



Scaling N from 1 to 1,000,000: Application of the Generalized Matching Law to Big Data Contexts

David J. Cox^{1,2}  · Bryan Klapes³  · John Michael Falligant^{4,5}

Accepted: 3 May 2021 / Published online: 09 June 2021

© Association for Behavior Analysis International 2021

Abstract

The generalized matching law (GML) has been used to describe the behavior of individual organisms in operant chambers, artificial environments, and nonlaboratory human settings. Most of these analyses have used a handful of participants to determine how well the GML describes choice in the experimental arrangement or how some experimental manipulation influences estimated matching parameters. Though the GML accounts very well for choice in a variety of contexts, the generality of the GML to all individuals in a population is unknown. That is, no known studies have used the GML to describe the individual behavior of all individuals in a population. This is likely because the data from every individual in the population has not historically been available or because time and computational constraints made population-level analyses prohibitive. In this study, we use open data on baseball pitches to provide an example of how big data methods can be combined with the GML to: (1) scale within-subjects designs to the population level; (2) track individual members of a population over time; (3) easily segment the population into subgroups for further analyses within and between groups; and (4) compare GML fits and estimated parameters to performance. These were accomplished for each of 2,374 individuals in a population using 8,467,473 observations of behavior-environment relationships spanning 11 years. In total, this study is a proof of concept for how behavior analysts can use data-science techniques to extend individual-level quantitative analyses of behavior to the population-level focused on domains of social relevance.

Keywords matching law · big data analytics · behavioral data science · baseball

Limited time and computational resources have typically required behavioral researchers to choose between collecting many data from a few research participants or few data from many participants. When choosing, Skinner (1966) recommended,

✉ David J. Cox
david@bhcoe.org; dcox@endicott.edu

“instead of studying a thousand rats for one hour each, or a hundred rats for ten hours each . . . study one rat for a thousand hours” (p. 21). The notion is that within-subject research designs—which are focused on the analysis of individual behavior—are more likely to accurately describe the mechanisms responsible for behavior change (Sidman, 1960). So, forced to choose, researchers interested in understanding the mechanisms responsible for behavior change are better off studying one rat for 1,000 hr. But, what if resource limitations no longer forced this choice? What if we could study 1,000 rats for 1,000 hr (Hagopian, 2020)? Or, better still, study every individual, using every applicable datapoint generated by that individual, within an entire population? What might we learn about behavior if we could scale N of 1 research designs up to N of millions?

Tools required to thoroughly analyze the complexity of human behavior must be precise and applicable at scale. Herein lies a key challenge with attempts to apply behavior-analytic concepts and principles to large-scale applied issues. Research with some socially significant topics (e.g., population-level health behavior, social issues) require methods that allow researchers to practically identify the variables controlling behavior for all (or most) members of a population while preserving the analysis of individual behavior (Hagopian, 2020). Such an approach requires the efficient analysis of data from millions of people, in many choice contexts, and with unique response options and reinforcers for each person and in each context. Constraints with data availability and efficient analysis have made scaling behavior-analytic research questions and applications a challenging endeavor.

Recent advances in computational storage, processing, and analytic techniques (i.e., “big data” methodologies, data science) now make scaling individual-level analyses practical. Large open datasets increase data availability for larger portions of the human population. And computer technology makes analysis of billions of data points possible within practical timeframes. The combination of increased data availability and analytic efficiency may permit the analysis of the variables controlling the behavior of each member of a group. This may eventually allow researchers to leverage data science to ask behavior-related questions of social significance at scale without sacrificing the methodological and conceptual rigor that is the hallmark of behavior analysis.

As one example, consider scaling the study of human choice and decision making in nonlaboratory contexts. Organisms tend to allocate behavior among concurrently available responses based on the relative reinforcement rates associated with each alternative (Herrnstein, 1970). Originally quantified by Herrnstein (1961), a modified version of the matching law—the generalized matching law (GML; Baum, 1974)—describes systematic deviations from strict equality of ratios of behavior with ratios of reinforcement. The GML predicts that, when log transformed, relative response allocation between two responses varies linearly with relative reinforcement rates. This relation is expressed as:

$$\log\left(\frac{B_1}{B_2}\right) = a*\log\left(\frac{R_1}{R_2}\right) + \log(b). \quad (1)$$

In Eq. 1, B_1 is the rate of responding on one alternative, B_2 is the rate of responding on the second alternative; R_1 and R_2 are the rates of acquired reinforcement for those

alternatives, respectively. The slope of the line (a) reflects sensitivity to changes in concurrent ratios of reinforcement rate, and the intercept [$\log(b)$] reflects bias for one of the responses when equality of reinforcement rate would predict indifference between the responses (e.g., Baum, 1974). If behavior matches reinforcement rate perfectly, the slope (a) equals one and the intercept [$\log(b)$] equals zero. Note that Eq. 1 has also been successfully used to describe behavior allocation across three or more alternatives in nonlaboratory settings (e.g., Cox et al., 2017; Falligant et al., 2020) via the following equation:

$$\log\left(\frac{B_i}{B_o}\right) = a * \log\left(\frac{R_i}{R_o}\right) + \log(b). \quad (2)$$

In Eq. 2, B_i and B_o represent rate of responding for a target behavior (i) and rate of all responding for all other behavior (o), respectively; R_i and R_o represent rates of acquired reinforcement for the target response and all other responses, respectively.

Extensive empirical support exists for the GML from laboratory and nonlaboratory research (see McDowell, 2013). For example, the GML accounts well for choice across an array of operants such as lever pressing in rodents (e.g., Boomhower & Newland, 2016), key pressing or mouse clicking in undergraduates (e.g., Klapes et al., 2020; Madden & Perone, 1999), play selection and performance in a number of amateur (e.g., Alferink et al., 2009; Romanowich et al., 2007; Rotta et al., 2020) and professional (e.g., Cox et al., 2017; Reed et al., 2006) sports contexts, gambit selection among expert chess players (Cero & Falligant, 2019), conversation allocation among young adults (Borrero et al., 2007), and much more.

Despite this wealth of empirical support, the GML has shown limited generality to large-scale applied issues of social importance (e.g., Pierce & Epling, 1995). Several studies have applied the GML to the behavior of a population (e.g., Alferink et al., 2009; Vollmer & Bourret, 2000). In these studies, the matching law described well the number of two-point and three-point shots taken by basketball teams as a function of the total number of points earned from those shots. However, only individuals behave, not groups. Fitting the GML to team shot selection makes it challenging to practically translate the findings in the above studies to understand the variables that control any one players' shot selection. To truly scale and test the generality of the GML at the population level, the GML should be fit to every individual's behavior in the population.

There are several benefits of scaling matching to all members of an individual population based on the resulting population distributions of sensitivity and bias parameters. First, scaling matching analyses to all members of a population can benefit those who want to apply the quantitative analyses of behavior to policy decisions. Quantifying the proportion of the population for which different outcomes likely serve as a reinforcer can help identify the conditional effectiveness of different reinforcers for different people and in different contexts. In turn, understanding the putative reinforcing effectiveness of different outcomes could inform policy and regulation which—by definition—are designed to change the behavior of a population through changes to antecedent and consequent events applied to the behavior of individuals.

Scaling matching analyses to all members of a population also can benefit behavior analytic practitioners working clinically. Where known bias and sensitivity parameters are predictive of better (or worse) performance, applied practitioners can determine where an individual's matching parameters fits against the known population distribution. Such normative comparisons would be similar to those performed by clinicians during standardized testing (e.g., assessment for learning disorders; testing for neurodegenerative diseases). It's possible that clinicians who know how their client's sensitivity to changing schedules of reinforcement or bias for particular responses compare with a larger, nonclinical group can better target intervention goals along these dimensions to improve performance. Stated differently, knowing the right degree of sensitivity and bias that corresponds with better performance provides the clinician a target to shape toward. However, the utility of this approach is unknown unless population-level distributions of matching parameters are obtained.

A third benefit to obtaining population distributions of matching parameters could be for the more judicious allocation of scarce resources. Clinicians can see only a limited number of clients and have a limited amount of money to spend on intervention materials and staff. In situations where resources limit access to clinical services in some way, practitioners have to determine how they will allocate their time and attention to maximize the overall benefit to everyone involved. In these situations, applied practitioners may use the population distributions of matching parameters to identify the "most-at-risk individuals" at the tail-ends of the distributions to focus their intervention efforts (e.g., Fletcher & Vaughn, 2009; Foltz et al., 2012).

In sum, where the right data exists, data-science techniques allow behavior analysts to scale within-subject analyses to the population level. For this proof of concept study, we chose to focus on scaling GML analyses given its robust empirical history, and to illustrate how combining data-science methodologies with the GML can provide answers to questions about population-level individual human behavior. In general, we sought to demonstrate how behavioral data science can allow researchers and practitioners to: (1) easily scale within-subjects designs to the population level; (2) easily segment the population into subgroups for further analyses within and between groups; (3) easily track individual members of a population over time; and (4) more directly analyze the utility of the GML for understanding choice in one applied context: pitching in professional baseball.

Methods

Data

We chose to illustrate how the GML can be applied using data-science techniques using existing open datasets from professional baseball. We chose an open dataset because there are millions of free-to-access open datasets on the internet (e.g., Google dataset search) pertaining to many areas of social importance. This allows behavior analysts to replicate these methods across many different areas of social significance where the right data exist. We chose professional baseball because previous matching research has demonstrated the generality of the GML to baseball-related choice (Cox et al.,

2017; Falligant et al., 2020). Thus, using baseball data allowed us to focus on the novel application as opposed to demonstrating the validity of the datasets for GML analyses.

The data were obtained from two sources. The first source was Major League Baseball's (MLB) Statcast database using the Statcast Search application on BaseballSavant (Willman, 2021). This process entailed using the available dropdown menus to include all information about the game context, pitch type, pitch characteristics, and pitch outcome for every pitcher in MLB during the 2009–2019 MLB seasons. The second source was Rotowire's database of historical baseball statistics (Roto Sports, Inc, 2021). This source was used to obtain pitcher performance metrics for each MLB season.

Quantitative Model

We used the GML (Eq. 2) to describe each pitcher's allocation of pitches to the different pitch types. Based on past research by Cox et al. (2017) and Falligant et al. (2020), B_i was always the count of pitches per game that were of the fastball variety (e.g., four-seam fastball, two-seam fastball) and B_o was always the count of pitches per game that were of any other pitch type (e.g., curveball, slider, eephus). Fastballs were chosen for B_i to maintain analytic consistency across pitchers making parameter interpretation across pitchers easier. Of note, past researchers observed that Eq. 2 described MLB pitcher pitch selection with similar accuracy regardless of the pitch category used for B_i (see Figure 3; Cox et al., 2017).

Calculating Different Reinforcers

Strikes/outs were used as the assumed reinforcer for three reasons. First, strikes/outs are events that can follow each pitch allowing for a similarly scaled comparison of rates of reinforcement to rates of pitch types (Shimp, 2020). Second, based on the rules of baseball, strikes and outs are events considered in the pitcher's favor. Finally, past research on the GML and pitch type has found that using strikes/outs as assumed reinforcers led to quantitative descriptions of behavior–environment relations common to laboratory studies of the GML (Cox et al., 2017; Falligant et al., 2020).

To use strikes/outs as the assumed reinforcers, we counted the number of times per game that each pitch type resulted in a called or swinging strike, or the number of times that each pitch type was put into play and resulted in an out. The number of strikes/outs contacted in a game following a fastball were summed to create R_i , and the number of strikes/outs contacted in a game following all other pitches were summed to create R_o .

GML Scripts and Computational Environment

All programming was completed using the Python programming language (version 3.7) and run in Google Colaboratory (Colab). Colab is a free product from Google Research that allows anyone to write and execute Python code through any internet browser. Colab is a hosted Jupyter notebook service that provides the programmer with access to GPU or TPU computing resources that allows them to conduct computational analyses beyond what is possible on

most personal computers due to hardware or time constraints. The notebooks and scripts for this project are available for download at Cox (2020) and can be run on a personal computer (not recommended due to size) or using Colab. Table 1 shows a pseudocode algorithm outlining the logic to turn the dataset into the analytic data frame for the present study.

For each fit of the GML, sensitivity (a , Eq. 2) and log-bias [$\log(b)$, Eq. 2] were estimated using the *curve_fit* function from the SciPy Python package (Millman & Aivazis, 2011). R^2 values were calculated using the *r2_score* function from the Sklearn Python package (Pedregosa et al., 2011). All statistical analyses were conducted using the corresponding statistical function from the Statsmodels Python package (Seabold & Perktold, 2010) or the Pingouin Python package (Vallat, 2018). Interested readers are referred to McKinney (2018) for a more robust description and demonstration of the techniques used for this study.

Table 1 Pseudocode of model fitting process used in the present study. Readers are referred to the cited GitHub repository for full implementation using Python

Algorithm 1: Pseudocode to fit the GME to the dataset.

input: X is the dataset of all pitches from 2009 to 2019 in MLB, SR is a dictionary of pitchers labeled as a starter or reliever; W is the data frame with pitcher performance statistics from RotoWire

output: M is a data frame with the features of player, season, their position (starter or reliever) VAC from GME, bias from GME, sensitivity from GME, and the performance metrics from W

top:

Initialize $U \leftarrow$ list of the unique pitchers in X

Initialize $O \leftarrow$ list of all pitch outcomes labeled as a putative reinforcer

Initialize $Y \leftarrow$ list of the unique years in X

Initialize $M \leftarrow$ empty data frame to store GME fits as we loop through the data set

loop:

for $u_i \in U$ **do**

for $y_i \in Y$ **do**

$x_{uy} \leftarrow$ subset of X containing pitcher u from season y

 Initialize $D \leftarrow$ empty data frame to temporarily store season data for pitcher u during season y

 Initialize $G \leftarrow$ list of unique games for pitcher u during season y

for $g_n \in G$ **do**

$hs_straight$ thrown \leftarrow count of hard-straight thrown by pitcher u in game g from season y

hb_break thrown \leftarrow count of hard-break thrown by pitcher u in game g from season y

$ss_straight$ thrown \leftarrow count of slow-straight thrown by pitcher u in game g from season y

$hs_reinforce$ \leftarrow count of hard-straight for o_i if in O for pitcher u in game g from season y

$hb_reinforce$ \leftarrow count of hard-break for o_i if in O for pitcher u in game g from season y

$ss_reinforce$ \leftarrow count of slow-straight for o_i if in O for pitcher u in game g from season y

$sb_reinforce$ \leftarrow count of slow-break for o_i if in O for pitcher u in game g from season y

$D \leftarrow$ logged ratio of each pitch to all other pitch types

$D \leftarrow$ logged ratio of putative reinforcement contacted by each pitch type

end for

$fit \leftarrow$ fit Equation 2 using data in D for pitcher u during season y

$position \leftarrow$ get value from dictionary SR of whether the pitcher was a starter or reliever

$M \leftarrow$ pitcher, year, position, GME VAC, GME bias, and GME sensitivity from pitcher u during season y

$M \leftarrow$ merge pitcher performance stats from W with M using pitcher name and season as primary keys

end for

end for

return M

Results

Population Descriptions and GML Fits

Figure 1 shows the R^2 (VAC), log-bias, and sensitivity values obtained by fitting Eq. 2 for each season (2009–2019) and for all 2,374 unique pitchers in the dataset for a total of 6,687 unique pitcher-season markers. Each marker represents the GML fit or parameter value for a single pitcher and for a single season. The unique color of each marker represents a unique pitcher in the data set and remains consistent across each MLB season. The median R^2 , log-bias, and sensitivity values ranged between 0.72 and 0.75, 0.02 and 0.05, and 0.73 and 0.75 across the 11 MLB seasons (Table 2 shows the mean values with 95% confidence intervals). The median R^2 values suggest that the GML described pitch choice relatively well for at least half of the pitchers. The median estimated log-bias parameters were slightly above 0.00, suggesting slightly more pitchers were biased toward fastballs compared to other pitch types. Finally, the median estimated sensitivity parameters indicated undermatching (value less than 1.00), which is common in laboratory and nonlaboratory studies with humans and nonhumans (Baum, 1979).

As would be expected with population level data, we observed functionally significant variability in R^2 values. Some pitchers' season data were described well by the

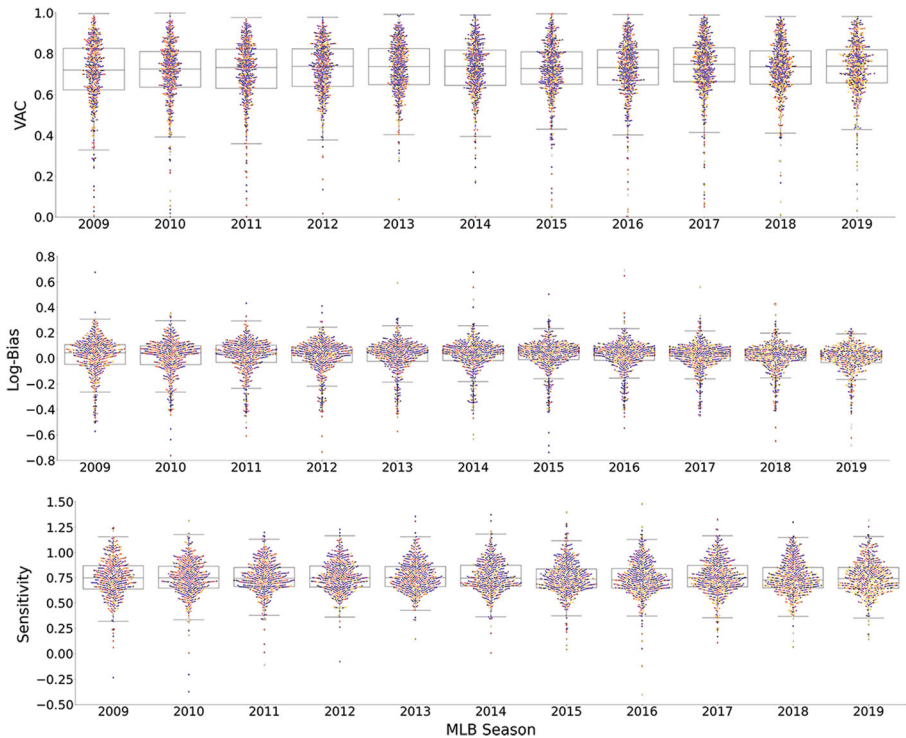


Fig. 1 Box-and-Swarm Plots of the Variance Accounted for (VAC; top row), Estimated Log-Bias (middle row), and Estimated Sensitivity (bottom row) from the GML for Each Pitcher and for Each Season in this Dataset. Each marker in a given year is a unique color and corresponds to a unique pitcher

Table 2 Number of Pitchers Included, Means and 95% Confidence Intervals for GML Parameter Estimates and GML R^2 Value for Each Season

Season	<i>Pitchers</i>	Sensitivity (<i>a</i>)	Log-Bias [$\log(b)$]	R^2
2009	568	0.75 [0.73, 0.76]	0.01 [0.00, 0.03]	0.71 [0.69, 0.72]
2010	544	0.75 [0.73, 0.76]	0.02 [0.00, 0.03]	0.71 [0.69, 0.72]
2011	565	0.75 [0.74, 0.76]	0.02 [0.01, 0.03]	0.71 [0.70, 0.72]
2012	584	0.76 [0.75, 0.77]	0.01 [0.00, 0.02]	0.72 [0.71, 0.73]
2013	590	0.76 [0.75, 0.78]	0.02 [0.01, 0.03]	0.73 [0.72, 0.74]
2014	583	0.76 [0.75, 0.78]	0.02 [0.01, 0.03]	0.72 [0.71, 0.74]
2015	631	0.75 [0.74, 0.76]	0.02 [0.01, 0.03]	0.72 [0.71, 0.73]
2016	623	0.74 [0.72, 0.75]	0.02 [0.01, 0.03]	0.72 [0.71, 0.73]
2017	648	0.76 [0.75, 0.78]	0.02 [0.01, 0.03]	0.73 [0.72, 0.74]
2018	660	0.75 [0.74, 0.76]	0.01 [0.00, 0.02]	0.72 [0.71, 0.74]
2019	691	0.75 [0.74, 0.76]	0.00 [-0.01, 0.01]	0.73 [0.72, 0.74]

GML with 1978 (30%), 531 (8%), and 123 (2%) R^2 values being greater than or equal to 0.80, 0.90, and 0.95, respectively. However, some pitchers' season data were not described as well with R^2 values below 0.60 and 0.50 for 1,039 (16%) and 409 (6%) pitcher seasons, respectively. This variability suggests that GML would be idiosyncratically helpful for managers, coaches, or trainers who wanted to use GML parameters to describe and understand pitcher performance. In total, the use of data-science techniques allowed us to create Figure 1, which provides population-level distributions and normative comparison for how well the GML describes any single professional pitcher's pitch allocation.

We also observed functionally significant variability in the estimated log-bias and sensitivity parameters. For log-bias, 4,561 (68%) pitcher seasons had estimated log-bias parameters between -0.10 and 0.10, indicating little-to-no bias toward fastballs or toward nonfastball pitches. However, 1,271 (19%) pitcher seasons showed a log-bias greater than 0.10 indicating a bias toward fastballs, and 855 (13%) pitcher seasons showed a log-bias less than -0.10 indicating bias toward nonfastball pitches. For sensitivity, 5,422 (81%) pitcher seasons had estimated sensitivity parameters less than 0.90 indicating undermatching. But 139 (2%) pitcher seasons had estimated sensitivity parameters greater than 1.10, indicating overmatching, and 1,126 (17%) had estimated sensitivity parameters between 0.90 and 1.10, indicating near perfect matching.

In sum, the use of data-science techniques allowed us to create a population-level distribution of log-bias and sensitivity parameters for professional baseball pitchers. This distribution is the first known distribution of matching parameters for all members of a population and allows for normative comparison of any single pitcher's behavior sensitivity to changing ratios of reinforcement and bias in pitch allocation.

GML Parameters by Subgroups of Population

In some situations, researchers might be interested in how smaller subgroups within a population compare to each other and whether distinguishing between

subgroups is practically relevant. Once data-science techniques have been leveraged to efficiently create an analytic data frame, Table 3 demonstrates how easily different segments of a population can be separated and analyzed relative to GML fits. The output of these three lines of code and 16 words can be seen in Figure 2. In particular, Figure 2 shows the R^2 , log-bias, and sensitivity values when the population is segmented by starting and relief pitchers. Each marker represents a single player-season combination.

We observed a statistically significant difference in R^2 values (mean, standard deviation) between starting pitchers (0.77, 0.13) and relief pitchers (0.68, 0.14; $t_{6685} = 25.02$, $p < 0.001$). We also observed a statistically significant difference in estimated log-bias parameters (mean, standard deviation) between starting pitchers (0.02, 0.106) and relief pitchers (0.01, 0.149; $t_{6685} = 3.02$, $p = 0.003$). However, no difference was observed in the estimated sensitivity parameters (mean, standard deviation) between starting pitchers (0.75, 0.16) and relief pitchers (0.75, 0.18; $t_{6685} = 0.17$, $p = 0.87$).

In sum, the GML described starting pitchers' choice of pitch type better than the GML described relief pitchers' choice of pitch type. And starters showed a statistically significant greater bias toward fastballs with no difference in estimated sensitivity parameters being observed between the two different types of pitching roles. It is important to note that the use of data-science techniques allowed us to create an analytic data frame allowing for easy segmentation and comparative analysis of the two subgroups.

Tracking Players over Time

Table 4 shows the relative ease with which individual members of a population can be tracked and compared to the population at-large. Figure 3 shows the output of this code, which is the population distributions of R^2 , log-bias, and sensitivity as box-and-whisker plots for six of the MLB seasons during which the players recorded games in MLB. Two players were chosen as they were in the top five of greatest increasing slopes (Player 1) and decreasing slopes (Player 2) in log-bias and sensitivity parameters over time. Player 3 was randomly selected.

For players 1 and 3, the variance accounted for by the GML increased as the years of experience in professional baseball increased. However, Player 1 was in the upper echelon of VAC by the GML whereas Player 3 remained primarily within the middle 50% of the distribution (the box of the box-and-whisker). For both of these players, bias trended toward (Player 1) or remained around 0.00 (Player 3), and sensitivity increased for both throughout these six seasons. In

Table 3 Code to generate Fig. 2 that highlights how easy it is to analyze subgroups from the population. The text in bold and italics shows the text to swap out to analyze any subgroups in the dataset with known labels

Code to generate figure 2.
<pre>import seaborn for column in ['VAC', 'bias', 'sensitivity']: seaborn.swarmplot(x_values = data['<i>starter_reliever</i>'], y_values=data[column])</pre>

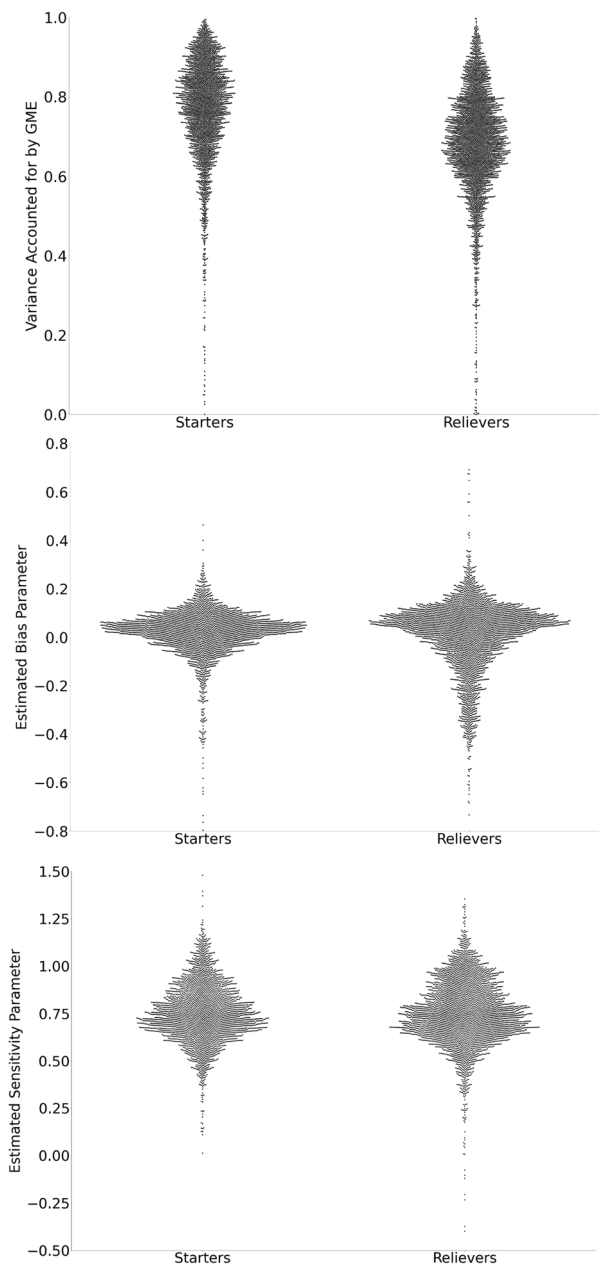


Fig. 2 Fits and Estimated GML Parameters Separated by Starters and Relievers. This was generated with three lines of code and 16 words demonstrating the ease with which population subgroups can be separated and analyzed

contrast, the VAC by the GML decreased over time for Player 2, log-bias shifted further and further from 0.00 toward non-fastballs, and sensitivity decreased. It is intriguing that Player 2 was no longer in professional baseball following their fifth season.

Table 4 Code to generate Fig. 3 and highlighting how easy it is to track any single player within the dataset relative to the larger population. The text in bold and italics shows the text to swap out to analyze any individual in the dataset

Code to generate figure 3.
<pre>player_one = data[data['Player']=='<i>Player 1 Name</i>'] player_two = data[data['Player']=='<i>Player 2 Name</i>'] player_three = data[data['Player']=='<i>Player 3 Name</i>'] def season_track(y_val): sns.boxplot(x="Season", y=y_val, data=player_fits) sns.swarmplot(x="Season", y=y_val, data=player_one) sns.swarmplot(x="Season", y=y_val, data=player_two) sns.swarmplot(x="Season", y=y_val, data=player_three) plt.show() for column in ['*VAC', 'Bias', 'Sensitivity']: season_track(y_val=column)</pre>

Analyzing the Large-Scale Utility of the GML

Utility Beyond Pitch Topography

To have applied utility, the GML should provide information beyond what is already known about pitching in professional baseball and current approaches to analyzing pitcher choice. For example, it is possible that batter behavior relative to the different pitch types is what led to the pitches contacting strikes and outs—not pitcher allocation of pitch types per se. Perhaps the faster someone throws a fastball the more likely a batter is to swing and miss or to hit the ball poorly resulting in an out. And, perhaps the further a breaking pitch breaks and the more slowly a slow pitch is thrown, the more likely they are to result in a strike/out. Stated differently, it is possible that pitchers do not need to be sensitive to how well different pitch types contact strikes and outs in a game. Rather, they only need to focus on throwing fastballs faster, slow pitches more slowly, and increasing spin rate to make breaking balls break further. If true, this would reduce the utility of the GML in this applied context. Figures 4 and 5 were aimed at addressing this possibility.

Figure 4 shows histograms of Pearson’s correlation coefficients calculated for each pitcher season and between the number of strikes/outs contacted by each pitch type and the raw speed of the pitch (left column) or the difference in speed from the pitchers’ fastball (right column).¹ Median (minimum, maximum) values for correlations of strikes and outs with raw pitch speeds were 0.10 (-0.98, 1.00), 0.26 (-0.99, 1.00), 0.44 (-0.96, 1.00), and 0.34 (-0.91, 1.00) for hard-and-straight, hard-and-breaking, slow-and-straight, and slow-and-breaking pitch types, respectively. Median (minimum, maximum) values for correlations of strikes and outs with speed differentials from the pitchers’ fastball were 0.01 (-0.98, 0.88), -0.04 (-1.00, 0.96), and -0.04 (-1.00, 0.99) for

¹ Relative speed was analyzed because it is possible the effectiveness of a slow and straight pitch is determined by how much slower it is compared to the pitcher’s fastball rather than its absolute speed. For example, perhaps an 85 MPH changeup is more likely to be perceived by a hitter as similar to an 88 MPH fastball compared to a 98 MPH fastball. Likewise, the effectiveness of a slider might improve the more its speed is similar to a fastball but then also breaks.

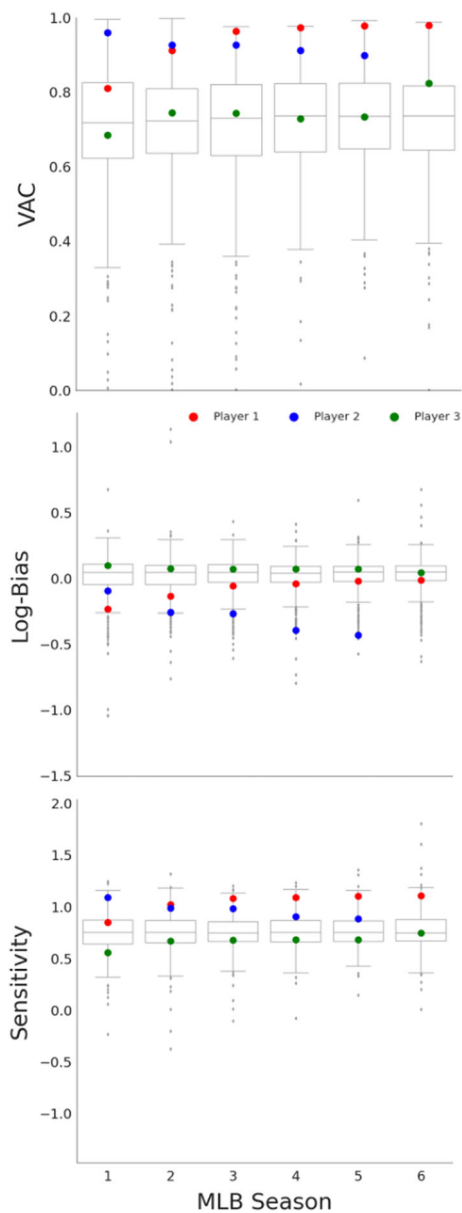


Fig. 3 Fits and Estimated GML Parameters for the Entire Population of Professional Pitchers (boxplots) and for Three Individual Pitchers (colored markers). This figure demonstrates the ease of tracking individuals within the population over time for normative comparison

hard-and-breaking, slow-and-straight, and slow-and-breaking pitch types, respectively. The data in Figure 4 indicate that using raw speed values centered the distribution of correlations away from 0.00. However, overall, strong correlations (> 0.70 or < -0.70) were observed in only 298 (4%), 433 (6%), 462 (7%), and 413 (6%) pitcher seasons using raw pitch speed for hard-and-straight, hard-and-breaking, slow-and-straight, and

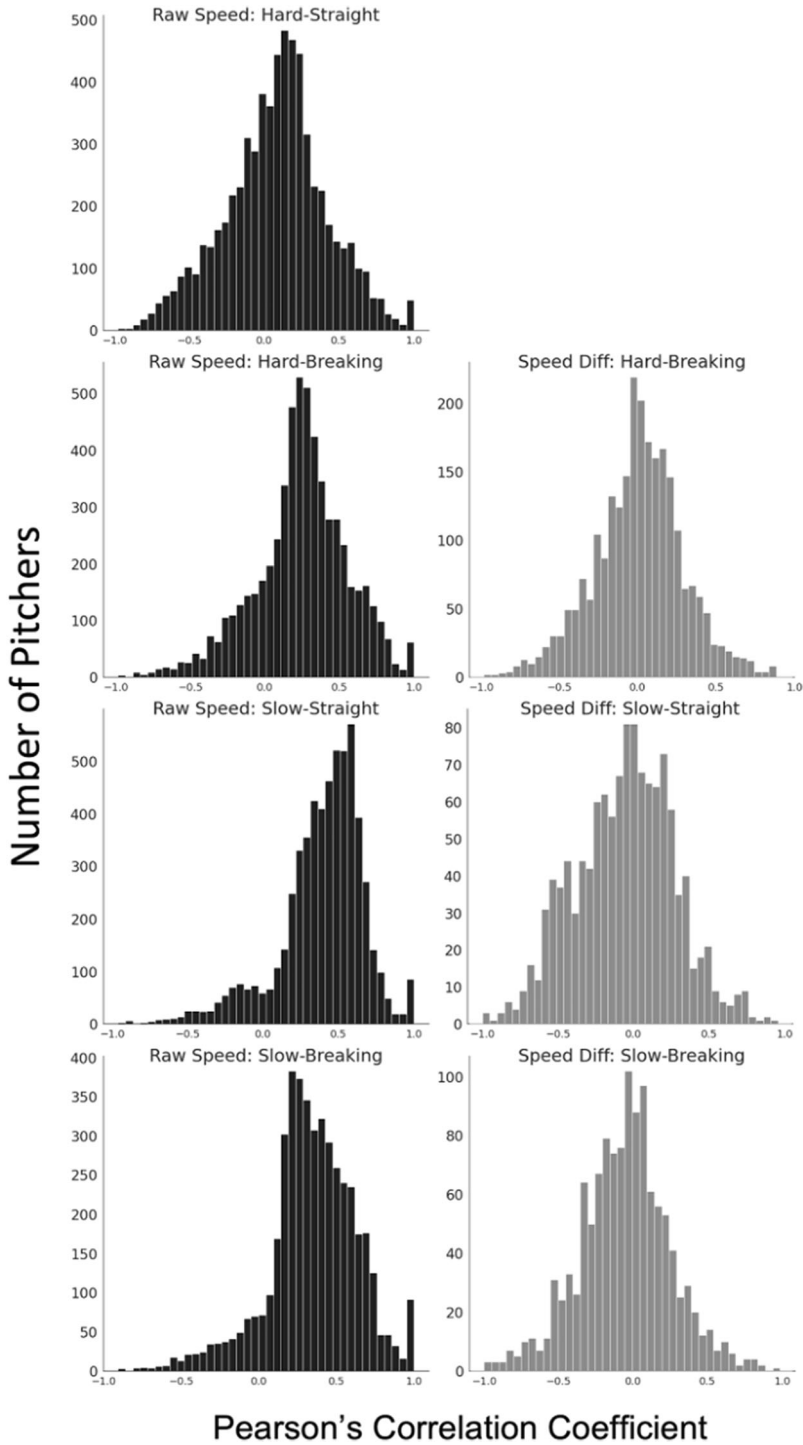


Fig. 4 Histograms of Pearson's Correlations between the Number of Strikes/Outs Contacted by Each Pitch Type and the Raw Speed of Pitch Types (left column) or Difference in Speed from the Pitchers' Fastball (right column)

slow-and-breaking pitches, respectively. Strong correlations were observed in 61 (< 1%), 46 (< 1%), and 43 (< 1%) using the difference in speed between hard-and-straight pitches and hard-and-breaking, soft-and-straight, and soft-and-breaking pitches, respectively.

Figure 5 shows the same analysis of Pearson correlation coefficients but between the distance of break for pitches categorized as hard-and-breaking or slow-and-breaking and the number of strikes/outs each pitch type contacted. The top row shows histograms using the amount of break in the horizontal direction; the middle row shows histograms using the amount of break in the vertical direction; and the bottom row shows histograms using the total break distance. Unlike pitch speed, the distributions of raw break distances (black histograms) and relative break distances (gray histograms) were not noticeably different. For horizontal break distances, strong correlations were observed using raw/difference distances for 277/120 (4%/2%) and 311/116 (4%/2%) of pitchers for hard-and-breaking and slow-and-breaking pitch categories, respectively. For vertical break distances, strong correlations were observed using raw/difference distances for 243/116 (4%/2%) and 226/75 (3%/1%) of pitchers for hard-and-breaking and slow-and-breaking pitch categories, respectively. For total break distances, strong correlations were observed using raw/difference distances for 243/150 (4%/2%) and 226/113 (3%/2%) of pitchers for hard-and-breaking and slow-and-breaking pitch categories, respectively.

The data presented in Figures 4 and 5 suggest that the raw characteristics of pitches are unlikely to be correlated with the number of strikes/outs contacted by each pitch

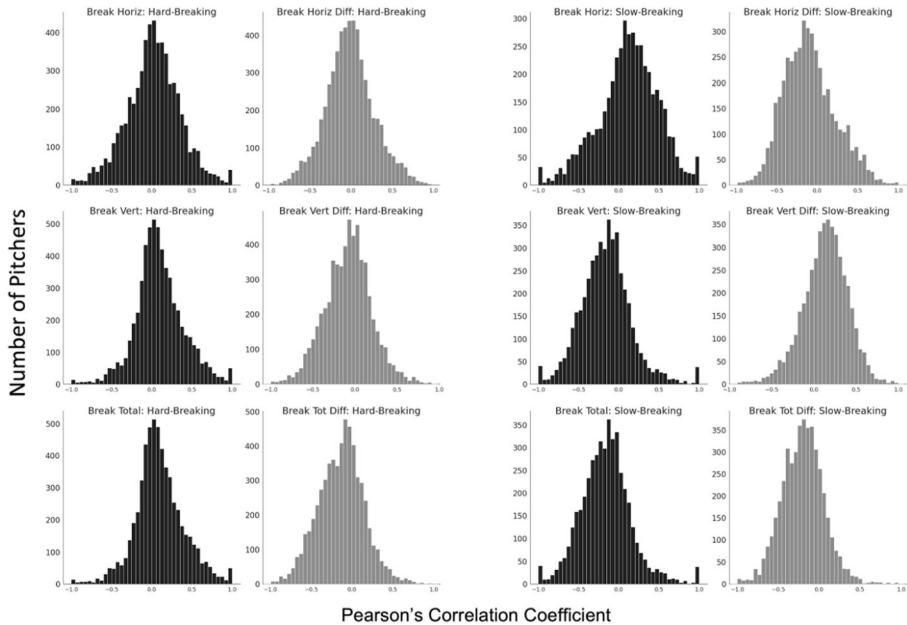


Fig. 5 Histograms of Pearson's Correlations between the Number of Strikes/Outs Contacted by Breaking Pitches and the Breaking Distance for Each Pitch Type. Black histograms are for correlations with the raw length of break in the horizontal direction (top row), vertical direction (middle row), and total (bottom row). Gray histograms are for correlations with the difference in break from the pitcher's fastball in the horizontal direction, vertical direction, and overall break

type. That is, pitching coaches or managers may gain additional predictive information about what leads to strikes/outs by analyzing pitch allocation via the GML rather than focusing only on the topographical characteristics of different pitch types. However, it's possible that bias for fastballs (an estimated measure of relative pitch allocation) is influenced by how fast pitchers throw in addition to the relative rate that fastballs contact strikes/outs compared to other pitches. Figure 6 shows a histogram (left panel) and corresponding box-and-swarmplot (right panel) of individual player Pearson correlation coefficients between their average fastball speed per season and their estimated log-bias parameter per season. Strong positive correlations ($r \geq 0.70$) were observed for 40% of the pitchers and strong negative correlations ($r \leq -0.70$) were observed for 25% of the pitchers. Thus, Figure 6 suggests that the speed of a pitcher's fastball likely has some relation to the estimated bias toward throwing a fastball, though not in a unidirectional way.

Relation to Common Pitching Performance Metrics

Figure 7 shows a correlation matrix of the estimated matching parameters (VAC, bias, sensitivity) with common metrics of pitching performance in professional baseball. The lower diagonal shows the raw estimated parameter values or pitching metrics plotted against each other. The upper diagonal shows the Pearson correlation coefficient value (r) for the corresponding comparison as a bubble plot. Each bubble is colored based on whether it is a positive (red) or negative (blue) correlation and also sized relative to the size of the correlation coefficient (larger bubbles indicate stronger correlation). The graphs that comprise the diagonal are density plots showing the distribution of that estimated parameter or performance metric.

The upper three rows of Figure 7 show correlations for the fit and parameters of the GML. The highest observed correlation was weak ($r = 0.30$) and occurred twice; once between VAC and the number of innings pitched and once between VAC and the

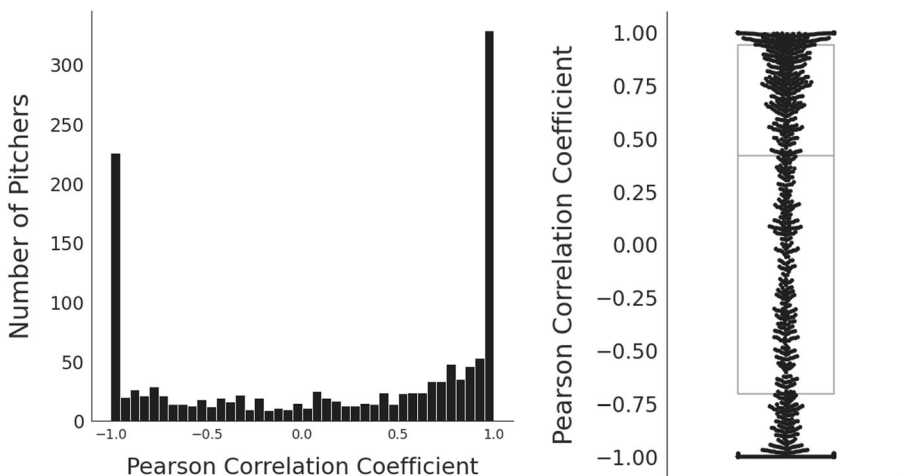


Fig. 6 Histogram (left panel) and Box-and-Beeswarm plot (right panel) Showing Pearson's Correlation Coefficients for Individual Players Calculated between the Estimated Bias Parameter for Throwing Fastballs and the Average Speed of the Player's Fastball across All Seasons Available in Our Dataset

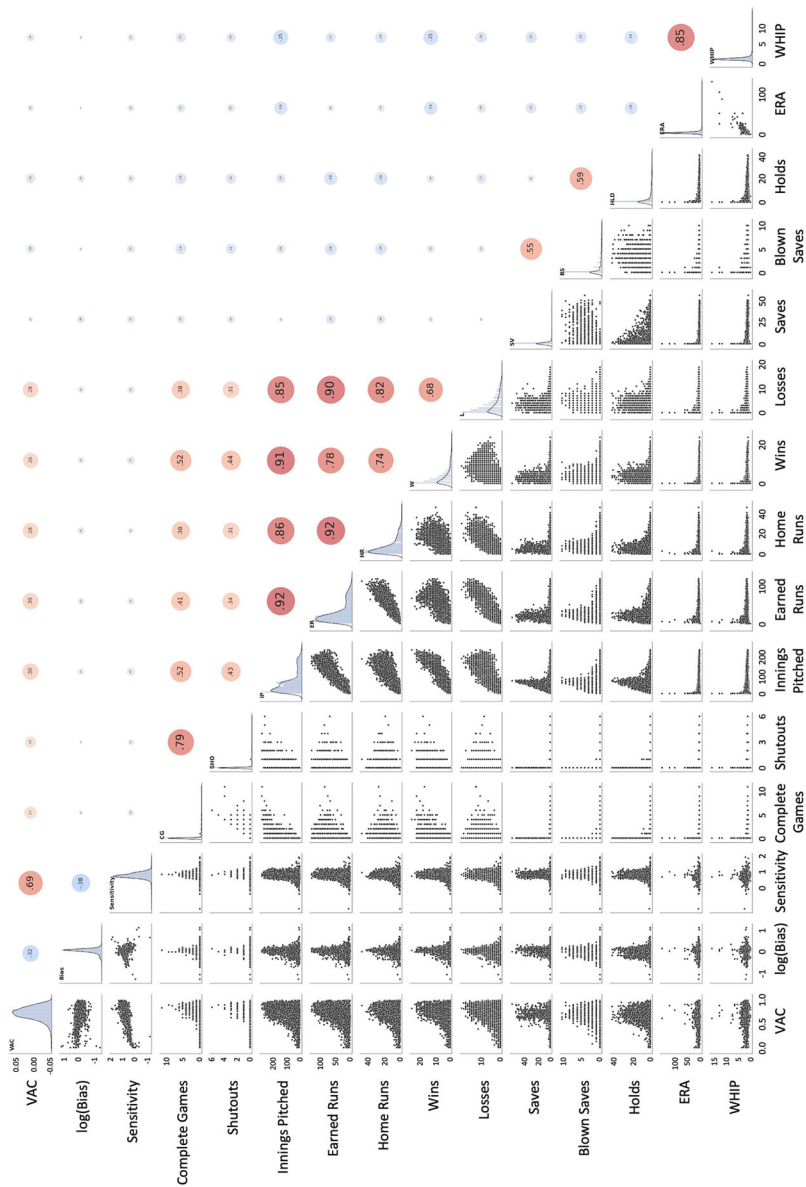


Fig. 7 Correlation Matrix of Estimated Matching Parameters for Each Player (i.e., VAC, bias, sensitivity) and Common Statistical Descriptions of Pitcher Performance in Baseball. CG = complete games; SHO = shutouts; IP = innings pitched; ER = earned runs; W = wins; L = losses; SV = saves; BS = blown saves; HLD = hold; ERA = earned run average; WHIP = (walks + hits) / (innings pitched)

number of earned runs. Additional weak correlations were observed between VAC and home runs (HR; $r = 0.28$), losses (L; $r = 0.28$), and wins (W; $r = 0.26$). No significant correlations were observed between the estimated GML parameters of log-bias or sensitivity and any of the common pitch performance metrics.

Discussion

Behavioral researchers have traditionally been forced to choose how to allocate scarce time and computational resources. Do they prioritize collecting many data for a few participants, or few data for many participants? Recent advances in technology make data collection, storage, and analysis easier and cheaper than in the past. In this study, we demonstrated how behavior analysts can use open data sets and data-science techniques to ask questions about human behavior at the population level without sacrificing individual-level analyses. This study contributes to the conversation on the application of quantitative analyses of behavior in two ways. First, by demonstrating how data-science techniques allow researchers and practitioners to easily scale the application of quantitative models of behavior from one to millions of individuals. Second, when scaled in this way, data-science techniques allow researchers and practitioners to ask questions not previously possible. We discuss both of these through the example of testing the generality and utility of the GML to the applied context of professional baseball pitching.

Scaling the GML Fits from One to Millions

Identifying Putative Reinforcers and Punishers

Past researchers have examined how well the GML describes pitch allocation in a small sample of professional baseball pitchers (Cox et al., 2017; Falligant et al., 2020). Similar to studies extending the GML to other nonlaboratory contexts (e.g., Alferink et al., 2009; Reed et al., 2006; Romanowich et al., 2007; Rotta et al., 2020), these studies found that: the GML described pitch allocation well (i.e., VACs often greater than 0.80); estimated parameters were similar to laboratory studies on the GML; and estimated parameters changed in ways logically consistent with game contexts. But nonlaboratory settings are much more complex than laboratory settings. To fully understand the utility of a quantitative model of behavior to nonlaboratory contexts, researchers should apply the model to an entire population to determine the conditions under which the model performs well or poorly.

We used data-science techniques to collect, aggregate, and fit the GML to the behavioral patterns of every individual player within the population of MLB pitchers across 11 years and using 8,467,473 observations of behavior–environment relationships. By using these data-science techniques, the historical tradeoff between internal and external validity was no longer an issue or choice. We could simply look at all relevant individuals. In doing so, several intriguing results arose relative to how well the GML described pitch allocation.

The most immediate observation was just how unique participants from past research were. Past pitchers studied fell at the higher end of how well the GML describes pitch selection. That is, using strikes/outs as reinforcement, the GML was a good description of a subset of all pitchers. This suggests that other variables are required to accurately predict the pitch allocation of the other pitchers. In retrospect, this should be obvious. The reinforcers maintaining one organism's behavior may not be the reinforcers maintaining another organism's behavior. But this is often not the communicated finding in past research fitting the GML to behavior in nonlaboratory contexts (e.g., Alferink et al., 2009; Cox et al., 2017; Falligant et al., 2016; Reed et al., 2006; Romanowich et al., 2007; Rotta et al., 2020). The findings of this study raise the question of whether the participants in previous studies applying the GML to nonlaboratory contexts were also uniquely special in the extent the GML described their choices.

Leveraging data-science techniques also allowed us to explore differences in how well the GML described different subgroups of pitchers. It takes 400 ms for a 90-MPH fastball to move from the pitcher's hand to cross home plate. And a batter's swing takes approximately 150 ms from onset to contact with a baseball (if made). Thus, in fewer than 0.25 s, the batter must visually spot the pitch, its speed and trajectory, and choose whether and where to swing their bat. Any information that improves a hitters' chance of guessing the pitch type and location can lead to a significant advantage and better outcomes for the hitter (see Vigdor, 2020, for a recent incident in MLB demonstrating this importance). Though more advanced analytics allow hitters increasing access to data on the trends and tendencies of individual pitcher behavior, these data are not available for every pitcher.

We highlighted one example for how pitchers might be analyzed by subgroups so that hitters could better predict whether in-game contact with strikes/outs predicts pitch allocation or, all else equal, whether the pitcher is more likely to throw a fastball or non-fastball pitch. In this population, the GML described pitch allocation better in starters compared to relievers, and starters tended toward a higher bias toward fastballs. Hitters privy to this information might be able to alter their approach for an at-bat accordingly. This simple demonstration does not negate previous demonstrations for how the complexity of different baseball game situations relate to changes in pitch allocation (Cox et al., 2017; Falligant et al., 2020). But this simple demonstration highlights how easy it is to use open data sets to parse and analyze any of the dozens-to-hundreds of other possible subgroupings. Most important, data-science techniques allow researchers to efficiently study these many subgroupings in isolation or in combination based on their research interests or the socially significant behaviors they are analyzing.

The ability for researchers to flexibly and efficiently fit the GML to all members of a population and analyze varying subgroups of individuals or antecedent contexts may be useful to those working outside of professional baseball. Researchers interested in policy, law, or other situations wherein they are trying to regulate the behavior of large groups of individuals necessarily create one (or perhaps a few) contingencies that are then applied to all members of the group. A robust finding is that the reinforcer or punisher for one individual may not be a reinforcer or punisher for another individual (e.g., Fisher et al., 1992; Piazza et al., 1996; Verriden & Roscoe, 2019). But individuals working in policy, law, and regulation are unlikely to have the resources to conduct

reinforcer or punisher assessments for every individual in a population. Instead, they must change the behavior of as many individuals as possible using a limited understanding of the consequences that are functionally related to each individuals' behavior. Using data-science techniques with the right big data sets, behavior analytic researchers can contribute to this conversation by scaling the N of 1 analyses they're familiar with up to N s of millions. We demonstrated how this could be accomplished using the GML. But there's no reason this could not be done with alternative quantitative models of human behavior or behavior-analytic approaches to behavior–environment relations where data exist.

Understanding population level allocation in responding may also be useful to behavior analysts who are interested in setting more objective criteria for what counts as socially appropriate responding. For example, the GML describes behavior in academic contexts (e.g., Reed & Martens, 2008) and manipulating response effort and reinforcer quality changes behavior in expected directions (e.g., Neef et al., 1992; Reed & Martens, 2008). In addition, rates of delay discounting are associated with unhealthy behavioral patterns related to substance use (e.g., Bickel & Marsch, 2001; Johnson et al., 2007), food consumption (e.g., Appelhans et al., 2011), physical activity (e.g., Daugherty & Brase, 2010), and behavioral addictions (e.g., Negash et al., 2016). But what is the distribution of these measures in the population of individuals who emit “typical” or “healthy” patterns of responding? What kind of variability is observed in “healthy” patterns of responding? And how do different measures of behavior (e.g., discounting, demand, matching) with different reinforcers (e.g., physical activity, drugs, television) coalesce into population distributions of “healthy” behavioral repertoires?

Deciding what is socially appropriate behavior can be subjective. Determining the amount of behavior that is socially appropriate is easier for some responses as they should probably occur only in rare circumstances (e.g., killing other humans) or most circumstances (e.g., covering a cough or sneeze). But, for many responses, most people only agree that more or less responding is socially appropriate. For example, students should probably spend more time studying compared to doing something else. But, surely, they should not study 24 hours a day. What's the right amount of time? And how does relative time studying associate with other measures of social significance (e.g., grades, fluency)? Knowledge of the population distribution of behavior allows practitioners and researchers to establish cutoff points and determine the extent to which deviations objectively quantify an individual's behavior as atypical. Researchers and practitioners can then communicate how their clients' behavior fits within that population distribution (e.g., Figure 3) and then listen to their client as to whether any differences from “typical” responding should be addressed or celebrated.

The Utility of Quantitative Models of Behavior in Nonlaboratory Contexts

Understanding, fitting, and leveraging quantitative models of behavior in applied contexts requires effort and skill. For example, one must understand how and what data to collect, how to fit the model, how to interpret model parameters, and how to translate the model parameters into an effective behavior change program. Before attempting to incorporate any quantitative model of behavior into an applied context, applied behavior analysts should ask whether other logical relations between behaviors

and reinforcers might provide sufficient information for their goals. Further, many quantitative models of human behavior require a fair amount of data. Before practitioners begin to spend time and resources trying to apply quantitative models of behavior to their clinical environments, it is important for them to ask, “What is gained by adding this model to the analytic milieu for this client?”

Data-science techniques allow practitioners to determine the probability that simpler, logical analyses provide sufficient practical information on the behavior-environment relations of interest. For example, for professional baseball pitchers, it is possible that simply throwing fastballs faster, throwing slow pitches more slowly, and increasing spin rate on breaking pitches will lead to higher rates of strikes and outs. No math is necessary to make these recommendations. Figures 4, 5 and 6 were aimed at the question of whether the GML provides useful information beyond logical relations between pitch topography and strikes/outs. For several dozen pitchers, the speed or break of the pitches thrown were nearly perfectly correlated with the number of strikes or outs several dozen pitchers contacted. For practitioners working with these pitchers, the utility of the GML is likely low as they can simply coach and condition the pitcher to throw fastballs faster and their breaking pitches with greater break.

However, putting in the effort to understand and apply the GML would likely be useful for practitioners working with the remaining ~2,000 pitchers. For these pitchers, data-science techniques allowed us to determine there was not a strong correlation between raw or relative pitch speed and strikes, nor for raw or relative break distance of breaking pitches. However, the decent fits of the GML to most pitchers’ behavior suggest that reframing the analysis of pitching around relative allocation of pitch types provided useful information toward predicting the ratio that pitch types contacted strikes and outs. Stated differently, the GML analyzes pitch types relative to one another rather than in isolation. The present study indicates that—for the applied context of professional baseball pitching—the GML provides useful information beyond analyzing the topographical characteristics of pitches in isolation. This was further supported by Figure 6 wherein bias toward fastballs (a relative measure pitch allocation) was strongly correlated with fastball speed in 65% of the pitchers. Using the GML reframes the coaching conversation around the relative use of fastballs and other pitch types rather than focusing on the specific topographies of any one pitch type.

To make practical use of the GML in nonlaboratory contexts, the estimated parameters for each individual should likely relate in some way to measures of good or poor performance. In this study, we found that estimated parameters from the GML were not strongly correlated with common pitching performance metrics. Further, where weak correlations were observed, it was between the VAC by the GML and metrics suggesting opposing success—wins and losses, innings pitched, earned runs allowed, and home runs allowed. This suggests that the GML likely describes pitcher allocation better simply as a function of having more data on a pitcher. This observation is further supported by the better fits of the GML to starting pitchers compared to relief pitchers. In total, these findings suggest that—though the GML may describe pitch allocation well for some individuals and in some contexts, its current status as a useful metric of pitcher behavior in professional baseball might be questionable.

Perhaps most important, we would not have obtained the above findings without using techniques from data science. By way of data-science techniques, we were able to analyze the GML for each individual in the professional pitching population. And,

through this analysis, we could determine that there are likely many additional variables that need to be considered to accurately describe, predict, and control the pitch allocation of professional pitchers across varying contexts, roles, and points in each individual's career. Without the current study, the findings from previous research (Cox et al., 2017; Falligant et al., 2020) may have inaccurately led behavior analysts, pitching coaches, and managers to assume the broad generality and utility of the GML to professional baseball pitching.

Conclusion

Researchers have traditionally had to choose whether to allocate resources to obtain a lot of data from a few individuals, or to obtain few data from a lot of individuals. Advances in data collection, storage, and analytic efficiencies now allow researchers and practitioners to get a lot of data for a lot of individuals. Stated differently, behavioral data-science techniques allow researchers and practitioners to combine the benefits of within-subject methods with an understanding of population distributions for behaviors of interest. We demonstrated how behavior scientists can use data-science techniques to analyze response–environment relations for every individual in a population. Data-science techniques make it easy to analyze population distributions of behavior–environment relations, segment and compare subgroups within the larger population, and track individuals within the population for normative comparisons (see Appendix for an applied example).

For quantitative models of behavior, analyses like the present study allow researchers to understand the generalizability and utility of using the model to describe and predict behavior in nonlaboratory contexts of social significance. For us, the socially important context was pitch allocation in professional baseball. Leveraging data-science techniques to analyze population distributions of GML fits and estimated parameters allowed us to learn several things. First, the overall variance accounted for by the GML is less than observed with the small samples used in past research (Cox et al., 2017; Falligant et al., 2020). Second, analyzing relative pitch types and outcomes relative to other pitch types and outcomes provides useful information beyond analyzing each pitch type in isolation (i.e., the theoretical assumptions of the GML appear useful in practice). Finally, the fit and estimated parameters of the GML were not correlated with commonly used metrics of pitching performance. In total, these three findings suggest that pitcher choice in the complex and dynamic game of baseball is more complex than previously assumed. Researchers who learn to leverage behavioral data-science techniques will be able to ask questions about this complex space more efficiently than past tools would have allowed.

Data Availability Data are available for download at Willman (2021)

Declarations

Conflicts of Interest/Competing Interests None to Declare

Availability of code Notebooks to replicate the current study are available at Cox (2020)

Appendix

Example Application: Extending a Matching-Law Analysis of Delinquent Behavior

McDowell and Caron (2010) described the verbal behavior of 13- and 14-year-old antisocial boys using data from the Oregon Youth Study conducted by the Oregon Social Learning Center (OSLC). The study entailed pairing each boy with a friend around the same age, sitting the pair in front of a video camera, introducing a topic, and then leaving them alone to discuss the topic. OSLC researchers then coded videotapes of the boys' conversations into two categories: rule-break talk (i.e., discussion of breaking social norms, customs, or laws) and normative talks (i.e., any other speech that did not meet the rule-break talk criteria). Frequency of verbal bouts served as the target behavior and positive social consequences from the friend (e.g., smiling, nodding, laughing) were coded as the putative reinforcer. OSLC researchers also calculated child deviance scores from the National Youth Survey measure (Elliot et al., 1985), which McDowell and Caron used to stratify the boys into four groups by quartiles and fit the GML to the data for all boys within each group. In comparing the parameter estimates across these four GML fits, McDowell and Caron (2010) found that log-bias parameter estimates towards rule-break talk scaled directly with the child deviancy scores (i.e., 1st quartile had the least bias towards rule-break talk, whereas the 4th quartile had the greatest bias towards rule-break talk).

Although impressive in its aim to quantitatively describe complex social behavior, McDowell and Caron's (2010) approach to describing delinquency via the matching law was challenging conceptually. The boys were exposed to only one reinforcement schedule (i.e., only one datum was collected for each boy), which precluded individual GML fits. This is challenging because the GML is not designed to describe data aggregated across multiple organisms (Caron, 2013). As shown in this study, individual GML parameter estimates fall in a distribution of (currently) unknown size and shape. Pooled analyses obfuscate this distribution. Further, McDowell and Caron's analysis is tautological by showing that boys with higher self-reported delinquency scores showed a preference for discussing delinquent behavior. A more powerful test of the GML would be to use log-bias parameter estimates derived for each boy to predict each boy's delinquent behavior outside the laboratory (e.g., school truancy, unlawful behavior). Related to this, the reinforcers analyzed may not have been directly related to actual delinquent behavior because the study measured sensitivity to reinforcement for discussing delinquent behavior in a laboratory setting—not sensitivity to reinforcement for delinquent behavior in nonlaboratory contexts.

Using the methodological techniques described in this study, researchers could leverage data obtained from schools implementing School-Wide Positive Behavior Support (SWPBS; Sugai et al., 2000; Sugai & Horner, 2006) to better understand the utility of GML descriptions of delinquent behavior. In SWPBS, teachers collect data regarding prosocial/productive (e.g., completing homework assignments, listening to teacher directions) and delinquent/deviant (e.g.,

disruptive comments, inappropriate and/or distracting actions) classroom behaviors using a School-Wide Evaluation Tool (SET; Horner et al., 2004). SETs also often include a place to record the consequences for prosocial and delinquent behaviors. These data are recorded for each student, in each of their classes, across the entire school. Thus, each student contacts multiple consequence “schedules” (i.e., one from each class) and has multiple corresponding response rates. Using the approach in the current study, the GML could be fit to these data. The result would be school-wide distributions of GML fits and parameter estimates that relate prosocial and delinquent behavior to the putative reinforcer and punisher schedules contacted across classrooms.

The advantages to this prospective study would be threefold. First, fits of the GML would be to individual-level data rather than pooled across multiple individuals. Second, the distribution of log-bias parameter estimates for the entire school would provide a direct and objective measure of preference for prosocial/delinquent behavior and how “at-risk” students compare to the larger population in their sensitivity to reinforcers that are manipulable in the school context. Finally, the data would be directly related to the behaviors of interest as opposed to verbal statements about those behaviors. And all this could be done using data already being collected. Taken together, this approach would result in a more theoretically sound and comprehensive analysis of the utility of the GML in describing and predicting delinquent behavior. In practice, implementing these analyses in schools with SWPBS may allow administrators to target interventions and resources more effectively to students at-risk for delinquent behavior.

References

- Alferink, L. A., Critchfield, T. S., Hitt, J. L., & Higgins, W. J. (2009). Generality of the matching law as a descriptor of shot selection in basketball. *Journal of Applied Behavior Analysis*, 42, 595–608. <https://doi.org/10.1901/jaba.2009.42-595>.
- Appelhans, B. M., Woolf, K., Pagoto, S. L., Schneider, K. L., Whited, M. C., & Liebman, R. (2011). Inhibiting food reward: delay discounting, food reward sensitivity, and palatable food intake in overweight and obese women. *Obesity*, 19(11), 2175–2182. <https://doi.org/10.1016/j.appet.2012.07.009>.
- Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, 22, 231–242. <https://doi.org/10.1901/jeab.1974.22-231>.
- Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior*, 32(2), 269–281. <https://doi.org/10.1901/jeab.1979.32-269>.
- Bickel, W. K., & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction*, 96(1), 73–86. <https://doi.org/10.1046/j.1360-0443.2001.961736.x>.
- Boomhower, S. R., & Newland, M. C. (2016). Adolescent methylmercury exposure affects choice and delay discounting in mice. *Neurotoxicology*, 57, 136–144. <https://doi.org/10.1016/j.neuro.2016.09.016>.
- Borrero, J. C., Crisolo, S. S., Tu, Q., Rieland, W. A., Ross, N. A., Francisco, M. T., & Yamamoto, K. Y. (2007). An application of the matching law to social dynamics. *Journal of Applied Behavior Analysis*, 40(4), 589–601. <https://doi.org/10.1901/jaba.2007.589-601>.
- Caron, P. O. (2013). On applying the matching law to between-subject data. *Animal Behaviour*, 85(4), 857–860. <http://www.sciencedirect.com/science/article/pii/S000334721300050X>. <https://doi.org/10.1016/j.anbehav.2013.01.022>

- Cero, I., & Falligant, J. M. (2019). Application of the generalized matching law to chess openings: A gambit analysis. *Journal of Applied Behavior Analysis*. Advance online publication, 53(2), 835–845. <https://doi.org/10.1002/jaba.612>.
- Cox, D. J. (2020). Man vs. machine. *GitHub repository*. <https://doi.org/10.5281/zenodo.4088070>
- Cox, D. J., Sosine, J., & Dallery, J. (2017). Application of the matching law to pitch selection in professional baseball. *Journal of Applied Behavior Analysis*, 50(2), 393–406. <https://doi.org/10.1002/jaba.381>.
- Daugherty, J. R., & Brase, G. L. (2010). Taking time to be healthy: Predicting health behaviors with delay discounting and time perspective. *Personality and Individual Differences*, 48(2), 202–207. <https://doi.org/10.1016/j.paid.2009.10.007>.
- Elliot, D. S., Huizing, D., & Ageton, S. S. (1985). *Explaining delinquency and drug use*. Sage.
- Falligant, J. M., Boomhower, S. R., & Pence, S. T. (2016). Application of the generalized matching law to point-after-touchdown conversions and kicker selection in college football. *Psychology of Sport & Exercise*, 26, 149–153. <https://doi.org/10.1016/j.psychsport.2016.07.006>.
- Falligant, J. M., Cero, I., Kranak, M. P., & Kurtz, P. F. (2020). Further application of the generalized matching law to multialternative sports contexts. *Journal of Applied Behavior Analysis*. <https://doi.org/10.1002/jaba.757>.
- Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis*, 25(2), 491–498.
- Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, 3(1), 30–37. <https://doi.org/10.1111/j.1750-8606.2008.00072.x>.
- Foltz, J. L., May, A. L., Belay, B., Nihiser, A. J., Dooyema, C. A., & Blanck, H. M. (2012). Population-level intervention strategies and examples for obesity prevention in children. *Annual Review of Nutrition*, 32, 391–415. <https://doi.org/10.1146/annurev-nutr-071811-150646>.
- Hagopian, L. P. (2020). The consecutive controlled case series: Design, data-analytics, and reporting methods supporting the study of generality. *Journal of Applied Behavior Analysis*, 53(2), 596–619. Advance online publication. <https://doi.org/10.1002/jaba.691>
- Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 4, 267–272. <https://doi.org/10.1901/jeab.1961.4-267>.
- Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior*, 13, 243–266. <https://doi.org/10.1901/jeab.1970.13-243>.
- Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, R., & Bolland, J. B. (2004). The School-Wide Evaluation Tool (SET): A research instrument for assessing School-Wide Positive Behavior Support. *Journal of Positive Behavior Interventions*, 6(1), 3–12. <https://doi.org/10.1177/10983007040060010201>.
- Johnson, M. W., Bickel, W. K., & Baker, F. (2007). Moderate drug use and delay discounting: a comparison of heavy, light, and never smokers. *Experimental & Clinical Psychopharmacology*, 15(2), 187–194. <https://doi.org/10.1037/1064-1297.15.2.187>.
- Klapes, B., Calvin, O. L., & McDowell, J. J. (2020). A discriminated rapid-acquisition laboratory procedure for human continuous choice. *Journal of the Experimental Analysis of Behavior*, 114(1), 142–159. <https://doi.org/10.1002/jeab.612>.
- Madden, G. J., & Perone, M. (1999). Human sensitivity to concurrent schedules of reinforcement: Effects of observing schedule-correlated stimuli. *Journal of the Experimental Analysis of Behavior*, 71(3), 303–318. <https://doi.org/10.1901/jeab.1999.71-303>.
- McDowell, J. J. (2013). On the theoretical and empirical status of the matching law and matching theory. *Psychological Bulletin*, 139, 1000–1028. <https://doi.org/10.1037/a0029924>.
- McDowell, J. J., & Caron, M. C. (2010). Bias and undermatching in delinquent boys' verbal behavior as a function of their level of deviance. *Journal of the Experimental Analysis of Behavior*, 93(3), 471–483. <https://doi.org/10.1901/jeab.2010.93-471>.
- McKinney, W. (2018). *Python for data analysis*. O'Reilly Media.
- Millman, K. J., & Aivazis, M. (2011). Python for scientists and engineers. *Computing in Science & Engineering*, 13, 9–12.
- Neef, N. A., Mace, F. C., Shea, M. C., & Shade, D. (1992). Effects of reinforcer rate and reinforcer quality on time allocation: Extensions of matching theory to educational settings. *Journal of Applied Behavior Analysis*, 25(3), 691–699. <https://doi.org/10.1901/jaba.1992.25-691>.

- Negash, S., Sheppard, N. V. N., Lambert, N. M., & Fincham, F. D. (2016). Trading later rewards for current pleasure: Pornography consumption and delay discounting. *Journal of Sex Research*, 53(6), 689–700. <https://doi.org/10.1080/00224499.2015.1025123>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning*, 12, 2825–2830.
- Piazza, C. C., Fisher, W. W., Hagopian, L. P., Bowman, L. G., & Toole, L. (1996). Using a choice assessment to predict reinforcer effectiveness. *Journal of Applied Behavior Analysis*, 29(1), 1–9.
- Pierce, W. D., & Epling, W. F. (1995). The applied importance of research on the matching law. *Journal of Applied Behavior Analysis*, 28(2), 237–241. <https://doi.org/10.1901/jaba.1995.28-237>.
- Reed, D. D., Critchfield, T. S., & Martens, B. K. (2006). The generalized matching law in elite sport competition: Football play calling as operant choice. *Journal of Applied Behavior Analysis*, 39, 281–297. <https://doi.org/10.1901/jaba.2006.146-05>.
- Reed, D. D., & Martens, B. K. (2008). Sensitivity and bias under conditions of equal and unequal academic task difficulty. *Journal of Applied Behavior Analysis*, 41(1), 39–52. <https://doi.org/10.1901/jaba.2008.41-39>.
- Romanowich, P., Bourret, J., & Vollmer, T. R. (2007). Further analysis of the matching law to describe two and three-point shot allocation by professional basketball players. *Journal of Applied Behavior Analysis*, 40, 311–315. <https://doi.org/10.1901/jaba.2007.119-05>.
- Roto Sports, Inc. (2021). MLB Player Stats [Data set]. *Rotowire*. <https://www.rotowire.com/baseball/stats.php>
- Rotta, K., Kranak, M. P., & Poling, A. (2020). Spikes, tips, and points: Matching college volleyball attacks? *Behavior Analysis: Research & Practice*, 21(1), 42–50. Advance online publication. <https://doi.org/10.1037/bar0000197>
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference*. <https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf>
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Authors Cooperative.
- Skinner, B. F. (1966). Operant behavior. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 12–32). Appleton-Century-Crofts.
- Shimp, C. P. (2020). Molecular (moment-to-moment) and molar (aggregate) analyses of behavior. *Journal of the Experimental Analysis of Behavior*, 114(3), 392–429. <https://doi.org/10.1002/jeab.626>.
- Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull III, H. R., Wickham, D., Wilcox, B., & Ruef, M. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions*, 2(3), 131–143. <https://doi.org/10.1177/109830070000200302>.
- Sugai, G., & Horner, R. H. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review*, 35(2), 245–259. <https://doi.org/10.1080/02796015.2006.12087989>.
- Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, 3(31), 1026. <https://doi.org/10.21105/joss.01026>.
- Verriden, A. L., & Roscoe, E. M. (2019). An evaluation of a punisher assessment for decreasing automatically reinforced problem behavior. *Journal of Applied Behavior Analysis*, 52(1), 205–226. <https://doi.org/10.1002/jaba.509>.
- Vigdor, N. (2020). The Houston Astros' cheating scandal: Sign-stealing, buzzer intrigue and tainted pennants. *The New York Times*. <https://www.nytimes.com/article/astros-cheating.html>. Accessed 16 Jul 2021.
- Vollmer, T. R., & Bourret, J. (2000). An application of the matching law to evaluate the allocation of two- and three-point shots by college basketball players. *Journal of Applied Behavior Analysis*, 33(2), 137–150. <https://doi.org/10.1901/jaba.2000.33-137>.
- Willman, D. (2021). *Major League Baseball (MLB) Statcast database* [Data set]. BaseballSavant. https://www.baseballsavant.mlb.com/statcast_search