Major League Baseball

Chase and Miss: Won't Get Fooled Again

July 26, 2023

Chase and Miss: Won't Get Fooled Again

Chris Johnson, Ph.D., Brandon Ally, Ph.D., Scott Wylie, PhD

Key Findings

Compared to MLB players in Low and Average S2 Instinctive Learning groups, players in the High S2 Instinctive Learning group showed,

- Better Control of the Zone - Significantly lower strikeout to walk (SO/BB) ratio.
- Lower Miss Rates - Significantly less swing-and-miss for Fastballs up in the zone, and for Curveballs in the zone.
- Lower Chase Rates - Significantly lower Pre2K Chase, Critical Count Chase, and Advantage Count Chase.

Baseball is a game of high uncertainty. Efforts to reduce uncertainty at the league level include a lengthy six-month season, the highest number of games in any sport, and recurring team matchups grouped together in multigame series. High uncertainty is also inherent to hitting. Hitters don't know what pitch will be thrown, when it will be thrown, or where it will be thrown. Successful adaptation to high uncertainty is made exponentially more difficult under compressed time (see Issue #1: *Looking for Guys Who See Fast*).

Given the high degree of uncertainty with each pitch, why do some hitters seem to have a better sense or feel for what to anticipate? Part of the answer may be found in individual differences in capacity to detect patterns and learn sequences. Complex calculations and working through mental simulations for all pitch types, velocities, and trajectories, is cumbersome and time consuming. In contrast, there are advantages for brains that can quickly learn cognitive shortcuts, patterns, and rough rules-of-thumb (i.e., *heuristics*) that facilitate learning when information is scarce, and uncertainty is high. The capacity to learn informal rules through patterns and associations confers advantages because they are simple to use, and they reduce information overload (i.e., *noise*). Examples of informal cognitive shortcuts in hitting include, '*don't swing if pitch is above eye level*', '*thin wrist breaking ball, wide wrist fastball*', and '*swing at top of fastballs, bottom of breaking pitches*'. Hitters with

greater capacity to detect cues and connect subtle patterns may also be quicker at learning the strike zone with each new umpire, recognizing pitch sequences, and identifying tipped pitches before others do.

Precise mechanisms of cognitive shortcuts and associative learning are beyond the scope of this paper. However, for those seeking additional context please see the section at the end of this paper titled *'Expanded Background'*. The focus of the current paper is identification of potential performance differences in hitters with high cognitive capacity for learning subtle patterns and cues that facilitate decision making at the plate.

Individual differences in the ability to recognize patterns, tendencies, and sequences is assessed with S2's Instinctive Learning task. Two metrics for examining hitter differences in patterns and sequences are swinging at pitches outside the strike zone (Chase%), and swing-and-miss (Miss%).
Evidence from our analysis of minor league hitters suggests that prospects with higher Instinctive Learning scores are more proficient making adjustments as they progress through each level (i.e., A, AA, and AAA leagues). What remains unclear is whether Instinctive Learning is associated with performance differences for hitters facing the unique complexities of big league '*stuff*'.

Before looking at individual differences, we first examined implications of Chase% and Miss% on major league team performance. MLB 2022 team Chase% and Miss% rankings are presented in Figure 1 below. Of eight teams that competed in post-season Division Series, only PHI and ATL did not show up in the top 5 ranks. Interestingly, ATL ended the 2022 season with the 4th highest run margin in baseball (Runs Scored - Runs Allowed = RMGN

RANK	Miss% FB Up	Miss% FB→BrkBall	Chase% Adv Count	Chase% Pre2K	Chase% BrkBall
1	CLE	CLE	HOU	LAD	MIL
2	WSH	STL	MIL	MIL	ARI
3	HOU	TOR	SD	ARI	NYY
4	PIT	BAL	SEA	SEA	TOR
5	SEA	HOU	CWS	SD	SEA
26	TB	PIT	STL	BAL	PHI
27	LAA	LAD	ATL	PIT	ATL
28	ATL	ATL	TEX	ATL	TEX
29	ATL	SF	CHW	CHW	CHW
30	CHC	MIA	DET	DET	DET
Top - Bottom Δ	12 %	12 %	11 %	8 %	11 %

180). They were also the only team in the bottom 5 across all five Chase% and Miss% metrics highlighted below. Similarly, PHI ranked in the bottom 5 on two of the metrics. PHI was 26th in the league for Chase% on Breaking Balls, and 27th in the league in Pre2K Chase%. PHI made it to the World Series, losing eventual champion Houston Astros. In contrast, HOU was ranked 3rd best in the league for connecting with fastballs up in the zone, 5th best in contact with breaking balls sequenced after fastballs, and 1st in the league for resisting the impulse to chase outside the zone in advantage counts.

Next, we examined how individual differences in S2 Instinctive Learning relate to hitter performance differences in MLB Chase% and Miss%. During the 2022 MLB regular season 130 players had 500 or more plate appearances (MLB PA $\geq$ 500). S2 Instinctive Learning scores were available for 63 of those players, comprising roughly half of the population of everyday position players on active rosters (48%). Hitters were categorized into three groups according to S2 Instinctive Learning score:

(1) *High Instinctive Learning* (S2 $\geq$ 80)
(2) *Average Instinctive Learning* (S2 40 - 79)
(3) *Low Instinctive Learning* (S2 $\leq$ 39)

Table 1 below shows group characteristics, along with average number of pitches seen and plate appearances for the 2022 season. One point of emphasis here is that MLB players comprise the extreme right-hand tail of a hitting performance distribution, with only 130 being good enough for consistent and regular playing time. Moreover, this examination was limited to roughly half of that already select group. However, this group was also selected for large numbers of plate appearances and pitches seen, significantly increasing the number of observations - rather than individuals - under consideration.

Table 1. **S2 Instinctive Learning** and MLB Plate Appearance Descriptives.

MLB S2 Group	n	Age	S2 Instinctive Learning	Pitches (P)	Plate Appearances (PA)	P/PA	SO/BB *
Low Instinct. Learn	20	26.4	24th %ile	4797	1242	3.87	3.40
Avg Instinct. Learn	19	27.1	60th %ile	4652	1175	3.95	3.55
High Instinct. Learn	24	27.5	**91st %ile**	**5299**	**1320**	**3.98**	**2.62**

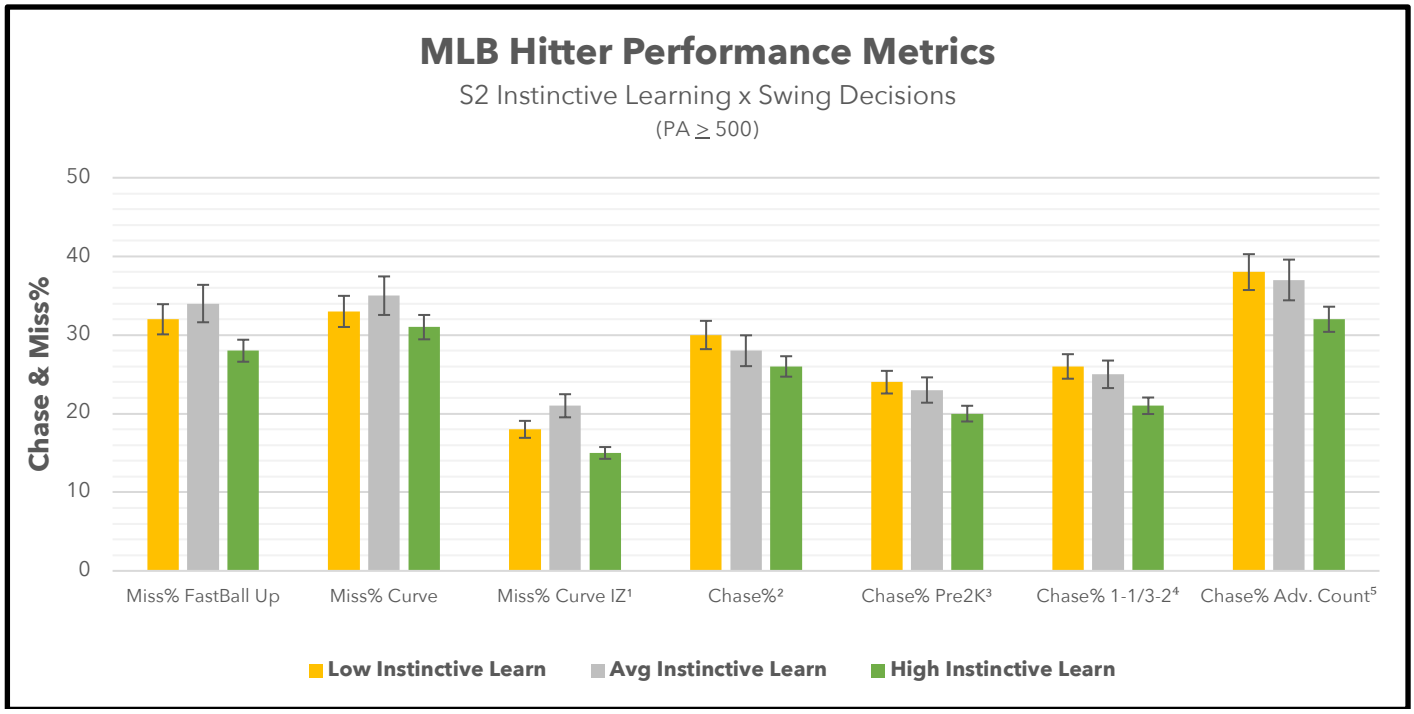
*$F = 3.53$, $p < .05$, $\omega^2 = 0.74$

Results

There were no differences among Instinctive Learning groups in age, plate appearances, or pitches seen. There was a marginal trend for the High Instinctive Learning group to see more pitches per plate appearance (P/PA). Where the groups did differ was in strikeout-to-walk ratio (SO/BB*), with the High Instinctive Learning group significantly outperforming Low and Average groups.

Group averages for Chase% and Miss% categories are shown in Figure 2 below. Groups did *not* differ on overall Miss% for curveballs. There were significant group differences for all other outcomes, including Miss% for Fastballs Up in the zone, as well as Miss% for Curveballs

in Zone. For each outcome type, the High Instinctive Learning group performed better than the other Average and Low groups.



[1] F = 3.80, p < .05; [2] F = 3.10, p = .05; [3] F = 3.72, p < .05; [4] F = 4.16, p < .05; [5] F = 3.17, p < .05

Magnitude of group differences (i.e., *effect size*) was greatest for Chase% in 1-1/3-2 counts. Effect size is used to classify the obviousness of a finding (small, medium, large) by categorizing how readily apparent observed differences are. For these analyses effect size for Chase% 1-1/3-2 was between moderate to large ($\omega^2 = .091$). Similarly, effect sizes for Chase%, Chase% Pre2K, and Chase Advantage Count were slightly smaller but also in the moderate range. One implication of these results is that even at highest level of performance (everyday MLB players), hitters with greater Instinctive Learning capacity appear to get fooled less in terms of miss and chase.

Insights

Key findings are that MLB players with High S2 Instinctive Learning Scores (> 80th %ile) had significantly lower SO/BB ratio, and the size of that effect was not small. They also had lower miss rate for Fastballs up in the zone and for Curveballs in the zone.

Higher S2 Instinctive Learning capacity was also related to lower chase rates across several situational counts, most notably in critical counts (i.e., 1-1/3-2 and Pre2K). These results suggest that hitters with greater S2 Instinctive Learning capacity are better at detecting patterns and sequences at game speed. Greater instinctive learning capacity may also confer advantages in identifying which pitcher behaviors and tendencies are meaningful, and which can be ignored. Moreover, instinctive learning may facilitate proficiency in learning contextual cues and adopting flexible, informal rules to guide decision making in milliseconds. As learned associations and responses become more accurate, they are likely to be further

reinforced by reward and success when they work. In summary, these data suggest that high instinctive learners are less likely to get fooled at the plate.

Limitations of Analysis

The primary limiting factor of these analyses is small sample size. For both MLB-level analyses and the Top 100 Draft picks, sample sizes are small and group sizes are unequal. Several analytical methods were utilized to reduce potential effects of these limitations.

1(a) Analysis of Variance (ANOVA) was the preferred method of analysis. Compared to common big-data techniques like machine learning and data mining that have steep sample size demands, ANOVA is hypothesis driven and less influenced by error common to overweighting models. ANOVA is also robust to large differences in sample size because analysis is based on group differences in grand means (or weighted means when samples sizes are unequal) and considers variation within each group as well as between them.

1(b) Theory-driven analysis. Big-data approaches utilize the law of large numbers to reduce unexplained error. Such methods include running simulations (i.e., simulating observations) that increase sample size. Data mining is often an exploratory approach when there are no *a priori* decisions regarding the focus of analysis and is most powerful as a discovery tool when it's not known what's being looked for. However, these approaches are not without drawbacks. In early stages of analysis, overfitting large datasets is inherent, and new variables are often overweighted. An alternative to machine learning and AI approaches is to utilize methods based on the type of data and the kinds of questions being asked. In datasets like the one analyzed here, statistical analysis with clearly defined analysis pathways and discrete outcome variables identified beforehand can be better suited to discovery real differences when they exist.

1(c) Utilizing repeated measures (i.e., 2A, 3A, and MLB metrics) and repeated trials (i.e., large numbers of pitches and plate appearances) for each player reduces within sample error.

2. Comparing group differences in a sub-sample of S2 scores to group differences at MLB level to see if pattern observed in the smaller samples corresponds to what is seen in the larger population of major league players.

3. Instead of making prediction the aim of analysis, indicators of cognitive capacity were used in classification for group analyses. Group classification analysis and evaluation of capacity can be preferred in conditions where complex human behavior does not meet assumptions necessary for probability and risk-based analysis that require identification of all possible outcomes, minimal variability, high certainty, and maximum control of external factors.

Expanded Background

S2 Instinctive Learning assesses capacity for recognizing meaningful but less obvious patterns that emerge during performance. Instinctive Learning is also understood as intuitive or implicit learning. The 'instinctive' aspect of Instinctive Learning is part of an advanced cognitive system that notices tendencies and patterns through experience. Our brains are constantly building linkages between what we see, what we do, and what consequence is produced. We start to build connections that help optimize our expectations and predictions of what will happen next. Error detection mechanisms in the brain continuously monitor differences between what we predict and real outcomes. As our predictions more closely match reality, we build new insights that optimize performance. If our predictions frequently do not match reality, our brains tell us update and adapt to new cues and start the process again. For hitters, this dynamic process helps them pick up on a pitcher's tendencies, such as tipping or tells, and promotes anticipation of pitches with higher degrees of confidence. Instinctive learning also underlies a hitter's capacity for detecting patterns in pitch sequences and counts.

The human brain is highly specialized for this type of intuitive learning, and cognitive neuroscience research suggests learning often occurs outside our full awareness (e.g., see classical and operant conditioning). Put another way, the brain knows and learns more than it can explain. Knowing something in the absence of a cogent explanation of how or why it's known often leads to misinterpretation of a response as coming from *gut instincts* or *intuition*. However, based on what we know about how the brain learns, stores, and retrieves information, intuition is not an abstraction void of data. In addition to building insights about the linkages between situations, actions, and consequences, there is growing evidence suggesting the brain continuously monitors intricate details of body sensations, including gut feelings, emotions, and time perception, all part of a process known as interoception.

From a scouting and coaching perspective, it may be helpful to think of each hitter as having their own internal dictionary of pitches comprising different pitch types seen, which pitches are thrown in certain counts, variations in pitcher arm slots and deliveries, changes in velocity and timing, and different trajectories of break. How does a hitter develop actionable instinctive insights about a specific pitcher standing sixty feet away? The brain develops heuristics and templates for based on what they have learned and experienced as they face each pitcher and each pitch. Results from our S2 assessments underscores individual differences in instinctive learning exist, even among major leaguers, and those differences are associated with some moderate differences in performance. Similarly, pitchers are also likely to differ in how easily their tendencies are to discern. Hitters with greater intuitive capacity for deciphering these idiosyncrasies have an advantage for both approach and adaptive execution.