., Sidman, 1989) and outcome variability (Kahneman & Tversky, 1979; Thaler & Sunstein, 2008) in the

everyday world, an outside observer might be forgiven for viewing the resulting account of behavior as somewhat limited. This underscores the essential role of translational research in revealing both the relevance and the frontiers of behavior principles as they currently are understood (Mace, 1994).

REFERENCES

Alamar, B. C. (2006). The passing premium puzzle. *Journal of Quantitative Analyses of Sport, 2(4)*, Article 5. Download from http://www.bepress.com/jqas/vol2/iss4/5.

Alferink, L. A., Critchfield, T. S., Hitt, J. L., & Higgins, W. J. (2009). Generality of the matching law as a descriptor of basketball shot selection. *Journal of Applied Behavior Analysis, 42*, 592–605.

American Football Coaches Association (Ed.). (1995). *Football coaching strategies.* Champaign, IL: Human Kinetics Press.

Allen, P. (2002). Passing on third down. In American Football Coaches Association (Eds.). (2002). *Football coaching strategies* (pp. 74–75). Champaign, IL: Human Kinetics Press.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 20*, 313–327.

Baron, A., & Derenne, A. (2002). Quantitative studies of single-subject studies: What do group comparisons tell us about individual performances? *The Behavior Analyst, 73*, 293–301.

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22*, 231–242.

Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior, 32*, 269–281.

Billington, E. J., & DiTommaso, N. M. (2003). Demonstrations and applications of the matching law in education. *Journal of Behavioral Education, 12*, 91–104.

Borrero, J. C., Crisolo, S. S., Tu, Q., Rieland, W. A., Ross, N. A., Francisco, M. T., et al. (2007). An application of the matching law to social dynamics. *Journal of Applied Behavior Analysis, 40*, 589–601.

Bryant, P. (1999). Executing in the critical zones. In American Football Coaches Association (Eds.). (1999). *Offensive football strategies* (pp. 287–292). Champaign, IL: Human Kinetics Press.

Bulow, P., & Meller, P. (1998). Predicting teenage girls' sexual activity and contraception use: An application of matching law. *Journal of Community Psychology, 26*, 581–596.

Buskist, W., & Degrandpre, R. J. (1995). Schedule-controlled responding of two persons under a single schedule of reinforcement. *The Psychological Record, 45*, 183–205.

Chomsky, N. (1959). A Review of B. F. Skinner's *Verbal Behavior. Language, 35*, 26–58.

Critchfield, T. S., Haley, R., Sabo, B., Colbert, J., & Macropoulis, G. (2003). A half century of scalloping in the work habits of the United States Congress. *Journal of Applied Behavior Analysis, 36*, 465–486.

Critchfield, T. S., & Kollins, S. H. (2001). Temporal discounting: Basic research and the analysis of socially important behavior. *Journal of Applied Behavior Analysis, 34*, 101–122.

Critchfield, T. S., Paletz, E. M., MacAleese, K. R., & Newland, M. C. (2003). Punishment in human choice: Direct or competitive suppression? *Journal of the Experimental Analysis of Behavior, 80*, 1–27.

Critchfield, T. S., & Reed, D. D. (2009). What are we doing when we translate from quantitative models? *The Behavior Analyst, 32*, 339–362.

Davison, M. (1988). Concurrent schedules: Interaction of reinforcer frequency and reinforcer duration. *Journal of the Experimental Analysis of Behavior, 49*, 339–349.

Davison, M., & Baum, W. M. (2003). Every reinforcer counts: Reinforcer magnitude and local preference. *Journal of the Experimental Analysis of Behavior, 80*, 95–129.

Davison, M., & Hogsden, I. (1984). Concurrent variable-interval schedule performance: Fixed versus mixed reinforcer durations. *Journal of the Experimental Analysis of Behavior, 41*, 169–182.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review.* Hillsdale, NJ: Erlbaum.

Davison, M., & Nevin, J. A. (1999). Stimuli, reinforcers, and behavior: An integration. *Journal of the Experimental Analysis of Behavior, 71*, 439–482.

Dube, W. V., & McIlvane, W. M. (2002). Quantitative assessments of sensitivity to reinforcement contingencies in mental retardation. *American Journal on Mental Retardation, 107*, 136–145.

Edwards, L. (2002). The passing game. In American Football Coaches Association (Ed.). (2002). *The football coaching Bible* (pp. 227–239). Champaign, IL: Human Kinetics Press.

Elliffe, D., Davison, M., & Landon, J. (2008). Relative reinforcer rates and magnitudes do not control concurrent choice independently. *Journal of the Experimental Analysis of Behavior, 90*, 169–185.

Farley, J., & Fantino, E. (1978). The symmetrical law of effect and the matching relation in choice behavior. *Journal of the Experimental Analysis of Behavior, 29*, 37–60.

Fulmer, P. (2002). Fast break, two-minute offense. In American Football Coaches Association (Ed.). (2002). *Football coaching strategies* (pp. 80–82). Champaign, IL: Human Kinetics Press.

Goodell, R. (2008). *Official playing rules of the NFL 2008.* Chicago, IL: Triumph Books.

Grace, R. C. (1995). Independence of reinforcement delay and magnitude in concurrent chains. *Journal of the Experimental Analysis of Behavior, 63*, 255–276.

Grace, R. C. (1999). The matching law and amount-dependent exponential discounting as accounts of self-control choice. *Journal of the Experimental Analysis of Behavior, 71*, 27–44.

Graft, D. A., Lea, S. E. G., & Whitworth, T. L. (1977). The matching law in and within groups of rats. *Journal of the Experimental Analysis of Behavior, 27*, 183–194.

Green, L., Rachlin, H., & Hanson, J. (1983). Matching and maximizing with concurrent ratio-interval schedules. *Journal of the Experimental Analysis of Behavior, 40*, 217–224.

Grott, R., & Neuringer, A. (1974). Group behavior of rats under schedules of reinforcement. *Journal of the Experimental Analysis of Behavior, 22*, 311–321.

Hamblin, R. L., & Miller, H. L. (1977). Matching as a multivariate power law: frequency of behavior versus frequency and magnitude of reinforcement. *Learning and Motivation*, *8*, 113–125.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, *4*, 267–272.

Herrnstein, R. J., & Heyman, G. M. (1979). Is matching compatible with reinforcement maximization on concurrent variable interval, variable ratio? *Journal of the Experimental Analysis of Behavior*, *31*, 209–223.

Herrnstein, R. J., & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior*, *24*, 107–116.

Hitt, J. L., Alferink, L. A., Critchfield, T. S., & Wagman, J. B. (2007). Choice behavior expressed in elite sport competition: Predicting shot selection and game outcomes in college basketball. In L. A. Chiang (Ed.), *Motivation of Exercise and Physical Activity* (pp. 79–91). Hauppauge, NY: Nova Science.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kehres, L. (2006). Redzone: Opponent's 25-yard line to goal line. In B. Mallory & D. Nehlen (Eds.), *Football offenses and plays* (pp. 179–188). Champaign, IL: Human Kinetics.

Kyonka, E. G. E., & Grace, R. C. (2008). Rapid acquisition of preference in concurrent chains when alternatives differ on multiple dimensions of reinforcement. *Journal of the Experimental Analysis of Behavior*, *89*, 49–69.

LaBounty, C. E., & Reynolds, G. S. (1973). An analysis of response and time matching to reinforcement in concurrent ratio-interval schedules. *Journal of the Experimental Analysis of Behavior*, *19*, 155–166.

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior*, *84*, 555–579.

Landon, J., Davison, M., & Elliffe, D. (2003). Concurrent schedules: Reinforcer magnitude effects. *Journal of the Experimental Analysis of Behavior*, *79*, 351–365.

Levy, M. (1999). Beating the clock. In American Football Coaches Association (Ed.). (1999). *Offensive football strategies* (pp. 281–286). Champaign, IL: Human Kinetics Press.

Lunneborg, C. E. (1994). Modeling experimental and observational data. Belmont, CA: Duxbury Press.

MacDonall, J. S. (1988). Concurrent variable-ratio schedules: Implications for the generalized matching law. *Journal of the Experimental Analysis of Behavior*, *50*, 55–64.

Mace, F. C. (1994). Basic research needed for stimulating the development of behavioral technologies. *Journal of the Experimental Analysis of Behavior*, *61*, 529–550.

Mace, F. C., Lalli, J. S., Shea, M. C., & Nevin, J. A. (1992). Behavioral momentum in college basketball. *Journal of Applied Behavior Analysis*, *25*, 657–663.

Magoon, M. A., & Critchfield, T. S. (2008). Concurrent schedules of positive and negative reinforcement: differential-impact and differential-outcomes hypotheses. *Journal of the Experimental Analysis of Behavior*, *90*, 1–22.

Mazur, J. (1991). Choice. In I. H. Iversen & K. A. Lattal (Eds.), *Experimental analysis of behavior, Part 1* (pp. 219–250). New York: Elsevier.

McCorduck, E. (1998). *Understanding American football*. Chicago: NTC Publishing.

McDowell, J. J., & Caron, M. L. (2010a). Matching in an undisturbed natural environment. *Journal of the Experimental Analysis of Behavior*, *93*.

McDowell, J. J., & Caron, M. L. (2010b). Bias and undermatching in delinquent boys' verbal behavior as a function of their level of deviance. *Journal of the Experimental Analysis of Behavior*, *93*.

Milliken, G. A., & Johnson, D. E. (2002). *Analysis of messy data, Volume III: Analysis of covariance*. Boca Raton, FL: Chapman & Hall/CRC.

Motulsky, H., & Christopoulis, A. (2006). *Fitting models to biological data using linear and nonlinear regression: A practical guide to curve fitting*. San Diego: GraphPad Software.

Reed, D. D., Critchfield, T. S., & Martens, B. K. (2006). The generalized matching law in elite sport competition: Football play calling as operant choice. *Journal of Applied Behavior Analysis*, *39*, 281–297.

Reed, D. D., & Martens, B. K. (2008). Effects of task difficulty on children's relative problem completion rates according to the generalized matching law. *Journal of Applied Behavior Analysis*, *41*, 39–52.

Rider, D. P. (1979). Concurrent ratio schedules: Fixed vs. variable response requirements. *Journal of the Experimental Analysis of Behavior*, *31*, 225–237.

Rockerbie, D. W. (2008). The passing premium puzzle revisited. *Journal of Quantitative Analysis in Sport*, *4*, 1–11.

Romanowich, P., Bourret, J., & Vollmer, T. R. (2007). Further analysis of the matching law to describe two- and three-point shot allocation by professional basketball players. *Journal of Applied Behavior Analysis*, *40*, 311–315.

Rubenstein, R. Y., & Kroese, D. P. (2008). *Simulation and the Monte Carlo method* (2nd ed.). Hoboken, NJ: Wiley.

Sanders, R. M. (1968). Note on magnitude of a money reinforcer in a concurrent schedule of reinforcement. *Psychological Reports*, *22*, 407–414.

Savastano, H., & Fantino, E. (1994). Human choice in concurrent ratio-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, *61*, 453–463.

Schmitt, D. R. (1974). Effects of reinforcement rate and reinforcer magnitude on choice behavior of humans. *Journal of the Experimental Analysis of Behavior*, *21*, 409–419.

Shahan, T. A., Podlesnik, C. A., & Jimenez-Gomez, C. (2006). Matching and conditioned reinforcement rate. *Journal of the Experimental Analysis of Behavior*, *85*, 167–180.

Shimp, C. P. (1966). Probabilistically reinforced choice behavior in pigeons. *Journal of the Experimental Analysis of Behavior*, *9*, 443–455.

Shurtleff, D., & Silberberg, A. (1990). Income maximizing on concurrent ratio-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, *53*, 273–284.

Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.

Sidman, M. (1989). *Coercion and its fallout*. Boston: Authors Cooperative.

Skinner, B. F. (1953). *Science and human behavior.* New York: Free Press.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Skinner, B. F. (1991). Why we are not acting to save the world. In W. Ishaq (Ed.), *Human behavior in today's world* (pp. 19–29). New York: Praeger.

Sokolowski, M. B. C., Tonneau, F., & Freixi I Baque, E. F. (1999). The ideal free distribution in humans: An experimental test. *Psychonomic Bulletin & Review, 6,* 157–161.

Stilling, S. T., & Critchfield, T. S. (in press). Using the matching law to evaluate situation-specific tendencies in football play selection: A preliminary report. *Experimental Analysis of Human Behavior Bulletin.*

Teaff, G. (1999). Moving the chains. In American Football Coaches Association (Ed.). (1999). *Offensive football strategies* (pp. 293–297). Champaign, IL: Human Kinetics Press.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New York: Penguin.

Todorov, J. C., Olivera Castro, J. M., Hanna, E. S., de Sa, M. C. N. B., & Barreto, M. (1983). Choice, experience, and the generalized matching law. *Journal of the Experimental Analysis of Behavior, 40,* 99–111.

Tranquil, G. (2006). Two-minute and no-huddle situations. In B. Mallory & D. Nehlen (Eds.), *Football offenses and plays* (pp. 201–210). Champaign, IL: Human Kinetics.

Tressel, J., & Bollman, J. (2000). Putting it in the end zone. In American Football Coaches' Assocation (Ed.). (2000). *Offensive football strategies* (pp. 301–305). Champaign, IL: Human Kinetics Press.

Tressel, L. J. (2000). Coaching for goal line success. In American Football Coaches' Assocation (Ed.). (2000). *Offensive football strategies* (pp. 315–320). Champaign, IL: Human Kinetics Press.

Vollmer, T. R., & Bourret, J. (2000). An application of the matching law to evaluate the allocation of two-and three-point shots by college basketball players. *Journal of Applied Behavior Analysis, 33,* 137–150.

Westering, F. (2002). Offensive play calling in key game situations. In American Football Coaches Association (Ed.). (2002). *The football coaching Bible* (pp. 315–326). Champaign, IL: Human Kinetics Press.

Wolff, P. C., Burnstein, D. D., & Cannon, D. L. (1964). The use of schedules of reinforcement to regulate a collective team response rate. *The Psychological Record, 14,* 57–70.

Wurster, R. M., & Griffiths, R. R. (1979). Human concurrent performances: Variation of reinforcer magnitude and rate of reinforcement. *The Psychological Record, 29,* 341–354.

Zar, J. H. (1999). *Biostatistical analysis* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

APPENDIX

Results of Paired Comparisons Among Levels of Game-Situation Variables

Down

	Sensitivity			Bias		
	F	df	p	F	df	p
1st vs. 2nd	0.40	1,60	.532	19.98	1,61	<.0001**
1st vs. 3rd	0.20	1,60	.660	55.76	1,61	<.0001**
2nd vs. 3rd	0.01	1,60	.924	26.48	1,61	<.0001**

Yards Needed to Earn New Downs

	Sensitivity			Bias		
	F	df	p	F	df	p
>10 vs. 5–10	0.42	1,60	.522	21.14	1,61	<.0001**
>10 vs. 1–4	1.06	1,60	.308	44.54	1,61	<.0001**
5–10 vs. 1–4	3.26	1,60	.076	23.37	1,61	<.0001**

Time Remaining in Half

	Sensitivity			Bias		
	F	df	p	F	df	p
>15:00 vs. 2:01–15:00	0.53	1,56	.466	1.10	1,57	.299
>15:00 vs. ≤2:00	0.89	1,56	.350	35.54	1,57	<.0001**
2:01–15:00 vs. ≤2:00	0.17	1,56	.687	29.05	1,57	<.0001**

Score

	Sensitivity			Bias		
	F	df	p	F	df	p
Winning vs. Tied	0.15	1,58	.702	5.67	1,59	.021
Wining vs. Losing	0.01	1,58	.906	27.42	1,59	<.0001**
Tied vs. Losing	0.04	1,58	.841	13.40	1,59	.001**

Field Position

	Sensitivity			Bias		
	F	df	p	F	df	p
83–99 vs. 9–82	4.82	1,48	.033*			***
83–99 vs. 1–8	0.74	1.48	.392	9.50	1,49	.0003**
9–82 vs. 1–8	1.32	1,48	.256	12.51	1.49	.0009**

* Statistically significant with alpha = .05. Note this outcome was not considered as evidence of a sensitivity effect due to a nonsignificant omnibus ANCOVA outcome.

** Statistically significant with alpha = .017 (.05/3).

*** No bias comparison conducted following sensitivity comparison with $p < .05$. See text for explanation.