

Multilevel analysis of matching behavior

PIER-OLIVIER CARON

UNIVERSITÉ TÉLUQ

Multilevel modeling has been considered a promising statistical tool in the field of the experimental analysis of behavior and may serve as a convenient statistical analysis for matching behavior because it structures data in groups (or levels) to account simultaneously for the within-subject and between-subject variances. Heretofore, researchers have sometimes pooled data erroneously from different subjects in a single analysis by using average ratios, average response and reinforcer rates, aggregation of subjects, etc. Unfortunately, this leads to loss of information and biased estimations, which can severely undermine generalization of the results. Instead, a multilevel approach is advocated to combine several subjects' matching behavior. A reanalysis of previous data on matching behavior is provided to illustrate the method and point out its advantages. It illustrates that multilevel regression leads to better estimations, is more convenient, and offers more behavioral information. We hope this paper will encourage the use of multilevel modeling in the statistical practices of behavior analysts.

Key words: choice, matching behavior, matching law, multilevel modeling, pooled data, statistical analysis

The purpose of the science of behavior is to find order; that is, to identify certain events that stand in lawful relation to others (Skinner, 1953). By manipulating conditions specified by a phenomenon, we can predict that an event will occur. Precise measurement and mathematical models provide a common framework for understanding behavioral phenomena (Mazur, 2006) and, hopefully, for predicting behavior. Mathematical modeling allows us to formulate theories precisely, unambiguously, and clearly by using straightforward formal language. It also allows comparison between opposing models and the development of new and more powerful models. Although it may be difficult to acquire such language, it was the cornerstone for the development of the quantitative analysis of behavior and the matching law (Mazur, 2006). These exemplify the effectiveness of mathematical models for discovering and exploring the laws that govern behavior.

I would like to thank Anne-Josée Piazza and Matthew Retino for their revisions on preliminary versions of this paper. I would like to thank Alexandre Gellen-Kamel with the extraction of the data. Finally, I am in debt to Michal E. Young, Greg Jensen, Matti Vuorre and an anonymous reviewer for their thoughtful comments on my manuscript.

Conflict of interest: none.

Address correspondence to: Pier-Olivier Caron, Université TÉLUQ, Département des Sciences humaines, Lettres et Communications, 5800, Rue Saint-Denis (Bureau 1105) Montréal (Québec, Canada), H2S 3L5 (e-mail: pocaron19@gmail.com)

doi: 10.1002/jeab.510

The Matching Law

There is a long tradition within behavior analysis of studying a subject's choices between two options. The concurrent schedule is defined as the well-known schedule of reinforcement in which behavior analysts can study decision processes. In such situations, the subject has to choose between two options, each of which are associated with its own differential reinforcer rate, generally a variable interval or a random interval schedule of reinforcement. Over numerous sessions, the subject can choose continuously between alternatives. In this procedure, a behavioral pattern known as matching behavior emerges. This is well described by a simple equation, often referred to as the matching law:

$$\frac{b_1}{b_2} = \frac{r_1}{r_2}, \quad (1)$$

where b refers to response rate, r to reinforcer rate, and the indices (1 and 2) specify the two options. The matching law states that the ratio of two response rates matches the ratio of their respective reinforcer rates and will tend toward an equilibrium (Herrnstein, 1961). Accordingly, the equilibrium will remain even if the reinforcer rate ratio changes. In this case, the subject's behavior will adjust to the new ratio. Over the course of five decades, there have been a tremendous number of studies on potential explanations for why the subject's behavior reaches such an

equilibrium (Davison & McCarthy, 1988; de Villiers & Herrnstein, 1976; McDowell, 2013). In parallel, it has been found that a subject's response allocation systemically deviates from Equation (1). To account for the deviation, Baum (1974) proposed the generalized matching law (GML). It takes the following form:

$$\frac{b_1}{b_2} = c \left(\frac{r_1}{r_2} \right)^a, \tag{2}$$

where a refers to sensitivity to reinforcement, or the degree to which an organism adjusts its response ratio according to the reinforcer ratio, and c refers to the bias, or the behavioral preference for one response (the numerator) over the other (the denominator) when the reinforcement rates are equal. For instance, subjects' behavior may follow Equation (1), that is, match, $a \approx 1$; or undermatch, $a < 1$; or overmatch, $a > 1$; or may have a strong bias for either alternative. The subject's behavior thus might not strictly match reinforcer ratios, but it should do so given the parameters a and c .

The GML is generally presented in a log form:

$$\log\left(\frac{b_1}{b_2}\right) = a \log\left(\frac{r_1}{r_2}\right) + \log c, \tag{3}$$

a linear form of Equation (2) that can be easily analyzed using regression analysis. The GML has been found to account for a variety of behaviors, species, and operant situations (Baum, 1979; Davison & McCarthy, 1988; McDowell, 2013). This equation is generally preferred because it can be fitted by a least square regression to obtain the bias, sensitivity, and explained variance (R^2). It also graphs itself intuitively and elegantly. Figure 1 depicts Equation (2), the power function, in the top panel and Equation (3), the linear function, in the bottom panel. Sensitivity is the degree of the slope in the bottom panel and the rate of growth in the top panel. Bias is clearly perceptible in the linear function as the intercept, the point where the line crosses the abscissa. If the bias is positive, the preference is for the behavior in the numerator; if it is negative, the preference is for the denominator. In the power function, bias and sensitivity are not as easy to interpret in the curvilinear slope. Still, the results remain the same for both functions; sensitivity, variance accounted for, and bias are equal: $\log c = \log(1.286) = .109$, a bias toward b_1 .

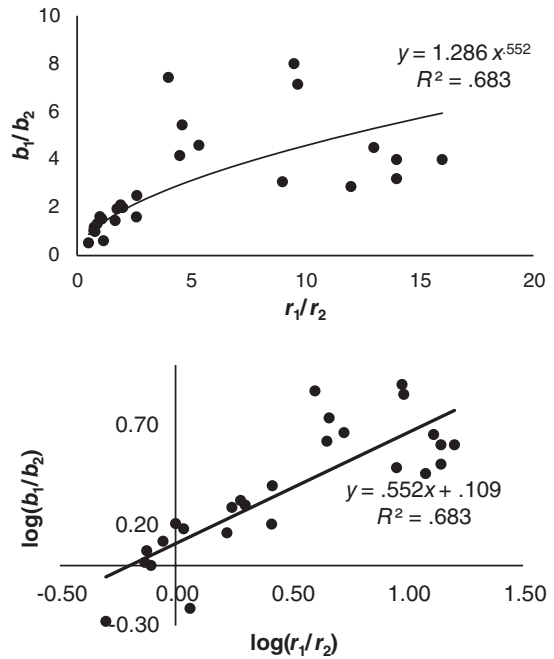


Fig. 1. Depiction of Equation (2) and Equation (3). Top panel: Equation (2) is the power function representation of the GML. The variable y refers to the response rate ratios and x corresponds to the reinforcer rate ratios. Bottom panel: Equation (3) is the linear representation. The variable y refers to the log response rate ratios and x corresponds to the log reinforcer rate ratios.

Challenges That Arise When Pooling Individual Models

Matching behavior is a single subject-orientated model; that is, parameters of the GML are *unique* to a given subject (Caron, 2013). These parameters can be different from one subject to the other. Thus, one of the main challenges is to make general claims regarding matching behavior by considering how a specific subject behaves in the concurrent schedule (the within-subject variance), while accounting for the differences between subjects' matching behavior (the between-subjects variance). In other terms, the challenge is how to generalize the information known on a few but extensively studied subjects over a long period of time in order to build a global and accurate portrait of matching behavior. For instance, in a seminal paper by Baum and Rachlin (1969), six pigeons had different sensitivities ($a = 1.09, 1.15, 1.29, .98, .84, .63$) and different biases ($\log c = -.29, -.49, -.06, -.27, -.25,$

-.12), even though they were all placed in the same operant situations (a concurrent schedule in which being on the left or right side of the apparatus was associated with a variable-interval schedule). In the same operant situation, two subjects' behavior undermatched, whereas three others overmatched, and the degree to which they did varied. Thus, while it would be simple to claim that the group's parameter estimate is the simple average of the subject parameter estimates (that is $a = .997$ and $\log c = -.247$), this approach is misleading because it ignores the characteristics of each subject's data that affect the precision of each estimate.

One strategy to support a general claim using a within-subject design is to pool subjects. However, pooling has led to troublesome conclusions, since the use of direct aggregation of data across subjects or averaged parameters could not take into account both within- and between-subjects variances (Caron, 2013). Depending on the case, researchers often (a) pooled data across ratios so that each ratio was analyzed independently; (b) pooled ratios across subjects, reducing the between-subjects variances and eliminating the variance induced by the ratios; (c) averaged subjects' behaviors in a single mean for each ratio, thus eliminating the within-subject variances and potentially overemphasizing unusual matching behavior (Caron, 2013). Inappropriately pooling data can conceal the number of different ratios used, the number of data points by ratio, and the variability (and consistency) between individuals, all of which are critical pieces of information about the precision of the slope (sensitivity) and the intercept (bias) that, when lost, severely limit the results' generalization.

Multilevel Regression of Matching Behavior

A more recent statistical technique that has been either unavailable or not widely disseminated in the field, which is probably explained by the uneasy relationship between behavior analysis and statistics (Young, 2017b), is multilevel modeling. Multilevel modeling (see de Leeuw & Meijer, 2008; Finch, Bolin & Kelley, 2014; Gelman & Hill, 2006; Hox, 2010) is relevant when (a) the observations are correlated or clustered over a certain dimension (spatial, nonspatial, or, for the current discussion, temporal); (b) the studied processes operate simultaneously at more than one level (behavior,

individual, phylogenetic, etc.); and (c) there is an interest in describing the variability and heterogeneity in the population. This statistical analysis has been recognized already as a promising tool for the experimental analysis of behavior, such as discounting (Young, 2017a, 2018).

Multilevel modeling structures data in groups or levels. Such data is the bread and butter of behavior analysts, since behavioral data are nested within subjects. Within-subject design may also arise as the result of a specific research design, operant situations, pre- and postintervention observations, different topographical responses, or different reinforcement schedules. In within-subject design, though, the average correlation between reinforcers and behaviors of a subject may be different from that of another one. When this assumption of the independence of observations is violated, the estimates of standard errors of traditional statistics (i.e. regressions) decrease, which then increase the outlying errors to reveal spurious results (Hox, 2010). If this is the case, it is necessary to use a multilevel regression. The most interesting property of multilevel modeling for matching behavior is that it accounts simultaneously for the within-subject and the between-subjects variances. In matching behavior, this involves estimating the model's parameters (a and $\log c$) for all subjects, but also a general estimate shared among subjects simultaneously. This results in more stable estimates of subjects' matching behavior because the complete data set (all behavior ratios and reinforcer ratios of all subjects) is considered when estimating both the group and subject parameters.

Modeling repeated measures data can be envisioned on a continuum between nonpooling and complete pooling of data. At the nonpooling pole, each subject's behaviors are analyzed independently, that is, a regression by subject, which is generally what has been done in the past for matching behavior (in Fig. 1, for instance). Comparatively, complete pooling gathers all response and reinforcer ratios together, regardless of the source of behavior. This analysis is carried in a single regression where all subjects' behaviors are pooled or combined together. These two ideas are presented in Figure 2. Dashes, circles, and crosses correspond to data sets from three different subjects. In the top panel, each subject's behavior was fitted using Equation (3) which

corresponds to three nonpooled regressions (each subject is analyzed independently). There is a variety of matching behavior: One subject's behavior slightly undermatched and showed no preference, while the behavior of the other two strongly overmatched and showed a strong preference for the first option (b_1). This analysis did not consider the between-subjects variance; the variability potentially shared among subjects' behavior was ignored. The bottom panel (dashed line) shows the pooled estimates that could be obtained if all subjects' behavior was analyzed in a single regression, that is, all data is merged in a single data set regardless of who emitted which responses. It represents a rough average of matching behavior among subjects but does not consider the within-subject variance, the precision of the subject's matching behavior, and it assumes incorrectly that each observation is independent.

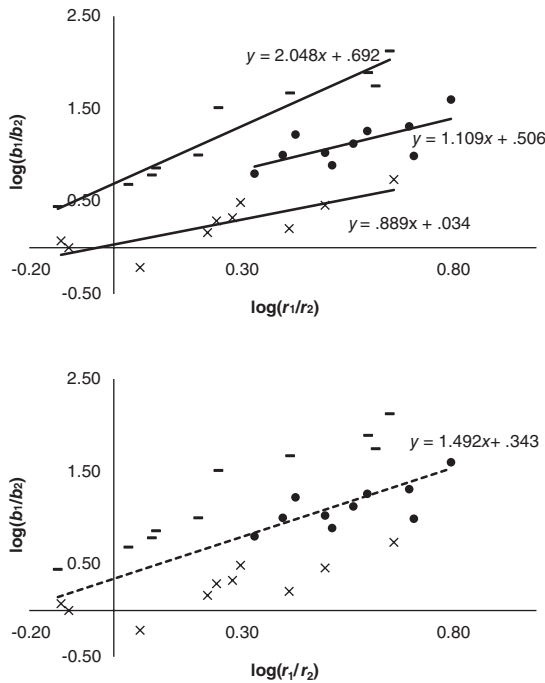


Fig. 2. Illustration of the nonpooled and completely pooled regressions. In equations, the variable y refers to the log response rate ratios and x corresponds to the log reinforcer rate ratios. Top panel: The nonpooled regressions (full lines) correspond to three regressions for each subject (dashes, circles, and crosses). Bottom panel: The completely pooled model (dash line) shows the regression when all response and reinforcer ratios are combined into one data set regardless of the subject.

Multilevel modeling constitutes a step further in the capacity to combine several subjects' behavior, as it stands between these two poles. The parameters of all subjects are considered in order to infer group characteristics using both the behavioral (within) and the subject (between) levels simultaneously (Gelman & Hill, 2006). Using maximum likelihood (or Bayesian estimation methods, depending on the software used), the analysis solves a series of equations that estimate the model's parameters by treating all subjects' matching equations and pooled estimates together.

A brief digression here to acknowledge, first, that Equation (3) can be rewritten as

$$\log\left(\frac{b_{1,s}}{b_{2,s}}\right) = a_s \log\left(\frac{r_{1,s}}{r_{2,s}}\right) + \log c_s \quad (4)$$

in which b_1 and b_2 indicate the number of behaviors emitted by a subject s (indices for sessions are not added to Equation [4]), to emphasize that sensitivity (a_s) and bias ($\log c_s$) are random variables fixed within an individual (specific to a given subject), but parameterized in the population (vary across subjects in a random, but predictive manner). We may theorize that these variables vary between individuals according to a Gaussian distribution with population means μ_a and $\mu_{\log c}$ and their corresponding variances–covariances σ_a^2 , $\sigma_{\log c}^2$, and $\rho\sigma_a\sigma_{\log c}$, such that

$$\begin{pmatrix} a_s \\ \log c_s \end{pmatrix} \sim \begin{pmatrix} \mu_a \\ \mu_{\log c} \end{pmatrix}, \begin{pmatrix} \sigma_a^2 & \rho\sigma_a\sigma_{\log c} \\ \rho\sigma_a\sigma_{\log c} & \sigma_{\log c}^2 \end{pmatrix} \quad (5)$$

This is an assumption for the multilevel model and not for the subjects themselves: This could be empirically falsified or theorized otherwise. From Equation (5), slopes (a_s) and intercepts ($\log c_s$) are random coefficients that are free to vary across subjects and differ according to their given means and variance–covariance matrix. According to the central limit theorem, we expect that with a sufficiently large number of subjects and behavioral observations, sensitivity and bias should tend toward μ_a and $\mu_{\log c}$ and their corresponding variance toward σ_a^2 and $\sigma_{\log c}^2$. The parameters can be interpreted from a phylogenetic

perspective (matching behavior could be stable within a given species), or by the analysis of different operant situations (matching behavior can be stable within specific schedules of reinforcement). This could be developed in a research program. We shall now close this digression to pursue the discussion on the continuum between nonpooling and complete pooling of data.

Multilevel modeling uses a weighted average of the nonpooling estimates of a subject and the population mean (a hyperparameter similar to complete pooling) to shrink (defined loosely, to improve an estimate by combining it with other information) the parameters (Gelman & Hill, 2006). Shrinkage leads to arguably better estimates and more accurate predictions than individual averages as they are closer to the true parameter in the population (Copas, 1983; Efron & Morris, 1977). The weight is sensitive to the number of data by subject n_s , the variance of the dependent variable (i.e., the response ratio, $\sigma^2_{\log(\frac{b}{a})}$), and the population variance σ_a^2 , that is, the variance of the hyperparameter (population sensitivity). The weight is a general case distinguishing the complete pooling estimate (when $\sigma_a^2 \rightarrow 0$), that is, when all subjects are pooled, from the nonpooling model (when $\sigma_a^2 \rightarrow \infty$). Shrinkage is carried relative to how accurate our guesses appear to be (Greenland, 2000). The method accounts for the consistency of matching behavior (notably, sample size and the response ratio variance), as well as the convergence toward the subjects' average (Gelman & Hill, 2006). There is less shrinkage when the variance of the hyperparameter is small (subjects show similar pattern of matching behavior), more shrinkage when there is more noise in the response rate ratios within subjects, and more shrinkage for subjects with fewer data points (potentially less consistent matching behavior).

There are other strong benefits to considering multilevel modeling for matching behavior (see for instance Young, 2017a, 2018; for a comprehensive overview, but also handbooks on multilevel modelling cited earlier). First, since it relies on maximum likelihood (or Bayesian estimation), it compensates for missing values and design imbalance. It can thus increase power (the probability of rejecting the absence of an effect when there is a true effect) by minimizing the rejection of

subjects and increasing the accuracy of estimates. Second, multilevel analysis weighs subjects according to the consistency of the subject's behavior. The behavior of subjects observed frequently (high number of data points) will have more weight in the estimation than subjects providing noisy, improbable, or limited data, or those with matching behavior severely departing from the sample. Third, it leads to more power by combining all observations from all subjects. Rather than having a certain amount of data points for each regression, it considers all data points from all the regressions at the same time. It also decreases the type I error rate (finding a significant relation when there is none) by reducing the number of tests. Instead of realizing a regression by subject (and a pooled regression), a single instance of analysis is conducted rather than the number of subjects plus one, thus not inflating the type I error rate deliberately. Finally, multilevel analysis is easily implemented in several statistical programs such as R, Python, Matlab, Mplus, and SPSS. It is worth noting that the statistical software handles all technical computations automatically; the user must prepare the data and implement the correct syntax to produce the desired results.

Comparing Nonpooled, Pooled and Multilevel Regressions of Matching Behavior

To illustrate the use of multilevel modeling, data¹ collected by Davison and Hogsden (1984) was used. The current analysis will allow for direct comparison of individual (nonpooled) matching behavior and matching behavior analyzed via multilevel modeling.² The purpose of the study was to investigate mixed versus fixed reinforcers' durations. Six pigeons were trained on 34 different concurrent variable-interval schedules of reinforcement. The schedule associated with the left key was unusual; it was associated with two states differentially reinforced,

¹The rdata file is available in the supplementary material.

²We use the data from Davison and Hogsden (1984) reported in the appendix (and available in the supplementary files), even though it would be more interesting to have the complete data set. The data reported are averages of the last five sessions; thus, some within-subject variance has been lost. Still, the data is interesting due to its well-carried design.

Table 1

Summary of subjects' matching behavior in Davison and Hogsden (1984)

Id	Nonpooled regressions		
	a (<i>s.e.</i>)	$\log c$ (<i>s.e.</i>)	R^2
1	.723 (.052)	.248 (.072)	.980
2	.726 (.049)	.039 (.067)	.982
3	.901 (.027)	.248 (.045)	.997
4	1.049 (.051)	.305 (.077)	.991
5	.923 (.070)	-.108 (.091)	.978
6	.873 (.072)	.313 (.113)	.973
Average	.866 (.054)	.174 (.078)	.983*
Pooled regression	.873 (.032)	.171 (.047)	.956

Id	Multilevel regressions		
	a (<i>s.e.</i>)	$\log c$ (<i>s.e.</i>)	R^2
1	.758 (.051)	.220 (.070)	
2	.753 (.051)	.061 (.070)	
3	.897 (.044)	.233 (.070)	
4	1.016 (.048)	.285 (.071)	
5	.899 (.054)	-.040 (.070)	
6	.876 (.046)	.281 (.070)	
Fixed	.867 (.050)	.173 (.070)	.978

Note. *s.e.* = Standard error.

*Because r and r^2 are not distributed normally, we computed this value as the square of the arctangent of the average of the inverse hyperbolic tangent of the correlations among subjects.

which were not discriminatively signaled, and determined probabilistically. This was the mixed component. The duration of the right key was fixed. Please refer to Table 1 of the target paper (p. 172) for more information on the conditions. The study was also divided into five parts. Parts 1 and 3 were the conditions for comparing the mixed versus the fixed reinforcers' durations. Part 2 maintained constant probability between the states of the left key, and reinforcer durations in state 2 were varied. Parts 4 and 5 were control conditions in which, respectively, only the variable interval schedules and only the reinforcers' durations were varied. Because manipulation of reinforcer rates were isolated in Part 4 (without probabilities or duration of reinforcement being varied which introduce additional complexities), only this part will be analyzed.

All the analyses were carried in R (R Core Team, 2017), using the usual `lm` function for non-pooled and pooled regression and the `lmer` function in the `lme4` package (Bates, Maechler, Bolker & Walker, 2015) for the multilevel modeling. The standard errors for multilevel regression

were computed using the `arm` package (Gelman & Su, 2015). The variance accounted for was computed using the `MuMIn` package (Bartoń, 2018), which determines the conditional variance explained by the model (but see Rights & Sterba, 2018, for a discussion on explained variances in multilevel models). The regressions were performed on the log response ratios and the log reinforcer ratios. The numerator was associated with the right key and the denominator with the left key. The function `summary`, a generic function producing result summaries of various model fitting functions such as `lm` and `lmer`, was used. The nonpooled regressions were carried out for each subject independently of the others. The pooled analysis consisted of a regression including all subjects regardless of their identifier (complete pooling). The multilevel analysis was conducted by following this syntax with the Wilkinson's notation: `Multilevel = lmer(logRespRateRatio~logReinRateRatio+(logReinRateRatio+1|id))`. The variable `logRespRateRatio` refers to the log response rate ratios, and `logReinRateRatio` corresponds to the log reinforcer rate ratios. The "`|`" specifies the level (the random effect) which is the subjects' identifier, `id`, in this case. The output `Multilevel` contains the results of the multilevel regression which can be extracted with the command `coef`, `se.coef` and `r.squaredGLMM`. The R syntax for implementing all the analyses shown herein is provided in the appendix.

A summary of the results is presented in Table 1. Multilevel regression suggests that all pigeons approximately undermatched their responses' rate ratios as a function of reinforcer rate ratios. The exception is bird 4, whose behavior overmatched. The six pigeons were sensitive to reinforcement, and data suggest that they might undermatch ($a \approx .867$) response rate ratios to reinforcer rate ratios. All but the fifth pigeon presented a bias for the first behavior (b_1), the right key; four pigeons showed a strong bias toward the right key. It is worth noting that multilevel modeling allows for a direct comparison between subjects and their average. Standard errors at both the within- and between-subject levels showed the quality of those estimates, as usual, but now they expressed, at the between-subject level, the expected deviance among subjects. For instance, sensitivity was .867 CI[.769; .965] and each subject's sensitivity fell between these bounds with

the exception of pigeon 4. The hyperparameter for bias was .173 CI[.036; .310], with pigeon 5 differing statistically from other pigeons.

Regarding the different regression methods, there were slight differences between the averages of nonpooled, completely pooled, and multilevel regressions. The parameter estimates were very similar across techniques. This suggests that multilevel estimates are in agreement with pooled and nonpooled data, which should reassure unconvinced readers. They should not point to contradictory outcomes, only to better estimation. For the current analysis, since there were a lot of data, and that matching behavior was consistent for each subject, shrinkage was minimal. There were some differences in the standard error of the the group parameter estimates, especially for the completely pooled regression. For complete pooling, the standard errors were artificially smaller because the regression does not account for the dependencies between responses of subjects.

In accordance with previous conclusions reached by Davison and Hogsden (1984), the data from Part 4 support the GML in the designed experiment. Explained variance was high (97.8%), and in the higher range than what is generally expected in studies on matching behavior, namely, 80% (Caron, 2017; McDowell, 2013). Even though comparisons of variances accounted for can be troublesome for multilevel models (see Rights & Sterba, 2018) and they should be done with caution, multilevel modeling leads to marginally better results (97.8%) than the pooled regressions (95.6%) and slightly lower ones than nonpooled regressions (98.3%). This can be explained by a better accounting for the within-subject variance (when compared to pooled regression) and for the between-subject variance (when comparing the nonpooled regressions). Multilevel modeling was sensitive enough to capture both the within- and between-subject variance, thus neither underestimating nor overestimating the results.

With only fairly small differences between the three regression methods in Table 1, we might ask which should be trusted more. We should start by acknowledging that we do not know the answer to this question, as we do not know the population parameters. However, we can argue four reasons in favor of a multilevel modeling approach. First, multilevel modeling requires a single analysis to produce all the desired parameter estimates, whereas the others required

minimally seven (six nonpooled plus one pooled) regressions. More analyses are needed if the behavior of subjects is to be compared in these regression methods but not in multilevel regression. Multilevel modeling thus reduces the type I error rate. Second, during the estimation process, multilevel modeling used all the information provided, like within- and between-subject variances, but also all subjects' data, to carry the computation, whereas the nonpooled and pooled regressions only used subsets of variables and data. A multilevel approach thus produces a more thorough portrait of the data at a single glance. Third, if we had access to more observations by condition (the appendix only shares the average of the last five sessions), we could have accounted for a second layer of analysis, that is, the within-conditions (or within-part) variances, more exhaustively. This last consideration explains the general convergence between regression models. Last, we can also recall other strong benefits mentioned previously, such as taking into account missing values, design imbalance, the increase in power, and weight of consistent and inconsistent matching behavior. To sum up, we are much more confident in the multilevel model and in extending these results to other subjects. If sensitivity and bias are constant within a population, then multilevel modeling estimates will enable better educated guesses regarding matching behavior for comparable subjects.

Conclusion

The purpose of the current paper is to advocate for better statistical inquiry of matching behavior using multilevel analysis. Based on the evidence showed in Table 1 the case for multilevel analysis may not seem terribly consequential, but it is just a better and simpler method to analyze behavioral data. It provides a theoretically grounded basis on which to estimate matching behavior across several subjects and to further generalize the results. It incorporates the differential validity of each subject's data regarding consistency, sample size, and unusual behavioral patterns. The method has the advantages of not necessitating the elimination of subjects, of increasing the power of the analysis, and of being readily implemented in many statistical software programs. As a cautionary reminder, though, it is worth emphasizing that strong statistical analyses do not supersede strong experimental

control; they only generate better educated guesses regarding parameter estimates.

Although many researchers may not be well acquainted with multilevel modeling, its potential superiority regarding the handling of repeated measures, as well as the substantial increase of its appearance in research in the past years, should encourage its use, especially among behavior analysts. Multilevel analysis provides an interesting avenue for pooling subjects or operant situations. This is especially relevant with regard to matching behavior, since it permits better estimations of sensitivity and bias. The examples carried herein support the idea that multilevel analysis leads to better estimations, is more robust regarding the treatment of data, and can give more information regarding behavior. It is a promising tool to add to the arsenals of behavior analysts. We hope this paper will encourage the use of this statistical technique within behavior analysis, in addition to stimulating statistical research into problems specific to the field.

References

- Bartoń, K. (2018). Package 'MuMin', from <https://CRAN.R-project.org/package=MuMin>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48. <https://doi.org/10.18637/jss.v067.i01>
- 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*, 269-281. <https://doi.org/10.1901/jeab.1979.32-269>
- Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior*, *12*, 861-874. <https://doi.org/10.1901/jeab.1969.12-861>
- Caron, P.-O. (2013). On applying the matching law to between-subject data. *Animal Behaviour*, *85*, 857-860. <https://doi.org/10.1016/j.anbehav.2013.01.022>
- Caron, P.-O. (2017). Sur la loi de l'appariement [On the matching law]. *Psychologie Française*, *62*, 29-55. <https://doi.org/10.1016/j.psfr.2015.10.003>
- Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society, Series B*, *45*, 311-354.
- Davison, M., & Hogsden, I. (1984). Concurrent variable-interval schedule performance: Fixed versus mixed reinforcer durations. *Journal of the Experimental Analysis of Behavior*, *41*, 169-182. <https://doi.org/10.1901/jeab.1984.41-169>
- Davison, M., & McCarthy, D. (1988). *The matching law: A research review*. Hilldale, NJ: Erlbaum.
- de Leeuw, J., & Meijer, E. (2008). *Handbook of multilevel analysis*. New York, NY: Springer-Verlag.
- de Villiers, P. A., & Herrnstein, R. J. (1976). Toward a law of response strength. *Psychological Bulletin*, *83*, 1131-1153. <https://doi.org/10.1037/0033-2909.83.6.1131>
- Efron, B., & Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, *236*, 119-127.
- Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R*. Boca Raton, FL: Chapman and Hall/CRC.
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York (NY): Cambridge University Press.
- Gelman, A., & Su, Y.-S. (2015). Package 'arm', from <https://CRAN.R-project.org/package=arm>
- Greenland, S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology*, *29*, 158-167
- 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>
- Hox, J. J. (2010). *Multilevel analysis: techniques and applications*. Mahwah, NJ: Routledge.
- Mazur, J. E. (2006). Mathematical models and the experimental analysis of behavior. *Journal of the Experimental Analysis of Behavior*, *85*, 275-291. <https://doi.org/10.1901/jeab.2006.65-05>
- 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>
- R Core Team. (2017). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.Rproject.org/>
- Rights, J. D., & Sterba, S. K. (2018). Quantifying explained variance in multilevel models : An integrative framework for defining R-squared measures. *Psychological Methods*, *23*, 434-457. <https://doi.org/10.1037/met0000184>
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Young, M. E. (2017a). Discounting : A practical guide to multilevel analysis of indifference data. *Journal of the Experimental Analysis of Behavior*, *108*, 97-112. <https://doi.org/10.1002/jeab.265>
- Young, M. E. (2017b). A place for statistics in behavior analysis. *Behavior Analysis: Research and Practice*, *18*, 193-202. <https://doi.org/10.1037/bar0000099>
- Young, M. E. (2018). Discounting : A practical guide to multilevel analysis of choice data. *Journal of the Experimental Analysis of Behavior*, *109*, 293-312. <https://doi.org/10.1002/jeab.316>

Received: December 19, 2018

Final Acceptance: January 31, 2019

Editor in Chief: Michael Young

Associate Editor: Greg Jensen.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website.

Appendix

```

#Preliminary steps
#load data
load("DAVISON_HOGSDEN.RDATA")
# Data found in the appendix of Davison & Hogsden (1984)
# in JEAB with the following column:
# "id", "Part", "Condition", "Res_L", "Res_R", "Time_L",
# "Time_R", "Rei_L_S1", "Rei_L_S2", "Rei_R",

# Creating the response and response rate ratios variables.
data$logReinRateRatio = log(Rei_R/(Rei_L_S1+Rei_L_S2))
data$logRespRateRatio = log(Res_R/Res_L)

# Choosing Part 4 only
mydata <- subset(data, Part == 4)

#Relevant libraries for multilevel modeling
require(lme4)
require(MuMIn)
require(arm)

#####
#Pooled regression
Pooled = lm(logRespRateRatio~logReinRateRatio, data=mydata)
summary(Pooled)

#####
#Nonpooled regressions
by(mydata, mydata$id, function(subjectData)
summary(lm(logRespRateRatio~logReinRateRatio, data=subjectData)))

#####
#Multilevel regression

MultiLevel = lmer(logRespRateRatio~logReinRateRatio+(logReinRateRatio+1|id), data=mydata)

#Results
MultiLevel
summary(MultiLevel)
coef(MultiLevel)
#estimate
se.coef(MultiLevel)
#standard error
r.squaredGLMM(MultiLevel)
#explained variance

```
