ABSTRACT

Effects of Pitch Location and Count on Professional Baseball Umpires' Ball–Strike Decisions

Aaron R. Baggett, Ph.D.

A. Alexander Beaujean, Ph.D., Co-Chairperson
Jaeho Shim, Ph.D., Co-Chairperson

In baseball, home plate umpires' perceptual-cognitive skills are tested with each pitch as they are required to judge, with accuracy, whether the ball passed through the imaginary region above home plate known as the strike zone. Home plate umpires must visually track the flight of a pitched ball as it leaves the pitcher's hand and travels over the home plate region in order to accurately determine whether a pitch should be called a strike or a ball.

This study applied literature related to judgment and decision making among expert sport performers to the professional baseball umpire population. Using generalized linear mixed modeling with secondary data generated by the *PITCHf/x* pitch tracking system, umpires' ball–strike decisions were measured over the course of the 2013 season. Emphasis was placed on accounting for the effects of pitch location and ball–strike count on umpires' accuracy in making ball–strike decisions. During the 2013 MLB season, umpires were responsible for deciding the outcome of approximately 149 pitches per game on average. Results indicate umpire accuracy rates range from 90% to 95%.

To test for the effects of pitch location and ball–strike count on the probability of umpires' accuracy in judging pitch outcomes, a multilevel model with interactions between fixed and random effects was estimated. Results indicate predicted probabilities of accurate umpire decisions for pitches located in the inner region of the strike zone appear to be noticeably lower compared to predicted probabilities of accurate umpire decisions for pitches located in the middle or outer regions. Specifically, pitchers appear to be at a distinct disadvantage, compared to when the ball–strike count is either neutral or favors the batter.

Effects of Pitch Location and Count on Professional Baseball Umpires'
Ball–Strike Decisions

by

Aaron R. Baggett, B.B.S., M.Ed.

A Dissertation

Approved by the Department of Educational Psychology

Marley W. Watkins, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

A. Alexander Beaujean, Ph.D., Co-Chairperson

Jaeho Shim, Ph.D., Co-Chairperson

Grant B. Morgan, Ph.D.

Eric L. Robinson, Ph.D.

Gregory D. Speegle, Ph.D.

Accepted by the Graduate School
December 2014

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My sincere thanks to my committee members, Dr. Alex Beaujean, Dr. Joe Shim, Dr. Grant B. Morgan, Dr. Eric Robinson, and Dr. Greg Speegle. Without their expertise, collegial spirit, and belief in this project it would not have developed into the combination of interdisciplinary literature and quantitative methods it is.

Many thanks to the following family members for their love, support, and encouragement while completing this project: My sisters, Adriene and Abby; their husbands, Bryan and Josh; my grandparents, Vernon and Faye Kidd; my father- and mother-in-law, Dr. James E. and Ruth Williams; my siblings-in-law, Brian and Kristen Williams, Shannon and Tami Brett, and Kyle and Audra Walsh.

In particular, I want to thank my parents, Kelly and Nancy Baggett. Without their unending support, encouragement, and gracious love throughout my life I would not have been able to reach this milestone.

Finally, to my wife, Lisa R. Baggett, M.D., and daughter, Rhett Elisabeth: *Your* love, support, and encouragement have transcended what I could have ever imagined. I love you both more every day, my loves.

DEDICATION

To my dad, Kelly R. Baggett

CHAPTER ONE

Introduction

One of the fundamental responsibilities of a Major League Baseball (MLB) home plate umpire is to judge the outcome of all pitches delivered during a game in which the batter does not swing or contact. For home plate umpires, the visual and cognitive process involved in accurately judging pitch outcomes are similar to those utilized by batters. Home plate umpires must visually track the flight of a pitched ball as it leaves the pitcher's hand and travels over the home plate region in order to accurately determine whether a pitch should be called a strike or a ball. Umpires must also contend with pitchers' tactics in which pitches are routinely delivered with unpredictable movements, at varying speeds, and to areas that may be difficult for batters to reach.

Although researchers from several disciplines have investigated performance-related variables among sports officials, a considerable portion of the evidence focuses on competitive team participants and individual performers (see Krane & Williams, 2010). Consequently, there exists a noticeable lack of comprehensive application to the domain of sports officiating, even less to the baseball umpire population (e.g., Bar-Eli, Plessner, & Raab, 2011; Ste-Marie, 2003).

Therefore, the purpose of the current study is to answer a set of research questions using data generated by the *PITCH/fx* (Sportvision, Inc., 2013) pitch tracking system to measure the association of pitch- and umpire-related factors on the outcomes of umpires' ball–strike judgment and decision making patterns during the course of an entire Major League Baseball (MLB) season. With the advent of real-time pitch tracking technology in MLB, researchers can now measure with considerable accuracy a variety of game- and situation-context variables in more empirical, less invasive forms.

First, I introduce the literature on domain-specific perceptual-cognitive skills among expert performers as applied to the professional baseball umpire population. Literature examining domain-specific perceptual-cognitive skills among sport performers is plentiful (see Mann, Williams, Ward, & Janelle, 2007; Williams, Ward, & Smeeton, 2004; Williams, Ford, Eccles, & Ward, 2011; Williams & Ericsson, 2005 for a review). Conversely, research applied to sport officials—particularly baseball umpires–is lacking, possibly because gaining access to expert-level sport officials is difficult. Since sport-performance research is often conducted in laboratory settings, findings may lack ecological validity. For example, using a two-dimensional display to test for the effects of visual search patterns baseball batters utilize while tracking the flight of a pitched ball may limit researchers' full understanding of the visual search process. Similarly, Shim, Chow, Carlton & Chae (2005) demonstrated limitations in experimental methods typically used in laboratory settings when measuring perceptual-cognitive skills associated with patterns of judgment and decision making among sport performers. Instead, experiments and observations in the field may be more reliable.

In Chapter Two, particular attention is paid to the ways in which superior performance is demonstrated and characterized among expert sport performers–especially sport judges, referees, and game officials. Each section includes a cumulative overview of relevant research literature related to sport performance followed by a thorough application to the sport official population. Since the particular focus of this study is devoted to a professional baseball umpire population, all research related to this sample is addressed, and when not available, convincing conclusions are be drawn from other sporting domains as necessary. In Chapter Three, I outline the research questions and hypotheses of the study as well as provide information on the data analyzed as well an overview of procedures used to analyze multilevel models with a categorical repeated measures outcome. In Chapter Four, I present descriptive statistics and model summary results as well as principle findings. In Chapter Five I discuss the implications of the results by making connections to the current literature related to expertise in judgment and decision making among expert sport performers.

CHAPTER TWO

Literature Review

Expertise

The genesis of the field of expertise studies can be traced back to three domains during the early to mid-twentieth century. Artificial intelligence, psychology, and educational theory each played a pivotal role in the field's development (Feltovich, Prietula, & Ericsson, 2006). Artificial intelligence (AI) addressed issues of human intelligence and thought by computation and the use of computers and computer programs as formal models of human cognition (Ernst & Newell, 1969; Feltovich et al., 2006; Newell & Simon, 1956; Samuel, 1959). Consequently, current theories of cognitive science and the information-processing model originated through early AI work (Feltovich et al., 2006). During the reign of behaviorism in the early part of the twentieth century, alternatives to the pervasive stimulus-response model began to appear with increasing regularity. Cognitive psychology was one such alternative, which was gaining momentum among researchers interested in exploring human information processing and problem solving (Buss, 2012). Work related to cognitive psychology during the time by de Groot (1965) and Chase and Simon (1973) stimulated the work of other researchers to explore similar constructs, which necessarily led to the formation of a field of expertise studies (Hodges, Huys, & Starkes, 2007).

Like psychology at the time, educational theory and practice were primarily informed by traditional behavioristic models (Feltovich et al., 2006). The use of teaching machines and programmed learning, which were designed to reinforce and stabilize learning connections as well as identify errors, were extensively utilized in educational settings. The programmed learning methodology bears a striking resemblance to a well-defined concept in expertise studies known as deliberate practice (see Ericsson, 2003, 2006a for a review). The purpose of deliberate practice, and programmed learning for that matter, is to provide clear goal-directed objectives, repeated practice, and meaningful feedback. A detailed definition and explanation of the role deliberate practice plays in expertise is provided in a later section of this chapter.

Research related to the scientific study of expertise and expert performance has become increasingly robust (Ericsson, 2006a). De Groot (1965), and Chase and Simon (1973) are generally credited with pioneering the field with their respective investigations of recall, pattern recognition, and performance among master chess players (Ericsson, 2003; Feltovich, Prietula, & Ericsson, 2006; Ste-Marie, 2003). De Groot (1965) examined concurrent verbalizations during move selection among highly skilled and lesser skilled players. His results demonstrated a noticeable difference between players' performance on a perceptual and short-term memory task. Master chess players outperformed novice players in reconstructing a game board arrangement after viewing the original position for only five seconds. However, when the game board was randomly rearranged, master and novice players performed similarly poor (Chase & Simon, 1973). This finding indicates an important distinction between experts and novices: Experts must be more adept at observing and perceiving familiar patterns or meaningful constellations, which are subsequently easier and more quickly retrievable (de Groot, 1965).

Chase and Simon (1973) attempted to isolate and characterize the perceptual structures identified by de Groot (1965). In general, findings from Chase and Simon's (1973) memory and perception tasks reveal insights into the sophisticated and cognitively complex memory and chunking patterns utilized by experts (Ste-Marie, 2003). Moreover, the most distinguishing factor consistently found among experts, particularly in task selection or recall paradigms, was a pattern of systematic organization of accumulated domain-specific knowledge (Ackerman & Beier, 2006; Chi, 2006; Ericsson, 2006c; Horn & Masunaga, 2006). Compared to non-experts, this knowledge, combined with experience, allows for greater speed and accuracy in pattern recognition and recall tasks (Abernethy & Russell, 1987).

A substantial number of studies related to expertise have concentrated on the bifurcation of expert-novice differences in performance and problem solving (see Hodges et al, 2007). A number of salient, generalizable characteristics have been found to exist among experts across multiple domains (Ericsson, 2006a). For example, experts consistently demonstrate efficiency in cognitive effort during domain-specific knowledge retrieval, skill execution becomes automatized (Janelle & Hillman, 2003), and pattern detection and recognition is achieved with greater speed and accuracy compared to non-experts (Ste-Marie, 2003). Likewise, during problem solving tasks, experts demonstrate ability to consistently and reliably generate strategies and solutions with greater speed and accuracy than novices, even under time constraints (Chase et al., 1973; Ericsson, 2003). In sport performance settings, experts demonstrate significant sensitivity to anticipatory cues (Williams, Ward, & Smeeton, 2004), advanced knowledge of situational probabilities (Abernethy, Maxwell, Masters, Van Der Kamp, & Jackson, 2007), and efficient visual search strategies (Abernethy, 1987a, 1987b).

Despite the above ways in which experts excel, they also are susceptible to errors (Chi, 2006). For example, as was demonstrated by de Groot and Chase and Simon, as well as others since (e.g., Mann, Williams, Ward, & Janelle, 2007), the knowledge and skills experts demonstrate are limited significantly to the domain in which they excel. Taken outside of his or her specific domain, experts appear to perform on average as well as novices. This finding has been repeatedly demonstrated in numerous studies evaluating domain-specific representative tasks (Ericsson, 2006b). Likewise, during pattern detection and recognition tasks, for example, experts are vulnerable to bias and functional fixedness whereby thinking becomes routinized and inflexible, thus inhibiting novel problem solving skills (Kotovsky, 2003). Bias in expert judgment and decision-making is explored in a later section of this chapter.

That said, experts do tend to possess more domain-related knowledge (e.g., Williams & Ericsson, 2005), demonstrate superior speed in judgment and decision-making (e.g., Williams & Ward, 2007), and are often more experienced than novice performers (e.g., Abernethy, 1989). Less common are studies related to the comprehensive investigation of relevant knowledge structures and cognitive skills exhibited by expert performers (Krane & Williams, 2010). Studies of this kind are believed to be valuable and necessary in order to fully capture the dynamics of expertise as well as further develop current theoretical and methodological understanding of expertise and expert performers (Abernethy, Farrow, & Berry, 2003).

Since the early work of de Groot and Chase and Simon, efforts have been made to explore the nature of expertise and the dynamics of expert performance from a variety of domain-related perspectives, including music and dance (e.g., Ericsson, 2003), the physical and biological sciences (e.g., Ericsson, 2003), sport (e.g., Hodges, Huys, & Starkes, 2007), and more recently, sport judges, referees, and game officials (e.g., Ghasemi, Momeni, Jafarzadehpur, Rezaee, & Taheri, 2011; Helsen & Bultynck, 2003; Ste-Marie, 2003; Ward & Williams, 2003).

Expertise in Sport

Expertise in sport has been defined as the possession of specialist knowledge and skills, which are necessarily accompanied by the accumulation of extensive hours of intentional practice and meaningful experience (Ericsson, 2006a; Janelle & Hillman, 2003; Moran, 2009). An important qualifier in defining and characterizing expertise in this regard is the presence of the ability to consistently reproduce superior performance within a particular domain over an extended period of time (Abernethy, 1989; Starkes, 1993). Janelle and Hillman (2003) provide three generalizable areas representative of expertise in sport. Technical expertise involves sensory motor coordination and the eventual automatization of refined and efficient movement patterns. Cognitive expertise involves tactical and strategic knowledge as well as perceptual and decision-making skills. Emotional expertise involves the ability to self-regulate and monitor levels of arousal and anxiety, for example.

More specifically, research literature related to expertise among sport performers consistently identifies a number of representative complex cognitive skills such as pattern recognition (Abernethy, 1991), superior memory and recall ability (Chase & Simon, 1973) as well as advance cue anticipation and visual search behavior (Hoffman & Lintern, 2006). These skills have been widely demonstrated across sports as diverse as baseball (Smith & Christensen, 1995), basketball (MacMahon & Plessner, 2008), golf (Boucher & Crews, 1987; Short et al., 2002), soccer (Ward & Williams, 2003), and tennis (McPherson & Kernodle, 2003).

Measurement and Representation of Expertise

It has been widely demonstrated that expert sport performers, compared to novice or less experienced performers, consistently achieve superior performance when factors such as perceptual and cognitive expertise (Mann, Williams, Ward, & Janelle, 2007; Williams, Ward, & Smeeton, 2004;), visual search and selective attention strategies (Abernethy, 1991; Davids & Williams, 1998; Takeuchi & Inomata, 2009), and anticipation and decision-making (Abernethy, 1989; Houlston & Lowes, 1993; Williams & Ward, 2007; Tenenbaum, Sar-El, & Bar-Eli, 2000; Houlston & Lowes, 1993) are empirically assessed. Following is a brief review of methodologies, measures, and other materials used in previous work evaluating domain-specific expertise.

In a recent review and evaluation of research in expert sport performance, Abernethy (2008) classifies several forms of expression, which are key to the expert's performance advantage. First, expert sport performers have consistently demonstrated superiority in pattern recognition and recall. The use of interactive video training or scenario simulation and still frame images from game sequences have most often been used in order to examine pattern recognition and recall. For example, Fadde (2006) used interactive video to measure perceptual decision-making in baseball. Others have used two-dimensional point-light displays to measure ball location prediction in tennis (Williams & Ward, 2007), still frame images from soccer match game film to measure players' knowledge of situational probabilities (Ward, Williams & Ericsson, 2003), information recognition implemented by squash players to predict forth-coming action (Abernethy, 1989; Abernethy, 1990), and infraction detection among basketball referees (MacMahon, Starkes, & Deakin, 2007).

Expert sport performers have also demonstrated abilities to selectively focus their attention, multi-task, and automatize output motor responses (e.g., Abernethy, 2008; Davids & Williams, 1998). In order to understand the dynamics of attention control, researchers would assign participants tasks involving a primary motor response such as that which is typically performed in their respective domain coupled with a secondary task designed to compete for attentional demands (e.g., Abernethy, 2008). Expert performers demonstrate an ability to automatize relevant movement and motor responses after prolonged experience and practice, freeing attention from a primary task in order to allocate attention to concurrent demands (Abernethy, 2001).

Measures of visual search strategies have been widely implemented in sport expertise literature, including eye movement and gaze fixation to measure timing and decision making in baseball batters (Abernethy, 2008; Paull & Glencross, 1997; Takeuchi & Inomata, 2009), proficiency in anticipation of ball landing positions among expert and non-expert wicketkeepers in cricket (Houlston & Lowes, 1993), selective attention in soccer (Davids & Williams, 1998), as well as other fast ball sports (Abernethy, 1991). Tenenbaum, Sar-El, and Bar-Eli (2000) employed similar methodology when examining developmental perspectives of ball location anticipation among tennis players in experimental groups ranging in ages from 8–11, 11–14, 14–18, to 18 and up. Older, more experienced players demonstrated greater ability to anticipate ball location through the use of an opponent's postural orientation such as racquet angle and other body parts prior to the ball-racquet contact. Other studies have demonstrated age effects in characteristics of expertise. For example, perceptual-cognitive skills and visual search behavior are typically developed over time, will peak, and subsequently decline with age (Abernethy, 1988; Bradbury, 2009; Krampe & Charness, 2006).

Specifically, a number of representative skills among expert sport performers have been assessed, leading to the effective representation of domain-specific expertise (Williams, Ward, & Smeeton, 2004). For example, Paull and Glencross (1997) examined what visual cue information elite baseball batters use and when such information is recognized. In the first experiment, expert and novice batters were presented with an interactive video simulation of neutral and strategic pitches delivered by an elite baseball pitcher. The experiment had two components. In the first component, batters were required to demonstrate swing-timing skill while predicting the pitch's location within the strike zone where the batter believed the ball would cross during pitch delivery. Bats were fitted with a strain gauge, which, when depressed, would record time of swing decision as well as pause further video playback. Expert batters were not only able to decide the pitch's trajectory through the strike zone earlier in ball flight, but were also significantly more accurate in their predictions than were the novice batters.

In the second component, batters were presented with a comprehensive game scenario card depicting a random game situation, including pitch count, number of outs, bases occupied, if any, game score, inning, and the types and locations of pitches previously delivered. Batters were then asked to predict the most likely pitch type and location given the game scenario. Again, expert batters demonstrated greater speed and accuracy in predicting both pitch type and location.

These findings support Abernethy's (2008) classifications of sport expertise related the role domain-specific knowledge plays in perceptual-cognitive skills in sport. Similarly, experts who recorded earlier decision times in pitch trajectory and location observation are believed to demonstrate an ability to form alternative decision schema on which they could rely if preparatory decisions are discovered to be flawed (Paull & Glencross, 1997). In short, the ability to recognize cue information faster allows expert performers more time to adapt output responses if necessary.

As biotechnology has become cost effective and more easily accessed, researchers are incorporating these methods in evaluations of expertise in sport (Fadde, 2006). Moran (1996) provides a thorough review of attentional measures through the use of psychophysiological methods such as heart rate monitoring and electroencephalogram technology; dual-task, self-report, and concurrent verbalizations; as well as psychometrically based tests and inventories of attention and other psychological skills. The question then of how performers acquire specialized, domain-specific expertise is addressed in the following section.

Acquisition of Expertise

The process of acquiring expertise in sport has been widely examined from the perspective of a variety of domains. Expertise is developed gradually and purposefully over the course of one's life (de Groot, 1946; Chase & Simon, 1973; Csikszentmihalyi, 2010; Ericsson, 2003). Ericsson (2003) conceptualizes the acquisition of expert performance as a form of problem solving. Throughout development, performers are faced with progressively more challenging problems that require a number of knowledge, skills, and abilities to solve. In order to gain a complete understanding of the characteristics and acquisition of expert performance, evaluation should necessarily take place within the specific domain where performance regularly occurs, involve tasks that are representative of superior performance within the domain, and should likewise be reproducible (Ericsson, 2006b). French and McPherson (1999) argue that the initial acquisition of perceptual skill in sport begins in conjunction with motor skill development.

Goodale and Milner (1992; 2005) have introduced a growing body of research proposing a dual-pathway model of visual anticipation. Instead of a single, utility-like neural mechanism responsible for regulating both perception and action, Goodale and Miner and subsequent others (e.g., Abernethy & Mann, 2008; Mann, Abernethy, & Farrow, 2010; Mann, Ho, De Souza, Watson, & Taylor, 2007; Milner & Goodale, 2008; Shim et al., 2005; Van der Kamp, Rivas, Van Doorn, & Savelsbergh, 2008) have found considerable evidence for the existence of separate but interacting cortical pathways. According to Goodale and Milner (2005) the role of perceptual representations is to help the organism arrive at a decision to act in a particular way. This division of labor (vision-for-perception and vision-for-action) is observed between the ventral visual pathway and the dorsal visual pathway located in the occipito-temporal cortex and posterior parietal cortex of the human brain, respectively (Goodale & Westwood, 2004).

Although the ventral and dorsal streams process information about the structure of objects and about their location in the environment, they appear to transform that information for output differently. For example, it is the ventral stream, or vision-for-perception system, which provides rich and detailed contextual representations required for cognitive actions such as recognition, identification, and planning (Goodale & Westwood, 2004). The dorsal stream, or vision-for-action system, on the other hand calculates absolute metrics of target objects essential for programming and executing an action response.

Role of experience. The demands of expertise necessitate a period of preparation during which an individual devotes years of training and deliberate practice toward mastery of a domain (Ericsson, 2006; Weisberg, 2006). Extensive domain-specific experience is necessary in order for performers to achieve expert status (Chase & Simon, 1973; Ericsson, 2006). With the exception of some prodigious children, expertise most often occurs in adulthood (Ericsson, 2006a). In a study of the effects of aging on peak performance in professional baseball, Bradbury (2009) discovered that players' respective skill- and task-demands were the deciding factor in determining age at peak performance. For example, age at peak performance among pitchers who were considered power pitchers and who accumulated a large number of strikeouts per season was 23. Age at peak performance among pitchers who demonstrated perhaps better control over pitch location and who subsequently allowed opponents to score fewer runs on average was 29.

For many performers it takes thousands of hours of accumulated experience in order to reach expert levels (Bryan & Harter, 1899; Chase & Simon, 1963; Ericsson, 2003; Ericsson, 2006a). It should be noted, however, that cumulative domain-related experience acquired by proxy is not by itself sufficient for the attainment of expertise (Ericsson, 2006b). This view supports literature citing a developmental trajectory from novice to expert status over the course of about ten years (e.g., Chase & Simon, 1973; Sternberg, 1996). More than experience, deliberate, methodological, and efficient practice has proven to be the best predictor of expertise (Ericsson, 2006a).

Deliberate practice is a highly structured process requiring the completion of sequentially prescribed tasks, which are initially beyond the performers ability, but can be mastered relatively quickly through repetition, monitored guidance, and meaningful feedback (Ericsson, 2003; Ericsson, 2006). For example, MacMahon, Helsen, Starkes, and Weston (2007) examined decision-making skills and performance improvement of soccer referees through the use of deliberate practice, finding that anticipatory cues, pattern recognition, and knowledge of situational probabilities increased among referees who participated in focused sessions of deliberate practice. Their findings reinforce the importance of regular and informative feedback from a coach or training figure during practice sessions. Among the referees, the mean number of years required to achieve expert status within the international soccer refereeing organization was 16 years; well above the traditional ten-year rule proposed by Csikszentmihalyi (1996). MacMahon et al. (2007), attribute the prolonged career trajectory to the lack of formal deliberate practice. These findings illustrate the influence deliberate practice plays compared to experience.

Similarly, the use of imagery and the mental rehearsal of a performance or task-related skill have been shown to be an effective method of deliberate practice (Nordin, Cumming, Vincent, & McGrory, 2006; Vealy & Greenleaf, 2010). Imagery and mental performance rehearsal are methods often employed by coaches and sport psychologists to train athletes or performers in sustaining prolonged concentration (Martin & Hall, 1995; Short et al., 2002). The assumption is that using one's senses to create or recreate an experience in the mind helps the brain interpret the imagery as identical to the actual stimulus (Jeannerod, 1994; Vealy & Greenleaf, 2010; Weinberg, 2008).

On the contrary, simply performing repetitions of an already acquired skill serves only to maintain proficiency (Cote & Fraser-Thomas, 2008). In this regard, deliberate practice is most effective in the acquisition and development of expertise as it allows for concentration on the continual adaptation, modification, and implementation of performance-related goals (Ericsson, 2006a). Furthermore, deliberate practice allows for the modification of the complex cognitive mechanisms, which mediate and allow for increases in performance level.

Role of cognitive and associative skills. Ericsson (2003) proposed a correlational model of experience and performance. The everyday skills people have acquired throughout the course of their lives serve as a functional foundation for the development of other skills. Likewise, they are relatively easily acquired and quickly pass through an associative phase toward autonomy. Once individuals pass through the cognitive and associative phases, performance can be generated automatically.

As new skills are eventually acquired, already existing skills are merely changed, transformed, or modified in a way that will mediate the new demands. Ericsson (2003) described the process of skill acquisition and performance as an ever-changing series of cognitive mechanisms, which the individual regulates and controls in order to facilitate increased skill level. Eventually, skills, which have been previously acquired will become automatic and therefore require little to no cognitive regulation during their execution. Several investigators have discussed the potential detriment that automatization of skill execution can have on performance development (Abernethy, 2008). During this phase, individuals adopt a particular method or skill representation to which they become committed (Ericsson, 2003).

Conversely, expert performers have managed to successfully avoid arrested development, or the plateau effect, and develop increasingly complex cognitive representations in order to sustain expert performance within their domain. The process of continual skill acquisition and performance improvement requires conscious, deliberate action on the part of the performer (Ericsson, 2006). Intrapersonal performance evaluation and guided practice with the help of an objective coach or trainer are essential to stimulating growth in skill development and performance (Ericsson, 2003). A review of psychological characteristics expert sport performers demonstrate will follow.

Perceptual-Cognitive Skills

Thus far, a case has been made for the relative importance that experience and goal-directed, concentrated practice play in the representation of expertise in sport performance. Similarly, a number of perceptual-cognitive skills and complex knowledge structures have been investigated and consistently associated with sport expertise. Following is a comprehensive review of research literature revealing ways in which perceptual-cognitive skills such as visual search behavior, pattern recognition, advance cue utilization, and concentration and attention are characterized in expert sport performance. In addition, a similar review of general knowledge structures such as declarative knowledge, procedural knowledge, and knowledge of situational probabilities will follow.

Perceptual-cognitive skill refers to the ability to recognize and adopt information within a particular context, which can then be integrated with existing knowledge to form and execute a response (Mann, Williams, Ward, & Janelle, 2007). Anticipation, perception, and decision-making are essential to the success of competitive athletes. Research related to these topics as well as a variety of relevant others is robust (see Mann, Williams, Ward, & Janelle, 2007; Williams, Ward, &

Smeeton, 2004; Williams, Ford, Eccles, & Ward, 2011; Williams & Ericsson, 2005 for a review). Furthermore, a significant portion of prior research studies have been devoted to examining individual differences between and among athletes who are considered experts and those who are considered novices (e.g., Ericsson, 2006; Moran, 2009; Paull & Glencross, 1997). Expertise in sport has been defined as the possession of specialist knowledge and skills, which are necessarily accompanied by the accumulation of extensive hours of intentional practice and meaningful experience (Ericsson, 2006; Moran, 2009).

In recent past, researchers have begun to regard perceptual and cognitive differences among elite athletes as a more precise discriminator of skill level and successful performance—regardless of their physiological or muscular features (Abernethy, Maxwell, Masters, van der Kamp, & Jackson, 2007; Ghasemi, Momeni, Jafarzadehpur, Rezaee, & Taheri, 2011; Moran, 2009; Williams & Ward, 2007; Williams, Ward, & Smeeton, 2004). This type of research is important for sport psychologists, coaches, and other athletic personnel responsible for improving the performance of competitive athletes. Conversely, very little research exists related to the perceptual and cognitive skills of officials in fast-ball sports such as baseball or soccer, to name a few (Ghasemi, et al., 2011).

Visual Search Behavior

Several researchers have thoroughly investigated the role visual skills play in certain aspects of fast-ball sports like baseball batting (e.g., Paull & Glencross, 1997; Takeuchi & Inomata, 2009). In the case of baseball, even the casual fan is familiar with the speed at which pitches are thrown and the limited amount of time a batter has to react. Therefore, it would seem natural to assume that the scientific study of visual skills would yield the most meaningful answers regarding how it is possible for a player to hit a baseball traveling upwards of 90 miles per hour with a bat no wider than 2.75 inches in diameter (Adair, 2002). Moreover, visual skills are perhaps better understood in terms of visual search patterns. Eye movements among expert performers have been demonstrated to be controlled by a search strategy, allowing the performer to use time more efficiently while analyzing the display (Williams, Davids, & Williams, 1999). These strategies include procedural knowledge, advance cue utilization, pattern recall and recognition, and knowledge of situational probabilities (Ghasemi, et al., 2011; Paull, et al., 1997; Takeuchi, et al., 2009; Williams, et al., 2004; Williams, et al., 2009).

As Van der Kamp et al. (2008) noted, there exists a pronounced limitation in the current body of visual anticipation research, which overlooks the contribution of the dorsal stream. Many of the aforementioned experiments have examined characteristics related to only the ventral stream. Shim et al. (2005), however, stressed the importance of a perception-action coupling response, and thus the dorsal stream, whereby highly skilled tennis players were able to obtain significantly more information from live-hitter conditions compared to projected visual conditions. Furthermore, much of the dual-pathway literature has focused on the visual anticipation among sport performers who, for example, in some fast-ball sports may be required to intercept a ball in a time-constrained context as in a return in tennis or swing in baseball. More work is needed in this area within the context of sport officiating–particularly professional baseball umpires who are responsible to judge and evaluate the type and precise location of a pitched ball, but are not constrained by time, as is the case for a batter.

Advance Cue Utilization

Advance cue utilization is the process of recognizing relevant domain-specific information within a given context for use in the formation of a response to forthcoming action. Furthermore, this skill is essential to expert performance within fast-ball sports as the pace of play is often enhanced and dictates the necessity for decisions to be made in advance (Williams, Ward, & Smeeton, 2004). Studies have incorporated visual occlusion techniques as well as point-light displays of an opponent's postural orientation in order to evaluate advance cue usage among expert sport performers (Williams & Ericsson, 2005). During visual occlusion procedures, participants are often presented with a filmed recording of a relevant, domain-specific event. At specified times during the presentation, researchers will remove certain visual patterns of body parts or occlude the scene entirely. The participant's task then is to anticipate forthcoming action at each stage of occlusion. Compared to novices, experts have consistently demonstrated superiority in predicting forthcoming action more accurately and with greater speed (Williams et al., 2005).

Declarative and Procedural Knowledge

Paull and Glencross (1997) discuss the differences between declarative knowledge and procedural knowledge by applying adaptive control of thought theory to perception and decision making in baseball. From the results they discovered that expert batters are able to string together a group of separate task-related items to form an efficient output. Furthermore, over time expert batters' units of knowledge tend to become sophisticatedly organized in terms of particular sport-specific domains. For example, as opposed to the novice baseball batter who tends to rely strictly upon declarative knowledge—or factual, concrete awareness involving the rudimentary construction of the essential components necessary for a reaction–the expert "...establishes greater numbers of links between related knowledge units to provide for reduced time to activate a node into working memory" (Paull, et al., 1997, p. 37). In other words, procedural knowledge suggests that the more an expert batter sees or experiences the rotation of a pitcher's breaking-ball, for example, the less time he needs to process what his reaction should be and instead concentrate on deciding if and when to swing. This is what Paull and Glencross (1997) call anticipatory cue utilization or anticipatory processing.

So in the case of an expert baseball umpire, instead of passively waiting for a possible play to ensue, he can anticipate any number of situations, which might unfold given the particular game situation or scenario regardless of where they may occur. This act of anticipatory processing thus reduces his reaction time while allowing him more time to consider how to obtain the proper distance and angle from which to make the call depending on the situational outcome.

Knowledge of Situational Probabilities

Knowledge of situational probabilities are essential to the success of a baseball umpire. Many of the decisions an umpire makes during the course of a game require flash bulb judgment. For example, if a speedy batter hits a slow ground ball toward the shortstop's position requiring him to charge the ball rather than wait for it to arrive, the ensuing play at first base is likely to be close, such that the first base umpire cannot determine which arrived first, the runner or the ball. In this case, often times the umpire is required to listen for the distinct sounds of the ball popping the leather of the first baseman's mitt and the thud of the runner's foot landing on top of first base. In this case, an expert umpire's knowledge of situational probabilities would indicate the call is likely to be a close one, and as the action unfolds, he can then begin to establish accurate expectations of likely events (Williams, et al., 2004).

Paull and Glencross (1997) argue that expert perception is a consequence of cognitive and knowledge structures instead of visual acuity alone. For example, an umpire's continuous accumulation of knowledge regarding a variety of visual cues allows for the eventual habituation of unnecessary stimuli such as crowd noise or movement of players in the umpire's periphery. Secondly, procedural knowledge allows for, even facilitates, the categorization of a potential outcome. In other words, when a batter or umpire recognizes certain visual cues there is already an output motor response available. Of course this can fail temporarily and cause the umpire to err in his judgment.

Finally, a cognitive or conceptual network allows for the priming of knowledge, which can allow for anticipation of forthcoming action. For example, an umpire can anticipate a number of likely sequences by quickly surveying the defense's alignment, the posturing of base runners, anticipate a pitcher's type of pitch and subsequent location, or even a team's offensive strategy for sending base runners to steal the next base while the batter sacrifices a bunt. Although neither a batter nor an umpire can be consistently successful without a certain level of visual acumen, it is their ability to store and build elaborate knowledge structures pertaining to a variety of outcomes that separate them from novices.

Judgment and Decision Making in Sport Performers and Officials

Accurately judging relevant visual and perceptual information is a primary task sport performers are required to complete during competition (Bar-Eli et al., 2011). Likewise, transforming judgments of visual and perceptual information into an effective and appropriate decision is also essential to sport performance with consequences being either success or failure (Bar-Eli et al., 2011).

Judgment involves differentiation between objects or stimuli in an environment with particular attention paid to the characteristics of each (Bar-Eli et al., 2011). Furthermore, judgments are primarily concerned with the appraisal of domain-specific information and may not necessarily involve consideration of subsequent consequences. However, judgments are often connected to decisions, which carry a spectrum of consequences. A decision involves commitment to a particular action with consequences intended to be satisfying for specified individuals, such as players, coaches, or teams (Yates & Tschirhart, 2006).

Although both judgment and decision-making are often used interchangeably to describe the same human process (e.g., Ross, Shafer, & Klein, 2006), current research distinguishes between the two. Judgments are believed to involve a process of evaluating and inferring from information readily available in order to be used in making a decision (Bar-Eli et al., 2011). Furthermore, a judgment is considered separate from the consequences of most decisions. Conversely, decisions involve a careful, calculated choice with frequently crucial consequences. For example, the primary responsibility of home plate umpires is to judge, with accuracy, the outcome of pitches. However, in certain situations, a home plate umpire may be required to evaluate, or judge, a combination of actions.

For example, a batter, in his attempt to hit the ball, may be struck in the hand by the pitch, causing the ball to enter the field of play. In this situation, the play is to be ruled "dead" as a result of the ball contacting the batter. However, since the batter swung at the pitch, the result would be a strike awarded to the pitcher. By viewing all action during this play, a home plate umpire must judge: (a) whether the pitch passed through the strike zone, (b) whether the batter's swing was deemed to be a legal attempt to make contact with the ball, and (c) whether the pitch contacted either the player or the bat. Thus, within the judgment and decision making framework, a home plate umpire compiles all information accumulated during the judgment phase in order to make a proper decision as to the play's ruling.

A number of studies have been conducted investigating judgment and decision-making in sport (e.g., Ford, Gallagher, Lacy, Bridwell, & Goodwin, 1999; MacMahon, Helsen, Starkes, & Weston, 2007; Newell, 1974; Paull & Glencross, 1997). Because the application of judgment and decision-making expertise to the sport domain has occurred relatively recently, much of the current understanding is gleaned from studies related to perception, knowledge, and decisions within a particular sport domain (Bar-Eli et al., 2011). Therefore, it seems logical to incorporate with other sections findings related to the judgment and decision making of sport performers. Theories and current perspectives of judgment and decision-making are provided in the next section, followed by a review of research on biased judgment and decision-making in sport.

Bias in Judgment and Decision Making

Sport performers and judges are susceptible to bias in their judgment and decision-making processes. From an evolutionary perspective, Buss (2012) proposes a reconceptualization of judgment bias. Instead of adhering to traditional views of human judgment, which suggest cognition, by its very nature, is rife with propensities for error, Buss supports a hypothesis that errors and biases are somewhat rare in the real world and when they occur, are related to frequency representations and judgment under uncertainty. According to this view, people remember the number of events occurring after a judgment or decision was made. In addition, frequency representations allow for an on-going process of updating long-term memory to accommodate any new information or events encountered. Similarly, they allow a person to modify and update their cognitive database in order to preserve memory of the event and its occurrence for future reference. This perspective offers interesting insight when considering bias in judgment and decision-making from the perspective of sport judges, referees, and officials. In the case of baseball umpires, each unhit pitch delivered to a batter during a game is, in essence, a judgment or decision task. According to Buss (2012), if an umpire makes a decision with some uncertainty in mind, he should either begin a frequency representation or add to an already existing one with the goal being avoidance of any future errors.

Others have investigated the nature of bias in sports officiating (e.g., Parsons, Sulaman, Yates, & Hamermesh, 2011; Rainey & Larsen, 1988; Rainey, Larsen, & Stephenson, 1989; Rainey, Larsen, Stephenson, & Coursey, 1989). In a study examining foul calls in college basketball, referees were observed to have called fouls on the team with the fewest number, suggesting an attempt among referees to balance team foul differential. Evidence even suggests that baseball umpires will demonstrate bias toward a defensive team during routine double plays by giving an infielder the benefit of the doubt when his foot is not in contact with the base while in possession of the ball (Rainey, Larsen, Stephenson, & Olson, 1993). This is customarily known as a phantom-tag. Additionally, effects of a pitcher's reputation on an umpire's ball-strike ratio have been observed (Rainey et al., 1989). Some research suggests that baseball umpires will demonstrate bias in the form of racial/ethnic preferences (Parsons, Sulaman, Yates, & Hamermesh, 2011). In addition to a variety of potential biases in judgments and decisions, sport performers are also susceptible to a variety of other external distractors, which, when left unmanaged, can have a detrimental effect on performance.

Researchers have also pointed to both perceived and experienced stress among sports officials as variables of distraction (e.g., Abernethy, Maxwell, Masters, van der Kamp, & Jackson, 2007; Rainey, 1994; Stewart, Ellery, Ellery, & Maher, 2004). Participants in both studies reported experiencing mild to moderate stress with frequent spikes in high stress. Conversely, given the number of pitches thrown during a nine inning baseball game, all of which the home plate umpire is expected to judge correctly, an umpire may become susceptible to various forms of judgment bias (Larson, et al., 1991; MacMahon & Starkes, 2008).

Objective and Subjective Judgment and Decision Making Among Sport Officials

Sport judges, referees, and other game officials are an essential element of competition and performance in sport. Their task is to objectively judge the application and enforcement of playing rules, recognize penalties or game infractions, as well as manage the collective flow and progression of a game or sporting event (Bar-Eli, Plessner, & Raab, 2011). Ste-Marie (2003) argues that the characteristics found in expertise studies involving athletes can be extrapolated and applied to the domain of sport judgment and officiating. This is important, as there are far fewer reviews of the domain-specific expertise demonstrated by sport judges, referees, and game officials.

Knowledge and Enforcement of Playing Rules

Umpires and other sport judges and referees are responsible for maintaining a comprehensive, working knowledge of the playing rules of their domain. Utz (1989) likens the rules of baseball to the rules of law as they both require, forbid, and permit people from behaving in various ways. There are 10 different rules in the Official Baseball Rules that outline and meticulously describe various divisions of the game (The Office of the Commissioner of Baseball, 2012). Rules one through five outline the objectives of the game including the regulation parameters of the playing field, definition of important terms, instruction on regulations related to the starting and ending of a game, as well as information on the concepts of a ball which is either in play (i.e., live ball) or not in play (i.e., dead ball) as well as instruction on how a dead ball should be put back in play.

Rules six through eight specifically address the batter, runner, and pitcher respectively. Each rule is thoroughly detailed as to what constitutes legal or illegal action as it relates to a game. Many of the rules and game action are definitive and often do not require subjective judgment on the part of an umpire. For example, when a base runner collides with a fielder who is attempting to make a play on live ball he is guilty of obstruction and, by rule is called out and any other members of the offensive team occupying a base are required to return to the base they occupied at the time of the pitch immediately before the infraction occurred (Rule 7.06[a]). This could be considered an objective ruling. Little interpretation is necessary in order to judge the action described.

Other rules require umpires to exercise subjective judgment and incorporate prior knowledge with game-action interpretation to form an appropriate response. For example, Rule 6.05(k) describes the parameters of the legal width of a base runner's running lane. Umpires are required to subjectively judge any potential violation of this rule during game-action, which is happening at a very fast pace. The rule describes that a batter is out when:

> In running the last half of the distance from home base to first base, while the ball is being fielded to first base, he runs outside (to the right of) the three-foot line, or inside (to the left of) the foul line, and in the umpire's judgment in so doing interferes with the fielder taking the throw at first base, in which case the ball is dead; except that he may run outside (to the right of) the three-foot line or inside (to the left of) the foul line to avoid a fielder attempting to field a batted ball;
>
> Rule 6.05(k) Comment: The lines marking the three-foot lane are a part of that lane and a batter-runner is required to have both feet within the three-foot lane or on the lines marking the lane. The batter-runner is permitted to exit the three-foot lane by means of a step, stride, reach or slide in the immediate vicinity of first base for the sole purpose of touching first base. (The Office of the Commissioner of Baseball, 2012, p. 50)

Rule 9 of the Official Baseball Rules (The Office of the Commissioner of Baseball, 2012) is devoted entirely to umpires. In this section of the rulebook, the overall jurisdiction for the umpiring crew is described and outlines the authority with which the umpires govern a contest. The rulebook also provides several amusing, yet practical tips for umpires to keep in mind throughout the course of a game. For example, Rule 9.05–General Instructions to Umpires is intended to remind umpires to:

> Keep your eye everlastingly on the ball while it is in play. It is more vital to know just where a fly ball fell, or a thrown ball finished up, than whether or not a runner missed a base. Do not call the plays too quickly, or turn away too fast when a fielder is throwing to complete a double play. Watch out for dropped balls after you have called a man out. (The Office of the Commissioner of Baseball, 2012, p. 83)

Physical and Quantitative Properties in Professional Baseball

The speed at which professional baseball is played may at times act as an external distractor for home plate umpires whose job is to make ball-strike decisions on pitches which can travel upwards from 90 miles per hour. For example, after reviewing data from the 2011 MLB season, pitchers threw 5,383 fastballs at an average starting velocity of 93.4 miles per hour. The amount of time it takes a baseball to travel the 60.5 feet from the pitcher's mound to the front edge of home plate is approximately 400 milliseconds (Adair, 2002). The primary task of a home plate umpire is to visually track the baseball as it leaves the pitcher's hand and travels toward home plate. If unhit by the batter, once the ball reaches the catcher's glove the umpire must then make a determination as to whether or not the pitch crossed through the strike zone.

Conversely, the task of a pitcher is to combine a series of pitches ranging in type and velocity so as to prevent a batter from either predicting which type of pitch the pitcher will throw, or his ability to accurately time his swing with the arrival of the pitched ball (Adair, 2002). Generally, professional pitchers use four types of pitches. The most common is a four-seam fastball, which is designed to reach maximum velocity while providing the pitcher with the ability to better control the ball's trajectory. The two-seam fastball is designed to achieve relatively identical velocity as the four-seam, but feature slightly less predictability in its trajectory. Depending on the grip, this pitch from a right-handed pitcher will approximately move one to three inches horizontally either toward or away from a batter depending on his handedness (Adair, 2002).

A curveball travels at significantly slower speeds compared to a fastball. In addition, the curve ball is designed to approach the batter on a plane similar to a fastball, however as it reaches home plate it falls downward by up to 14 inches (Adair, 2002). A changeup is another pitch typically thrown by a pitcher often called an off-speed pitch. The change up is designed to be released from approximately the same point as a fastball, on approximately the same plane, yet travel significantly slower than an average fastball. This type of pitch can be effective when a batter predicts the pitcher will throw him a fastball and is prepared to time his swing in conjunction with the average velocity. However, because the changeup travels much slower, the batter, in this case, will likely initiate his swing too early and therefore miss the ball (Adair, 2002).

When we consider the primary task of home plate umpires, the difficulty in determining whether a pitched ball traveled through the strike zone should be apparent. The strike zone is defined as "...that area over home plate the upper limit of which is a horizontal line at the midpoint between the top of the shoulders and the top of the uniform pants, and the lower level is a line at the hollow beneath the kneecap. The strike zone shall be determined from the batter's stance as the batter is prepared to swing at a pitched ball" (The Office of the Commissioner of Baseball, 2012). From this definition, we can attempt to parameterize both the edges and upper and lower limits of the strike zone. The width of a regulation-size home plate is 17 inches (The Office of the Commissioner of Baseball, 2012). Therefore, according to baseball rules, the strike zone is uniform in its width. The upper and lower limits, however, are less definitive.

Because the upper and lower limits of the strike zone are determined once the player assumes his stance in the batter's box and is prepared to swing, we can deduce that what constitutes the top and bottom of the strike zone will vary significantly according to players' height and stance mechanics. In theory, this makes differentiating between pitches inside and outside of the strike zone considerably more complicated for the home plate umpire. Furthermore, an umpire will observe on average 291 pitches per regulation-length nine-inning game. Of those 291, the umpire is responsible to make the sole decision on the pitch's location on approximately 149 of them. Adding to the difficulty in differentiating between pitches located inside and outside of the strike zone is the fact that the umpire will typically crouch down before every pitch behind the catcher in a position where he can view the pitched ball through the area between the batter's body and the catcher's stationary position. In the pre-pitch state, umpires are surrounded by a myriad of contextual and domain-related information, which can be used to construct potential sequences of forthcoming action as well as any other preparations that might need to be made such as positioning behind the catcher.

With the advent of pitch tracking technology in use by MLB in 2007, academic and amateur researchers are able to quantitatively measure and evaluate not only the parameters of the professional strike zone, but also the types of pitches thrown, the subsequent location of the pitch, as well as the umpire's decision. The *PITCHf/x* technology, as it is known, is the proprietary creation of California-based sports technology company SportVision. A series of camera and video monitoring systems are installed in all 30 MLB stadiums, which can quantitatively represent, within 0.5 inches, a number of variables generated with each pitch thrown (Nathan, 2008). Subsequently, MLB makes all of the data generated by the *PITCHf/x* system freely available for download from an online XML directory.

Since 2007, a number of reports from baseball and sport analytics websites have been issued from various individuals who are either employed in the professional baseball industry or are enthusiasts with the requisite skill and technical aptitude to analyze statistical data (e.g., Fast, 2011a, 2011b, 2011c, 2011d; Gassko, 2007; Goldblatt, 2011; Hale, 2007; Walsh, 2007;). Several reports have briefly critiqued the parameters of the MLB strike zone as well as the accuracy rates and tendencies demonstrated by umpires. For example, Fast (2011a) addressed the topic of individual batter zones based on the variability in the top and bottom of each player's strike zone. He reviewed the strike zone as determined by the frequency of calls made by home plate umpires on a number of batters whose height represent the upper and lower maximum and minimum respectively. He concluded that, across all batters, there may be a typical strike zone area from which to evaluate umpire accuracy.

Similarly, Walsh (2010) analyzed the MLB strike zone given different ball-strike counts against the batter. His findings demonstrated that in a three ball no strike count, umpires are more likely to call a pitch a strike, even when the pitch is normally outside of the strike zone parameters. Conversely, when a batter is faced with a zero ball two strike count, an umpire is more likely to call a pitch a ball, even when the pitch is marginally close to one of the four edges of the strike zone. According to Walsh (2010), this demonstrates a tendency among umpires to display bias toward a player when he is theoretically at a disadvantage.

In a review of MLB umpire performance between 2007 and 2010, Goldblatt (2011) ascribed a variety of traditional baseball statistics to umpires. For example, the R/9, BB/9, K/9, and the K/BB are all statistics most commonly associated with pitchers to reflect the number of runs allowed per nine-innings, the number of walks allowed per nine-innings, the number of strikeouts recorded per nine-innings, and the ratio of strikeouts to walks, respectively. In the case of MLB umpire number 20, according to Goldblatt's calculation, teams scored on average 9 runs per nine-innings, walked 6 batters per nine-innings, and struck out 13 batters per nine-innings.

Rationale for the Current Study

More empirical research is needed in the area of MLB umpire performance evaluation. As previously noted, an abundance of literature related to the acquisition of expertise among sport performers and perceptual-cognitive skills employed by sport performers is available. However, a comparable body of knowledge related to the baseball umpire population is needed. With the availability of *PITCHf/x* data, there has been a recent trend in scholarly contributions related to the effects of MLB umpires' ball–strike decisions (e.g., Marchi and Albert, 2014; Mills, 2014). Thus, the current study will add to this growing body of literature by statistically modeling the effects of pitch location and count on MLB umpires' ball–strike decisions.

It is clear that umpires, sport judges, and officials are responsible for a variety of often intricate and subtle playing rules. In addition, these individuals are expected to objectively and without bias maintain governance of the contest or performance. Likewise, sport judges utilize a variety of perceptual, cognitive, and visual skills to perform at a peak level. Similar to the research previously described related to the development and acquisition of domain-specific expertise, umpires and other referees typically spend many years developing the requisite skills and knowledge base necessary for elite levels.

Research Questions

The purpose of the current study is to answer the following research questions related to the effects of pitch location and count on professional baseball umpires' ball–strike decisions during the course of an entire MLB season. To do so, the current study uses data generated by the *PITCH/fx* (Sportvision, Inc., 2013) pitch tracking system:

(1) Do MLB umpires differ in their accuracy of judging pitch outcomes?

(2) Is pitch location associated with MLB umpires' accuracy of judging pitch outcomes?

(3) Is the ball–strike count associated with MLB umpires' accuracy of judging pitch outcomes?

(4) After controlling for ball–strike count, is pitch location associated with MLB umpires' accuracy of judging pitch outcomes?

(5) After controlling for pitch location, is ball–strike count associated with MLB umpires' accuracy of judging pitch outcomes?

(6) Is umpire expertise (as measured by years of experience) associated with MLB umpires' accuracy of judging pitch outcomes?

(7) Does umpire expertise moderate the relationship between the accuracy of judging pitch outcomes and either pitch location or ball–strike count?

Hypotheses

The following hypotheses were stated in conjunction with the research questions above. The principle hypothesis directing the current study is that umpires should not greatly effect pitch outcomes, and, instead, be interchangeable.

(1) MLB umpires will not differ in their accuracy of judging pitch outcomes.

(2) Pitch location will be associated with MLB umpires' accuracy of judging pitch outcomes. Specifically, compared to pitches located within the middle and outer regions of the strike zone, umpires will demonstrate less accurate decisions when pitches are located within the inner region of the strike zone.

(3) Ball–strike count will be associated with MLB umpires' accuracy of judging pitch outcomes. Specifically, compared to neutral and batter-favored ball–strike counts, umpires will demonstrate less accurate decisions when the ball–strike count favors the pitcher.

(4) After controlling for ball–strike count, pitch location will be associated with MLB umpires' accuracy of judging pitch outcomes. Specifically, umpires will demonstrate greater accuracy when pitches are located either in the middle or outer regions of the strike zone.

(5) After controlling for pitch location, ball–strike count will be associated with MLB umpires' accuracy of judging pitch outcomes. Specifically, umpires will demonstrate greater accuracy when ball–strike counts are either neutral or favor the batter.

(6) Experience in MLB will be associated with umpires' accuracy of judging pitch outcomes. Specifically, compared to umpires with less experience, umpires with more experience will demonstrate less accurate decisions.

(7) Umpire experience will not moderate the association between umpire accuracy and either pitch location or ball–strike count.

CHAPTER THREE

Methods

Design

Secondary Data Analysis Methods

Using secondary data allows researchers to investigate a range of phenomena within populations they may not otherwise be able to access (Andersen, Prause, & Silver, 2011). For example, educational, health-related, and survey data collected from a variety of large-scale federal research projects are made available for researchers. These projects may contain relatively large, nationally-representative samples, which are beneficial for investigators interested in the generalizability of their research (Andersen et al., 2011; Stewart, 2012).

Despite its advantages, there are several factors to consider when utilizing secondary data. For example, knowing the variables actually collected, how the data were collected, and when can influence the methods with which researchers pose questions and investigate hypotheses (Stewart, 2012). Further, when multiple sources of related secondary data are available, effort should be taken to evaluate the consistency of the information across sources. Secondary data sets may also contain missing data, which can be problematic for researchers (Andersen, et al., 2011).

PITCHf/x Data

Data for the current study are generated by a camera system, known as *PITCHf/x* (Sportvision, Inc., 2013), installed inside each MLB team's stadium for the purpose of capturing and measuring the flight path of pitched baseball trajectories (Nathan, 2008; Sievert, 2014). Cameras track the flight of each baseball beginning with the point at which the pitcher releases the ball as he is positioned on the pitching rubber and ending with the point at which the pitch passes over the front edge of home plate. The regulation distance between the front edge of the pitching rubber and the front edge of home plate is a uniform 60 feet 6 inches (MLB, 2014). With each pitch the *PITCHf/x* system tracks, it generates 50 variables containing information unique to all pitcher–batter match ups over all ball–strike counts for both teams in all innings of play. *PITCHf/x* data are available in Extensible Markup Language (XML) format from the MLB Advanced Media (2014) Gameday application programming interface.

PITCHf/x data consist of all pitch-outcome decisions made by home plate umpires over the course of a single season. Umpires typically are responsible for calling pitches behind home plate every fourth game due to the rotation system employed among the four-person umpire crews assigned to officiate the game. Therefore, an umpire's decisions may be accumulated over the course of 20 or more games in a single season. An umpire may observe approximately 300 pitches during the course of a 9-inning game. However, not all pitches require a decision from the home plate umpire.

Data collection, cleaning, and classification. *PITCHf/x* data for the current study were collected using the pitchRx package (Sievert, 2014) in the R statistical computing software application (R Core Team, 2014) representing all pitches delivered over the course of the 2013 season. No missing values were present in the data set. pitchRx contains functions designed to access the Gameday XML data and store it in a SQL database. XML data are structured around a set of encoding rules which enable users to parse information according to identifiers, or variables (Agreste, De Meo, Ferrara, & Ursino, 2014). R syntax used to construct the *PITCHf/x* database to be used in the current analysis is available in Appendix B.

A total of 754,362 pitches were delivered to batters during the course of the 2013 MLB season. Several steps were taken during the data collection phase in order to prepare the data for the current analysis. First, all pitches not requiring the home plate umpire to decide the outcome were removed. These include all pitches in which the batter swung and missed, hit the ball into foul territory, hit the ball into fair territory, or was struck by the pitch.

Second, I calculated batters' average upper and lower strike zone limits. This involved creating new variables for the season-mean top and season-mean bottom parameters which were then used as reference values when determining whether a pitch was located inside our outside of a batter's strike zone. Much of the process for collecting *PITCHf/x* data is automated by the camera system installed in each stadium. One exception is adjusting each batter's unique strike zone parameters. A batter's strike zone can be defined as an imaginary rectangular area above home plate that varies in length given a batter's height and batting stance. Thus, the top and bottom parameters of each batter's strike zone are manually adjusted by a *PITCHf/x* operator at each MLB stadium.

Third, I corrected the width of the strike zone. Home plate measures 17" wide across the front portion. However, Rule 2.00 of the MLB Rule Book defines a strike as occurring "if any part of the ball passes through any part of the strike zone" (The Office of the Commissioner of Baseball, 2012, p. 21). Since the *PITCHf/x* system collects coordinates of pitches by using the center point of the baseball as it crosses the front edge of home plate, the length of the radius of a regulation size baseball was added to each side of home plate. This expands the uniform width of the strike zone from 17" to 17" + (1.57"×2) = 20.14". Figure 3.1 displays the transformed width of the strike zone area.

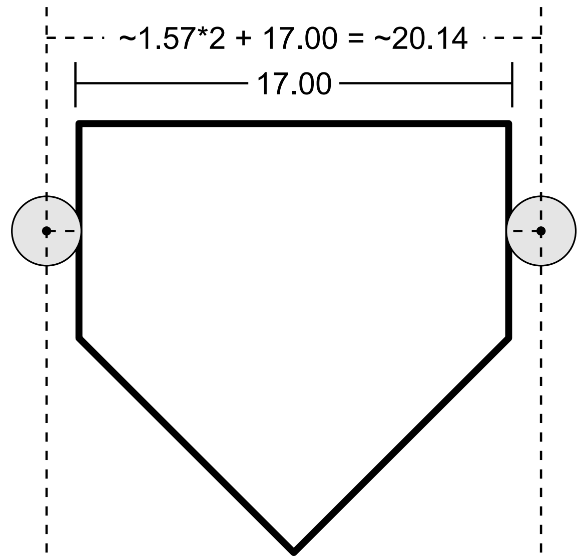


Figure 3.1: Strike zone width after adding the radius of a regulation size baseball to both sides.

Variables

The current analysis includes all pitches from the 2013 MLB season in which the home plate umpire had to judge if it was a ball or strike. All variables for the current study are provided in Table 3.1 along with with their corresponding definitions. During the game the home plate umpire positions himself behind the catcher, who is positioned behind or in the vicinity of home plate, in order to view pitches delivered by the pitcher. One of the home plate umpire's primary responsibilities is to judge whether an unhit ball delivered by the pitcher passed through the batter's strike zone or was located outside of the batter's strike zone. Depending on the outcome of the pitch, either a ball is awarded to the batter or a strike is awarded to the pitcher. The accumulation of balls and strikes in a single plate appearance is known as the ball–strike count.

Table 3.1: *Variables Used in Analysis.*

Variable Name	Definition	Level of Analysis
Ball–strike count	Specifies whether the ball–strike count favors the batter, pitcher, or is neutral	Level-1
Strike zone location	Region of the zone in which pitch was located given batter handedness	Level-1
Umpire	Randomly assigned umpire ID number	Level-2
Umpire experience	Home plate umpire's experience in number of seasons of full-time employment in MLB	Level-2
Umpire decision	Dichotomous outcome indicating whether the home plate umpire's decision was correct or incorrect	Level-1

Note. Level of analysis refers to the location of each variable within the multilevel, or nested, data structure. Level-1 variables are those at the pitch level in the data, while Level-2 variables are those at the umpire level.

Pitch-Related Variables

Two variables in the current study are related to pitches: (a) ball–strike count and (b) pitch location. Among the pitch-related variables available in the *PITCHf/x* data, ball–strike count and pitch location are hypothesized as being related to umpire decisions. Previous research has demonstrated evidence for the presence of effects of pitch-related variables on umpire decisions (Fast, 2011).

There are a maximum of 12 ball-strike combinations that can occur in any given batter-pitcher matchup. With each plate appearance, the batter and pitcher start with a 0 ball 0 strike count. The maximum number of balls pitchers are allowed to throw before a batter is awarded first base is four, while the maximum number of strikes a batter is allowed to obtain before striking out is three. This variable was transformed according to Marchi and Albert's (2014) coding method.

Marchi and Albert (2014) calculated expected run values for each of the 12 ball–strike counts in order to determine which were most likely to favor batters, which were most likely to favor pitchers, and which were most likely to be neutral. Table 3.2 displays the ball–strike count combinations and corresponding advantage classification. The process of transforming the 12 ball–strike count possibilities into one of three advantage–disadvantage scenarios (i.e., batter advantage, pitcher advantage, neutral) is common among baseball analytics research (e.g., Tango, Lichtman, & Dolphin, 2007; Thorn & Palmer, 1984). The ball–strike count predictor variable was categorized for the current study as: neutral, batter-, and pitcher-advantaged counts. Pitches delivered during neutral counts were treated as the reference group.

Table 3.2: *Ball–Strike Count by Advantage Classification.*

Ball–Strike Count Advantage		
Batter	Pitcher	Neutral
2-0	0-1	0-0
3-0	0-2	1-0
3-1	1-2	1-1
3-2	2-2	2-1

Note. Batter: pitches delivered to batters in which the ball–strike count favors the batter; Pitcher: pitches delivered to batters in which the ball–strike count favors the pitcher; Neutral: pitches delivered to batters in which the ball–strike count favors neither the batter nor the pitcher.

The locations of pitches tracked by the *PITCHf/x* system are bounded within one of several finite regions of the strike and ball zone areas around home plate. Conceptually, bounding pitches within certain regions of the strike and ball zone areas allows users of the *PITCHf/x* data to more easily classify pitch locations. For example, researchers interested in determining a particular batter's tendency to swing at certain pitches within the regions of the strike zone can take advantage of this classification system in order to model performance predictions and graphically illustrate offensive trends (Marchi and Albert, 2014).

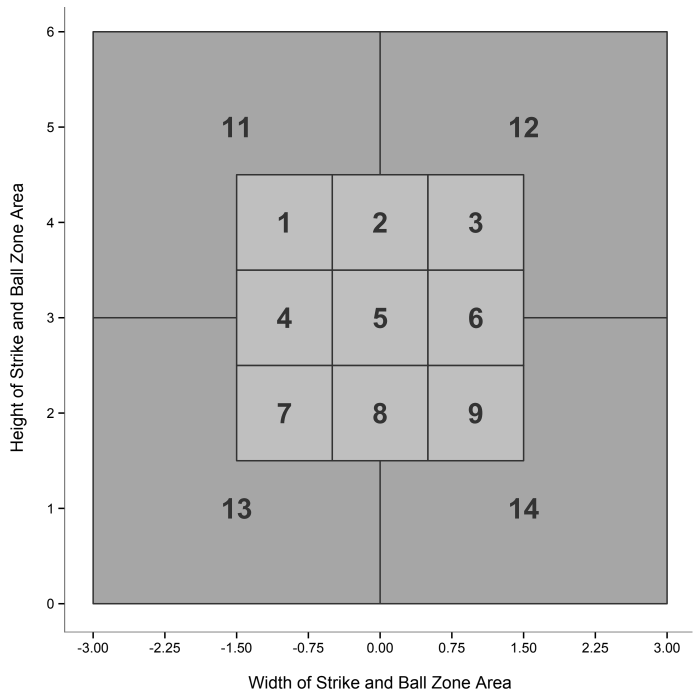
The pitch location variable indicates the region within the batter-relative strike zone area where each pitch crossed the front edge of home plate. The raw *PITCHf/x* zone variable divides the area of each batter's strike zone into nine regions. For the current study, pitch location was transformed into four regions to indicate whether the pitch was located within either the inner, middle, outer, or ball portions of the batter's strike zone by comparing the batter's handedness with the raw *PITCHf/x* location data. Figure 3.2a provides an example of the nine untransformed *PITCHf/x* zone regions.

Figure 3.2b provides an example of the pitch location variable transformation, taking into account batter handedness. From the umpire's vantage point, pitches located in zones 1, 4, and 7 were classified as being in the inner region for right handed batters and outer region for left handed batters. Pitches located in zones 2, 5, an 8 were classified as being in the middle region for both right and left handed batters. Pitches located in zones 3, 6, and 9 were classified as being in the outer region for right handed batters and inner region for left handed batters. There is no zone labeled 10 in the *PITCHf/x* data. Pitches located in zones 11, 12, 13, and 14 were classified as being the ball region. The pitch location predictor variable was categorized as: ball, inner, middle, and outer transformed regions of the strike zone. Pitches located in the ball region were treated as the reference group.

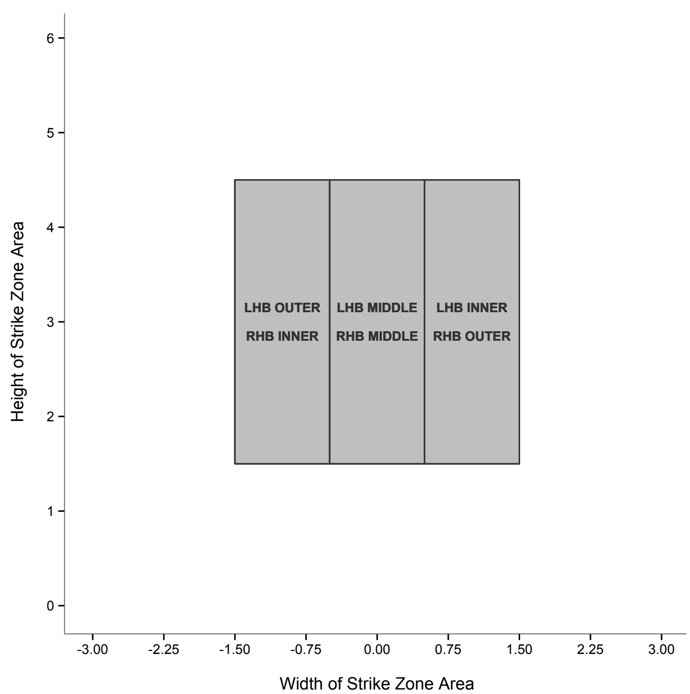
Umpire-Related Variables

Three variables in the current study are related to umpires: (a) the home plate umpire in each game; (b) the home plate umpire's decision, or call, on each pitch; and (c) the home plate umpire's full-time experience in years. As data from the same umpires are collected throughout the 2013 season, I assigned each umpire a random integer between 1100 and 1175 to identify them in the dataset.

The outcome variable, umpire decision, was dichotomized to indicate whether the home plate umpire's decision on each pitch was correct or incorrect. In order to determine home plate umpire accuracy, the coordinate-location and batter-relative strike zone parameters of each pitch were compared against the each umpire's



(a) Untransformed *PITCHf/x* strike and ball regions.



(b) Transformed *PITCHf/x* strike zone regions

Figure 3.2: *PITCHf/x* strike and ball zone regions from the umpire's perspective.

original call. Decisions made by home plate umpires who call a strike when a pitch is located outside of the strike zone parameters are considered incorrect, while decisions made by umpires who call a strike when a pitch is located inside the strike zone parameters are considered correct. Similarly, decisions made by home plate umpires who call a ball when a pitch is located outside of the strike zone parameters are considered correct, while decisions made by home plate umpires who call a ball when a pitch is located inside the strike zone parameters are considered incorrect.

Umpire experience is measured in years of full-time service in MLB prior to the start of the 2013. Once hired, umpires traditionally are promoted to MLB after several years of service through varying levels of Minor League Baseball (MiLB) (Weber, 2010). Umpires can spend between 5 and 10 years employed in MiLB before they are hired in MLB on a full-time basis. Once umpires reach the highest level of play in MiLB, they may be added to an active MLB reserve roster and be temporarily promoted to fill-in for full-time MLB umpires who are either injured or on leave. Umpire experience was coded as 0 for umpires in their first full-time season. For all other umpires, experience was treated as the number of full-time years completed in MLB.

Umpire experience was grand-mean centered and standardized when included in the model building process. This was done by subtracting the umpire experience sample mean (12.02) from each umpire's number of years experience, then dividing by the umpire experience sample standard deviation (8.42). This results in values which indicate the number of standard deviations an umpire's experience is above or below the grand-mean. For example, the grand-mean umpire experience in the sample was approximately 12 years with a standard deviation of approximately 9 years. Thus, for an umpire with $12 - 9 = 3$ years experience, his grand-mean centered and scaled experience would be -1. In terms of model estimation and interpretation, the effect of umpire experience can be interpreted as the resulting change in the predicted umpire accuracy after a one standard deviation change in umpire experience.

Data Analysis

The purpose of the current study is to answer a set of research questions related to the effects of pitch location and count on professional baseball umpires' ball–strike decisions during the course of an entire MLB season. To do so, the current study uses data generated by the *PITCHf/x* (Sportvision, Inc., 2013) pitch tracking system.

Model Specification

Data in the social and behavioral sciences frequently contain a multilevel, or nested, structure in which observations at a lower level (level-1) are grouped within units at a higher level (level-2) (Heck, et al. 2012; Gelman & Hill, 2007; Raudenbush & Bryk, 2002). Common examples include students who are nested within classrooms and employees who are nested within departments or organizations (Heck et al., 2012). As a result, a multilevel model (MLM) can account for the effects of predictor variables present at successive levels of data (Snijders & Bosker, 2012). Moreover, compared to traditional regression models, MLMs allow for both fixed and random coefficients (Raudenbush & Bryk, 2002). With fixed coefficients, an effect is the same across all observations; with random coefficients, however, an effect can vary across observations (Hox, 2010).

Multilevel data structures are suitable when questions of interest involve examining separate observations on some outcome which are nested within groups, classes, or, in the current study, umpires (Heck, et al., 2012). Data in the current study treat umpire decisions as repeated measurements nested within the umpire making the decision. Although time is not treated as a factor in the current study, MLMs are still appropriate for analyzing questions from data in which observations are clustered within individuals (Heck, et al., 2012; Hoffman, 2007). Pitches were treated as the separate measurements or observations at level 1 made by individual umpires at level 2. Umpire experience is located at level 2 in the multilevel modeling framework.

Multilevel models with a categorical repeated measures outcome. MLMs with continuous outcome variables assume (a) a linear relationship between regression coefficients, (b) that random effects are normally distributed (Raudenbush & Bryk, 2002). However, MLMs are not limited to continuous outcome variables. Models with categorical outcomes may also be estimated (Hox, 2010).

In the case of MLMs with categorical outcome variables, the assumptions of linearity and normality are not realistic (Raudenbush & Bryk, 2002). Additionally, observations in MLMs with categorical outcomes are assumed to be discrete. For example, the outcomes in the current study–umpire decisions–are dichotomous (i.e., correct or incorrect). As a result, a nonlinear, transformation is used. For the current study, I use the logit transformation, which is the natural logarithm (log) of the odds that an umpire's decisions is correct. Further specification of the logit transformation is given below. Likewise, in models with a dichotomous outcome, differences between predicted and observed values are assumed to be non-normally distributed and have unequal variance.

MLMs are often developed using a hierarchical sequence of models. The first model is typically a null model, which includes no predictor variables at either level. Subsequent models are then fit that include one or more variables at level-1 or level-2 (Heck et al., 2012; Raudenbush & Bryk, 2010). In general, a two-level multilevel models with a dichotomous outcome and no predictor variables takes the following form:

$$\text{Level-1:} \quad \eta_{ij} = ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0j}, \tag{3.1}$$

where η_{ij} is the predicted logit for the outcome event i within individual j, ln is the natural logarithm of the odds for the outcome event, π_{ij} is the probability that the outcome event occurs, $1 - \pi_{ij}$ is the probability of the outcome event not occurring, and β_{0j} is the model intercept term. The value inside the parentheses in Equation 3.1 is the odds, which is the probability of an event by the probability of the event no occurring (i.e., $\pi/1 - \pi$).

The level-1 equation (3.1) omits a residual, or error, variance term (σ^2) because residuals from MLMs with a dichotomous outcome can only take on two values: correct or incorrect. Thus, the residual variance is fixed (Heck et al., 2012). A common fixed value for the level-1 residual variance is 3.29, which is approximately the variance of the standard logistic distribution (Hox, 2002).

The level-2 equation is:

$$\text{Level-2:} \quad \beta_{0j} = \gamma_{00} + u_{0j}, \tag{3.2}$$

where γ_{00} is the intercept term. The variance parameter (u_{0j}) allows the level-1 intercept (β_{0j}) to vary between observations. When u is fixed at zero, the level-1 intercept is a fixed effect (β_0); when u is estimated, the level-1 intercept is a random effect (β_{0j}).

Replacing β_{0j} in the level-1 model with its definition in the level-2 model produces a combined model:

$$\text{Combined model:} \quad \eta_{ij} = \gamma_{00} + u_{0j}, \tag{3.3}$$

where η_{ij} represents the predicted logarithm of the odds (i.e., logit) resulting from the MLM (Heck, et al., 2012).

Interpreting Multilevel Models with Dichotomous Outcomes. Estimated coefficients from MLMs with a dichotomous outcome can be interpreted in several ways. Heck et al. (2012) noted that because logits are in units of log odds, they are difficult to interpret directly. Consequently, transforming them into an odds ratio can aid in interpretation. Odds ratios can be obtained by exponentiating (i.e., the inverse of the log function) the MLM coefficients (e.g., β, γ).

In the current study, odds ratios express the change in the odds of an umpire making a correct decision associated with a unit change in the predictor X, while controlling for any other variables in the MLM (Heck et al., 2012). The general formula for an odds ratio (OR) is given in Equation 3.4:

$$OR = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}, \tag{3.4}$$

where π_1 is the probability of success in condition 1 (i.e., the predictor variable is specified at one value) and π_2 is the probability of success in condition 2 (i.e., the predictor variable is specified at another value). Odds ratios range from 0 to infinity, with a value of 1 indicating equal odds of the event occurring under both conditions. Odds ratios greater than 1 indicate that the outcome has a higher odds of occurring in condition 1 than condition 2, while odds ratios less than 1 indicate that the outcome has a lower odds of occurring in condition 1 than condition 2 (Szumilas, 2010).

In addition to odds ratios, MLM coefficients can also be transformed into probabilities. For the current study, the probability is of an umpire making a correct decision (Grimes & Schulz, 2008). Whereas each coefficient in the MLM can be transformed to an OR, their is only one probability value for a given model.

MLM coefficients can be transformed to the probability metric using Equation 3.5:

$$\hat{\pi} = \frac{exp(\beta_0)}{1 + exp(\beta_0)}, \tag{3.5}$$

where $\hat{\pi}$ is the predicted probability of an accurate umpire decision, exp is the exponential function, and β_0 is the estimated intercept value from Equation 3.1. For models with predictor variables, the predicted probabilities are calculated by summing across all terms in the MLM. For a MLM with an intercept and a single predictor variable, the predicted probability of a correct umpire decision can be calculated using Equation 3.6:

$$\hat{\pi} = \frac{exp(\beta_0 + \beta X)}{1 + exp(\beta_0 + \beta X)}, \tag{3.6}$$

where β is the MLM coefficient associated with the predictor variable X.

Models Used in Current Study

Model 0: Null model. This model examines the overall accuracy of the umpires' calls across all umpires, games, and calls. No level-1 or level-2 variables are included in the null model, so the only coefficient estimated is a fixed level-1 intercept. The value of the null model intercept can be interpreted as the predicted log odds that a randomly selected umpire's decision is accurate (Heck et al., 2012). The null model equation is specified as:

$$\eta_{ij} = \beta_0, \tag{3.7}$$

where β_0 is the estimated intercept.

Model 1: Fully unconditional model. In this model, accuracy is allowed to vary across umpires. The estimated model intercept coefficient is interpreted as the predicted log odds that a given umpire's decision is correct across all games and calls.

At level-1, the unconditional model is specified as:

$$\eta_{ij} = \beta_{0j}, \tag{3.8}$$

where β_{0j} is the intercept term for umpire j.

At level-2, the presence of a variance parameter (u_{0j}) allows the level-1 intercept (β_0) to vary between umpires:

$$\beta_{0j} = \gamma_{00} + u_{0j}. \tag{3.9}$$

u_{0j} is assumed to follow a normal distribution with mean 0 and a constant variance (τ_{00}). Substituting equation 3.9 into equation 3.8 produces the following combined model:

$$\eta_{ij} = \gamma_{00} + u_{0j}. \tag{3.10}$$

Allowing accuracy to vary across umpires allows for an examination of how much of the total proportion of variance in decision accuracy is explained by differences between umpires. One way to estimate this proportion is to calculate an intraclass correlation (ICC), although Hox (2010) noted that ICC values should be interpreted with caution in MLMs–especially those with dichotomous outcomes. The ICC formula in this MLM framework is:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}, \tag{3.11}$$

where τ_{00} is the population variance estimate for umpires and σ^2 is the fixed variance estimate for level-1 units (Raudenbush & Bryk, 2002).

Model 2: Effects of pitch location on umpire decisions. This model examines the association of pitch location and umpire decision accuracy. The level-1 model includes the pitch location predictor variable, with the ball region of the strike zone as the reference category. In this model, interpretation of the estimated coefficient for each pitch location variable is made in comparison to the ball region. In addition, pitch location is treated as a fixed effect while the intercept is random. Thus, accuracy is allowed to vary by umpire, but pitch location has the same affect on accuracy for all umpires. The level-1 model is characterized as:

$$\begin{aligned}\eta_{ij} = \beta_{0j} &\\ &+ \beta_1(\text{PITCH LOCATION} = \text{INNER}_{ij}) \\ &+ \beta_2(\text{PITCH LOCATION} = \text{MIDDLE}_{ij}) \\ &+ \beta_3(\text{PITCH LOCATION} = \text{OUTER}_{ij}).\end{aligned} \tag{3.12}$$

The level-2 model is characterized as:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30}.\end{aligned} \tag{3.13}$$

Substituting Equation 3.13 into Equation 3.12 results in the following combined model:

$$
\begin{aligned}
\eta_{ij} = \gamma_{00} & \\
& + \gamma_{10}(\text{PITCH LOCATION} = \text{INNER}_{ij}) \\
& + \gamma_{20}(\text{PITCH LOCATION} = \text{MIDDLE}_{ij}) \\
& + \gamma_{30}(\text{PITCH LOCATION} = \text{OUTER}_{ij}) + u_{0j}.
\end{aligned}
\tag{3.14}
$$

Model 3: Effects of ball–strike count on umpire decisions. This model examines the association of ball–strike count and umpire decision accuracy. The level-1 model includes the ball–strike count predictor variable, with neutral counts as the reference category. In this model, interpretation of the estimated coefficient for each level of the ball–strike count variable is made in comparison to pitches delivered during neutral counts. In addition, pitch location is treated as a fixed effect while the intercept is random. Thus, accuracy is allowed to vary by umpire, but ball–strike count has the same affect on accuracy for all umpires. The level-1 model is characterized as:

$$
\begin{aligned}
\eta_{ij} = \beta_{0j} & \\
& + \beta_1(\text{COUNT ADVANTAGE} = \text{BATTER}_{ij}) \\
& + \beta_2(\text{COUNT ADVANTAGE} = \text{PITCHER}_{ij}).
\end{aligned}
\tag{3.15}
$$

The level-2 model is characterized as:

$$
\begin{aligned}
\beta_0 &= \gamma_{00} + u_{0j} \\
\beta_1 &= \gamma_{10} \\
\beta_2 &= \gamma_{20}.
\end{aligned}
\tag{3.16}
$$

Substituting Equation 3.16 into Equation 3.15 results in the following combined model:

$$\begin{aligned}\eta_{ij} = \gamma_{00} \\ &+ \gamma_{10}(\text{COUNT ADVANTAGE} = \text{BATTER}_{ij}) \\ &+ \gamma_{20}(\text{COUNT ADVANTAGE} = \text{PITCHER}_{ij}) + u_{0j}.\end{aligned} \tag{3.17}$$

Model 4: Effects of pitch location and ball–strike count on umpire decisions. This model estimates the association of pitch location and ball–strike count on umpire decision accuracy. The level-1 model includes two categorical pitch-related predictor variables: pitch location and ball–strike count advantage. With two predictor variables, interpretation is made by comparing the estimated coefficient to the reference category, controlling for the other variable. In Model 4, both pitch location and ball–strike count are treated as fixed effects. The level-1 model is characterized as:

$$\begin{aligned}\eta_{ij} = \beta_{0j} \\ &+ \beta_1(\text{PITCH LOCATION} = \text{INNER}_{ij}) \\ &+ \beta_2(\text{PITCH LOCATION} = \text{MIDDLE}_{ij}) \\ &+ \beta_3(\text{PITCH LOCATION} = \text{OUTER}_{ij}) \\ &+ \beta_4(\text{COUNT ADVANTAGE} = \text{BATTER}_{ij}) \\ &+ \beta_5(\text{COUNT ADVANTAGE} = \text{PITCHER}_{ij}).\end{aligned} \tag{3.18}$$

The level-2 model is characterized as:

$$\begin{aligned}
\beta_0 &= \gamma_{00} + u_{0j} \\
\beta_1 &= \gamma_{10} \\
\beta_2 &= \gamma_{20} \\
\beta_3 &= \gamma_{30} \\
\beta_4 &= \gamma_{40} \\
\beta_5 &= \gamma_{50}.
\end{aligned} \tag{3.19}$$

Substituting Equation 3.19 into Equation 3.18 results in the following combined model:

$$\begin{aligned}
\eta_{ij} = \gamma_{00} & \\
&+ \gamma_{10}(\text{PITCH LOCATION} = \text{INNER}_{ij}) \\
&+ \gamma_{20}(\text{PITCH LOCATION} = \text{MIDDLE}_{ij}) \\
&+ \gamma_{30}(\text{PITCH LOCATION} = \text{OUTER}_{ij}) \\
&+ \gamma_{40}(\text{COUNT ADVANTAGE} = \text{BATTER}_{ij}) \\
&+ \gamma_{50}(\text{COUNT ADVANTAGE} = \text{PITCHER}_{ij}) + u_{0j}.
\end{aligned} \tag{3.20}$$

Model 5: Effects of umpire experience on umpire decisions. This model examines the effects of umpire experience on umpire decision accuracy. As umpire-experience is a level-2 variable, there are no level-1 predictors included in Model 5. Thus, the level-1 equation for Model 5 is identical to the level-1 equation of Model 1:

$$\eta_{ij} = \beta_{0j}. \tag{3.21}$$

The level-2 model is characterized as:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{UMPIRE EXPERIENCE}_j) + u_{0j}. \tag{3.22}$$

Substituting equation 3.22 into equation 3.21 results in the following combined model:

$$\eta_{ij} = \gamma_{00} + \gamma_{01}(\text{UMPIRE EXPERIENCE}_j) + u_{0j}. \tag{3.23}$$

Model 6: Moderation of pitch location and ball–strike count by umpire experience on umpire decisions. This model examines if the relation between umpire accuracy and pitch location (level-1) or ball-strike count (level-1) changes as a function of umpire experience (level-2).

The level-1 model is:

$$\begin{aligned}\eta_{ij} = {} & \beta_{0j} \\ & + \beta_1(\text{PITCH LOCATION} = \text{INNER}_{ij}) \\ & + \beta_2(\text{PITCH LOCATION} = \text{MIDDLE}_{ij}) \\ & + \beta_3(\text{PITCH LOCATION} = \text{OUTER}_{ij}) \\ & + \beta_4(\text{COUNT ADVANTAGE} = \text{BATTER}_{ij}) \\ & + \beta_5(\text{COUNT ADVANTAGE} = \text{PITCHER}_{ij}).\end{aligned} \tag{3.24}$$

The level-2 model is characterized by:

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{UMPIRE EXPERIENCE}_j) + u_{0j} \\ \beta_1 &= \gamma_{10} + \gamma_{11}(\text{UMPIRE EXPERIENCE}_j) \\ \beta_2 &= \gamma_{20} + \gamma_{21}(\text{UMPIRE EXPERIENCE}_j) \\ \beta_3 &= \gamma_{30} + \gamma_{31}(\text{UMPIRE EXPERIENCE}_j) \\ \beta_4 &= \gamma_{40} + \gamma_{41}(\text{UMPIRE EXPERIENCE}_j) \\ \beta_5 &= \gamma_{50} + \gamma_{51}(\text{UMPIRE EXPERIENCE}_j) \\ \beta_6 &= \gamma_{60} + \gamma_{61}(\text{UMPIRE EXPERIENCE}_j).\end{aligned} \tag{3.25}$$

Substituting Equation 3.25 into equation 3.24, the following model is obtained, which is characterized by cross-level interactions between the level-1 pitch-related variables and the level-2 umpire experience predictor:

$$
\begin{aligned}
\eta_{ij} = \gamma_{00} &+ \gamma_{01}(\text{UMPIRE EXPERIENCE}_j) + u_{0j} \\
&+ \gamma_{10}(\text{PITCH LOCATION} = \text{INNER}_{ij}) \\
&+ \gamma_{11}(\text{UMPIRE EXPERIENCE}_j)(\text{PITCH LOCATION} = \text{INNER}_{ij}) \\
&+ \gamma_{20}(\text{PITCH LOCATION} = \text{MIDDLE}_{ij}) \\
&+ \gamma_{21}(\text{UMPIRE EXPERIENCE}_j)(\text{PITCH LOCATION} = \text{MIDDLE}_{ij}) \\
&+ \gamma_{30}(\text{PITCH LOCATION} = \text{OUTER}_{ij}) \\
&+ \gamma_{31}(\text{UMPIRE EXPERIENCE}_j)(\text{PITCH LOCATION} = \text{OUTER}_{ij}) \\
&+ \gamma_{40}(\text{COUNT ADVANTAGE} = \text{BATTER}_{ij}) \\
&+ \gamma_{41}(\text{UMPIRE EXPERIENCE}_j)(\text{COUNT ADVANTAGE} = \text{BATTER}_{ij}) \\
&+ \gamma_{50}(\text{COUNT ADVANTAGE} = \text{PITCHER}_{ij}) \\
&+ \gamma_{51}(\text{UMPIRE EXPERIENCE}_j)(\text{COUNT ADVANTAGE} = \text{PITCHER}_{ij}).
\end{aligned}
\tag{3.26}
$$

Software Used for Analysis

All statistical models were estimated using the `lme4` package (Bates, Maechler, Bolker, & Walker, 2014) in R.

Model Evaluation

Models were evaluated using measures of model fit. In general, lower values of model fit indices indicate a better fit to the data (Hosmer, Lemeshow, & Sturdivant, 2013). A common measure of model fit is the model deviance. The formula for calculating deviance is given in Equation 3.27:

$$Deviance = -2 \times ln(L), \tag{3.27}$$

where L is the log-likelihood of the model. The likelihood is the probability of obtaining the collected data if the model were true. For more information about likelihood estimation, see Enders (2005). Deviance values cannot be directly interpreted. Instead, they are only useful when comparing multiple models.

Another measure of model fit is the Akaike information criterion (AIC). (Hosmer, et al., 2013). The formula for calculating AIC is given in Equation 3.28:

$$AIC = -2 \times ln(L) + 2k, \tag{3.28}$$

where L is the log-likelihood of the model and k represents the number of predictors included in the model. Like deviance, AIC values are only useful when comparing multiple models. Lower values of AIC are preferred to larger values when determining model fit between competing models. Unlike deviance, the AIC imposes a penalty for model complexity–$2k$ term. Thus, AIC favors models that are parsimonious.

AIC values can be re-scaled (AIC Δ) to account for differences between models in sample size and arbitrary constants (Burnham, Anderson, & Huyvaert, 2011). The formula for calculating AIC Δ is given in Equation 3.29:

$$\text{AIC}\Delta_i = AIC_i - AIC_{min}, \tag{3.29}$$

where AIC_{min} is the minimum of the different AIC values for the compared models. Further, AIC Δ values can be re-scaled so they sum to 1, called AIC weights (w_i). The formula for calculating w_i is given in Equation 3.30:

$$w_i = \frac{\exp(-\text{AIC}\Delta_i/2)}{\sum_{i=1}^{M} \exp(-\text{AIC}\Delta_i/2)}, \tag{3.30}$$

where M_i is the particular model being evaluated. Values of w_i can be interpreted as the probability that M_i is the best model (Wagenmakers & Farrell, 2004).

Another measure of model fit is the Bayesian information criterion (BIC) estimate. The formula for calculating BIC is given in Equation 3.31:

$$BIC = -2ln(L) + ln(n)k, \tag{3.31}$$

where ln is the natural logarithm, L is the log-likelihood of the model and k represents the number of predictors included in the model. Lower values of BIC are preferred to larger values when determining model fit.

CHAPTER FOUR

Results

Descriptive Statistics

A total of 754,362 pitches were delivered to batters during the course of the 2013 regular and post-season. In addition to pitches in which the umpire was responsible to judge, this total also includes pitches in which the batter swung and missed, made contact with the ball, or when the batter was struck by the pitch (n = 392,210). Umpires were required to make a total of 362,152 decisions, of which 123,966 (34%) were called strikes and 238,186 (66%) were called balls. On average, home plate umpires made decisions on approximately 149 pitches per game (SD = 5.89).

A total of 76 umpires were present in the data with experience ranging from 0 to 34 years of full-time MLB service (M = 12.02, SD = 8.42). Figure 4.1 presents a kernel density estimation curve for umpire experience. During the regular season, umpires are grouped in four-person umpiring crews, with umpires rotating home plate assignments every fourth day. During the 2013 season, umpires were behind the plate for a mean of 29 games (SD = 9.63). On average, home plate umpires observed approximately 291 pitches per game (SD = 9.71).

Umpire Decisions by Pitch Location

Pitch locations were categorized to indicate where, in one of four regions, pitches were delivered (See Figure 3.2b). Three of the four pitch location regions reflect pitches that were delivered within either the inner, middle, or outer areas of the strike zone. Pitches located outside of the strike zone were considered to be in the fourth, or ball, region. This transformation takes into account batter

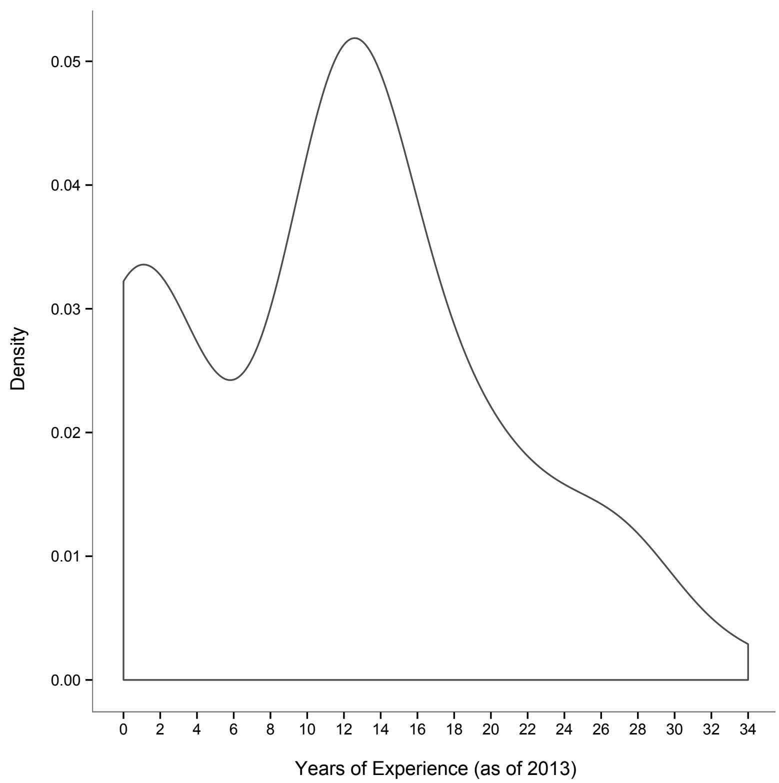


Figure 4.1: Kernel density estimation curve of umpire experience (in seasons).

handedness, such that what is classified as the the inner region for right handed batters would serve as the outer region for left handed batters. Descriptive statistics for umpire decisions by pitch location are reported in Table 4.1.

Table 4.1: *Number and Proportion of Umpire Decisions by Pitch Location.*

	Pitch Location				
Umpire Decision	Ball	Inner	Middle	Outer	Total
Ball	228,806	3,453	2,216	3,711	238,186
	(.85)	(.15)	(.07)	(.09)	(.66)
Strike	40,438	19,309	27,926	36,293	123,966
	(.15)	(.85)	(.93)	(.91)	(.34)
Total	269,244	22,762	30,142	40,004	362,152
	(.74)	(.06)	(.08)	(.11)	(1.00)

Note. Values inside parentheses are conditional proportions, except the Total columns and rows, which are unconditional proportions. Ball: pitches located outside of batter's strike zone; Inner: pitches delivered within the inner-most one-third of the strike zone area above home plate, taking into account batter handedness; Middle: pitches delivered within middle-most one-third of the strike zone area above home plate, taking into account batter handedness; Outer: pitches delivered within outer-most one-third of the strike zone area above home plate, taking into account batter handedness.

Table 4.2 presents the number and proportion of correct and incorrect umpire decisions by pitch location. Umpires made 333,412 (92%) correct decisions. For pitches located within the strike zone (i.e., inner, middle, outer), umpires are more accurate when pitches are delivered in either the middle (95%) our outer (93%) regions, compared to those in the inner region (88%). For pitches located in the ball region, umpires are approximately 92% accurate.

Umpire Decisions by Ball–Strike Count

Ball–strike count was categorized to indicate who, in the batter–pitcher matchup, holds the advantage. During situations in which neither the batter nor the pitcher held a distinct advantage over the other, the ball–strike count was

Table 4.2: *Number and Proportion of Correct and Incorrect Umpire Decision Outcomes by Pitch Location.*

Umpire Decision	Pitch Location Ball	Inner	Middle	Outer	Total
Incorrect	21,658 (.08)	2,788 (.12)	1,508 (.05)	2,786 (.07)	28,740 (.08)
Correct	247,586 (.92)	19,974 (.88)	28,634 (.95)	37,218 (.93)	333,412 (.92)
Total	269,244 (.74)	22,762 (.06)	30,142 (.08)	40,004 (.11)	362,152 (1.00)

Note. Values inside parentheses are conditional proportions, except the Total columns and rows, which are unconditional proportions. Ball: pitches located outside of batter's strike zone; Inner: pitches delivered within the inner-most one-third of the strike zone area above home plate, taking into account batter handedness; Middle: pitches delivered within middle-most one-third of the strike zone area above home plate, taking into account batter handedness; Outer: pitches delivered within outer-most one-third of the strike zone area above home plate, taking into account batter handedness.

categorized as neutral. Descriptive statistics for umpire decisions by ball–strike count are reported in Table 4.3.

Table 4.3: *Number and Proportion of Umpire Decisions by Ball–Strike Count.*

Umpire Decision	Ball–Strike Count Neutral	Batter	Pitcher	Total
Ball	126,778 (.58)	20,007 (.57)	91,401 (.84)	238,136 (.66)
Strike	91,442 (.42)	15,025 (.43)	17,499 (.16)	123,966 (.34)
Total	218,220 (.60)	35,032 (.10)	108,900 (1.00)	362,152 (.30)

Note. Values inside parentheses are conditional proportions, except the Total columns and rows, which are unconditional proportions. Neutral: pitches delivered to batters in which the ball–strike count favors neither the batter nor the pitcher; Batter: pitches delivered to batters in which the ball–strike count favors the batter; Pitcher: pitches delivered to batters in which the ball–strike count favors the pitcher.

Table 4.4 presents the number and proportion of correct and incorrect umpire decisions by ball–strike count. Umpire decisions were approximately equally accurate across neutral (91%), batter- (92%), and pitcher-advantaged counts (94%).

Table 4.4: *Number and Proportion of Correct and Incorrect Umpire Decision Outcomes by Ball–Strike Count.*

	Ball–Strike Count			
Umpire Decision	Neutral	Batter	Pitcher	Total
Incorrect	19,339 (.09)	2,910 (.08)	6,491 (.06)	28,470 (.08)
Correct	198,881 (.91)	32,122 (.92)	102,409 (.94)	333,412 (.92)
Total	218,220 (.60)	35,032 (.10)	108,900 (.30)	362,152 (1.00)

Note. Values inside parentheses are conditional proportions, except the Total columns and rows, which are unconditional proportions. Neutral: pitches delivered to batters in which the ball–strike count favors neither the batter nor the pitcher; Batter: pitches delivered to batters in which the ball–strike count favors the batter; Pitcher: pitches delivered to batters in which the ball– strike count favors the pitcher.

Table 4.5 presents proportions of correct umpire decisions by both pitch location and ball–strike count. Umpires were approximately equally accurate when pitches are located within the strike zone and the ball–strike count is either neutral (Inner: 89%; Middle: 96%; Outer: 94%) or favors the batter (Inner: 91%; Middle: 97%; Outer: 95%). However, umpires demonstrate somewhat lower accuracy rates on pitches within the strike zone when the ball–strike count favors pitchers, especially the inner region (Inner: 81%; Middle: 88%; Outer: 86%).

Model Results

A series of models were estimated in order to answer the research questions provided in Chapter Two. A summary of model rankings is presented in Table 4.6.

Table 4.5: *Proportion of Correct Umpire Decisions by Pitch Location and Ball–Strike Count.*

	Ball–Strike Count Advantage		
Pitch Location	Neutral	Batter	Pitcher
Ball	.90	.90	.95
Inner	.89	.91	.81
Middle	.96	.97	.88
Outer	.94	.95	.86

Note. Ball: pitches located outside of batter's strike zone; Inner: pitches delivered within the inner-most one-third of the strike zone area above home plate, taking into account batter handedness; Middle: pitches delivered within middle-most one-third of the strike zone area above home plate, taking into account batter handedness; Outer: pitches delivered within outer-most one-third of the strike zone area above home plate, taking into account batter handedness. Neutral: pitches delivered to batters in which the ball–strike count favors neither the batter nor the pitcher; Batter: pitches delivered to batters in which the ball–strike count favors the batter; Pitcher: pitches delivered to batters in which the ball– strike count favors the pitcher.

Table 4.6: *Summary of Model Fit Results.*

Models	Deviance	BIC	BIC Rank	AIC	AIC Δ	AIC Weight
Model 4b	196089.28	196511.68	2.00	196155.28	0.00	1.00
Model 4a	196158.45	196324.85	1.00	196184.45	29.17	0.00
Model 6	198522.51	198688.91	4.00	198548.51	2393.23	0.00
Model 5	200509.73	200548.13	8.00	200515.73	4360.44	0.00
Model 4	198530.39	198619.99	3.00	198544.39	2389.11	0.00
Model 3	199628.67	199679.87	6.00	199636.67	3481.38	0.00
Model 2	199545.95	199609.95	5.00	199555.95	3400.67	0.00
Model 1	200515.13	200540.73	7.00	200519.13	4363.85	0.00
Model 0	200777.71	200790.51	9.00	200779.71	4624.43	0.00

Note. Models 0, 1, 4, 4a, and 4b are discussed in this chapter. BIC: Bayesian Information Criterion; BIC Rank: Ranking based on BIC value (smaller values are ranked higher); AIC: Akaike Information Criterion; AIC Δ: Measure of each model relative to the best fitting model; AIC Likelihood: Exponentiated product of AIC $\Delta \times -0.5$; AIC Weight: Ratio of AIC likelihood and the sum of AIC likelihood; Cumulative AIC Weight: Cumulative sum of AIC weight. Models 4a and 4b were fit after the original research questions were posed in order to better account for the effects of pitch location and ball–strike on umpires' probability of decision accuracy.

A single-level null model with no predictors or random effects (Model 0) was first estimated. Results are displayed in Table 4.7 and indicate the predicted probability of an umpire making a correct decision is .92 (95% CI = .91, .92).

Table 4.7: *Summary of Model 0 (Null Model) for Predicting Probability of Accurate Umpire Decision, with No Random Effects.*

Predictor	*B*	*SE*	*OR*	95% CI *OR*		Random Variance
Intercept	2.45	.01	11.60	11.46	11.74	–

Note. *B*: Regression coefficient; *SE*: Standard error; *OR*: Odds ratio; CI: Confidence interval; Random Variance: Variance of random parameter.

Following estimation of the null model, Model 1 was estimated as a fully unconditional model with a random effect for umpires and no level-1 or level-2 predictors. Results of Model 1 are presented in Table 4.8 and Figure 4.2 presents individual umpire accuracy rates, which ranged from .90 to .95. This model shows that there is some, albeit small, variability between umpires in accuracy. By taking the Model 1 random variance and applying it to the formula for calculating the ICC (.016/[.016+3.290]), approximately 0.48% of the variability in pitch outcomes can be attributed to umpire effects. In terms of model fit, Model 1 was a better fit to the data, compared to Model 0. Thus, the random umpire effect is included in subsequent models.

Associations Between Pitch Location and Ball–Strike Count

Accounting for the effects of pitch location and count on umpires' ball–strike decisions is a major emphasis in the current study. It has been demonstrated that home plate umpires' decisions can be influenced by the location in which pitches are delivered (e.g., Fast, 2011) as well as the ball–strike count (e.g., Marchi and Albert, 2014). However, I posit that modeling umpire decisions in terms of the effects of

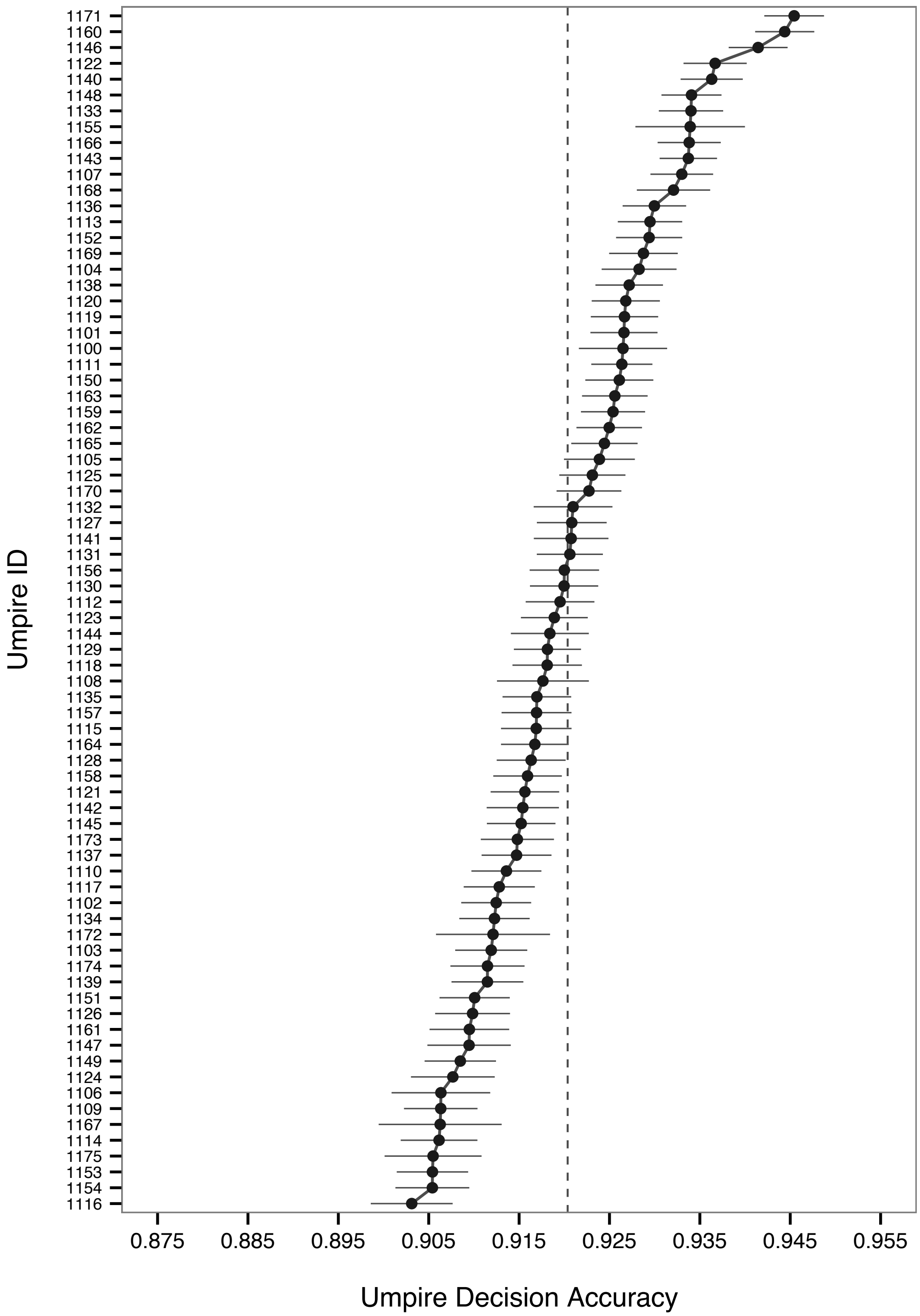


Figure 4.2: Umpire decision accuracy by transformed ball–strike count and pitch location.

Table 4.8: *Summary of Model 1 for Predicting Probability of Accurate Umpire Decision, with Random Intercept.*

Predictor	*B*	*SE*	*OR*	95% CI *OR*		Random Variance
Intercept	2.45	.02	11.62	11.28	11.99	0.016

Note. *B*: Regression coefficient; *SE*: Standard error; *OR*: Odds ratio; CI: Confidence interval; Random Variance: Variance of random parameter. Model intraclass correlation coefficient (ICC) = .0048.

both factors may provide a more comprehensive understanding of umpire decision accuracy. In the current study, three of the six models estimated (Models 2, 3, and 4) address effects of pitch location and count on umpires' ball–strike decisions. Among these three models, Model 4 was observed to best fit the data. Model 4 examines the association of both ball–strike count and pitch location on umpire decision accuracy. Table 4.9 presents a summary of Model 4.

Compared to pitches located outside of the strike zone (i.e., balls), umpires are less likely to call pitches located in the inner region of the strike zone accurately ($B = -.39$), controlling for ball–strike count. The odds ratio (OR) is 0.68 (95% CI = 0.65, 0.71), which suggests the odds of umpires correctly calling pitches in the inner region of the strike zone are about .68 that of pitches thrown in the *ball* region, controlling for ball–strike count. Conversely, the odds of umpires correctly calling pitches thrown in the ball region are $1/.68 = 1.47$ times greater than the odds of correctly calling a pitch thrown in the inner region of the strike zone, controlling for ball-strike count. The predicted probability ($\hat{\pi}$) of an umpire making an accurate decision when pitches are delivered in the inner region of the strike zone was .87.

Compared to pitches located outside of the strike zone (i.e., balls), umpires are more likely to call pitches located in either the middle ($B = 0.61$) or outer ($B = 0.24$) regions of the strike zone accurately, controlling for ball–strike count. For

Table 4.9: *Summary of Model 4 for Predicting Probability of Accurate Umpire Decision by Pitch Location and Ball–Strike Count, with Random Effect for Umpire.*

Predictor	*B*	*SE*	*OR*	95% CI *OR*		Random Variance
Intercept	2.29	.02	9.84	9.51	10.17	.02
Pitch Location						
Inner	−.39	.02	.68	.65	.71	–
Middle	.61	.02	1.84	1.74	1.94	–
Outer	.24	.03	1.28	1.22	1.33	–
Ball–Strike Count						
Batter	.07	.02	1.07	1.03	1.12	–
Pitcher	.47	.02	1.59	1.55	1.64	–

Note. *B*: Regression coefficient; *SE*: Standard error; *OR*: Odds ratio; CI: Confidence interval; Random Variance: Variance of random parameter. Pitch Location reference group: Ball (pitches located outside of batter's strike zone); Inner: pitches delivered within the inner-most one-third of the strike zone area above home plate, taking into account batter handedness; Middle: pitches delivered within middle-most one-third of the strike zone area above home plate, taking into account batter handedness; Outer: pitches delivered within outer-most one-third of the strike zone area above home plate, taking into account batter handedness. Ball–Strike Count reference group: Neutral (pitches delivered to batters in which the ball–strike count favors neither the batter nor the pitcher); Batter: pitches delivered to batters in which the ball–strike count favors the batter; Pitcher: pitches delivered to batters in which the ball–strike count favors the pitcher.

pitches located in the middle region of the strike zone, the OR is 1.84 (95% CI = 1.74, 1.94), which suggests the odds of umpires correctly calling pitches in the middle region are about 1.84 that of pitches thrown in the ball region, controlling for ball–strike count. The predicted probability of an umpire making an accurate decision when pitches are delivered in the middle region of the strike zone was $\hat{\pi}$ = .95. Likewise, for pitches located in the outer regions of the strike zone, the OR is 1.28 (95% CI = 1.22, 1.33), which suggests the odds of umpires correctly calling pitches in the outer region are about 1.28 that of pitches thrown in the ball region, controlling for ball–strike count. The predicted probability of an umpire making an accurate decision when pitches are delivered in the outer region of the strike zone was $\hat{\pi}$ = .93.

Regarding effects of the ball–strike count in Model 4, compared to pitches delivered during neutral counts, umpires are slightly more likely to call pitches during batter-advantaged counts accurately (B = .07), controlling for pitch location. The OR is 1.07 (95% CI = 1.03, 1.12), which suggests the odds of umpires correctly calling pitches during batter-advantaged counts are about 1.07 that of pitches thrown during neutral counts. The predicted probability of an umpire making an accurate decision when pitches are delivered during batter-advantaged counts was $\hat{\pi}$ = .91.

When pitchers hold the advantage over batters in the ball–strike count, compared to pitches delivered during neutral counts, umpires are more likely to call pitches accurately (B = .47), controlling for pitch location. The OR is 1.59 (95% CI = 1.55, 1.64), which suggests the odds of umpires correctly calling pitches during pitcher-advantaged counts are about 1.59 that of pitches thrown during neutral counts. The predicted probability of an umpire making an accurate decision when pitches are delivered during pitcher-advantaged counts was $\hat{\pi}$ = .94.

After plotting umpire-aggregated proportions of decision accuracy by pitch location and ball–strike count, an interaction between location and count was observed. Specifically, umpires' decisions appeared to be less accurate when pitchers held the advantage in the ball–strike count and the pitch was located in either the inner, outer, or middle region of the strike zone. Thus, following estimation of Model 4, two subsequent models (Models 4a and 4b) were fit in order to better account for both the main effects of and interaction effects between pitch location and ball–strike on umpires' probability of decision accuracy.

I posit that accounting for interaction effects between pitch location and ball–strike count will provide a better understanding of umpire decision accuracy as umpires' decisions are made during situations in which the ball–strike count may influence where pitchers deliver the ball (Marchi and Albert, 2014). Hence, the probability that an umpire will judge a pitch accurately depends on a pitch's location and corresponding ball–strike count. Models 4a and 4b are described below.

Model 4a predicts the probability of accurate umpire decisions by pitch location, ball–strike count, and pitch location $\times$ ball–strike count interactions, with a random effect for umpires. Model 4b builds on Model 4a by adding random effects for pitch location, and ball–strike count. The umpire random effect in Model 4a was also estimated in Model 4b. AIC values indicate Models 4a and 4b were both a better fit over Model 4 (AIC $\Delta = 2389.11$). Although Models 4a (AIC = 196184.45) and 4b (AIC = 196155.28) performed relatively similarly, Model 4b was preferred on the basis of model fit as judged by the AIC Δ (29.17, see equation 3.29), which accounts for differences between models in terms of sample size and arbitrary constants (Burnham et al., 2011).

The fixed and interaction effects included in Model 4a were also included in Model 4b. Model 4a fixed and interaction effects results are similar to those obtained in Model 4b. Thus, full description of results from Model 4a are deferred in order to fully explain Model 4b, as model fit estimates indicated Model 4b was a slightly better fit to the data.

Pitch Location and Ball–Strike Count Interactions. Figure 4.3 displays a noticeable interaction between predicted probabilities for pitch location and ball–strike count—particularly during counts in which the pitcher holds the advantage over the batter and pitches are located in either the middle, outer, or inner regions. Interaction effects between pitch location and ball–strike count were estimated in Model 4b. Full summary results of Model 4b are presented in Table 4.10.

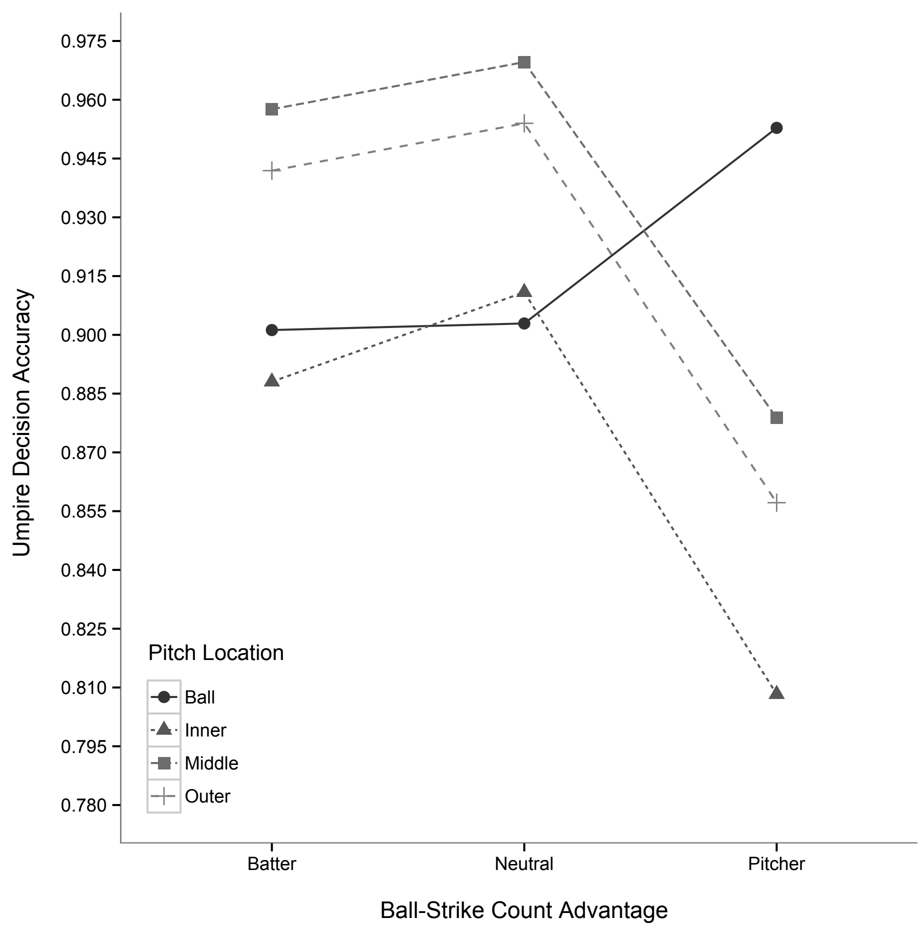


Figure 4.3: Model 4b predicted probabilities for umpire decision accuracy by ball–strike count and pitch location.

Table 4.10: *Summary of Model 4b for Predicting Probability of Accurate Umpire Decision by Pitch Location, Ball–Strike Count, and Pitch Location × Ball–Strike Count Interactions, with Random Effects for Umpire, Pitch Location and Ball–Strike Count.*

Predictor	*B*	*SE*	*OR*	95% CI *OR*		Random Variance
Intercept	2.21	.02	9.12	8.81	9.45	.02
Pitch Location						
Inner	−.14	.03	.87	.83	.92	.00
Middle	.90	.04	2.47	2.29	2.68	.04
Outer	.57	.04	1.77	1.65	1.91	.05
Ball–Strike Count						
Batter	.02	.03	1.02	.96	1.08	.02
Pitcher	.80	.02	2.21	2.14	2.29	.00
Pitch Location × Ball–Strike Count						
Inner × Batter	.23	.08	1.26	1.09	1.47	–
Middle × Batter	.33	.10	1.39	1.13	1.70	–
Outer × Batter	.23	.08	1.25	1.07	1.46	–
Inner × Pitcher	−1.43	.05	.24	.22	.27	–
Middle × Pitcher	−1.93	.06	.15	.13	.16	–
Outer × Pitcher	−1.79	.05	.17	.15	.18	–

Note. *B*: Regression coefficient; *SE*: Standard error; *OR*: Odds ratio; CI: Confidence interval; Random Variance: Variance of random parameter. Pitch Location reference group: Ball (pitches located outside of batter's strike zone); Inner: pitches delivered within the inner-most one-third of the strike zone area above home plate, taking into account batter handedness; Middle: pitches delivered within middle-most one-third of the strike zone area above home plate, taking into account batter handedness; Outer: pitches delivered within outer-most one-third of the strike zone area above home plate, taking into account batter handedness. Ball–Strike Count reference group: Neutral (pitches delivered to batters in which the ball–strike count favors neither the batter nor the pitcher); Batter: pitches delivered to batters in which the ball–strike count favors the batter; Pitcher: pitches delivered to batters in which the ball–strike count favors the pitcher.

A model covariance table is presented in Table 4.11 with random effect variances in rows 1-6 of column 3, covariances in rows 7-21 of column 3 and random effect correlations in rows 7-21 of column 4. Random effect variances are presented for each of the random effects included in Model 4b. Covariance values between two random effects are restricted to be 0 in a diagonal covariance matrix (Heck, et al., 2012).

Model 4b mean predicted logits and predicted probabilities for umpire accuracy by pitch location and ball–strike count are presented in Table 4.12. Across all pitch locations, predicted probabilities of accurate umpire decisions for pitches located in the inner region of the strike zone ($\hat{\pi} = .87$) appear to be noticeably lower compared to predicted probabilities of accurate umpire decisions for pitches located in the middle ($\hat{\pi} = .94$) or outer regions ($\hat{\pi} = .92$). When examining predicted probabilities of accurate umpire decisions for ball–strike counts when pitches are located in the inner region of the strike zone, pitchers appear to be at a distinct disadvantage ($\hat{\pi} = .81$) compared to when the count is either neutral ($\hat{\pi} = .89$) or favors the batter ($\hat{\pi} = .91$).

Umpire accuracy appears to be similar across pitches located in either the middle or outer regions of the strike zone. Again, when compared to ball–strike counts with a neutral or batter advantage, umpires appear to be less accurate when the ball–strike count favors the pitcher ($\hat{\pi}$ Middle $= .88$; $\hat{\pi}$ Outer $= .86$).

Table 4.11: *Covariances and Correlations of Model 4b Random Components*

Variable 1	Variable 2	VCov	Cor
Intercept	–	0.12	–
Inner	–	0.00	–
Middle	–	0.40	–
Outer	–	0.05	–
Batter	–	0.02	–
Pitcher	–	0.00	–
Intercept	Inner	−0.00	−1.00
Intercept	Middle	−0.00	−0.09
Intercept	Outer	−0.01	−0.20
Intercept	Batter	−0.00	−0.09
Intercept	Pitcher	−0.00	−0.50
Inner	Middle	0.00	0.09
Inner	Outer	0.00	0.20
Inner	Batter	0.00	0.09
Inner	Pitcher	0.00	0.50
Middle	Outer	0.03	0.76
Middle	Batter	0.00	0.10
Middle	Pitcher	0.01	0.79
Outer	Batter	−0.01	−0.24
Outer	Pitcher	0.00	0.39
Batter	Pitcher	0.00	0.40

Note. VCov: Model 4b random effect variances (rows 1-6) and covariances (rows 7-21); Cor: Model 4b random effect correlations (rows 7-21). Pitch Location reference group: Ball (pitches located outside of batter's strike zone); Inner: pitches delivered within the inner-most one-third of the strike zone area above home plate, taking into account batter handedness; Middle: pitches delivered within middle-most one-third of the strike zone area above home plate, taking into account batter handedness; Outer: pitches delivered within outer-most one-third of the strike zone area above home plate, taking into account batter handedness. Ball–Strike Count reference group: Neutral (pitches delivered to batters in which the ball–strike count favors neither the batter nor the pitcher); Batter: pitches delivered to batters in which the ball–strike count favors the batter; Pitcher: pitches delivered to batters in which the ball–strike count favors the pitcher.

Table 4.12: *Summary of Predicted Logits and Predicted Probabilities by Pitch Location and Ball–Strike Count from Model 4b.*

Ball–Strike Count	Pitch Location	Predicted Log Odds	Predicted Probability	95% CI LL	95% CI UL
Ball	Neutral	2.21	0.901	0.900	0.902
Ball	Batter	2.23	0.903	0.902	0.904
Ball	Pitcher	3.00	0.953	0.952	0.954
Inner	Neutral	2.07	0.888	0.887	0.889
Inner	Batter	2.32	0.911	0.910	0.912
Inner	Pitcher	1.44	0.808	0.807	0.809
Middle	Neutral	3.11	0.958	0.957	0.958
Middle	Batter	3.46	0.970	0.969	0.970
Middle	Pitcher	1.98	0.879	0.878	0.880
Outer	Neutral	2.78	0.942	0.941	0.943
Outer	Batter	3.03	0.954	0.953	0.955
Outer	Pitcher	1.79	0.857	0.856	0.858

Note. Pitch Location reference group: Ball (pitches located outside of batter's strike zone); Inner: pitches delivered within the inner-most one-third of the strike zone area above home plate, taking into account batter handedness; Middle: pitches delivered within middle-most one-third of the strike zone area above home plate, taking into account batter handedness; Outer: pitches delivered within outer-most one-third of the strike zone area above home plate, taking into account batter handedness. Ball–Strike Count reference group: Neutral (pitches delivered to batters in which the ball–strike count favors neither the batter nor the pitcher); Batter: pitches delivered to batters in which the ball–strike count favors the batter; Pitcher: pitches delivered to batters in which the ball–strike count favors the pitcher.

CHAPTER FIVE

Discussion

A growing area of research related to expert sport performance involves analyzing the outcomes of MLB home plate umpires' ball–strike decisions (e.g., Green & Daniels, 2014; Lindbergh, 2014; Marchi & Albert, 2014; Mills, 2014; Sievert, 2014). With the availability of *PITCHf/x* data, researchers can now quantitatively model home plate umpire decisions while examining patterns in judgment accuracy and the presence of effects of pitch- and umpire-related variables.

The purpose of the current study was to examine the association of pitch- and umpire-related factors on the accuracy of umpires' ball–strike decisions during the course of an entire MLB season. Umpire decisions were treated as any pitch on which the umpire was required to call a strike or ball. The following hypotheses were tested in the current study:

(1) MLB umpires will not differ in their accuracy of judging pitch outcomes.

(2) Pitch location will be associated with MLB umpires' accuracy of judging pitch outcomes. Specifically, compared to pitches located within the middle and outer regions of the strike zone, umpires will demonstrate less accurate decisions when pitches are located within the inner region of the strike zone.

(3) Ball–strike count will be associated with MLB umpires' accuracy of judging pitch outcomes. Specifically, compared to neutral and batter-advantaged ball–strike counts, umpires will demonstrate less accurate decisions when pitchers hold the advantage.

(4) After controlling for ball–strike count, pitch location will be associated with MLB umpires' accuracy of judging pitch outcomes. Specifically, umpires

will demonstrate greater accuracy when pitches are located either in the middle or outer regions of the strike zone.

(5) After controlling for pitch location, ball–strike count will be associated with MLB umpires' accuracy of judging pitch outcomes. Specifically, umpires will demonstrate greater accuracy when ball–strike counts are either neutral or favor when batters hold the advantage.

(6) Experience in MLB will be associated with umpires' accuracy of judging pitch outcomes. Specifically, compared to umpires with less experience, umpires with more experience will demonstrate less accurate decisions.

(7) Umpire experience will not moderate the association between umpire accuracy and either pitch location or ball–strike count.

The mean number of decisions made by umpires during the 2013 season was approximately 4,357 (SD = 1,432.69). Umpires made approximately 149 decisions per game during the 2013 season (SD = 5.89). Holding pitch location, ball–strike count, and umpire experience constant, Major League Baseball umpires demonstrated accuracy rates of approximately 90% to 95% ($M = 92\%, SD = 1\%$). To put this proportion in perspective, MLB umpires can be expected to make approximately 8–15 decision errors per game.

Results from the current study indicate umpires explain approximately 0.48% of the variability in pitch outcomes. Both pitch location and ball–strike count were observed to be associated with umpires' accuracy of judging pitch outcomes. Of the models estimated, Model 4b was observed to best fit the data. This model examined the effects of pitch location and ball–strike count on umpire decisions during the 2013 season. As predicted, umpires demonstrated greater accuracy during ball–strike counts in which the either batter held the advantage or the count was neutral when pitches were located in either the middle or outer regions of the strike zone.

Effects of Pitch Location and Ball–Strike Count on Umpire Decisions

Predicted probabilities of umpire decision accuracy should be interpreted in light of the dependencies between pitch location and ball–strike count. In other words, calculating the proportion of umpires' accurate decisions when, for example, pitchers hold the advantage in the ball–strike count will be misleading. Since each pitch delivered by a pitcher during this situation is conditional on a particular location in the strike zone, discretion should be used when interpreting results. Hence, the probability that an umpire will judge a pitch accurately depends on that pitch's location and corresponding ball–strike count. As was demonstrated in Model 4b, an interaction was observed between pitch location and ball–strike count, which suggested umpires are generally less accurate when pitchers hold the advantage in the ball–strike count and pitches are located within the strike zone.

Minimal differences were observed between the predicted probabilities of umpires making accurate decisions when the ball–strike count was either neutral or when batters held the advantage, and the pitch was located in either the middle or outer region of the strike zone. In order to help layreaders grasp the larger context of MLB umpires' decisions, it may be of interest to examine the mean predicted probabilities that umpires will make an accurate decision during each ball–strike count condition across all pitch location conditions. For example, when pitchers hold the advantage in the ball–strike count, across all pitch location conditions, umpires have a mean predicted probability of approximately 87%, compared to 93% and 92% when the ball–strike count is either neutral or when batters held the advantage, respectively. The implication is that, across all pitch locations, umpires appear to make less accurate decisions when a pitcher holds the advantage in the ball–strike count. This noticeable difference in umpire decision accuracy likely forces pitchers to be more cognizant of throwing the ball to precisely the area of the intended target, lest they shift the ball–strike to the batter's advantage.

Likewise, we can examine predicted probabilities that umpires will make an accurate decision during each pitch location condition across all ball–strike count conditions. When pitches are delivered to the inner region of the strike zone, across all ball–strike count conditions, umpires have a mean predicted probability of approximately 87%, compared to 92%, 93%, and 92% when pitches are delivered outside, in the middle, and in the outer regions of the strike zone, respectively. Several authors have made similar observations (e.g., Ford, Gallagher, Lacy, Bridwell, & Goodwin, 1999; Fast, 2007).

More specifically, during ball–strike counts in which the batter held the advantage, umpires were predicted to be 97% accurate when pitches were located in the middle region of the strike zone. During neutral counts, umpires were predicted to be approximately 96% accurate when pitches were located in the middle region of the strike zone. When batters held the advantage in the count and pitches were located in the outer region of the strike zone, umpires were predicted to be approximately 95% accurate. Likewise, umpires were predicted to be 94% accurate during neutral counts when pitches were located in outer region of the strike zone. Overall, umpires demonstrated relatively lower accuracy rates on pitches located in the inner region of the strike zone. In this condition, umpires were predicted to be 90% and 89% accurate when either batters held the advantage in the ball–strike count or when the count was neutral, respectively.

Conversely, umpires demonstrated noticeably lower accuracy rates across all pitch location conditions when pitchers held the advantage in the ball–strike count. Specifically, umpire accuracy rates were predicted to be 88% and 86% when pitches were located in the middle and outer regions of the strike zone, respectively.

Umpires were predicted to be approximately 81% accurate when pitches were located in the inner region of the strike zone. This suggests umpires may be more inclined to shrink, or tighten the strike zone when pitchers hold the advantage over batters in the ball–strike count.

In terms of pitch location, during the game pitchers may be more apt to try and deliver the ball away from the middle region of the plate as batters typically prefer to hit pitches located in this region. Thus, pitchers may concentrate on delivering the ball as as close to the inner and outer regions of the strike zone while also ensuring the pitch remains within the strike zone area. For example, during the 2013 season, pitchers threw to the inner and outer regions of the strike zone 68% of the time (n = 62,766) and only 32% to the middle region (n = 30,142).

As a result, umpires' decisions may be affected by pitchers' strategies. As pitchers concentrate on locating their pitches near the inner and outer edges of the strike zone, away from the middle, umpires' decisions become more critical. An expert umpire should, regardless of a pitch's location, demonstrate mastery of the strike zone by accurately distinguishing balls from strikes. Results obtained in the current study demonstrate that, during the 2013 season, MLB umpires exhibit expertise when making ball–strike decisions, even to the inner and outer regions.

Umpire Expertise

When making decisions, expert umpires appear to demonstrate an ability to utilize knowledge of situational probabilities when making decisions. For example, Millslagle, Hines, and Smith (2013) examined the visual search behavior of expert and near-expert baseball umpires, noting differences in quiet eye gaze behavior. Millslagle et al. (2013) define quiet eye gaze behavior as the period of visual or tracking fixation on a specified location or target. Their results indicated that, as pitchers released the ball, the quiet eye gaze of expert umpires, occurred earlier and was longer in duration, compared to near-expert umpires. This suggests that expert umpires appear to demonstrate an ability to gather more information about a pitch and its eventual location while tracking the ball, thus improving umpire decision accuracy. Similarly, eye movements among expert sport performers have been demonstrated to be controlled by a search strategy, allowing the performer to use time more efficiently while analyzing the display (Williams, Davids, & Williams, 1999). These strategies include procedural knowledge, advance cue utilization, pattern recall and recognition, and knowledge of situational probabilities (Ghasemi, et al., 2011; Paull, et al., 1997; Takeuchi, et al., 2009; Williams, et al., 2004; Williams, et al., 2009).

In a recent study evaluating contextual influences on ball–strike decisions in umpires, players, and control group participants, MacMahon and Starkes (2008) found that both umpires and players were significantly more accurate in their ball–strike judgments compared to the control group participants. In a direct information task MacMahon and Starkes (2008), pitches located on the periphery of the strike zone were presented through the use of interactive video followed by pitches of definite balls and definite strikes. After viewing a definite ball, participants were more likely to call subsequent pitches strikes than they were after viewing an obvious strike. Participants also demonstrated bias toward calling strikes when presented with scenarios of differing pitch counts. For example, when a batter was faced with a three-ball count (i.e., 3-0, 3-1, or 3-2), participants were more likely to call the subsequent pitch a strike. These findings suggest that standards for ball–strike decisions will fluctuate based on the context in which they were presented.

Relatedly, Marchi and Albert (2014) demonstrated that the size of the strike zone is observed to change noticeably during 0-2 (pitcher-advantaged), and 3-0 (batter-advantaged) ball–strike counts, compared to 0-0 counts. Walsh (2010) suggested a tendency in umpires to purposefully make a wrong decision may be related to an attempt to maintain competitive balance. In this case, an umpire may make decisions which favor a batter who is behind in the ball–strike count when a pitch is marginally close to the strike zone. This tendency may be more reflective of umpires' desire to refrain from being the deciding factor in a game. Green and Daniels (2014) found evidence for umpires making so-called strategic errors during ball–strike situations in which umpires consciously avoid making pivotal decisions. Relatedly, Bill Klem, one of only two umpires elected to the Baseball Hall of Fame, once quipped, "The best umpired game is the game in which the fans cannot recall the umpires who worked it" (Baseball Reference, 2012). While there appears to be empirical evidence which suggests that umpires may purposefully make wrong decisions in order to avoid unnecessary scrutiny, there may also be unwritten rules among umpires to remain as anonymous as possible during competition, which may compel them to make decisions resulting in less contention from players and managers.

Strike Zone Conceptualization

The MLB Rule Book (2014) definition of the strike zone was employed in order to determine umpire decision accuracy at the pitch-level. The strike zone is conceptualized as an imaginary, rectangular area with a standard width and upper and lower boundaries dependent upon the height and stance of a batter. Several specifications are noted.

First, the upper and lower limits of each batter's strike zone parameters were mean-aggregated across the 2013 season. This allowed each individual batter in the data to have a uniform upper and lower strike zone limit, while controlling for any differences between *PITCHf/x* operators' estimations of batters' strike zone parameters. Furthermore, since batters' height and stance are assumed to remain constant across all plate appearances, taking the mean of individual batter's upper and lower strike zone limits is reasonable.

Second, the standard width of the strike zone was determined to be approximately 20.14 inches. This value was obtained by adding the length of the radius of a regulation-size baseball (1.57 inches) to both sides of the 17 inch home plate. Per the MLB Rulebook (2014), a strike occurs when any part of the ball passes through any part of the strike zone. Readers are reminded that the x- and y-coordinates of pitches in the *PITCHf/x* data represent the center point of the ball. Thus, adding the length of the radius of a regulation-size baseball reconceptualizes the width of the strike to be 20.14 inches, not 17 inches. I posit that this allows for a more realistic and accurate estimation of umpire decision accuracy.

Umpire Positioning

Current umpire training curriculum and instruction materials recommend home plate umpires position themselves directly behind the inner region of home plate in order to view the area between the catcher's shoulder and the batter's hands (Moore, 2013). From the umpire's perspective this area would be the left side of home plate for right handed batters. For left handed batters, umpires would be positioned over the right side of home plate. It is possible that as batters try and position themselves to better reach pitches delivered to the outside region of home plate they obstruct the umpire's view of the inner region, thus preventing him from making an accurate decision.

Ford, Gallagher, Lacy, Bridwell, and Goodwin (1999) found that when umpires are repositioned from viewing pitches over the inner region to viewing pitches over the outer region, their accuracy rates are improved. Repositioning the umpire in this way may offer a better angle at which to view inside pitches. One limitation of this study is that Ford, et al. (1999) did not account for effects of ball–strike count on umpire decisions, nor did they account for the effect of visual search behavior among the umpire sample. Although I hypothesize that the visual search behavior of expert sport performers is similar among expert sport officials, more research in this area is needed. As noted in Chapter Two, numerous studies have been conducted in order to understand the visual search patterns of expert sport performers (e.g., Abernethy, 1991; Davids & Williams, 1998; Ghasemi, Momeni, Jafarzadehpur, Rezaee, & Taheri, 2011; Shim, Carlton, & Chow, & Chae, 2005). However, fewer studies have been designed in order to test the visual search patterns of baseball umpires.

I posit cognitive processes involved in umpire decision making can be observed by examining the differences between umpire accuracy rates during pitcher- and batter-advantaged ball–strike counts. In general, across all ball–strike count combinations, umpires' decisions are less accurate when pitches are located in the inner region of the strike zone. One explanation for this tendency may be related to the location in which an umpire positions himself when viewing pitches. However, more experimental trials are needed in order to better understand this trend.

General Conclusions

Over the course of an entire season, professional baseball umpires demonstrate accurate ball–strike judgment and decision making expertise. Before umpires earn a full-time job in MLB, they typically spend a number of years umpiring in the minor leagues (Weber, 2010). Furthermore, umpires typically are required to accrue a minimum of 300 MLB games before becoming eligible for full-time employment at the Major League level. Thus, given the scrutiny with which they are evaluated, umpires who reach MLB on a full-time basis can be assumed to possess domain-specific expertise.

Although noticeable differences were observed during certain ball–strike count and pitch location conditions, umpires appear to be able to accurately distinguish when a pitch is located inside our outside of the imaginary strike zone area. Moreover, umpires are required to make these decisions in a relatively rapid sequence of judgment and decision making, even in spite of the ways in which the strike zone parameters change from batter to batter. Umpires may also be forced to adjust to the varying conditions in which pitches are delivered. For example, pitchers routinely vary the speed, trajectory, and location of pitches, which may combine to make it difficult for umpires to accurately view and judge the pitch outcome.

Limitations

Data from a Single Season

The current study examined the effects of pitch location and count on professional baseball umpires' ball–strike decisions over the course of a single season. The assumption for analyzing data for a single season was based on the fact that that home plate umpires' performances are regularly reviewed and critiqued by MLB performance evaluators (Lindbergh, 2014). As a result, umpires receive post-game reports and feedback on all decisions made. Therefore, given the regular pattern of performance evaluation, I hypothesized that home plate umpires' performance would remain fairly consistent over the course of a single season. Thus, between-season effects were omitted.

Longitudinal Effects

Ignoring potential effects for time, or more specifically, season, could be a potential limitation of the current study. Umpires who, in one season, may have performed at a lower rate may, in the next season, improve. Accounting for growth over time may result in more accurate estimations of home plate umpire decision making expertise. Since data generated by the *PITCHf/x* system dates back to 2008, including a season-level effect may be beneficial for future studies. One such study supports this view. Analyzing *PITCHf/x* data for the 2008–2013 seasons, Mills (2014) observed reductions in MLB umpire performance variability, suggesting this improvement is related in part to increased umpire performance monitoring.

PITCHf/x Data Accuracy

Data from the *PITCHf/x* system are assumed to be accurate. Nathan (2008) demonstrated that the *PITCH/fx* system is calibrated to be accurate within 0.5 inches. Thus, pitches within this margin of error when calculating the parameters of the strike zone used to determine umpire accuracy may have been incorrectly coded as either correct or incorrect, depending on the location. Similarly, because the top and bottom parameters of each batter's strike zone are manually adjusted by a *PITCHf/x* operator at each MLB stadium, differences between stadium operator's strike zone settings may affect the strike zone parameters used in order to determine umpire decision accuracy. Future studies should account for potential stadium operator effects by measuring the differences in batter's top and bottom strike zone limits between stadiums.

Outcome Variable Categorization

For the current study, the outcome variable, umpire decision, was dichotomized to indicate whether the umpire's decision was either correct or incorrect. In terms of the research questions posed and hypotheses tested, dichotomizing the outcome variable was important in order to determine initially how accurate umpires are given the pitch-related predictor variables examined. Knowing more about the effects of pitch-related factors on umpire decision accuracy may lay the foundation for further investigating umpire decisions. However, dichotomizing the outcome variable may decrease understanding of the degree to which an umpire is accurate or inaccurate. For example, future studies should examine the effects of a pitch's proximity to the strike zone in order to determine degrees of error in umpire decisions. In other words, distance from the strike zone should be included in the outcome variable. A possible solution may involve weighting the difficulty in judging a pitch's outcome from a reference point, such as the center of the strike zone.

APPENDICES

APPENDIX A

IRB Approval

BAYLOR UNIVERSITY

INSTITUTIONAL REVIEW BOARD

One Bear Place #97310 Waco, TX 76798-7310 • (254) 710-3763 • FAX (254) 710-7309 • WEBSITE: www.baylor.edu/research/irb

DATE:	06/13/2014
TO:	Aaron Baggett, M.Ed.
FROM:	Office of the Vice Provost for Research, Research Compliance Baylor University Institutional Review Board
STUDY TITLE:	Estimating professional baseball umpire ball--strike judgment and decision making patterns using multilevel mixed-effects logistic regression models
IRB REFERENCE #:	590478
SUBMISSION TYPE:	New Project
ACTION:	DETERMINATION OF NON-HUMAN SUBJECT RESEARCH
DECISION DATE:	06/13/2014

Thank you for your research study submission. Your research has been determined to not meet the definition of human subject research under the purview of the IRB according to federal regulations at 45 CFR 46.102(d) & (f). Specifically, "identifiable private information" will not be obtained. The subjects are aware that their behavior is being recorded in a public forum; therefore, this does not meet the definition of "private information."

This determination is based on the protocol and/or materials submitted. If the research is modified, you must contact this office to determine whether your modified research meets the definition of human subject research.

If you have any questions, please contact Deborah Holland at (254) 710-1438 or Deborah_L_Holland@baylor.edu. Please include your study title and reference number in all correspondence with this office.

Sincerely,

Deborah L Holland

Deborah L. Holland, JD, MPH
Assistant Vice Provost of Research
Director of Compliance

APPENDIX B

Database Construction

In order to prepare and arrange the data for multilevel analysis, two file structures with a common identification variable were created. The presence of two file structures allows for the arrangement of data at Level-2 to be joined or merged with data at Level-1. Level-2 data consist of umpire-specific variables including umpire names and number of years experience in MLB. Using the R function `join()` from the `dplyr` library (Wickham, 2011), both file structures were combined by matching umpire names. In this case, the umpire names act as an ID variable when the `join()` function is called in R. R code used to construct the *PITCHf/x* database used in the current analysis is available below.

```
# ==============================================================
# PITCHf/x Database Construction
# ==============================================================

# --- Load package libraries (if necessary) --- #
library(pitchRx)
library(RSQLite)
library(RSQLite.extfuns)
library(devtools)
library(dplyr)

# --- Initialize SQLite database --- #
# *src_sqlite* creates a local database called *pfx_13.sqlite3*
pfx_13 <- src_sqlite("pfx_13.sqlite3", create = TRUE)

# --- Set XML files to collect --- #
# *inning_all.xml*, *players.xml*, and *miniscoreboard.xml* contain
# pitch/game variables, player IDs, and umpire IDs, respectively
# NOTE: Player IDs and umpire IDs are used for joining purposes
files <- c("inning/inning_all.xml", "players.xml", "miniscoreboard.xml")

# --- Collect 2013 PITCHf/x data --- #
scrape(start = "2013-01-01", end = "2013-12-31",
       suffix = files, connect = pfx_13$con)
```

```
# --- Verify *pfx_13* tables --- #
pfx_13

# --- Load SQLite table --- #
pfx_13_all <- src_sqlite("~/pfx_13.sqlite3")

# --- Convert *atbat*, *pitch*, and *umpire* to standalone tables --- #
atbat <- tbl(pfx_13_all, "atbat")
pitch <- tbl(pfx_13_all, "pitch")
umpire <- tbl(pfx_13_all, "umpire")

# --- Select, group, and collect variables from tables above --- #
# *num*: Variable for pitch sequence number
# *stand*: Batter handedness
# *gameday_link*: ID variable specifying teams and game start time
atbats <- atbat %>%
  select(num, stand, b_height, gameday_link) %>%
  group_by(gameday_link)
atbats <- collect(atbats)

# *call*: Renamed umpire desision variable (*des*)
# *sz_top*: Specifies top parameter of batter's strike zone
# *sz_bot*: Specifies bottom parameter of batter's strike zone
# *px*: X coordinate of pitch location
# *pz*: Y coordinate of pitch location
# *zone*: Specifies which region of the strike zone the pitch was located
# *num*: Variable for pitch sequence number
# *count*: Specifies ball-strike count at time of pitch
# *gameday_link*: ID variable specifying teams and game start time
pitches <- pitch %>%
  select(call = des, sz_top, sz_bot, px, pz, zone, num,
         count, gameday_link) %>%
  #filter(call == "Called Strike" | call == "Ball") %>%
  group_by(gameday_link)
pitches <- collect(pitches)

# *umpire*: Renamed umpire ID variable (*name*)
# *gameday_link*: ID variable specifying teams and game start time
# *position == home*: Selects only home plate umpires
umpires <- umpire %>%
  select(umpire = name, gameday_link) %>%
  filter(position == home) %>%
  group_by(gameday_link)
umpires <- collect(umpires)

# --- Join *atbat*, *pitch*, and *umpire* tables by *gameday_link* --- #
# Joins *pitches* and *atbats* data frames by *num* and *gameday_link*
```

```
ps_abs <- left_join(pitches, atbats, by = c("num", "gameday_link"))
# Joins *ps_abs* data frame with *umpires* data frame by *gameday_link*
ps_abs_us <- left_join(ps_abs, umpires, by = "gameday_link", copy = TRUE)

# Rename ps_abs to pfx_13 and collect all rows
# *dplyr* package uses *n = -1* to collect all rows
pfx_13 <- tbl_df(as.data.frame(ps_abs_us, n = -1))

# --- Exclude pitches with missing X and Y pitch coordinates --- #
pfx_13 <- na.omit(pfx_13)

# --- Add player sz limits by *b_height* --- #
# *player_sz_bot*: New variable that adds mean bottom parameter of
  # batter's strike zone accross all pitches
# *player_sz_top*: New variable that adds mean top parameter of
    # batter's strike zone accross all pitches
pfx_13 <- pfx_13 %>%
  group_by(b_height) %>%
  mutate(player_sz_bot = mean(sz_bot)) %>%
  mutate(player_sz_top = mean(sz_top))

# --- Create *u_test* variable for umpire's decision [1 = correct] --- #
# If pitch coordinates are located inside strike zone and umpire
    # called pitch a strike, *u_test* = 1
# If pitch coordinates are located outside strike zone and umpire
    # called pitch a strike, *u_test* = 0
# If pitch coordinates are located inside strike zone and umpire
    # called pitch a ball, *u_test* = 0
# If pitch coordinates are located outside strike zone and umpire
    # called pitch a ball, *u_test* = 1
# px limits account for addition of ball radius to strike zone width
# Ball radius = ((1.57^2 + 17) / 12) / 2
pfx_13$u_test <- with(pfx_13,
    ifelse(call == "Ball" & px < -0.8110375 | px > 0.8110375 |
      pz < player_sz_bot | pz > player_sz_top, 1,
    ifelse(call == "Called Strike" & pz > player_sz_bot &
             pz < player_sz_top &
      px >= -0.8110375 & px <= 0.8110375, 1,
    ifelse(call == "Ball" & pz > player_sz_bot &
             pz < player_sz_top &
      px > -0.8110375 & px < 0.8110375, 0,
    ifelse(call == "Called Strike" & px < -0.8110375 | px > 0.8110375 |
      pz < player_sz_bot | pz > player_sz_top, 0, 99)))))

# --- Rearrange variables in preferred order --- #
pfx_13 <- pfx_13 %>%
  select(c(count, sz_top, sz_bot, player_sz_top, player_sz_bot,
```

```
    px, pz, zone, b_height, stand, call, umpire, u_test))

# Ensure all *u_test* outcomes are either 1 or 0
with(pfx_13, mean(as.numeric(u_test)))
table(pfx_13$u_test)

# --- Specify *count* advantages --- #
# 1. Add *bs_count* variable to indicate who has the advantage (p vs. b)
# Based on Marchi & Albert, 2014
pfx_13$bs_count <- with(pfx_13,

  # 1.1 Neutral
  ifelse(count == "0-0" | count == "1-0" |
         count == "1-1" | count == "2-1", "neutral",

  # 1.2 Batter
  ifelse(count == "2-0" | count == "3-0" |
     count == "3-1" | count == "3-2", "batter",

  # 1.3 Pitcher
  ifelse(count == "0-1" | count == "0-2" |
         count == "1-2" | count == "2-2", "pitcher", 99)
)))

# 2. Convert *count* to factor
pfx_13$bs_count <- as.factor(pfx_13$bs_count)

# --- Respecify *zone* regions --- #
pfx_13$zone_reg <- with(pfx_13,

  # 1. RHBs
  ifelse(stand == "R" & zone == "1" |
         stand == "R" & zone == "4" |
         stand == "R" & zone == "7", "inner",

  ifelse(stand == "R" & zone == "2" |
         stand == "R" & zone == "5" |
         stand == "R" & zone == "8", "middle",

  ifelse(stand == "R" & zone == "3" |
         stand == "R" & zone == "6" |
         stand == "R" & zone == "9", "outer",

  # 2. LHBs
  ifelse(stand == "L" & zone == "1" |
         stand == "L" & zone == "4" |
         stand == "L" & zone == "7", "outer",
```

```
  ifelse(stand == "L" & zone == "2" |
         stand == "L" & zone == "5" |
         stand == "L" & zone == "8", "middle",

  ifelse(stand == "L" & zone == "3" |
         stand == "L" & zone == "6" |
         stand == "L" & zone == "9", "inner", "ball")))))))

# 3. Convert *zone* to factor
pfx_13$zone_reg <- as.factor(pfx_13$zone_reg)

# --- Relevel *zone_reg* to make "ball" reference group --- #
pfx_13$zone_reg <- relevel(pfx_13 $zone_reg, ref = "ball")

# --- Relevel *bs_count* to make "neutral" reference group --- #
pfx_13 $bs_count <- relevel(pfx_13 $bs_count, ref = "neutral")

# --- Rearrange variables --- #
pfx_13 <- pfx_13 %>%
  select(c(count, bs_count, sz_top, sz_bot, player_sz_top, player_sz_bot,
    px, pz, zone, zone_reg, b_height, stand, call, umpire, u_test))

# --- Read in umpire-level --- #
ump_df <- read.csv("umpire_level_data.csv", header = TRUE)
ump_df$umpire <- as.character(ump_df$umpire)

# --- Join pfx_13 with ump_df by *umpire* --- #
pfx_13 <- left_join(pfx_13, ump_df, by = "umpire", copy = TRUE)
```

APPENDIX C

R Scripts for Data Analyses

```
# --- Load pacakges (if necessary) --- #
library(lme4, quietly = TRUE)
library(dplyr, warn.conflicts = FALSE)
library(ggplot2)
library(xtable)
library(VGAM)
library(descr)

# =================================================================
# Model Descriptive Statistics
# =================================================================

# --- Load SQLite table --- #
pfx_13_all <- src_sqlite("~/pfx_13.sqlite3")

# --- Total number of pitches thrown during each game in 2013 --- #
pfx_13_all %>%
  group_by(gameday_link) %>%
  summarize(N = length(call))

pfx_13_all %>%
  group_by(gameday_link) %>%
  summarize(N = length(call)) %>%
  summarize(min = min(N), max = max(N), mean = mean(N), sd = sd(N))

# --- Mean/SD number of pitches per game in 2013 --- #
msd_pitches <- pfx_13_all %>%
  group_by(gameday_link) %>%
  summarize(N = length(call)) %>%
  summarize(mean = mean(N), sd = sd(N))

# --- Mean/SD number of decisions per game in 2013 --- #
msd_decisions <- pfx_13_all %>%
  group_by(gameday_link) %>%
  filter(call == "Called Strike" | call == "Ball") %>%
  summarize(N = length(call)) %>%
  summarize(min = min(N), max = max(N), mean = mean(N), sd = sd(N))

# --- Create data frame for pitches and decisions --- #
names <- c("Total Pitches", "Umpire Decisions")
```

```
N <- c(n_pitches, n_decisions)
M <- c(msd_pitches[, 1], msd_decisions[, 1])
SD <- c(msd_pitches[, 2], msd_decisions[, 2])
pitches_tbl <- data.frame(N, M, SD, row.names = names)
xtable(pitches_tbl)

# --- Umpire decisions by pitch location --- #
zones <- CrossTable(pfx_13$call, pfx_13$zone_reg, prop.r = FALSE,
  prop.c = FALSE, prop.t = TRUE, prop.chisq = FALSE)

# XTABLE
xtable(zones, caption = "Number and proportion of umpire
  decisions by pitch location", digits = 2)

# --- Umpire decision outcomes by pitch location --- #
zones.acc <- CrossTable(pfx_13$u_test, pfx_13$zone_reg, prop.r = FALSE,
  prop.c = FALSE, prop.t = TRUE, prop.chisq = FALSE)

# XTABLE
xtable(zones.acc, caption = "Number and proportion of umpire
  decision outcomes by pitch location", digits = 2)

# --- Umpire decisions by ball-strike count --- #
counts <- CrossTable(pfx_13$call, pfx_13$bs_count, prop.r = FALSE,
  prop.c = FALSE, prop.t = TRUE, prop.chisq = FALSE)

# XTABLE
xtable(counts, caption = "Number and proportion of umpire
  decisions by ball--strike count", digits = 2)

# --- Umpire decision outcomes by ball-strike count --- #
counts.acc <- CrossTable(pfx_13$u_test, pfx_13$bs_count, prop.r = FALSE,
  prop.c = FALSE, prop.t = TRUE, prop.chisq = FALSE)

# XTABLE
xtable(counts.acc, caption = "Number and proportion of umpire
  decision outcomes by ball--strike count", digits = 2)

# --- Proportion of correct decisions by location and count --- ##
ump_dd <- pfx_13 %>%
  group_by(zone_reg, bs_count) %>%
  summarize(u_test = mean(u_test))
```

```
# ================================================================
# Multilevel Model Summary Code
# ================================================================

# --- Read in multilevel data frame *pfx_13* --- #
load("pfx_13.rda")

# --- Relevel "zone_reg" to make "ball" reference group --- #
pfx_13$zone_reg <- relevel(pfx_13$zone_reg, ref = "ball")

# --- Relevel "bs_count" to make "neutral" reference group --- #
pfx_13$bs_count <- relevel(pfx_13$bs_count, ref = "neutral")

# --- Inverse Logit Function --- #
inv.logit <- function(x){
  1 / (1 + exp(-x))
}

# --- Standard error function --- #
se <- function(var, length){
  sqrt(var) / sqrt(length)
}

# ================================================================
# Model 0
#
# Empty logistic regression model for predicting probability of
# accurate umpire decision
# ================================================================

m0 <- glm(u_test ~ 1, data = pfx_13, family = binomial)

# Print model output
summary(m0)

# Extract Model 0 coefficients
coef <- m0$coefficients

# Calculate odds ratio for Model 0
exp_coef <- exp(coef)

# Obtain 95% CI for odds ratio
exp_ci <- exp(confint(m0))

# Calculate inverse logit of exponentiated coefficient
ll <- exp_ci[1] / (1 + exp_ci[1])
ul <- exp_ci[2] / (1 + exp_ci[2])
```

```
# ================================================================
# Model 1
#
# Predcits probability of accurate umpire decision with
# random effect for umpire
# ================================================================

m1 <- glmer(u_test ~ 1 + (1 | umpire), data = pfx_13,
  family = binomial, nAGQ = 0)

# Extract umpire variance
m1_vc <- VarCorr(m1)
print(m1_vc, comp = c("Variance", "Std.Dev"), digits = 3)

# Transform m1 fixed effects
# 1. Extract fixed effects
(m1_fe <- fixef(m1))

# 2. Obtain odds ratios
(m1_odds <- exp(m1_fe))

# 3. Obtain probability
(m1_prob <- inv.logit(m1_fe))

# 4 Extract residual variance
m1_var <- summary(m1)$varcor$umpire[1, 1]

# 4.1 Calculate m1 ICC
(ump_icc <- m1_var / (m1_var + (pi^2) / 3))

# 5. m1 fixed effect CIs
m1_se <- sqrt(diag(vcov(m1)))
(m1_cis <- data.frame(Est = fixef(m1),
  LL = fixef(m1) - 1.96 * m1_se,
  UL = fixef(m1) + 1.96 * m1_se))
(m1_cis_exp <- exp(m1_cis))

# 6. m1 random effect CIs
m1_s2 <- m1@theta
m1_n <- nrow(ranef(m1)$umpire)
(m1_sd_cis <- data.frame(Est = m1_s2,
  LL = (m1_n - 1) * m1_s2 / qchisq(0.975, df = m1_n - 1),
  UL = (m1_n - 1) * m1_s2 / qchisq(0.025, df = m1_n - 1)))
```

```r
# ==============================================================
# Model 2
#
# Predicts probability of accurate umpire decision with fixed
# effects for pitch location (*zone_reg*), with random effects
# for umpire
# ==============================================================

m2 <- glmer(u_test ~ factor(zone_reg) + (1 | umpire),
  data = pfx_13, family = binomial, nAGQ = 0)

# ==============================================================
# Model 3
#
# Predicts probability of accurate umpire decision with fixed
# effects for ball-strike count (*bs_count*), with random effects
# for umpire
# ==============================================================

m3 <- glmer(u_test ~ factor(bs_count) + (1 | umpire),
  data = pfx_13, family = binomial, nAGQ = 0)

# ==============================================================
# Model 4
#
# Predicts probability of accurate umpire decision with fixed
# effects for pitch location (*zone_reg*) and ball-strike count
# (*bs_count*), with random effects for umpire
# ==============================================================

m4 <- glmer(u_test ~ factor(zone_reg) + factor(bs_count) + (1 | umpire),
  data = pfx_13, family = binomial, nAGQ = 0)

# ==============================================================
# Model 4a
#
# Predicts probability of accurate umpire decision with fixed
# effects and interactions for pitch location (*zone_reg*)
# and ball-strike count (*bs_count*), with random effects for
# umpire
# ==============================================================

m4a <- glmer(u_test ~ factor(zone_reg)*factor(bs_count) +
  (1 | umpire), data = pfx_13,
  family = binomial, nAGQ = 0)
```

```
# ================================================================
# Model 4b
#
# Predicts probability of accurate umpire decision with fixed
# effects and interactions for pitch location (*zone_reg*)
# and ball-strike count (*bs_count*), with random effects for
# pitch location, ball-strike count, and umpire
# ================================================================

m4b <- glmer(u_test ~ factor(zone_reg)*factor(bs_count) +
       (factor(zone_reg) + factor(bs_count) | umpire),
       data = pfx_13, family = binomial, nAGQ = 0)

# Extract random components for variance/covariance table
m4b_vc <- as.data.frame(VarCorr(m4b))
names(m4b_vc) <- c("Group", "V1", "V2", "VCov", "SD/Cor")
xtable(m4b_vc[, -1])

# Model 3d predicted probabilities table
pfx_13.pred <- pfx_13
X <- model.matrix(terms(m4b), data = pfx_13.pred)
b <- fixef(m4b)
pred.logit <- X %*% b

pred.prob  <- logit(pred.logit, inverse = TRUE)

pfx_13.pred2 <- data.frame(cbind("pred.logit" = pred.logit,
  "pred.prob" = pred.prob, zone = pfx_13.pred$zone_reg,
  count = pfx_13.pred$bs_count))

pfx_13.pred2$zone <- as.factor(pfx_13.pred2$zone)
levels(pfx_13.pred2$zone) <- c("Ball", "Inner", "Middle", "Outer")

pfx_13.pred2$count <- as.factor(pfx_13.pred2$count)
levels(pfx_13.pred2$count) <- c("Neutral", "Batter", "Pitcher")

colnames(pfx_13.pred2)[1:2] <- c("pred.logit","pred.prob")
(pred_prob_table <- unique(pfx_13.pred2[order(pfx_13.pred2$zone,
  pfx_13.pred2$count), ])
)
```

```
# ===============================================================
# Model 5
#
# Predicts probability of accurate umpire decision with grand-mean
# centered fixed effect for umpire experience and random effect
# for umpire
# ===============================================================

m5 <- glmer(u_test ~ scale(yr_exp) + (1 | umpire), data = pfx_13,
  family = binomial, nAGQ = 0)

# ===============================================================
# Model 6
#
# Predicts probability of accurate umpire decision with fixed
# effects for pitch location (*zone_reg*) and ball-strike count
# (*bs_count*), with interactions between pitch location, ball-
# strike count and umpire experience (*yr_exp*) with random
# effect for umpire
# ===============================================================

m6 <- glmer(u_test ~ factor(zone_reg)*scale(yr_exp) +
  factor(bs_count)*scale(yr_exp) +
  (1 | umpire), data = pfx_13, family = binomial, nAGQ = 0)


# ===============================================================
# Model Rankings
#
# Calculates model fit indices for each model estimated, stores
# results in new data frame, *results*, and ranks each model
# by AIC index
# ===============================================================

# Create lists of model summary output for each model
pfx_models <- list( )
pfx_models[[1]] <- m0
pfx_models[[2]] <- m1
pfx_models[[3]] <- m2
pfx_models[[4]] <- m3
pfx_models[[5]] <- m3a
pfx_models[[6]] <- m3b
pfx_models[[7]] <- m3c
pfx_models[[8]] <- m3d
pfx_models[[9]] <- m4
pfx_models[[10]] <- m5
pfx_models[[11]] <- m6
```

```r
# Vector of model names for *pfx_models*
model.names <- c("Model 0", "Model 1", "Model 2", "Model 3",
  "Model 3a", "Model 3b", "Model 3c", "Model 3d", "Model 4",
  "Model 5", "Model 6")

# Add  *model.names* to *pfx_models* list
names(pfx_models) <- model.names

# Create new data frame, *results* to contain output from
# model fit indices
results <- data.frame(models = model.names)
results$bic.val <- unlist(lapply(pfx_models, BIC))
results$bic.rank <- rank(results$bic.val)
results$aic.val <- unlist(lapply(pfx_models, AIC))
results$aic.delta <- results$aic.val-min(results$aic.val)
results$aic.likelihood <-  exp(-0.5* results$aic.delta)
results$aic.weight <- results$aic.likelihood/sum(results$aic.likelihood)
results <- results[rev(order(results[, "aic.weight"])), ]
results$cum.aic.weight <- cumsum(results[, "aic.weight"])

xtable(results)
```

REFERENCES

Abernethy, B. (1987a). Review: Selective attention in fast–ball sports I: General principles. *The Australian Journal of Science and Medicine in Sport, 19*(4), 3–6.

Abernethy, B. (1987b). Review: Selective attention in fast ball sports II: Expert novice differences. *The Australian Journal of Science and Medicine in Sport, 19*(4), 7–15.

Abernethy, B. (1987c). Expert–novice differences in an applied selective attention task. *Journal of Sport Psychology, 9*, 326–345.

Abernethy, B. (1988). The effects of age and expertise upon perceptual skill development in a racquet sport. *Research Quarterly for Exercise and Sport, 59*, 210–221.

Abernethy, B. (1989a). Expert–novice differences in perception: How expert does the expert have to be? *Canadian Journal of Sport Sciences, 14*(1), 27–30.

Abernethy, B. (1989b). Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sports Sciences, 8*, 17–34.

Abernethy, B (1990). Expertise, visual search, and information pick–up in squash. *Perception, 19*, 63–77.

Abernethy, B. (1991). Visual search strategies and decision–making in sport. *International Journal of Sport Psychology, 22*, 189–210.

Abernethy, B. (2008). Introduction: Developing expertise in sport, how research can inform practice. In D. Farrow, J. Baker, & C. MacMahon (Eds.), *Developing sport expertise: Researchers and coaches put theory into practice* (pp. 1–14). New York, NY: Routledge.

Abernethy, B., Farrow, D., & Berry, J. (2003). Constraints and issues in the development of a general theory of expert perceptual–motor performance. In J.L. Starkes & K.A. Ericsson (Eds.). *Expert performance in sports: Advances in research on sport expertise* (pp. 349–370). Champaign, IL: Human Kinetics.

Abernethy, B., Maxwell, J.P., Masters, R.S.W., Van der Kamp, J., & Jackson, R.C. (2007). Attentional processes in skill learning and expert performance. In Tenenbaum, G. & Eklund, R.C. (Eds.). *Handbook of sport psychology* (pp. 245–263). San Francisco: John Wiley & Sons, Inc.

Abernethy, B. & Mann, D. (2008). Dual pathways or dueling pathways for visual anticipation?: A response to van der Kamp, Rivas, van Doorn & Savelsbergh (2007). *International Journal of Sport Psychology, 39*, 131–141.

Ackerman, P. L., & Beier, M. E. (2006). Methods for studying the structure of expertise: Psychometric approaches. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Ackerman, P. L., & Lohman, D. F. (2006). Individual differences in cognitive functions. In P. A. Alexander & Winne, P. H. (Eds.). *Handbook of educational psychology* (pp. 139–161). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Adair, R. K. (2002). *The physics of baseball.* New York, NY: Harper–Collins Publishers Inc.

Agreste, S., De Meo, P., Ferrara, E., & Ursino, D. (2014). XML Matchers: Approaches and challenges. *Knowledge-Based Systems, 66*, 190–209.

Andersen, J. P., Prause, J., & Silver, R. C. (2011). A step-by-step guide to using secondary data for psychological research. *Social and Personality Psychology Compass, 5*(1), 56–75.

Bar–Eli, M., Plessner, H., & Raab, M. (2011). *Judgment, decision making and success in sport.* West Sussex, UK: Wiley–Blackwell.

Baseball-Reference. (2012). *Bill Klem.* Retrieved from: `http://www.baseball-reference.com/bullpen/Bill_Klem`

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4.* R package version 1.1-6. http://CRAN.R-project.org/package=lme4

Bradbury, J. C. (2009). Peak athletic performance and ageing: Evidence from baseball. *Journal of Sports Sciences, 27*, 599–610.

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological review, 6*, 345–375.

Burnham, K. P., & Anderson, D. R. (2004). *Sociological Methods & Research, 33*(2), 261–304.

Burnham, K. P., Anderson, D. R. & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65*(1), 23–35.

Buss, D. (2012). *Evolutionary psychology: The new science of the mind.* Boston: Pearson.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.). *Visual information processing.* Oxford, England: Academic.

Chi, M. T. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Christian, T., Beaujean, A. A., & Wright, W. (2014). *Individual differences in affective states during meditation.* Manuscript submitted for publication.

Cote, J., & Fraser–Thomas, J. (2008). Play, practice, and athlete development. In D. Farrow, J. Baker, & C. MacMahon (Eds.), *Developing sport expertise: Researchers and coaches put theory into practice.* New York, NY: Routledge.

Csikszentmihalyi, M. (1996). *Creativity: flow and the psychology of discovery and invention.* New York, NY: HarperCollins.

Csikszentmihalyi, M. (2010). Implications of a systems perspective for the study of creativity. In Sternberg, R.J. (Ed.). *Handbook of creativity.* New York: Cambridge University Press.

Davids, K., & Williams, A. M. (1998). Visual search strategy, selective attention, and expertise in soccer. *Research Quarterly for Exercise and Sport, 69*(2), 111–126.

De Groot, A. D. (1946/1978). Thought and choice in chess. Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In Davidson, J. E. & Sternberg, R. J. (Eds.). *The psychology of problem solving.* New York: Cambridge University Press.

Enders, C. K. (2005). Maximum likelihood estimation. In B. Everitt & D. C. Howell (Eds.), *Encyclopedia of behavioral statistics* (pp. 1164–1170). West Sussex, England: Wiley.

Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In J. E. Davidson & R. J. Sternberg (Eds.). *The psychology of problem solving* (pp. 31–85). New York, NY: Cambridge University Press.

Ericsson, K. A. (2006a). An introduction to Cambridge handbook of expertise and expert performance: Its development, organization, and content. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Ericsson, K. A. (2006b). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Ericsson, K. A. (2006c). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Ernst, G. W. and A. Newell (1969). *GPS: A case study in generality and problem–solving.* New York: Academic Press.

Fadde, P. J. (2006). Interactive video training of perceptual decision–making in the sport of baseball. *Technology, Instruction, Cognition, and Learning, 4*, 237–255.

Fast, M. (2011, February 16). *Spinning yarn: The real strike zone.* Baseball Prospectus. Retrieved from http://www.baseballprospectus.com/article.php?articleid=12965

Fast, M. (2011, September 7). *Spinning yarn: Home plate umpire positioning.* Baseball Prospectus. Retrieved from http://www.baseballprospectus.com/article.php?articleid=14951

Fast, M. (2011, September 24). *Spinning yarn: Removing the mask encore presentation.* Baseball Prospectus. Retrieved from http://www.baseballprospectus.com/article.php?articleid=15093

Fast, M. (2011, October 8). *BP unfiltered: NLCS umpire charts and data.* Baseball Prospectus. Retrieved from http://www.baseballprospectus.com/article.php?articleid=15269\&mode=print\&nocache=1334171022

Feltovich, P. J., Prietula, J. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Ford, G. G., Gallagher, S. H., Lacy, B. A., Bridwell, A. M., & Goodwin, F. (1999). Repositioning the home plate umpire to provide enhanced perceptual cues and more accurate ball–strike judgments. *Journal of Sport Behavior, 2*(1), 28–44.

French, K. E., & McPherson, S. L. (1999). Adaptations in response selection processes used during sort competition with increasing age and expertise. *International Journal of Sport Psychology, 30*, 178–193.

Gassko, D. (2007, February 01). *The outside corner.* The Hardball Times. Retrieved from http://www.hardballtimes.com/main/article/the--outside--corner/

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press.

Ghasemi, A. Momeni, M., Jafarzadehpur, E., Rezaee, M., & Taheri, H. (2011). Visual skills involved in decision making by expert referees. *Perceptual and Motor Skills, 112*(1), 161–171.

Goldblatt, A. (2011). *Major League umpires' performance, 2007–2011: A comprehensive statistical review.* Jefferson, NC: McFarland & Company, Inc., Publishers.

Goodale, M. A., & Milner, A. D. (2005). *Sight unseen: An exploration of conscious and unconscious vision.* New York: Oxford.

Goodale, M. A., & Westwood, D. A. (2004). An evolving view of duplex vision: Separate but interacting cortical pathways for perception and action. *Current Opinion in Neurobiology, 14*, 203–211.

Green, E. & Daniels, D. P. (2014). Higher stakes, more mistakes: Decision bias and strategic error in Major League Baseball. Unpublished Manuscript. Retrieved from: http://www.etangreen.com/uploads/2/6/4/5/26458476/140320_green_daniels.pdf

Grimes, D. A., & Schulz, K. F. (2008). Making sense of odds and odds ratios. *Obstetrics & Gynecology, 111*(2), 423–426.

Hale, J. (2007, November 28). *A zone of the their own.* The Hardball Times. Retrieved from http://www.hardballtimes.com/main/article/a--zone--of--their--own/

Hamrick, J. & Rasp, J. (2013). The connection between race and called strikes and balls. *Journal of Sports Economics, 00*(0), 1-21.

Haynes, B. R. (2006). Forming research questions. *Journal of Clinical Epidemiology, 59*(9), 881–886.

Heck, R., Thomas, S., & Tabata, L. (2013). *Multilevel modeling of categorical outcomes using IBM SPSS.* Taylor & Francis.

Helsen, W., & Bultynck, J. B. (2004). Physical and perceptual–cognitive demands of top–class refereeing in association football. *Journal of Sports Sciences, 22*(2), 179–189.

Hodges, N. J., Huys, R., & Starkes, J. L. (2007). Methodological review and evaluation of research in expert performance in sport. In Tenenbaum, G. & Eklund, R. C. (Eds.). *Handbook of sport psychology* (pp. 161–183). San Francisco: John Wiley & Sons, Inc.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods, 39*(1), 101–117.

Hoffman, R. R., & Lintern, G. (2006). Eliciting and representing the knowledge of experts. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression.* Hoboken, NJ: Wiley.

Horn, J., & Masunaga, H. (2006). A merging theory of expertise and intelligence. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Houlston & Lowes (1993). Anticipatory cue–utilization processes amongst expert and non–expert wicketkeepers in cricket. *International Journal of Sport Psychology, 24*, 59–73.

Hox, J. (2010). *Applied multilevel analysis*. Routledge.

Jackson, D. L. (2010). Reporting results of latent growth modeling and multilevel modeling analyses: some recommendations for rehabilitation psychology. *Rehabilitation Psychology, 55*(3), 272–285.

Jackson, S. A., Kimiecik, J. C., Ford, S. K., & Marsh, H. W. (1998). Psychological correlates of flow in sport. *Journal of Sport & Exercise Psychology, 20*, 358–378.

Janelle, C. M., & Hillman, C. H. (2003). Expert performance in sport: Current Perspectives and Critical Issues. In K. A. Ericsson & J. Starkes (Eds.), *Recent Advances in Research on Sport Expertise* (pp. 19–47). Human Kinetics: Champaign, IL.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences, 17*, 187–202.

Kotovsky, K. (2003). Problem solving–large/small, hard/easy, conscious/nonconscious, problem–space/problem–solver: The issue of dichotomization. In J. E. Davidson & R. J. Sternberg (Eds.). *The psychology of problem solving* (pp. 373–384). New York, NY: Cambridge University Press.

Krampe, R. T., & Charness, N. (2006). Aging and expertise. In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Krane, V., & Williams, J. M. (2010). Psychological characteristics of peak performance. In J. M. Williams (Ed.), *Applied sport psychology: Personal growth to peak performance*. New York, NY: McGraw–Hill.

Lindbergh, B. (October 10, 2014). Rung Up: Are Postseason Umpires Actually Baseball's Most Accurate?. Retrieved from: *http://grantland.com/the-triangle/postseason-umpires-mlb-accurate-joe-west/*

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology, 16*(1), 149–157.

MacMahon, C., & Plessner, H. (2008). The sports official in research and practice. In D. Baker, C. Farrow, J. Baker, & C. MacMahon (Eds.). *Developing elite sports performers: Lessons from theory and practice* (pp. 172–192). London, UK: Routledge.

MacMahon, C. & Starkes, J. L. (2008). Contextual influences on baseball ball–strike decisions in umpires, players, and controls. *Journal of Sports Sciences, 26*(7), 751–760.

MacMahon, C., Helson, W. F., Starkes, J. T., & Weston, M. (2007). Decision–making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences, 25*(1), 65–78.

MacMahon, C., Starkes, J., & Deakin, J. (2007). Referee decision making in a video–based infraction detection task: Application and training considerations. *International Journal of Sports Sciences & Coaching, 2*(3), 257–265.

Major League Baseball (2011). *The 2011 MLB umpire media guide*. New York: The Office of the Commissioner of Baseball.

Major League Baseball Advanced Media [MLB Advanced Media] (2014). Gameday Data. Retrieved from http://gd2.mlb.com/components/game/mlb/

Mann, D. L., Ho, N. Y., De Souza, N. J., Watson, D. R., & Taylor, S. J. (2007). Is optimal vision required for the successful execution of an interceptive task? *Human Movement Science, 26*, 343–356.

Mann, D. L., Williams, A. M., Ward, P., & Janelle, C.M. (2007). Perceptual–cognitive expertise in sport: A meta–analysis. *Journal of Sport & Exercise Psychology, 29*(4), 457–478.

Mann, D. L., Abernethy, B., & Farrow, D. (2010). Action specificity increases anticipatory performance and the expert advantage in natural interceptive tasks. *Acta Psychologica, 135*, 17–23.

Maas, C. J. M., & Snijders, T. A. B. (2003). The multilevel approach to repeated measures for complete and incomplete data. *Quality & Quantity, 37*, 71–89.

Marchi, M., & Albert, J. (2013). *Analyzing baseball data with R*. Taylor & Francis.

Maxwell, S., & Delaney, H. (2004). *Designing experiments and analyzing data: A model comparison perspective*. Lawrence Erlbaum Associates.

McGuire, W. J. (1997). Creative hypothesis generating in psychology: some useful heuristics. *Annual Review of Psychology, 48*, 1–30.

McPherson, S. L., & Kernodle, M. W. (2003). Tactics, the neglected attribute of expertise: Problem representations and performance skills in tennis. In J. L. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 137–168). Champaign, IL: Human Kinetics.

Mills, B. M. (2013) Social pressure at the plate: Inequality aversion, status, and mere exposure. *Managerial and Decision Economics, 35*(6), p. 387–403.

Millslagle, D. G., Hines, B. B., & Smith, M. S. (2013). Quiet eye gaze behavior of expert, and near-expert, baseball plate umpires. *Perceptual & Motor Skills, 116*(1), 69–77.

Milner, D., & Goodale, M. (2008). The two visual streams: In the right ballpark? *International Journal of Sport Psychology, 39*, 131–135.

Moore, M. (2013). *Balls and strikes: Every pitch counts.* Franksville, WI: Referee Enterprises, Inc.

Moran, A. (2009). Cognitive psychology in sport: Progress and prospects. *Psychology of Sport and Exercise, 10*(4), 420–426.

Moran, A. P. (1996). *The psychology of concentration in sport performers.* East Sussex, UK: Psychology Press, Publishers.

Nathan, A. M. (2008). *A Statistical Study of PITCHf/x Pitched Baseball Trajectories.* Unpublished Manuscript.

Nordin, S. M., Cumming, J., Vincent, J., & McGrory, S. (2006). Mental practice or spontaneous play? Examining which types of imagery constitute deliberate practice in sport. *Journal of Applied Sport Psychology, 18*(4), 345–362.

Osborne, E. (2010). Why do some kinds of stars get the calls? *Journal of Sports Economics, 11*(2), 203-213.

Parsons, C. A., Sulaeman, J., Yates, M. C., & Hamermesh, D. S. (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review 101*(4) (June 2011), 1410–1435.

Paull, G., & Glencross, D. (1997). Expert perception and decision making in baseball. *International Journal of Sport Psychology, 28*(1), 35–56.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85–112.

Powers, D. A. (2012). Multilevel models for binary data. *New Directions for Institutional Research* (154), 57–75.

Quene, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication, 43*(1–2), 103–121.

Rainey, D. W. (1994). Magnitude of stress experienced by baseball and softball umpires. *Perceptual and Motor Skills, 79*(1), 255–258.

R Core Team (2014). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Rainey, D. (1995). Sources of stress among baseball and softball umpires. *Journal of Applied Sport Psychology, 7*(1), 1–10.

Rainey, D. W., Larsen, J.D., & Stephenson, A. (1989). The effects of a pitchers reputation on umpires calls of balls and strikes. *Journal of Sport Behavior, 12*(3), 139–150.

Rainey, D. W. & Larsen, J. D., Stephenson, A., & Coursey, S. (1988). Accuracy and certainty judgments of umpires and nonumpires. *Journal of Sport Behavior, 12*(1), 12–22.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods.* SAGE Publications.

Rodriguez, G., & Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society. Series A, 58*(1), 73–89.

Ross, K. G., Shafer, J. L., & Klein, G. (2006). Professional judgments and "naturalistic decision making." In K. A. Ericsson, N. Charness, P. Feltovich, and R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.

Russell, J.S. (1997). The concept of a call in baseball. *Journal of the Philosophy of Sport, XXIV*, 21–37.

Office of the Commissioner of Baseball, The (2012). *Official baseball rules.* New York, NY: Major League Baseball.

Sage, G. (1984). *Motor learning and control.* Dubuque, IA: Brown.

Sheskin, D. (2003). *Handbook of parametric and nonparametric statistical procedures.* CRC Press.

Shim, J., Carlton, L. G., Chow, J. W., & Chae, W. (2005). The use of anticipatory visual cues by highly skilled tennis players. *Journal of Motor Behavior, 37*(2), 164–175.

Short, S. E., Bruggeman, J. M., Engel, S. G., Marback, T. L., Wang, L. J., Willadsen, A., & Short, M. W. (2002). The effect of imagery function and imagery direction on self–efficacy and performance on a gulf–putting task. *The Sport Psychologist, 16*(1), 48–67.

Sievert, C. (2014). Taming PITCHf/x data with `pitchRx` and `XML2R` *The R Journal, 6*(1). URL http://journal.r-project.org/archive/accepted/sievert.pdf.

Sievert, C. (2014). ***pitchRx**: Tools for harnessing MLBAM Gameday data and visualizing PITCHf/x.* R package version 1.4.

Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford University Press.

Smith, R. E., & Christensen, D.S. (1995). Psychological skills as predictors of performance and survival in professional baseball. *Journal of Sport & Exercise Psychology, 17*(4), 399–415.

Snijders, T., & Bosker, R. (n.d.). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*, 1999. Sage, London.

Sportvision, Inc. (2013). *PITCH/fx*. Retrieved from https://www.sportvision.com/baseball/pitchfx.

Sternberg, R. J. (1996). Costs of expertise. In K.A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*. Mahwah, NJ: Erlbaum.

Stewart, D. W. (2012). Secondary analysis and archival research: Using data collected by others. In H. Cooper, P. M. Camic, D. L. Long, & A. T. Panter (Eds). *APA handbook of research methods in psychology, Vol. 3: Data analysis and research publication*. Washington, DC: American Psychological Association.

Stewart, M. J., Ellery, P. J., Ellery, J., & Maher, L. (2004). Perceived psychological stress among high school basketball officials. *Perceptual and Motor Skills, 99*, 463–469.

Szumilas, M. (2010). Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry, 19*(3), 227-229.

Takeuchi, T., & Inomata, K. (2009). Visual search strategies and decision making in baseball batting. *Perceptual and Motor Skills, 108*(3), 971–980.

Tango, T., Lichtman, M., & Dolphin, A. (2007). *The book: Playing the percentages in baseball*. Potomac Books.

Tainsky, S., Mills, B. M. & Winfree, J. A. (2013). Further examination of potential discrimination among MLB umpires. *Journal of Sports Economics, 00*(0), 1-22.

Tenenbaum, G., Sar–El, T., & Bar–Eli, M. (2000). Anticipation of ball location in low and high–skill performers: A developmental perspective. *Psychology of Sport and Exercise, 1*, 117–128.

Thorn, J., Palmer, P., & Reuther, D. (1984). *The hidden game of baseball: A revolutionary approach to baseball and its statistics*. Doubleday.

Utz, S.G. (1989). The authority of the rules of baseball: The commissioner as judge. *Journal of the Philosophy of Sport, XVI*, 88–99.

Van der Kamp, J., Rivas, F., Van Doorn, H., & Savelsbergh, G. (2008). Ventral and dorsal system contributions to visual anticipation in fast ball sports. *International Journal of Sport Psychology, 39*, 10–130.

Vealy, R. S., & Greenleaf, C.A. (2010). Seeing is believing: Understanding and using imagery in sport. In J.M. Williams (Ed.), *Applied sport psychology: Personal growth to peak performance.* New York, NY: McGraw–Hill.

Wagenmakers, E. & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192–196.

Walsh, J. (2007, July 25). *The eye of the umpire.* The Hardball Times. Retrieved from `http://www.hardballtimes.com/main/article/the--eye--of--the--umpire/`

Walsh, J. (2010, April 7). *The compassionate umpire.* The Hardball Times. Retrieved from `http://www.hardballtimes.com/main/article/the--compassionate--umpire/`

Ward, P., & Williams, A. M. (2003). Perceptual and cognitive skill development in soccer: The multidimensional nature of expert performance. *Journal of Sport & Exercise Psychology, 25*, 93–111.

Ward, P., Farrow, D., Harris, K. R., Williams, A. M., Eccles, D. W., & Ericsson, K.A. (2008). Training perceptual–cognitive skills: Can sport psychology research inform military decision training? *Military Psychology, 20*(Suppl. 1), S71–S102.

Ward, P., Williams, A. M., & Ericsson, K. A. (2003). Underlying mechanisms of perceptual–cognitive expertise in soccer. *Journal of Sport and Exercise Psychology, 25*, S136.

Weber, B. (2010). *As they see 'em: A fan's travels in the land of umpires.* New York: Scribner.

Weisberg, R. W. (2006). Expertise and reason in creative thinking: Evidence from case studies and the laboratory. In J. C. Kaufman & J. Baer (Eds.). *Creativity and reason in cognitive development* (pp. 7–42). New York, NY: Cambridge University Press.

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software, 40*(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998) The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodoligical recommendations. *Development and Psychopathology, 10*, 395–426.

Williams, A. M., Davids, K., & Williams, J. G. (1999). *Visual perception & action in sport.* New York: Taylor & Francis.

Williams, A. M. & Ericsson, K.A. (2005). Perceptual–cognitive expertise in sport: Some considerations when applying the expert performance approach. *Human Movement Science, 24*(3), 283–307.

Williams, A. M. & Ward, P. (2007). Anticipation and decision making: Exploring new horizons. In Tenenbaum, G. & Eklund, R.C. (Eds.). *Handbook of sport psychology* (pp. 203–223). San Francisco: John Wiley & Sons, Inc.

Williams, A. M., Ford, P. R., Eccles, D. W., & Ward, P. (2011). Perceptual–cognitive expertise in sport and its acquisition: Implications for applied cognitive psychology. *Applied Cognitive Psychology, 25*(3), 432–442.

Williams, A. M., Ward, P., & Smeeton, N.J. (2004). Perceptual and cognitive expertise in sport: Implications for skill acquisition and performance enhancement. In A.M. Williams & , N. J. Hodges (Eds.), *Skill acquisition in sport: research, theory and practice*. Routledge: London.

Wong, G. Y., & Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association, 80*(391), 513–524.

Yates, F. J., & Tschirhart, M.D. (2006). In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman, R. R. (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge, UK: Cambridge University Press.