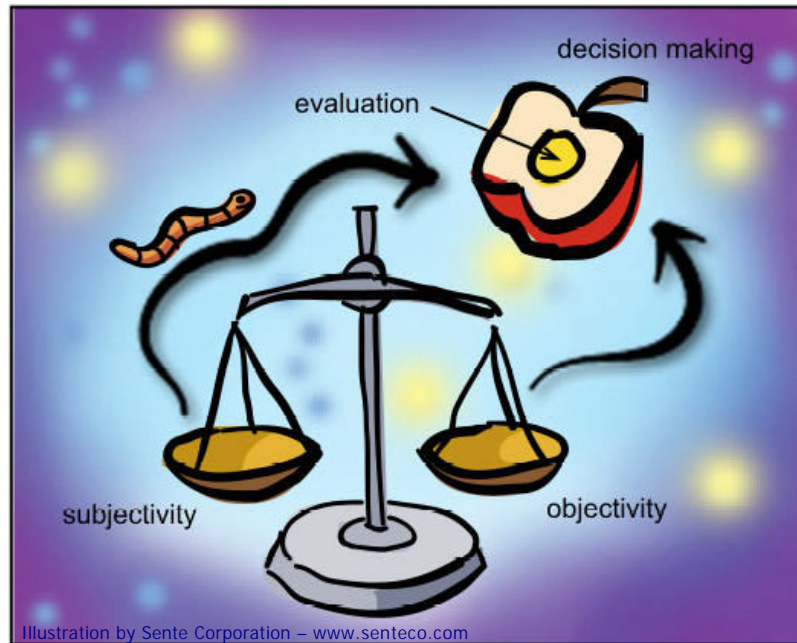# The Life of Game

Sports, Stats, and the Lessons they Teach Us

Baseball, football, and basketball possess the defining property of drama, which is tension and release—that is, uncertainty ultimately relieved by a definitive conclusion.

Michael Mandelbaum
The Meaning of Sports [1]

Illustration by Sente Corporation – www.senteco.com

- **Statistics provide a way to understand a game or situation that our intuition may not capture.**

- **By and large, intuition and tradition have played a larger role than statistics in shaping sports measurement, management, and play. The same is true in business and investing.**

- **Teams are stronger or weaker across a range of performance variables, or dimensions. Because teams have different strengths across dimensions, you can't really say there is one best team. The problem looks like a game of rock, paper, and scissors.**

- **The same psychological traps are at play in sports, business, and investing.**

- **Many sports and business organizations are too short-term oriented, and inappropriately focused on outcome versus process.**

Which Q for Your Eye?

People often describe sports as a microcosm of life. Sports capture everyday issues like discipline and disorder, planning and improvisation, cooperation and competition, skill and luck. Combine these with big money in professional sports, and it's no wonder the newspapers and airwaves are filled with who's hot and who's not.

But when trying to evaluate individual and team performance, sports team managers and pundits face a fundamental question: what's more reliable, quantitative statistics or qualitative intuition? While the right answer lies somewhere between the extremes, many in sports have a hard time striking an optimal balance.

In June 2004, the Santa Fe Institute hosted a one-day conference directly addressing the above question. While the discussion was ostensibly about sports, many of the frameworks and conclusions relate directly to business and investing.

What You See versus What's Happening

Most players, coaches, and fans evaluate games using perception and past practice. By and large, intuition and tradition assume a larger role than statistics in shaping sports measurement, management, strategy, and play.

However, there may be a difference between what we see when we watch a game and what is going on. To illustrate this point, economist Colin Camerer projected an image that flickered between two nearly identical scenes and asked the participants to spot the difference.[2] Many couldn't within the time allowed, even though the difference was clear once Camerer pointed it out.

The message: we don't always perceive all that's going on in a scene. Statistics may help us by providing ways to understand a game or situation that our intuition or perception may not capture.

Michael Lewis's bestseller, *Moneyball*, drilled home this point and created a stir in the sports and business worlds when it showed how the low-budget Oakland Athletics fielded very competitive teams by using statistics more effectively than their rivals. The A's relied not only on the "tools" that baseball traditionally prized (ability to hit, hit for power, throw, run, and field) but also on skills—the ability to get the job done. As a result, players that looked statistically attractive but didn't fit the ideal player mold were relatively cheap. Like good value investors, the A's scooped up a portfolio of undervalued players and won as many games as many big market teams at a fraction of the cost.

BC and CB (Bowl Championship and Colonel Blotto)

The conference's first session was about ranking teams. Owners, fans, coaches, and players all have an interest in determining which team is "best." Standings and tournaments are two ways to answer the question. Yet both speakers concluded that team ranking is extremely difficult and potentially very misleading. Many of these lessons apply to the business world—especially in markets for goods and services.

Mathematician Ken Massey opened by discussing challenges associated with ranking teams. Massey should know; his Massey Ratings form part of the Bowl Championship Series (BCS) Rankings that determine which college football teams will play in key bowl games. (See http://www.masseyratings.com/.)

Rating systems attempt to objectively measure each team's performance relative to the schedule it faces. They differ from polls, standings, or point systems. Massey noted a number of hurdles in creating rating systems, including:

- **A lack of transitivity.** Just because team A beats team B, and team B beats team C, doesn't mean that team A will beat team C. (See Exhibit 1.)

- **Disparate schedules.** The rankings must compare teams playing very different schedules. Comparing a losing team with a strong schedule to a winning team with a weak schedule is difficult.

- **Noise.** Lots of factors come into play that shape a team's performance versus its potential. These include the environment (venue, weather, crowd), physical (injuries, travel, elevation), and luck.

Exhibit 1: Transitivity Doesn't Hold in College Football (2003 Results)

Winner	Loser	Winner	Loser
Chowan 21	Randolph Macon 20	JMU 48	Liberty 6
Randolph Macon 10	Emory & Henry 7	Liberty 49	Hofstra 42
Emory & Henry 33	Methodist 30	Hofstra 34	Villanova 32
Methodist 37	Ferrum 34	Villanova 23	Temple 20
Ferrum 19	C. Newport 17	Temple 44	Mid TN State 36
C. Newport 16	Bridgewater 12	Mid TN State 27	Troy State 20
Bridgewater 58	Catholic 20	Troy State 33	Marshall 24
Catholic 32	LaSalle 31	Marshall 27	Kansas State 20
LaSalle 33	Marist 31	Kansas State 42	Univ. of California 28
Marist 33	Central CT 29	Univ. of California 52	Virginia Tech 49
Central CT 14	Monmouth 10	Virginia Tech 31	Miami FL 7
Monmouth 12	Georgetown 10	Miami FL 38	Univ. of Florida 33
Georgetown 17	Lafayette 10	Univ. of Florida 19	LSU 7
Lafayette 41	Columbia 27	LSU 21	Univ. of Oklahoma 14
Columbia 16	Harvard 13	Univ. of Oklahoma 59	UCLA 24
Harvard 28	Northeastern 20	UCLA 23	Univ. of California 20
Northeastern 41	JMU 24	Univ. of California 34	USC 31

Prediction: Chowan by 307 points over USC
Univ. of California would beat themselves by 89 points

Massey reviewed a number of mathematical techniques to rank teams, including a least squares, maximum likelihood, and Markov chains. While Massey ratings are designed to measure past performance, they also have predictive value. Massey noted that his best maximum likelihood model correctly predicts the outcome of roughly three-quarters of pro baseball games and about two-thirds of NFL games.

University of Michigan social scientist Scott Page opened his talk by challenging three widely held premises: there is a best team; you can rank teams; and settling it on the field is better than looking at statistics.

The issue of multi-dimensionality creates the main problem when determining the best team. Teams are stronger or weaker across a range of performance variables, or dimensions. Examples of performance variables in football include running offense, running defense, and special teams. Since teams of similar ability have different strengths across performance variables, the problem of determining which team is better is difficult. Page suggested the problem looks like the rock, paper, and scissors game.

Page introduced the group to Colonel Blotto, a two-player competitive game model. Here's a simple version:

- Both players get 100 playing pieces
- There are three locations to place the pieces
- The player who places the most pieces on a location wins that location
- The player who wins the most locations wins the game

Here's an illustration:

	L1	L2	L3
Player 1	42	32	26
Player 2	21	34	45

Player 2 wins two to one.

Page suggests that you can think about each location as a dimension. Given these rules, one can choose some really bad strategies (100, 0,0—akin to having a great quarterback with no offensive line) but by-and-large winning is random if two teams have similar abilities. (See Exhibit 2. [3]) Even if teams have uneven ability, it often takes a lot more ability to make a difference.

Exhibit 2: Colonel Blotto's Pinwheel: When Ability is Similar, Winning is Random
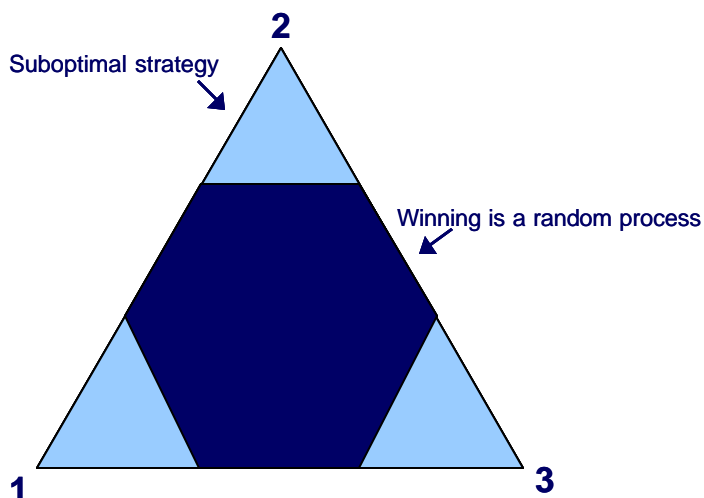
Source: Scott E. Page, "On the Possibility of Value in Sports", Presented at *A Complex Look at Sports* Conference, June 17, 2004. Used by permission.

The more dimensions a game has, the less predictable it becomes. We should expect (and see) more upsets in high dimension games like football than in low dimension games like tennis and wrestling. No surprise then that the betting line, which reflects the experience and intuitions of bettors, often doesn't match the computer rankings. Bettors may be able to aggregate dimensions better than the ranking models or polls can.

Page then discussed what he calls "the general manager's backpack" problem. General managers have to put together a team of players with varying dimensions within a finite (except for the Yankees) budget. This problem is extremely hard to solve, especially when the solution relies to some degree on what other teams are doing.

Paradoxically, Page noted that if you add complexity—for example, make the strength of one dimension contingent on another—the problem becomes simpler. To illustrate, the ability to run the fast break in basketball is more valuable with good defensive rebounding. Skillful general managers select teams to take advantage of these contingencies. This line of thinking might help explain why the Detroit Pistons "upset" the heavily favored Los Angeles Lakers in the 2004 NBA championship.

Go For It

The next pair of speakers, Cal Berkeley economist David Romer and USC football's offensive coordinator Norm Chow, discussed football play calling in general and kicking strategy in particular. This pair offered the clearest dichotomy between statistics-driven theory and intuition-and-habit-driven practice.

Romer's paper, "It's Fourth Down and What Does the Bellman Equation Say?", analyzes pro football play selection.[4] His data come from all NFL games in the 1998-2000 seasons, and provide expected points for various positions on the field and down situations. Romer's analysis shows that NFL coaches call plays too conservatively: they don't go for first downs frequently enough and too often settle for field goals when they should attempt to score a touchdown. [5] Exhibit 3 summarizes Romer's findings and recommendations.

Exhibit 3: To Kick or Not to Kick
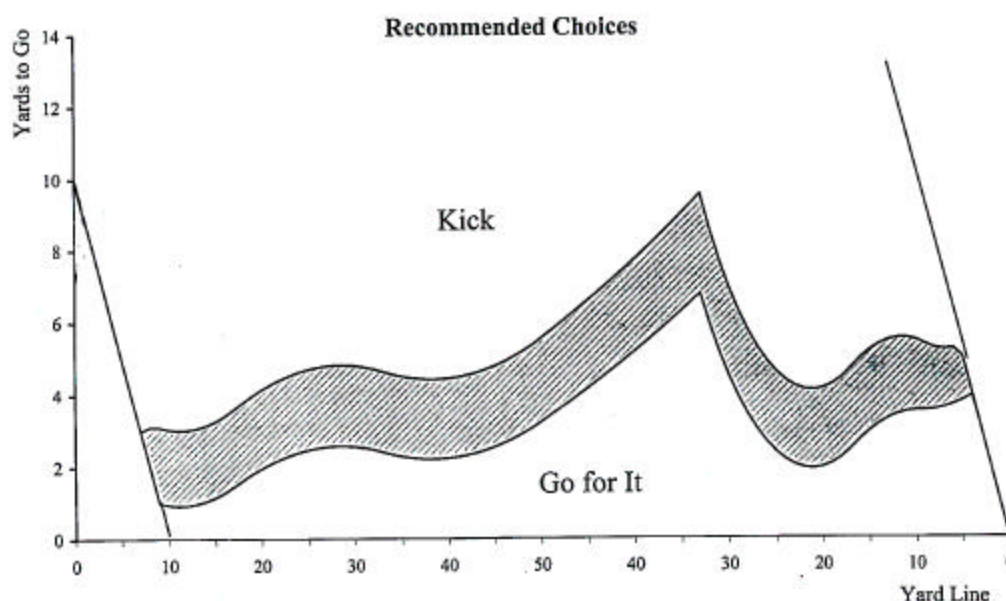
Source: David Romer, "It's Fourth Down and What Does the Bellman Equation Say?", *Working Paper,* February 2003. Used by permission.

For example, of the 532 fourth downs in the offense's half of the field where Romer's analysis suggests going for it, teams only went for it eight times. In the 183 fourth downs with five or more yards to go where the analysis indicates teams should go for it, they did so only 13 times. Romer notes that this sub optimal play calling might cost NFL teams one game per year.

Players, coaches, and fans often find Romer's work objectionable, pointing to factors like momentum, third versus fourth down plays, and selection bias (the data are true only for average teams). Romer addresses these concerns head on, making a case that the objections are either weak or invalid.

Take the issue of momentum. Convention holds that if a team stops its opponent's fourth-down attempt, it gains an energy and emotional edge. Romer's skepticism about this stems from two sources. First, while the analysis doesn't take into account the deflating downside of a fourth down failure, it also doesn't incorporate the lift from a successful fourth down play. Second, other studies of momentum found weak evidence for momentum effects in the data. Said differently, statistics reflect emotion.

In stark contrast to Romer, coach Chow emphasized the role of emotion. He clearly believes in momentum and streaks, notwithstanding that both Romer's analysis and a substantial body of evidence suggest sports outcomes are generally consistent with probabilities. [6]

Chow also underscored the risk aversion inherent in sports. Coaches generally eschew new strategies because they don't want to lose—at least not in an unconventional fashion. What John Maynard Keynes said about investing applies to sports: "worldly wisdom teaches that it is better for the reputation to fail conventionally than to succeed unconventionally."

Chow's take on statistical analysis brought Camerer's point about perception into sharp focus. The approach most coaches use is reminiscent of Fisher Black's famous 1986 paper "Noise" discussing how investors operate: [7]

> *Because there is so much noise in the world, people adopt rules of thumb. They share their rules of thumb with each other, and very few people have enough experience with interpreting noisy evidence to see that the rules are too simple.*

Like many practitioners, Chow relies heavily on rules of thumb and traditional approaches. That said, Chow and his teams have been very successful, so his intuition and pattern recognition skills are probably well honed.

The Art and Science of Baseball

Los Angeles Dodgers general manager Paul DePodesta, an advocate of using statistical techniques to build and manage baseball teams, discussed the limitations of statistics and common decision-making foibles. He emphasized the large amount of day-to-day noise, based on factors like personal issues, media attention, and fan reactions, mixed in with statistics.

What biases does DePodesta see? First, there are emotional biases—like overemphasizing most recent outcomes, oversimplifying complex situations, and adherence to conventional wisdom. These biases often encourage poor decisions.

Second, he noted the short-term orientation of many organizations. For example, half of major league baseball's 30 general managers have fewer than three years of tenure, eight have fewer than two years of service, and only one has been at his post for more than a decade. Similarly, a slew of new coaches joined the NBA in the 2003-04 season. As a result, various people might have different and conflicting time horizons even within an organization. For example, a field manager with a one-year contract may seek to field a squad very differently than a general manager with a five-year contract.

Finally, DePodesta emphasized the importance of process versus outcome in decision-making. He acknowledged the intense difficulty involved in avoiding this trap, but reinforced the importance of the distinction in good decisions.

Basketball and the Brain

Dean Oliver, author of *Basketball on Paper,* provided some insight on how to break down statistics in basketball. Basketball sits between baseball (lots of one-on-one interaction) and football (eleven-on-eleven) in complexity, or number of dimensions.

As Oliver described it, basketball boils down to number of possessions and how efficiently each team performs with their possessions. Taking it one step further, Oliver argues that four factors define the crucial aspects of basketball: [8]

1. Field goal shooting percentage
2. Offensive rebounds
3. Committing turnovers
4. Going to the foul line (and making the shots)

This approach allows Oliver to rate teams and players. For example, he has developed a software program, *Roboscout*, to help predict game outcomes. (See http://www.82games.com/.)

Basketball on Paper includes some analysis that applies well beyond basketball and sports. The first is another take on streak data. Stephen Jay Gould said that streaks are "luck imposed on skill." Oliver shows this with data in Exhibit 4. Streaks fall within the realm of statistics: probability shows that the longest streaks accrue to the best performers (in Oliver's example, teams with higher win percentages).

Exhibit 4: Chance of at Least One Winning Streak of the Shown Length in 82-Game Season

Win Streak	0.200 Team	0.300 Team	0.400 Team	0.500 Team	0.600 Team	0.700 Team	0.800 Team
3	47%	89%	99.50%	100.00%	100.00%	100.00%	100.00%
4	12%	47%	87%	99.39%	100.00%	100.00%	100.00%
5	2.5%	17%	55%	91.6%	99.82%	100.00%	100.00%
6	0.49%	5.5%	27%	70%	97.5%	99.99%	100.00%
7	0.10%	1.6%	12%	45%	88%	99.85%	100.00%
8	0.02%	0.49%	4.8%	25%	72%	98.8%	100.00%
9	0.00%	0.15%	1.9%	13%	53%	95.3%	100.00%
10	0.00%	0.04%	0.76%	6.9%	36%	88%	99.97%
11	0.00%	0.01%	0.30%	3.5%	23%	76%	99.84%
12	0.00%	0.00%	0.12%	1.7%	14%	63%	99.38%
13	0.00%	0.00%	0.05%	0.85%	8.7%	49%	98.1%
14	0.00%	0.00%	0.02%	0.42%	5.3%	37%	95.5%
15	0.00%	0.00%	0.01%	0.21%	3.1%	28%	91.2%

Source: Dean Oliver, *Basketball on Paper* (Washington, D.C: Brassey's, Inc., 2004), 70. Copyrighted material, used by permission.

Another form of analysis is reversion to the mean. Based on NBA data, Oliver shows that both losing and winning teams tend to revert back to average—a .500 win percentage—over time. These analyses clearly parallel aspects of business and investing.

Exhibit 5: Five Years After: Most Teams Approach 0.500
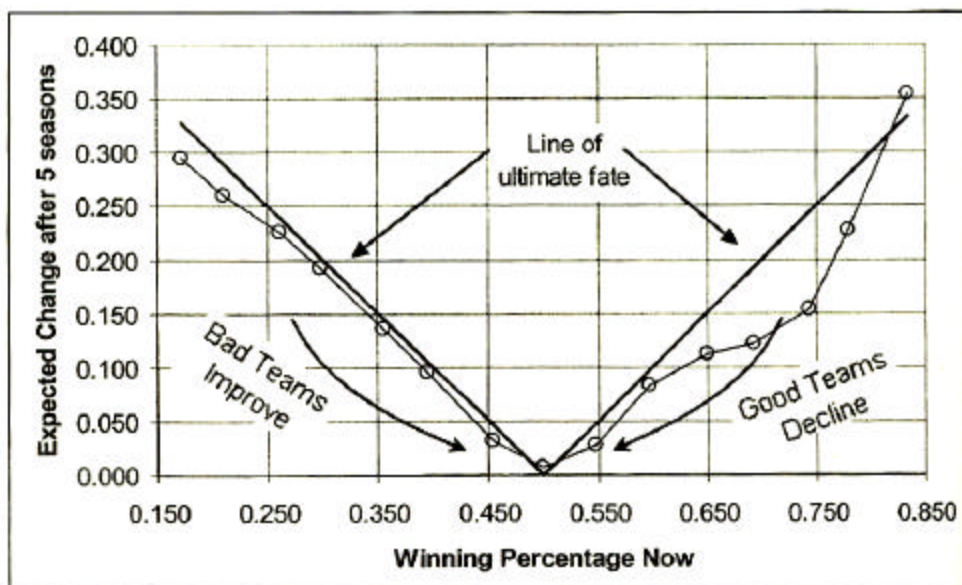
Source: Dean Oliver, *Basketball on Paper* (Washington, D.C: Brassey's, Inc., 2004), 112. Copyrighted material, used by permission.

Cal Tech economist Colin Camerer gave the most eclectic talk of the day. As noted, Camerer started by underscoring why statistics might help us "see" what's going on in sports. He then went on to document a case where a sports market was inefficient, followed by a case where the sports market was efficient.

The inefficiency Camerer described is based on the disposition effect—the notion that once we make a financial decision, we often refuse to reverse that decision until it works out. In stock market terms, the disposition effect predicts that people don't sell a stock if they have a loss but rather wait to get even or post a gain to dispose of the shares. Empirical studies of investor behavior support the theory. [9]

Camerer and his colleagues studied whether or not the disposition effect operated for NBA draft picks. Specifically, they determined whether high draft picks played more than they should based on their contribution. The study confirmed that indeed high draft picks played more minutes than warranted for about the first three years. Pro basketball managements are not immune to the disposition effect.

As an illustration of market efficiency, Camerer shared a case from horse race betting. Quite by accident, he discovered one day that bets could be cancelled. That gave him an idea: what would happen to the odds of two comparable horses if he placed a big bet on one of them and then cancelled it at the last minute? Would technical traders read the inflows and odds changes as likely asymmetric information, or would the fundamental bettors use the opportunity to change their betting strategy?

Camerer selected two horses with similar, relatively low probabilities of winning. He placed a large bet on one of them (determined via a coin toss) which he later cancelled just moments before betting closed. The experiment showed that while some bettors did follow the money trail, the odds the horse would win improved considerably. But after canceling the bet, he showed that the odds returned to their pre-manipulation levels. Enough people bet on the other horse (a relative bargain) to offset those who tried to take advantage of the perceived smart-money trend. His work demonstrates the large degree of efficiency in pari-mutuel markets. [10]

Another of Camerer's topics was how many steps people tend to look ahead. General equilibrium theory assumes individuals have a complete understanding of their preferences for their future—a clearly unrealistic assumption. But how far do people look out? Camerer's research shows that people tend to look out one or two steps—and rarely more.

The Lessons

Statistical techniques clearly factor more prominently in the world of business and investing today than they did a generation ago. Perhaps sports analysis is just catching up to these other markets. But it is worth noting that there are some significant challenges in intelligently using statistics in sports, business and investing. These include:

1. Predictions help shape the outcome. In capital markets, researchers and practitioners have identified many anomalies and trading strategies that delivered excess returns based on past results. But herein lies the paradox—in exploiting market inefficiencies, practitioners make the market more efficient.

Likewise, if your strategy on the field (based on past statistics) becomes predictable—you never bunt or you go for it in certain fourth down situations—and your opponent knows that, they will change their strategy to offset your moves. In a sense, past inefficiencies disperse, and the "market" becomes more efficient.

In some spheres, like weather forecasting, predictions and outcomes are independent. In markets and sports, there is at least the potential for significant interdependence: predictions help shape outcomes.

2. Statistics are context dependent. To derive meaning from statistical comparisons, the data need to exhibit sufficient stationarity—that is, the samples must be drawn from statistically similar populations. But in the real world, data are often nonstationary, making comparisons perilous or even nonsensical.

It's pretty easy to find nonstationarity in the world of investing. Take the ubiquitous price-earnings ratio. Pundits frequently compare today's multiple to the multiples of past periods. But significant items like inflation, taxes, and the composition of assets muddy these comparisons.

In the world of sports, it's hard to compare players across time due to changes in rules—for example, expansion of the lane and the shot clock in basketball—or location—some ball parks are more hitter or pitcher friendly than others. Naturally we can adjust for this context dependence, but that adds another challenge.

3. The role of probability. Most of the analysis of hot hands—for example, the idea that a basketball player will more likely make his or her next shot after making a bucket—finds that outcomes are consistent with the player's ability, or probability of success. [11] Said differently, one should expect hot hand streaks given a player's field goal percentage average. In reality, though, the vast majority of sports fans and players still perceive the hot hand to exist. This suggests we humans don't have a clear-cut sense of probabilities.

Working off probabilities is terrific if you're assured a large sample. But what if you have a limited sample size?

Economics suggests that you still need to think probabilistically, although we now have substantial evidence that humans don't make decisions this way. In fact, humans are risk adverse. This may explain why, for example, NFL coaches make conservative decisions.

As a result of this human psychological feature (as well as others), we tend to misspecify probabilities and outcomes, which leads to the next challenge and opportunity.

4. The role of psychology. Sports, business, and investing are all activities we do with others. Even though our understanding of psychological influences has grown measurably in recent decades—Daniel Khaneman and Amos Tversky's seminal work in Prospect Theory played a large role in that effort—we still operate in largely sub optimal ways. So the final challenge and opportunity is to understand psychological factors and to try to use them in our favor.

Evaluation horizon is a good example. Research shows that investors suffer from myopic loss aversion—frequent portfolio evaluation triggers aversion. As a result, long-term investors are willing to pay more for a risky asset than short-term investors.

You can think of these psychological pitfalls on two levels: first, the errors we each make as individuals—overconfidence, framing problems, etc.—second, the pitfalls related to collective behavior—the roles of influence and imitation. Both levels are vitally important to understand.

Legg Mason Capital Management

5. Circumstance- versus attribute-based thinking. In a very relevant article, Clayton Christensen and his colleagues discuss a three-step process for theory building. [12] First, you describe what you want to understand in numbers, then you classify the phenomena into categories based on similarities, and finally you build a theory that explains the behavior of the phenomena.

Once the theory is in place, researchers often find anomalies that force them to rethink and restate the descriptions and categories. Perhaps the paper's most important message is that good theories require proper categorization, and as theories improve, categories typically evolve from attribute-based to circumstance-based. Theories built on circumstance-based categories tell practitioners what to do in different situations. In contrast, attribute-based categories prescribe action based on the traits of the phenomena.

Most theories in business (management fads), investing (style boxes) and sports (kicking versus going for it) are attribute-based theories. In every case, there is room to evolve toward better, circumstance-based frameworks.

So these five issues are relevant for everyone—sports people, academics, business people and investors.

Endnotes

[1] Michael Mandelbaum, *The Meaning of Sports* (New York: PublicAffairs, 2004), 5.

[2] See http://www.cs.ubc.ca/~rensink/flicker/download/index.html

[3] The pinwheel relies on barycentric coordinates to represent various Colonel Blotto strategies. For example, the coordinate in the bottom left hand corner is 100, 0, 0. The top corner is 0, 100, 0, etc. As you move away from corner 1, the number in slot 1 declines proportionately until you reach the opposite side, where the value for slot 1 is 0. This holds for each corner. The point in the middle is 33 ?, 33 ?, 33 ?. The dark region shows where the winning strategy is random. The light areas in the corners show sub optimal strategies. See http://www.cut-the-knot.org/triangle/barycenter.shtml.

[4] See http://emlab.berkeley.edu/users/dromer/papers/nber9024.pdf.

[5] When asked about Romer's paper, New England Patriots coach Bill Belichick said: "I read it. I don't know much of the math involved, but I think I understand the conclusions and he has some valid points." See David Leonhardt, "Incremental Analysis, With Two Yards to Go," *The New York Times*, February 1, 2004.

[6] See http://www.hs.ttu.edu/hdfs3390/hothand.htm for a review of the literature.

[7] Fisher Black, "Noise," *Journal of Finance*, 1986.

[8] Dean Oliver, *Basketball on Paper* (Washington, D.C: Brassey's, Inc., 2004), 63.

[9] Terrance Odean, "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance,* 53, October 1998, 1775-1798; Colin Camerer and Martin Weber, "The Disposition Effect in Securities Trading: An Experimental Analysis," *Journal of Economic Behavior and Organization*, 33, 1998, 167-184.

[10] Colin Camerer, "Can Asset Markets be Manipulated? A Field Experiment with Racetrack Betting," *Journal of Political Economy,* June 1998, 457-482.

[11] See http://www.hs.ttu.edu/hdfs3390/hothand.htm.

[12] Clayton M. Christensen, Paul Carlile, and David Sundahl, "The Process of Theory-Building," *Working Paper, 02-016.* See http://www.innosight.com/template.php?page=research#Theory%20Building.pdf.