6

Of Models, Vectors, Matrices: Advancing Analyses of the Multiple Control of Behavior

David J. Cox

"Behavior is a difficult subject matter, not because it is inaccessible, but because it is extremely complex" (Skinner, 1953). But jumping wholesale into studying complexity is likely to leave gaps and require assumptions to make the facts fit neatly together. As a result, sciences often start as simple as possible. For behavior analysis, this involved studying how a single schedule of reinforcement influenced patterns of responding in an operant chamber (Ferster & Skinner, 1957; Skinner, 1938). But as sciences progress, more complex phenomena are studied that build on interactions among variables previously studied in isolation. For example, compare the law of effect published by Herrnstein (1970) to that of resurgence as choice in context (Shahan et al., 2020), or the schedules of reinforcement and stimulus presentations in Ferster and Skinner (1957) to any recently published study on derived stimulus relations (e.g., Shawler et al., 2022). Increasing the complexity of phenomena studied in a field seems to suggest maturity, so long as critical theoretical holes are few and far between.

Behavior is not unique in its complexity as a natural phenomenon. For example, consider current theories of quantum mechanics in physics that began with John Dalton's simple claim that all matter is made of tiny indivisible particles (1808) and now includes matrix mechanics, wave mechanics, and the non-relativistic Schrödinger equation (1926). As another example, consider evolutionary theory which began with natural selection as the only known mechanism (Mayr, 1996) and now contains mutation, genetic drift, and gene flow (Santangelo et al., 2018). Arguably, an advancing science is one that–over time–learns the conditions under which independent variables do and do not influence dependent variables; how phenomena interact that were initially studied in isolation; and the description, prediction, and control over increasingly complex phenomenon. To be static is to be dead (Critchfield & Shue, 2018).

As scientists add complexity, bookkeeping becomes more challenging as they try to keep track of how the many independent variables under study affect dependent variables. Models help scientists do this succinctly for reasons explained in greater detail below. Further, as we add variables that contribute to a holistic and empirical understanding of environment-behavior relations, it helps to have a consistent system to store data from our observations that is flexible in terms of scale and analytic possibilities. Vectors and matrices help us do this and are likely how many behavior scientists store their data already.

Formalizing environment-behavior relations as vectors and matrices has the primary benefit of being an agnostic analytic system that allows easy tests of competing behavioral theories. Further, as scientists scale the amount of data they collect on behavior-environment relations, open-source tools from data science and machine learning allow for the easy manipulation and analysis of massive matrices. In what follows, we discuss:

a) The types of models common in behavior science.
b) How keeping track of variables and their interactions get murky as models scale in attempts to capture complexity.
c) How matrices as the medium and data science as the tools offer one practical solution to keeping things tidy.
d) How the resulting framework allows for flexible scaling in degree of complexity and analytic ease.

In short, models built on vector and matrix thinking might be one route to analyzing complex environment-behavior relations that are likely to be of increasing interest to behavior scientists.

Models in Behavior Science

Models can be defined, in the most general sense, as some kind of representation of a person, thing, physical structure, or physical process (Merriam-Webster, 2022). Models are often smaller in scale than the physical thing they are meant to describe, where “smaller” might be defined physically or in the amount of detail included. For example, readers might be familiar with 3D physical models of cars or boats that would be impossible for a human to drive or sail. 3D physical models of bridges allow engineers to test various weight-bearing capacities before they spend millions of dollars building an overpass. And 3D physical anatomical models provide artists and physiologists alike the opportunity to visually see how the body might bend and move in space.

All the above are proportionally scaled, three-dimensional models. However, models certainly do not have to be in three dimensions nor be scaled proportionally with the things they are meant to describe. For example, ecological models describe energy cycles across animals, plants, and the atmosphere (e.g., Patten, 1985). Models of particle interactions in atoms help us describe the structure of molecules, which allow us to predict chemical reactions (e.g., Schrödinger, 1926). And, models of the expanding universe help us understand the laws of physics better in our very small corner of this vast cosmic arena (e.g., Heacox, 2015; Sagan, 1994). In science, models typically offer precision in describing mechanisms that allow for prediction (Dallery & Soto, 2013). Models certainly are not the only way to describe the world, but many of the triumphs of science suggest that model building and testing has been remarkably successful in uncovering a wide range of unsuspected possibilities including electromagnetic waves, engines, transistors, and antibiotics.

Models have also been around in behavior analysis from the beginning. For example, Skinner (1935, 1937, 1938) used arbitrary symbols and graphics to describe the difference between environment-behavior relations in respondent and operant conditioning[1] and these same models are used to train budding behavior analysts today (Cooper et al., 2020; pp. 31, 37). Lastly, as alternatives to linear models of environment-behavior relations, others have proposed feedback loops with embedded hierarchies of variables (Moxley, 1982). Each of these models describe environment-behavior relations in much less detail than the plethora of physical ongoings that occur as organisms move through time and space.

In short, models are an integral part of all sciences. At a basic level, models describe in fewer terms the important details of a physical system under study. If the models are good enough, they allow the model user to describe, predict, and (potentially) control the physical universe they contact and are interested in influencing. To help frame conversations about models of multiple control and complexity later in the manuscript, we next briefly review some of the basic forms and functions of models.

Model Forms

One way to classify models is based on the form that they take. Figure 6.1 shows some of the topographies that models can take, including

[1] See Figure 1 from Moxley, 1982 for side-by-side comparisons.

textual-vocal models[2], graphical models, quantitative models, and computational models. Perhaps the most used by researchers and practitioners are textual-vocal models. Textual-vocal models use words to describe the relations between the environment and some behavior of interest. For example, a practitioner might state, “In the presence of demands to engage in schoolwork, the client is more likely to hit and kick staff, which is often followed by removal of those demands.” Here, many instances of observed environment-behavior relations and the collection of data have been compressed to a set of textual stimuli carrying less detail than any one observation of aggression (i.e., it’s a model) but provide the listener with an understanding of why the client engages in aggressive behavior. Given past research on functional approaches to reducing aggression (e.g., Tiger et al., 2008), such a textual-vocal model might be all that is needed for describing the causes of behavior in many circumstances (i.e., audience matters; Dallery & Soto, 2013).

Graphical models use pictures to display visually the relationship between the environment and behavior. Often graphical models use shapes, lines, arrows, or other icons to show directions of influence or time, inclusion or exclusion (e.g., Venn Diagrams; Killeen, 2019), and to denote the different variables that comprise the model, which are ideally observed and measured (Moxley, 1982). Many behavior analysts are likely familiar with using a one-dimensional linear notation of the three-term contingency (Figure 6. 1; Graphical *a*). However, one-dimensional linear notations of the three-term contingency carry limitations (Moxley, 1982). Thus, some have offered alternative graphical displays (e.g., Figure 6.1; Graphical *b*) that more easily and visually differentiate respondent from operant contingencies and that contain multiple levels (Figure 3 from Moxley, 1982) and varying timescales (Figure 4 from Moxley, 1982) in a single graphical display.

[2] Textual-vocal models have been referred to as verbal models in previous writing (e.g., Dallery & Soto, 2013). However, since all models are instances of verbal behavior (e.g., Cox, 2019; Marr, 2015), the term textual-vocal model is used throughout to differentiate the use of words to describe environment-behavior relations as opposed to graphics, mathematical expressions, or computational expressions.

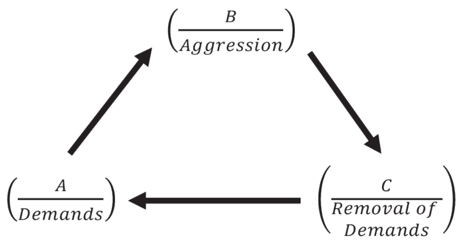
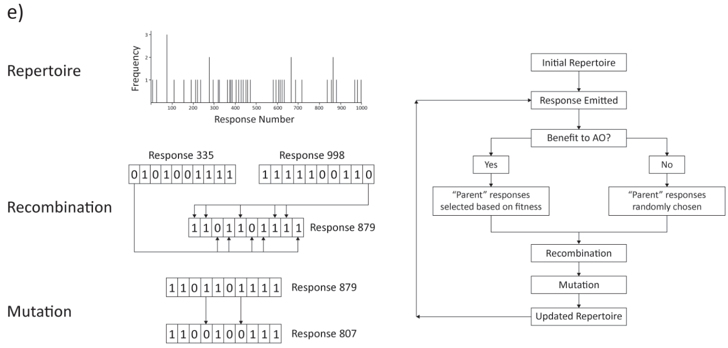
Model Form	Examples From Behavior Science
Textual-vocal	Person states: "In the presence of demands to engage in schoolwork, the client is more likely to hit and kick staff which is often followed by removal of those demands."
Graphical	a) $S^{D} \rightarrow B \rightarrow S^{R-}$
	b) 
Quantitative	c) $B = \dfrac{kR}{R + R_o}$
	d) $(Rate\ of\ Aggression) = \dfrac{(Maximum\ Aggression\ Possible)x(Rate\ of\ Reinforcement\ for\ Aggression)}{(Rate\ of\ Reinforcement\ for\ Aggression) + (Rate\ of\ All\ Other\ Reinforcement)}$
Computational	e) 

Figure 6.1. Examples of different model form that scientists have used within behavior analysis: **a)** Linear notation as one-dimensional graphical representation of the three-term contingency (Moxley, 1982). **b)** Modified triadic graphical display from Moxley (1982) representing the three-term contingency described by the textual-vocal model. **c)** Law of effect from Herrnstein (1969). **d)** Restatement of the law of effect relative to the textual-vocal example. **e)** Recast of the evolutionary theory of behavior dynamics from McDowell (2013b, 2019).

Quantitative models use mathematical expressions to describe the relations between environmental variables and their influence on behavior. For example, the law of effect (Figure 6.1; Quantitative *c*) describes how the absolute rate of a behavior in a specified environment is controlled by the absolute rate of reinforcement up to an asymptote (Herrnstein, 1970). The generalized matching equation describes the relative allocation of responding to two or more alternatives given the amount of reinforcement contacted by each response (e.g., Baum, 1974). The exponential demand equation describes how increasing the cost (e.g., number of responses, price in currency) to contact a reinforcer reduces the number of reinforcers earned (e.g., Hursh & Silberberg, 2008). And the hyperbolic discounting equation describes how much an organism will work (e.g., response rate, response allocation, response force; i.e., the reinforcer value) as some dimension of that reinforcer is systematically manipulated (e.g., delay to receipt, uncertainty; Rachlin, 2006).

Lastly, computational models use computers to simulate environment-behavior systems. Computational modeling typically involves transforming the variables that comprise a system into a numeric representation (Figure 6.1; Computational). Once specified, one or more algorithms then update the system across many discrete timesteps, and the researcher observes the resulting outcomes. Taking an experimental approach, the research then further tests the computational model by adjusting combinations of the number, types, and how variables are represented numerically, as well as the rules for updating the system (USDHHS, 2020). Examples of computational models in behavior analysis include the evolutionary theory of behavior dynamics (e.g., McDowell, 2019), computer simulations of the operant reserve (e.g., Berg & McDowell, 2011; Catania, 2005), neural-network interpretations of selection mechanisms of learning (e.g., Burgos, 2001; Donahoe et al., 1993), and reinforcement learning models that mimic patterns of choice (e.g., Pearson & Platt, 2013).

Model Function

We can also talk about the function, or reason why, a scientist is modeling environment-behavior relations. Figure 6.2 shows a common way that model function is described as a balance between relative difficulty (*x*-axis) and utility to the modeler (*y*-axis). Perhaps the easiest (but often least useful) model function is descriptive. Here, the researcher or practitioner simply tacts something they have observed in answering the question, “What happened?”. Note that descriptive models can take any form, such as textual-vocal, graphical, quantitative, or computational (Figure 6.1).

Associative models can lie somewhere between answering the questions: “What happened?” and “When will it happen?”. Here, scientists or practitioners start to identify how often and the conditions under which two (or more) variables are likely to co-occur. Associative models can also take various forms. For example, they might be graphical depictions such as AB plots that show changes in behavior from baseline to an intervention condition (e.g., Wolfe & Slocum, 2015) or scatterplots showing how changes in one continuous variable correspond to changes in a second continuous variable (e.g., Baum, 1974; Rachlin, 2006). Associative models might also take the form of quantitative descriptions such as Pearson’s correlation coefficient and Spearman’s rank correlation (e.g., Motulsky, 2014).

Predictive models answer the question, “When will it happen?” or “What will happen?”. The use of “baseline logic” (e.g., Cooper et al., 2020) is one form of textual-vocal predictive model where researchers would predict the trend, level, and variability of behavior will remain unchanged if no intervention is introduced. Similarly, researchers could textual-vocally predict that responding will increase/decrease when the intervention is introduced and support it with a graphical model of the assumed three-term contingency in effect. However, note that the predictions lack precision in both instances. Because of the lack of predictive precision with textual-vocal models, predictive models are most often quantitative or computational. For example, the generalized matching equation (e.g., Baum, 1974; McDowell, 2005) makes precise predictions about the exact ratio of responding between two or more responses if the ratio of contact with reinforcement for those responses is known. And, the computational predictive model of the evolutionary theory of behavior dynamics makes predictions that coincide with these predictions (e.g., McDowell, 2019).

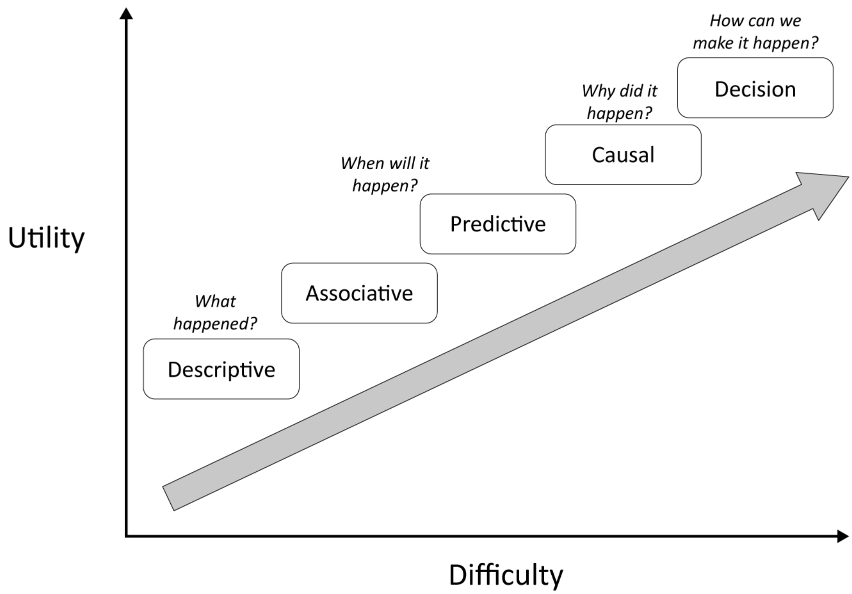


Figure 6.2. Visualization highlighting the relative tradeoff between the difficulty of building a good model (x-axis), the utility of the resulting model for the model user (y-axis), and terms common to other disciplines that describe the various locations in this graphical representation.

Causal models answer the question "Why did it happen?". Here, the model builder specifies precisely the physical mechanisms that lead one pattern of environment-behavior relations (independent variables) to a specific and different pattern of environment-behavior relations. Though causal models can be used to make predictions, not all predictive models are causal models, and vice versa. For example, the evolutionary theory of behavior dynamics makes predictions about behavior that coincide with what is observed in nonhuman animals across many contexts (McDowell, 2019). However, it seems unlikely the genetic algorithm used to make those predictions is, in fact, the biological mechanism whereby response repertoires are created and selected over time (i.e., it's not a causal model of biological behavior). Similarly, meteorologists have a good understanding of what atmospheric conditions cause tornados, hurricanes, and red tides. However, the complexity of collecting the requisite data to make precise predictions of when severe weather events will occur continues to elude this field.

Arguably, the most useful kinds of models are decision models. Here, the causal mechanisms allowing for prediction are so well known that the model tells the model user the optimal decision to make based on the facts they have at their disposal. Examples of decision models in behavior analysis include models for selecting measurement procedures (e.g., LeBlanc et al., 2016) and selecting evidence-based treatments (e.g., Geiger et al., 2010; Saini et al., 2017). Decision models are arguably the function of science generally: to create generalizable knowledge that allows humans to make better decisions within the situations they have found themselves in. However, as the examples in this paragraph indicate, decision models often require large amounts of evidence verifying causal environment-behavior relations before they can be developed and tested for whether they lead the model user to make optimal decisions.

Model Summary

Models are an integral part of the science of behavior and occur any time that a scientist uses verbal behavior to compress the totality of observed environment-behavior relations into fewer details. The set of verbal behavior tacted as models might be of the form textual-vocal, graphical, quantitative, or computational; and the function of the model might be descriptive, associative, predictive, causal, or to lead to an optimal decision. The specific combination of form and function will likely be determined by the function of the modeler's behavior (Marr, 2015) and the extent of past research and available evidence to which the researcher has access. This last point is critical because models from data science are typically quantitative or computational.

But the function of the model will be determined by factors outside the data scientist's model.

Modeling the Complexity of Multiple Control

Complexity has been defined in many ways across sciences and might be so general a term that it means something different to each speaker based on the field in which they are trained (Adami, 2002). Nevertheless, *complexity* can be grouped into three major divisions (Manson, 2001). *Algorithmic complexity* focuses on how difficult it is to describe the totality of a physical system, often described using mathematical complexity theory or information theory (e.g., Chaitin, 1992; Norton & Ulanowicz, 1992). *Deterministic complexity* focuses on how a few critical variables or rules can lead to physical systems that demonstrate both patterns and emergent discontinuities, often via chaos theory and catastrophe theory (e.g., Brown, 1995; Byrne, 1997; Lowe, 1985; Nijkamp & Reggiani, 1990; Renfrew & Cooke, 1979). Finally, *aggregate complexity* focuses on how many individual system components interact and work together or in opposition to determine the behavior of the physical system (e.g., Allen & Hoekstra, 1992; Andrade et al., 1995; Corrieg et al., 1997; Holland, 1992). Though algorithmic and deterministic complexity have relevance to behavior science (e.g., Killeen, 1994; McDowell, 2013a, 2013b), the remainder of this chapter will focus on aggregate complexity as a framework for the multiple control of behavior and how data science techniques might allow researchers to efficiently model such aggregate complexity.

In its simplest form, aggregate complexity seeks to understand how variables studied in isolation might interact to influence behavior. Such research certainly is not new in behavior analysis. For example, arguably, research on compound schedules of reinforcement (Catania, 2013) fits this mold wherein researchers ask how patterns of responding change when two or more simple schedules of reinforcement (FR, VR, FI, VI) are combined and operate successively (multiple, mixed), as a sequence (chained, tandem), simultaneously (concurrent, conjoint), or are dependent (alternative, conjunctive, interlocking). Such research has led to associative/predictive quantitative models such as behavioral momentum (e.g., Nevin et al., 1983), the generalized matching equation (e.g., Baum, 1974), and hyperbolic discounting (e.g., Rachlin, 2006).

Researchers in behavior analysis have also examined multi-dimensional aspects of compound schedules relative to aggregate complexity. For example, researchers have examined the independent effects of reinforcer rate and amount in concurrent schedules arrangements (e.g., Baum & Rachlin, 1969) and reinforcer rate, amount, and delay in concurrent schedules arrangements (e.g., Davison & McCarthy, 1988) which has led to derivation of the associative/predictive quantitative model of the concatenated matching law (e.g., Davison & McCarthy, 1988). Similarly, researchers have examined how delay and probability interact to influence choice in concurrent schedules arrangements (e.g., Blackburn & El-Deredy, 2013; Cox & Dallery, 2016; Killeen, 2009; Vanderveldt et al., 2015; Yi et al., 2006) and how multiple delayed and probabilistic outcomes that follow each response interact to influence choice in concurrent schedules arrangements (e.g., Cox & Dallery, 2018). The above lines of research again led to associative/predictive quantitative models for how these multiple dimensions interact to influence choice.

Studying aggregate complexity requires the researcher to conduct considerable bookkeeping. One example of the analytic challenges contacted by someone studying aggregate complexity is shown in Figure 6.3. The top panel in Figure 6.3 shows a set of data and the resulting plot, which many readers are likely familiar with. This panel shows a prototypical situation wherein a researcher is interested in how some modification of reinforcement contingencies will influence the rate of responding. Here, functional communication training (FCT; e.g., Tiger et al., 2008) is implemented and results in an increase in the rate of responding compared to baseline. Further, readers who graph their data in Excel, Prism, or other common software packages are likely familiar with the spreadsheet image on the left where the session number, condition, and rate of responses per minute are entered. In turn, storing data this way allows the graph on the right to be made.

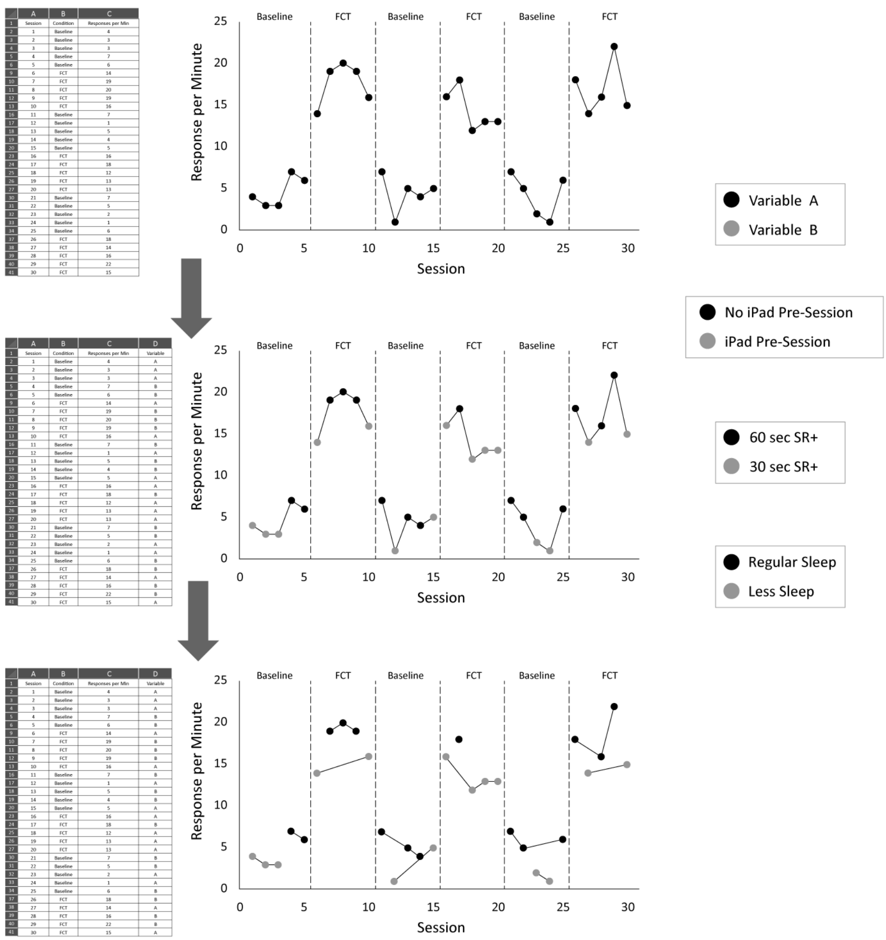


Figure 6.3. Complexity as additional columns in a structured data frame.

Studying aggregate complexity suggests something else is added to the mix beyond the intervention in isolation. The middle panel of Figure 6.3 shows how we might add an extra variable to the analysis by adding data collected on a third variable (i.e., Column D). Perhaps the added variable is whether the individual has access to the maintaining reinforcer as part of a closed economy (“No iPad pre-Session”) or open economy (“iPad Pre-Session”; e.g., Collier et al., 1992; Foster et al., 1997; Hall & Lattal, 1990; Zeiler, 2013); or whether the amount of reinforcement contacted with each reinforcer delivery influences behavior (“60 sec SR+” vs. “30 sec SR+”; e.g., Estle et al., 2006; Green et al., 2013; Rachlin et al., 2015); or the extent to which biological variation might influence responding across sessions (“Regular Sleep” vs. “Less Sleep”; e.g., Kirby & Kennedy, 2013). The bottom panel in Figure 6.3 shows how a researcher or practitioner might replot the middle panel. When data across the third variable are connected, the visual analyst might discover that the rate of responding is typically higher when Variable A is present compared to when Variable B is present, regardless of condition.

Figure 6.3 shows one way that a researcher might analyze the convergent multiple control of behavior where many different variables influence a single response (Michael et al., 2011). Each column highlights one measure of behavior or one measure of the environment surrounding behavior; and each row of the data frame tracks one instance of these observations of multiple control. When we realize that adding complexity is simply adding columns to our dataset, it’s easy to see that there is no reason we have to limit ourselves to adding just one variable. For example, the top panel in Figure 6.4 shows a dataset where we might be interested in the influence of many different variables on rate of responding such as the setting of session, the different technicians who work with the client, the rate of trial presentation, the exact schedule of reinforcement contacted by the client, and whether the client had pre-session access to the maintaining reinforcer.

Matrices as Medium & Data Science as Tools

The bottom panel in Figure 6.4 shows a more general form of the top panel. Just as in the top panel, there is a set of data on antecedent variables (a_{11}-a_{86}), rates of responding (b_{17}-b_{87}), and consequence variables (c_{18}-c_{89}); and where each row corresponds to one observation. Another name for this way to represent data is a “matrix” (e.g., Strang, 2003).

	A	B	C	D	E	F	G	H	I
1	Session	Condition	Responses per Min	Variable	Setting	RBT	Trials per Min	FR Schedule	Pre-Session SR+ Contact
2	1	Baseline	4	A	Home	Person A	8	1.19	Yes
3	2	Baseline	3	A	Home	Person A	10	2.05	Yes
4	3	Baseline	3	A	Home	Person A	9	1.00	Yes
5	4	Baseline	7	B	Home	Person A	10	1.54	Yes
6	5	Baseline	6	B	Home	Person A	10	1.90	Yes
9	6	FCT	14	A	Home	Person A	6	1.00	Yes
10	7	FCT	19	B	Home	Person A	11	1.37	Yes
11	8	FCT	20	B	School	Person A	7	2.25	Yes
12	9	FCT	19	B	School	Person A	8	1.46	Yes
13	10	FCT	16	A	Home	Person A	5	1.78	Yes
16	11	Baseline	7	B	School	Person A	10	1.86	Yes
17	12	Baseline	1	A	School	Person A	10	1.25	No
18	13	Baseline	5	B	School	Person A	11	1.33	No
19	14	Baseline	4	B	School	Person A	6	1.12	Yes
20	15	Baseline	5	A	School	Person A	10	1.46	No
23	16	FCT	16	A	School	Person A	4	1.12	No
24	17	FCT	18	B	Home	Person A	8	1.11	No
25	18	FCT	12	A	Home	Person B	6	1.39	No
26	19	FCT	13	A	Home	Person B	8	1.88	No
27	20	FCT	13	A	Home	Person B	14	1.29	Yes
30	21	Baseline	7	B	Home	Person A	12	1.30	Yes
31	22	Baseline	5	B	School	Person A	11	2.26	No
32	23	Baseline	2	A	Home	Person B	12	1.69	Yes
33	24	Baseline	1	A	School	Person B	6	1.02	No
34	25	Baseline	6	B	School	Person A	8	1.40	No
37	26	FCT	18	B	School	Person A	8	2.08	No
38	27	FCT	14	A	School	Person B	10	1.46	No
39	28	FCT	16	B	School	Person A	7	1.94	No
40	29	FCT	22	B	School	Person B	9	1.44	No
41	30	FCT	15	A	School	Person A	5	1.39	No

$$\downarrow$$

$$\begin{bmatrix}
a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} & b_{17} & c_{18} & c_{19} \\
a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & b_{27} & c_{28} & c_{29} \\
a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} & b_{37} & c_{38} & c_{39} \\
a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} & b_{47} & c_{48} & c_{49} \\
a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & b_{57} & c_{58} & c_{59} \\
a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} & b_{67} & c_{68} & c_{69} \\
a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & b_{77} & c_{78} & c_{79} \\
a_{81} & a_{82} & a_{83} & a_{84} & a_{85} & a_{86} & b_{87} & c_{88} & c_{89}
\end{bmatrix}$$

Figure 6.4. One example of a data frame that would allow a researcher to analyze the convergent multiple control of seven different variables.

Translated for our purposes, this matrix is a set of data on environment-behavior relations where each row corresponds to data collected at some moment in time (i.e., a vector); and each column corresponds to the physical state of some environmental variable antecedent or consequent to a behavior of interest, or corresponds to some dimension of behavior (e.g., rate, duration, intensity).

As our dataset grows to the right to include data on the multiple variables likely to influence responding (i.e., aggregate complexity of convergent multiple control), researchers need methods to efficiently analyze these larger datasets to identify the patterns that allow for the description, prediction, and control of responding.

Behavior analysts have shown significant variability when visually analyzing data along two dimensions such as time and intervention condition (e.g., Cox & Brodhead, 2021; cf. Diller et al., 2016; Kahng et al., 2010; Ninci et al., 2015). Figure 6.4 highlights how adding a third dimension can increase the difficulty of visual analysis. As behavior analysts collect data on dozens or hundreds of variables, they will need efficient analytic methods that can be constrained by basic operant and respondent principles. In these more complex modeling situations, thinking about our data as a matrix is useful because it allows us to take advantage of math that can handle these data sets: linear algebra which is one of the primary languages of machine learning and artificial intelligence.

Matrices and their manipulation via linear algebra has many practical advantages (e.g., Strang, 2003). For behavior scientists structuring datasets as matrices as in Figure 6.4, an intuitive way to think about such a dataset is as a system of linear equations. For example, taking a deterministic, aggregate complexity, convergent multiple control view of the data in Figure 6.4, the assumption is that the various combinations of antecedent and past consequences determine responding (i.e., allow us to predict responding). Restated as equations, the rows become:

$$a_{1,1}x_1 + a_{1,2}x_2 + \cdots + c_{1,n}x_n = b_1$$

$$\vdots$$

$$a_{m,1}x_1 + a_{m,2}x_2 + \cdots + c_{m,n}x_n = b_m$$

Here, *a* refers to the antecedent data, *c* refers to the consequent data, *b* refers to the behavior data, *x* refers to the relative influence of that environmental variable on behavior, *n* is the number of antecedent and consequence columns in the dataset, and *m* is the number of rows in the dataset.

Thinking of our data as a matrix and system of linear equations can be useful for several reasons. First, researchers have developed computational algorithms to find exact or approximate solutions to large systems of equations (e.g., Golub & Van Loan, 1996; Trefethen & Bau, 1997) which play a practical role in engineering, physics, chemistry, computer science, and economics. In our context, such a solution would provide a quantitative estimate of the approximate influence of each environmental variable on response rate. In turn, this would allow researchers to understand –and quantify– the relative importance of different environmental variables in the multiple control of behavior. *Variable importance* can then be tested experimentally by manipulating one or more independent variables and observing how close the observed behavior is to that predicted by the equation.

Second, technology increasingly allows behavior analysts to collect data on a multitude of environmental and behavioral variables (e.g., Cox et al., 2022; Dallery et al., 2015; Kurti & Dallery, 2014). However, as the number of rows or columns in our matrix increases, the complexity and difficulty of finding an approximate solution to a system of equations also increases. To make such problems tractable, researchers have developed algorithms to reduce the number of variables considered via combination or removal, and to improve the accuracy of the approximate solution. These algorithms are often referred to as dimensionality reduction techniques (e.g., Huang et al., 2019; Reddy et al., 2020). For our purposes, these techniques would allow researchers to identify variables that are redundant with each other or, when all other variables are considered, have less relative influence on behavior. In both instances, the researcher can learn what environmental variables can likely be removed from the analysis, subsequent data collection, or consideration for derived interventions.

Third, matrices can be considered entire objects and—as entire objects—can be transformed. Data transformations are often used productively in science. For example, log-log transformations turn power functions into straight lines (e.g., GME; Baum, 1974) and semi-log transformations turn exponential functions into straight lines (e.g., IRT analyses; Kessel & Lucke, 2008) making data in both cases much easier to work with. These examples involve the transformation of two variables relative to each other. But, by thinking of our datasets as matrices, we can do the same thing with an entire dataset. For example, we discussed above how decision models allow the model user to efficiently attain a specific output as a function of a set of inputs. Recast in terms of thinking about our data as a matrix, we start with one set of environment-behavior relations (i.e., one matrix) and we want our intervention to lead to a new set of environment-behavior relations (i.e., a second matrix). Matrix transformations describe how we can go from the first matrix to the second matrix and whether it is possible (e.g., Strang, 2003). Restated as a question and for our purposes, if one matrix describes how multiple variables interact to control responding, how might we have to change that entire system to bring about the environment-behavior system we desire?

Matrix & Data Science Summary

Many behavior analysts likely store their collected data in a structured format in spreadsheets. That is, each column corresponds to one measured variable related to the context, consequences, or behavior of interest; and each row corresponds to one observation (e.g., learning trial) or an aggregate of several observations (e.g., sessions). When structured this way, studying aggregate complexity and the multiple control of behavior simply means collecting and storing data in more columns. Visual analyses of data become much more difficult when many columns are added to the dataset. But, viewing each observation as a vector and the set of observations as a matrix unlocks tools from mathematics and statistics, computer science, and behavior science to quantify precisely how much each variable likely controls behavior.

An Example Analysis of the Multiple Control of Behavior

The above section may have sounded a bit abstract for readers new to linear algebra, data science, and machine learning. To make this more tractable, next, we work through analyses for a simulated dataset containing 1000 sessions where each session involved one value from each of the following dimensions: (1) an antecedent stimulus context that was either a “Red Light” or a “Green Light”; (2) one out of 14 possible rate of reinforcement delivery per minute (1-15; analog to a VI-4 s to VI-60 s); (3) one out of four different delays to the reinforcer once earned (1-4 seconds); and (4) the resulting response rate. Important to know is that the reinforcement rate was randomly set between 1-10 per min in the red condition and between 5-15 in the green condition. Table 6.1 shows five sample rows from the simulated dataset, and Figure 6.5 shows the general relations between each independent variable and response rates. All code to replicate the experiment is contained as an Appendix at the end of the chapter.

Table 6.1. Sample rows from the simulated dataset for the examples presented in the chapter

Session	Stimulus Context	Reinforcement Rate	Reinforcer Delay	Response Rate
1	Green Light	3	2	19.90
2	Red Light	5	1	67.95
3	Green Light	5	4	41.71
4	Red Light	5	3	58.59
5	Green Light	5	3	42.55
...	...	...	...	...
996	Green Light	5	4	35.41
997	Green Light	3	1	26.05
998	Green Light	3	3	14.55
999	Red Light	9	3	156.86
1000	Green Light	5	4	39.55

Throughout the section, I include a little bit of the math to try to highlight where and how vectors and matrices are being used. Understanding the specific details of the math is not needed to make use of this work with your own datasets given modern analytic programming languages (e.g., R, Python). You should still be able to get a good understanding of the benefits gained by skipping over the math. However, as any modeler knows, a nontrivial proportion of model building is knowing what tools are available and selecting the right one for the job based on the dataset you have, the way you think about environment-behavior relations, and the question you have. Taking the time to understand the approaches offered below may provide you with more tools in your toolset, different ways to think about environment-behavior relations, and the ability to ask and answer different questions.

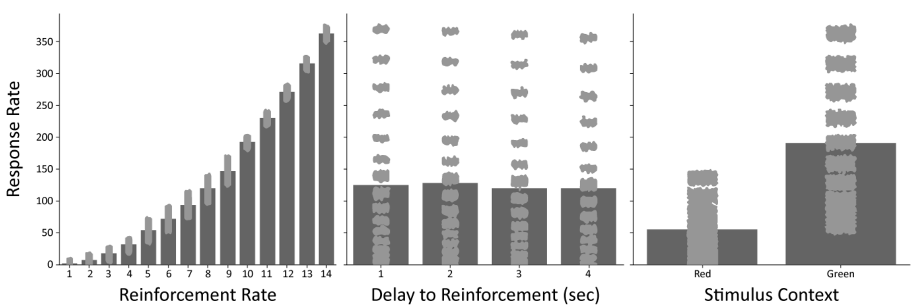


Figure 6.5. General characteristics of the simulated dataset. The black barplot shows the arithmetic mean of the response rates for that value of that independent variable across all sessions. Each gray marker represents data from a single session.

Linear Algebra for Analysis

We can use vectors and matrices[3] to represent our data set containing behavior-environment by writing it as:

[3] For the remainder of the chapter, any ***bold and italicized*** term represents a vector or matrix, which should be deducible from the context.

1 $$Y = Xb + \epsilon$$

Here, $\boldsymbol{Y}$ is the vector of the observed response rates ("Response Rate", Table 6.1); $\boldsymbol{X}$ is the matrix (1000 rows x 3 columns) containing data on the manipulated independent variables of stimulus context, reinforcement rate, reinforcer delay (Table 6.1); $\boldsymbol{\beta}$ is the vector of coefficients that describes the influence of each independent variable on response rates; and $\boldsymbol{\epsilon}$ is an error term. A potentially easier-to-visualize format for Equation 1 that says the same thing is:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{1000} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,stimulus} & x_{1,SR+rate} & x_{1,SR+delay} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1000,1} & x_{1000,2} & x_{1000,3} \end{bmatrix} \begin{bmatrix} \beta_{stimulus} \\ \beta_{SR+rate} \\ \beta_{SR+delay} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{1000} \end{bmatrix}$$

To estimate $\boldsymbol{\beta}$ we can use the normal equation[4]:

2 $$\beta = (X^TX)^{-1}X^TY$$

Doing so with the simulated dataset leads to $\boldsymbol{\beta}$ coefficients of stimulus context = 24.28, reinforcement rate = 21.96, and delay to reinforcement = –16.81. This is logical as we correctly identify that increasing the reinforcement rates increases response rates, changing from the red to green stimulus context increases response rates in parallel with the different programmed reinforcement schedules, and increasing

[4] There are other methods to estimate $\boldsymbol{\beta}$ of which readers might be more familiar (e.g., ordinary least squares regression). See further on in the Chapter, however, for why taking a matrix-based approach offers flexibility in scaling analytic options and provides an opportunity to do things beyond traditional statistical modeling approaches.

the delay to reinforcement decreases response rates as commonly observed in the literature (e.g., Hursh & Fantino, 1973).

One downside to using linear regression is that it assumes that each independent variable linearly increases or decreases response rates. This is clearly not the case for the influence of reinforcement rates or stimulus context on response rates in Figure 6.5. We also do not know if we have accounted for potential interactions among the independent variables in our dataset. Three matrix-based approaches might help us account for nonlinearity in the relationship among variables in the dataset[5]. Support Vector Regression (SVR) uses a kernel function to map inputs into a higher-dimensional space where nonlinear patterns can be captured. Gaussian Process Regression (GPR) takes a probabilistic approach, where covariance matrices model uncertainty and complicated or complex trends in the data. And, XGBoost constructs an ensemble of decision trees to iteratively refine predictions (i.e., a gradient-boosting method). Each of these methods relies on matrix-based computations –SVR through kernelized operations, GPR through covariance matrix inversion, and XGBoost through gradient-based optimization of decision tree ensembles (e.g., Bishop, 2006).

SVR was implemented using the scikit-learn package for Python (Pedregosa et al., 2011). SVR works by finding a function that fits the data within a specified margin of tolerance (e-insensitive loss). SVR is a kernel-based method meaning that it transforms the input data into a higher-dimensional space to capture nonlinear relationships (Platt, 1999). This transformation is computed using the kernel function $\boldsymbol{K}(\boldsymbol{X}_i, \boldsymbol{X}_j)$ to determine similarity between data points. The predictions in SVR are derived from a weighted sum of support vectors which can be represented mathematically as:

3

$$\hat{y} = \sum_{i=1}^{m} (a_i - a_i^*)\boldsymbol{K}(\boldsymbol{X}_i, \boldsymbol{X}) + b$$

[5] There are many approaches one could use in this scenario. I chose these three given their commonality in the literature and ease of implementation and communication. It should not be read that these are the best or only alternative approaches.

Here, a_i, a^*_i are the learned dual coefficients and *b* is the bias term. The optimization process uses matrix-based calculations, including vectorized dot products and matrix inversions. We can also estimate feature influence in SVR by examining the magnitude of learned support vector coefficients (e.g., Guyon & Elisseeff, 2003; Smola & Schölkopf, 2004).

GPR was also implemented using the scikit-learn package for Python (Pedregosa et al., 2011). GPR is a Bayesian non-parametric model that generalizes the concept of multivariate Gaussian distributions to functions, allowing it to model uncertainty in predictions (Rasmussen & Williams, 2005). GPR provides a distribution over possible functions, where the predicted mean and variance are derived from a multivariate Gaussian. Like linear regression and SVR, GPR represents input data as a matrix, but instead of explicitly learning parameters, it models relationships between data points using a covariance function (kernel function). The covariance structure between training points is represented by a Gram matrix $\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X})$ to make probabilistic predictions. The posterior predictive mean in GPR is given by:

$$\hat{\boldsymbol{y}} = \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}_*)^{\boldsymbol{T}}[\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{\sigma}^{\boldsymbol{2}}\boldsymbol{I}]^{-\boldsymbol{1}}\boldsymbol{Y} \tag{4}$$

Here, $\boldsymbol{K}(\boldsymbol{X},\boldsymbol{X})$ is the kernel matrix of training data; $(\boldsymbol{X}, \boldsymbol{X}_*)$ is the kernel function between training and test data; $\boldsymbol{\sigma^2 I}$ represents noise; and $\boldsymbol{Y}$ is the observed response variable. The covariance matrix inversion required for this computation is a core matrix operation (Rasmussen & Williams, 2005). Because GPR provides a full posterior distribution over predictions, it naturally quantifies uncertainty, unlike models that only produce point estimates. Feature importance in GPR is often inferred using length scales of the kernel function, which describe how much each feature contributes to predictive variance.

The third algorithm we can use is XGBoost, which trains decision trees sequentially using gradient boosting (i.e., each tree corrects the errors of the tree before it; Chen & Guestrin, 2016). XGBoost was implemented using the xgboost package for Python (Chen & Guestrin, 2016). Matrices are everywhere here, too. The core update step is based on the gradients of the loss function:

$$g_i = \frac{\partial L}{\partial f(x_i)}, h_i = \frac{\partial^2 L}{\partial f(x_i)^2} \tag{5}$$

These first-order and second-order gradients are stored as vectors and used in matrix-based optimization. XGBoost also performs Newton's method to update the Hessian matrix:

$$H = \begin{bmatrix} \frac{\partial^2 L}{\partial x_1^2} & \cdots & \frac{\partial^2 L}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 L}{\partial x_n^2} \end{bmatrix}$$

and sparse matrix storage to make matrix operations faster (Chen & Guestrin, 2016). XGBoost also uses parallel matrix operations to compute splits across multiple trees to take advantage of matrix representations of splits.

If the previous few paragraphs are a bit Greek, no worries. Figure 6.6 and Table 6.2 show the results from using linear regression, SVR, GPR, and XGBoost to predict response rates. Though the linear regression R^2 value was decent at 0.90, the root mean square error (RMSE) indicated that model predictions were off by an average of 32.63 response per minute (Table 6.2). Figure 6.6 also shows how the linear regression model failed to account for the nonlinear patterns in the dataset (which we would not expect it to without a bit of additional work). In contrast, the SVR regressor was off by an average of 4.29 responses per minute ($R^2 = 0.99$), the GPR regressor was off by an average of 1.95 responses per minute ($R^2 = 0.99$), and XGBoost was off by an average of 1.95 responses per minute ($R^2 = 0.99$). The exemplary performance of each matrix-based approach is shown by the predicted responses falling almost perfectly along the diagonal of "perfect prediction" in Figure 6.6.

Table 6.2. Fit metrics for the three modeling approaches

Model	RMSE	R^2	AIC
Linear Regression	32.63	0.90	98,074.10
SVR	4.29	0.99	29,123.77
GPR	1.95	0.99	13,339.80
XGBoost	1.95	0.99	13,540.50

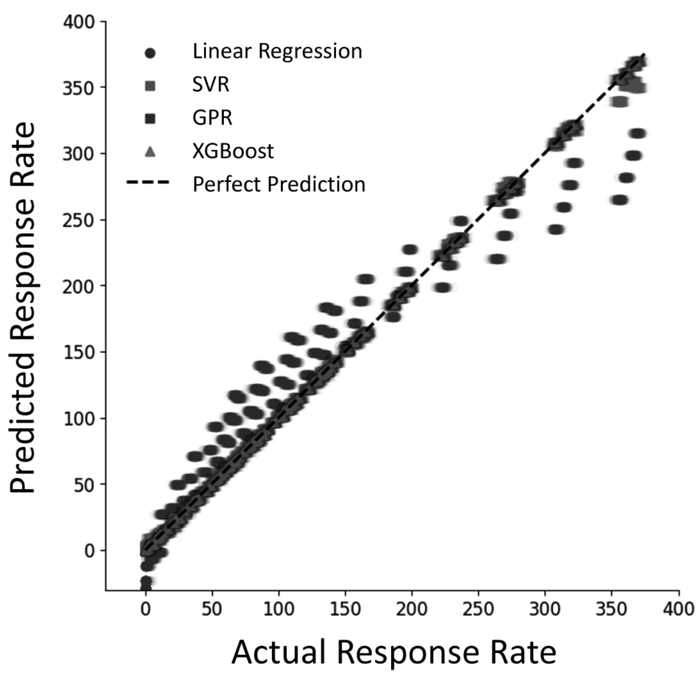


Figure 6.6. Response rate predictions of the linear regression, SVR, GPR, and XGBoost models for the synthetic dataset.

Figure 6.7 also show the Shapley additive explanations (SHAP) plots, which explain the contribution of individual features to a model's predictions (Lundberg & Lee, 2017). That is, similar to the β coefficients above for linear regression, SHAP values show the relative influence of that independent variable on the predicted response rate except it can vary for each observation by nature of the algorithms used. SHAP values are computed using the difference in predicted response rates when different combinations of features are included (i.e., vector subtraction). Each feature can be thought of as a dimension in a vector space, making the SHAP value for each feature being calculated as:

$$\phi_i = \sum\nolimits_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \tag{6}$$

Here, ϕ_i is the SHAP value for feature *i*, *F* is the set of all features, *S* is the subset of features excluding *i*, and *f*(*S*) is the model's output when only using features in *S*. For this particular implementation, Random Forest regressors and XGBoost are tree-based algorithms and, thus, TreeSHAP uses matrix decomposition (e.g., eigenvalues of Hessians) to compute feature importance (Lundberg & Lee, 2017).

Beyond Traditional Statistical Modeling

The above examples could have been accomplished using traditional statistical modeling techniques. That is, linear regression could have been conducted using ordinary least squares regression, and multivariable modeling techniques likely would have performed just fine to understand the influences and interactions among the independent variables in our dataset to influence response rates (Hidalgo & Goodman, 2013). However, linear algebra also allows you to identify the geometry, structure, and complexity of a dataset in ways that traditional multivariable statistical modeling often does not (Kutner et al., 2004; McInnes et al., 2018; Strang, 2003; Tenenbaum et al., 2000). If a dataset contains the learning history for an organism, then linear algebra using vector and matrix thinking allows researchers to identify the geometry, structure, and complexity of the relations between many environmental variables and behavior.

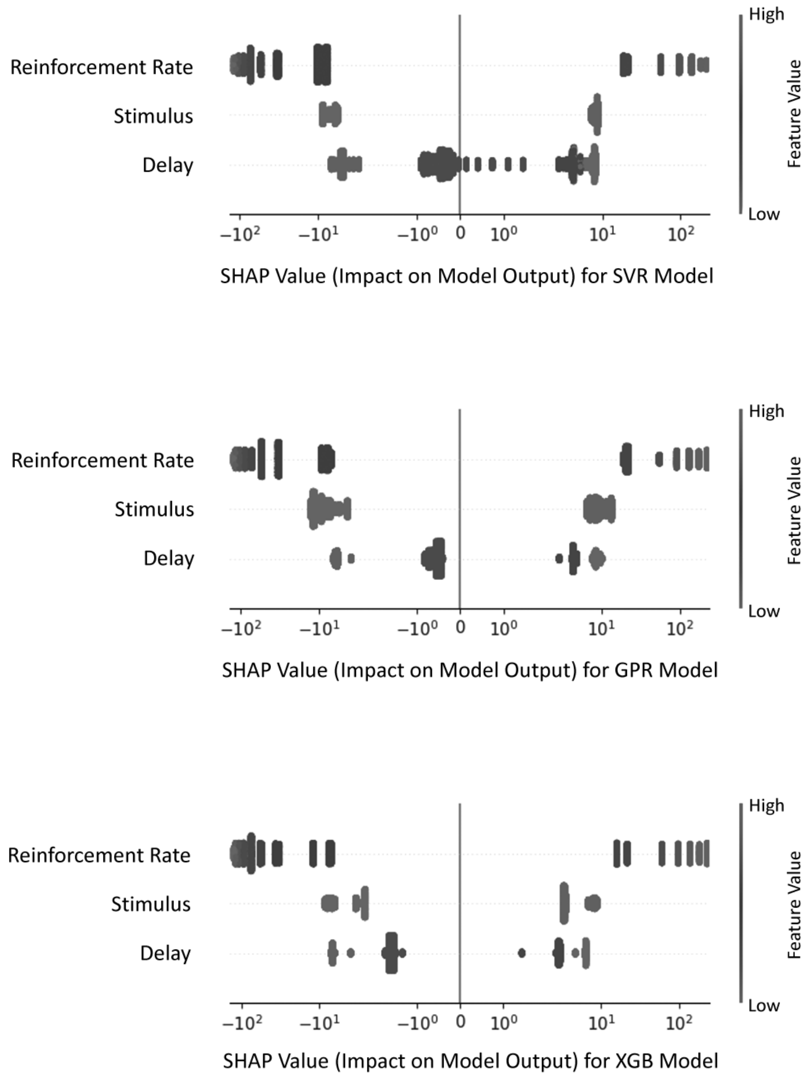


Figure 6.7. SHAP values returned for each of the matrix-based models described in text.

An example might help show this utility. Let's imagine we have the simulated data from above that captures an organism's learning history relative to the influence of context, reinforcer rate, and reinforcer delay on response rates. A behavior scientist might sometimes be interested in simply describing quantitatively an organism's learning history relative to specific environment behavior relations (e.g., matching law, discounting). However, behavior scientists are often more interested in how applying an environmental change might lead to a change in behavior. Or asked predictively and with precision, how difficult would it be to shape one learning history into a specific new one?

One way we can estimate a quantitative answer to this question is through singular value decomposition (SVD; e.g., Horn & Johnson, 2012). SVD is a matrix operation that decomposes a matrix (e.g., an organism's learning history, Table 6.1) into three specific matrices that capture the important structural properties of the original matrix. Specifically, SVD for an m x n matrix $\boldsymbol{A}$, is described as:

7

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{T}$$

where $\boldsymbol{U}$ is an m x m orthogonal matrix comprising the left singular vectors (column space basis); $\boldsymbol{S}$ is an m x n diagonal matrix with non-negative singular values (which determines the importance of each dimension); and $\boldsymbol{V}^{T}$ is an n x n orthogonal matrix comprising the right singular vectors (row space basis).

How does this help us? The information we get from SVD can help us in at least two ways based on the values of the singular values in each matrix. First, the number of non-zero singular values in $\boldsymbol{S}$ tells us the rank of the matrix (i.e., how many of the variables provide non-redundant, unique information about the dataset; Horn & Johnson, 2012). Mathematically, if two datasets describe an organism's baseline learning history and the learning history we want to have post-intervention or post-research, then knowing the rank of each learning history matrix lets us know whether one can be transformed into the other. If an organism's baseline learning history matrix is the same rank or a higher rank than the terminal learning history matrix one wants to create, then it is possible to mathematically transform the baseline learning history into the post-intervention learning history. If the post-intervention learning history is a higher rank than the baseline learning history, it is not possible to mathematically transform the baseline learning history into the post-intervention learning history (Fisher, 2011; Fleisch, 2011; Horn & Johnson, 2012). In the latter scenario, you would need to do something additional to the independent variables included in the dataset to successfully transform the baseline learning history into the post-intervention learning history.

The second way the singular values help us is by comparing the ratios of the singular values for each matrix. If the singular values are very similar, then their geometric properties are also similar, and the transformation is likely to be easy (Fisher, 2011; Horn & Johnson, 2012). In contrast, if the singular values are very different, then their geometric properties are likely very different, and it would take more effort to transform the baseline learning history into the post-intervention learning history –even if they are the same rank (Fisher, 2011; Horn & Johnson, 2012).

Figure 6.8 shows a concrete example of the previous paragraphs. The top plot shows the original dataset shown in Table 6.1. Let's say this is our baseline learning history dataset. The lower left plot shows a very similar dataset, though slightly different. For example, perhaps we want the individual's repertoire post-intervention to be very similar but for the conditional discriminations in their repertoire to be more fine-tuned. Both learning history matrices are ranked four, suggesting that each independent variable distinctly influences the response rate, and the matrices that comprise these images span the same subspace. Table 6.3 shows the singular values associated with each matrix (left columns) and how similar those singular values are for the two matrices (right columns). The singular values indicate stimulus context contributes the most to variability followed by reinforcement rate and then delay to reinforcement. The similarity ratios are also close to 1.00, suggesting that it would not be challenging to transform the organism's original baseline learning history into the new, similar set of environment-behavior relations we might like to see post-intervention.

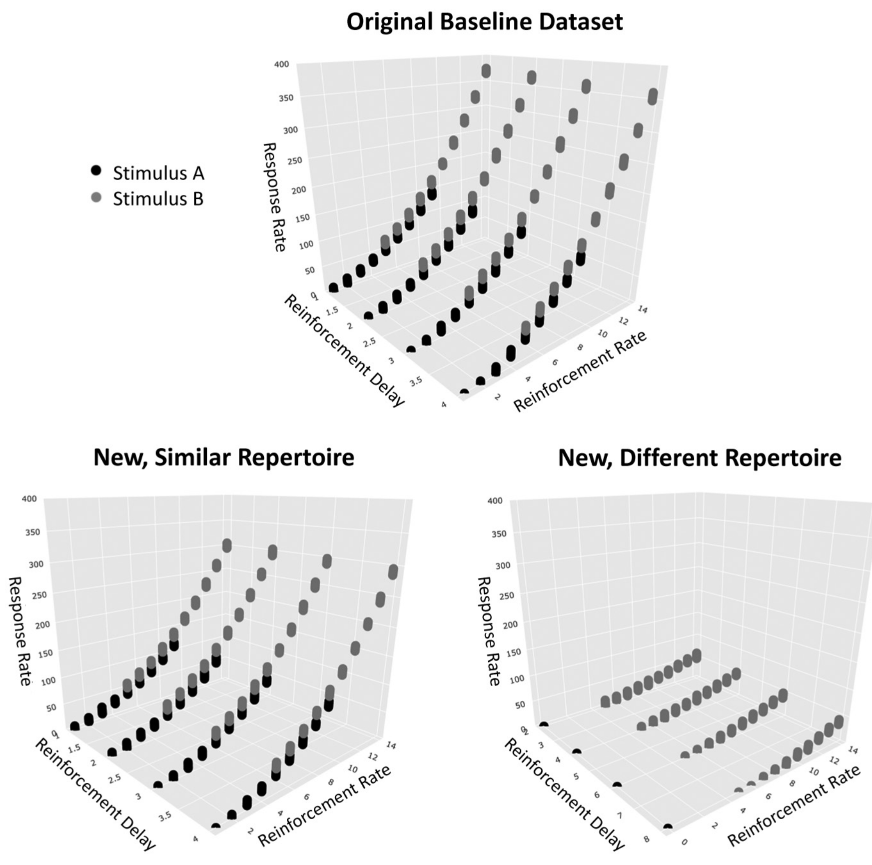


Figure 6.8. Visualizations of an organism's hypothetical learning histories are plotted from matrix form to highlight what different matrix views of learning histories might look like.

Table 6.3. Singular values of the datasets (left columns) plotted in Figure 7 and their ratio similarity to the baseline dataset (right columns)

Variable	Singular Values			Similarity Ratios	
	Baseline	New, Similar	New, Different	$\left(\frac{\text{Baseline}}{\text{New, Similar}}\right)$	$\left(\frac{\text{Baseline}}{\text{New, Different}}\right)$
Stimulus Context	16012.86	12991.72	2057.22	0.81	0.13
Reinforcement Rate	278.385	274.78	498.23	0.99	1.79
Reinforcement Delay	95.64	96.55	194.57	1.01	2.03
Response Rate	37.34	37.21	12.68	1.00	0.34

The bottom right panel of Figure 6.8 also shows an example where the environment-behavior relations we might want to see post-intervention differ significantly from baseline. For example, perhaps we are interested in responding occurring only in one stimulus context, at a lower rate, and with less variability. Both environment-behavior matrices are still ranked four, indicating they span the same subspace, and it is possible to transform the baseline environment-behavior relations into the other. The singular values of this new, different matrix are shown in Table 6.3. The same order of singular values occurs in each, though the stimulus context, reinforcement rate, and reinforcement delay are now closer to one another than in the baseline dataset. The similarity ratios, however, highlight the difficulty of the task at hand. All differ significantly from 1.00. Thus, though it is theoretically possible to transform the baseline environment-behavior relations into the post-intervention matrix of environment-behavior relations, the current set of environmental variables would make it more difficult to do than in the first example.

There was a lot of jargon in this section. Let's briefly summarize the high-level point being made. When an organism enters any learning environment (e.g., research experiment, clinical therapy session), we can collect baseline data on how they behave relative to a set of environmental variables. The total dataset we now have can be thought of as a matrix representing the organism's learning history relative to the environment-behavior contexts they have experienced. Similarly, as a researcher or practitioner, I might want my experimental manipulation or therapeutic intervention to result in a specifically different pattern of environment-behavior relations. We can use linear algebra to predict how easy it will be to transform the baseline learning history into the desired post-experiment or post-intervention learning history, and whether it is even possible without adding some else to the mix we may not have otherwise considered.

Chapter Summary

Advances in many sciences come by way of combining mathematical and computational models to analyze data and propose new ideas. Though this might sound intimidating to the uninitiated, behavior scientists already use all sorts of models in their everyday work. For example, they may use textual-vocal descriptions of the four-term contingency surrounding aggression, graphical models of the three-term contingency when explaining behavior analysis to someone new to the field, or the evolutionary theory of behavior dynamics to model resurgence. Similarly, the models we use might describe the things we have observed happen; they may help us guess when something will happen again; they may tell us why something happened; and they may even help us know the best decision to make. Models are everywhere in behavior science. Each one has a specific form and function.

Quantitative and computational models are particularly useful when behavior scientists want to study complicated or complex environment-behavior interactions. This is primarily because the precision of quantitative and computational models makes it easier to keep track of the precise values of environmental and behavioral variables at varying points in time and to estimate precisely how those data relate. One area where this might be useful is in the study of complex multiple control of behavior. That is, situations where two or more environmental variables interact to influence two or more behaviors. Influential and interacting multiple environmental variables might occur as an antecedent to responding (e.g., compound stimulus arrangements) or as a consequence of responding (e.g., two or more reinforcement rates in matching law experiments).

From a data perspective, beginning to study the complicated or complex multiple control of behavior does not require the behavior scientist to do anything different or unique. It simply requires adding more columns to your dataset to include more measures of the environment or more measures of behavior. Here, each row in the dataset would include the unique environmental and response values at the moment the observation is made. And, each column would represent one aspect of the environment or of behavior. Beyond two or three different independent variables, however, traditional visual analytic methods become difficult to make use of or potentially impossible for researchers interested in dozens, hundreds, or even thousands of variables we can now collect data on.

Traditional statistical modeling can help make sense of our data at this point. In many situations, it may often be good enough for the question at hand (e.g., discounting, matching). That is, many analytic techniques exist that allow us to identify the relative influence of each variable on behavior, as a whole, and many techniques allow us to manually account for interactions among variables in our dataset. However, as the number of variables grow in size, traditional statistical modeling techniques start to become less computationally performant and more difficult to account for many interactions simultaneously. And, though manually making assumptions about interactions among variables is technically possible, at a certain degree of complexity it might be more productive to let our data drive the conversation.

Linear algebra becomes useful in these situations. Linear algebra suggests we think about each observation recorded as a vector (i.e., a unique combination of all the environment-behavior components that comprise that moment in time) and that we think about the organism's learning history as a matrix by way of the dataset *in toto*. Consider that at any one point in time our sensory systems are impinged upon by thousands of different stimuli and we are constantly emitting dozens or hundreds of responses (e.g., breathing, coordinated finger movements on a keyboard, orientation of gaze at a computer screen). The behavior of organisms involves the integration of all of this within a dynamically changing stream of existence spanning the molecular granularity of our day to molar aggregated of our life. Researchers can now practically ask experimental questions about the integration of this information. It is a novel contemporary challenge we have the luxury to be confronted with. At a minimum, vectorized operations provide a computationally efficient means whereby researchers interested in the multiple control of multiple behaviors can conduct these analyses faster and with larger datasets.

Linear algebra is also one method whereby behavior scientists can more precisely describe and predict the complex multiple control of behavior. Many modern machine learning algorithms involve matrix-based computations as described via several examples above. The benefits of using machine learning to analyze environment-behavior data is only just beginning to be realized. However, behavior scientists interested in analyzing more complex environment-behavior relations can gain tremendous precision in their descriptions and predictions alongside transparency around what variables are most influential for each response emitted and how many different variables likely interact and dynamically influence behavior over time.

Finally, thinking about learning histories as matrices may allow behavior scientists to ask questions previously unconsidered outside theoretical writings. Organisms are integrated wholes and cannot be understood *in toto* by isolating individual responses relative to a few isolated environmental variables. Thinking about learning histories as matrices gives behavior scientists a method to talk about, analyze, and manipulate learning histories as a single, integrated object. This carries many interesting possibilities. One demonstrated in this chapter was that it allows behavior scientists to predict quantitatively the difficulty of transforming one behavioral repertoire into another. In clinical, educational, and other applied contexts, we often want to know what effort is required and how long it will take to change someone's baseline behavioral repertoire into a new behavioral repertoire where the clinical concern is absent, the things in need of learning have been learned, and so on. The mathematics of linear algebra tell us whether or not it will be possible and how difficult it might be given the environmental variables at play. This approach to thinking about behavioral repertoires is in its infancy and, thus, many potential applications likely exist that are more productive than the one offered in this chapter. But it provides a glimpse of what might be yet to come.

To return to the quote that opened the chapter, to handle the complexity required for the analysis of the behavior of whole organisms necessitates that behavior scientists have the tools in their toolbox to efficiently do so. Behavior scientists have a long history of using models to talk about environment-behavior relations. Advances in data collection tools, software, and hardware allow behavior scientists to collect an increasingly greater amount of information on environment-behavior relations. These data can be thought of as vectors (each observational point) and matrices (an organism's learning history). Linear algebra provides one mathematical approach that, when instantiated into computational algorithms, will help us to turn the complex nature of behavior from difficult into just a bit easier.

References

Adami, C. (2002). What is complexity? *BioEssays, 24,* 1085-1094. https://onlinelibrary.wiley.com/doi/pdf/10.1002/bies.10192

Allen, T. F. H., & Hoekstra, T. W. (1992). *Toward a Unified Ecology.* Columbia University Press.

Andrade, J. S., Wainer, I., & Moreira, J. E. (1995). Self-organized criticality in the El Nino southern oscillation. *Physica A: Statistical Mechanics and its Applications, 215*(33), 331-338. https://doi.org/10.1016/0378-4371(95)00004-Q

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior, 22*(1), 231-242. https://dx.doi.org/10.1901%2Fjeab.1974.22-231

Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior, 12*(6), 861-874. https://doi.org/10.1901%2Fjeab.1969.12-861

Berg, J. P., & McDowell, J. J. (2011). Quantitative, steady-state properties of Catania's computational model of the operant reserve. *Behavioral Processes, 87*(1), 71-83. https://doi.org/10.1016/j.beproc.2011.01.006

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. ISBN: 978-0-387-31073-2

Blackburn, M., & El-Deredy, W. (2013). The future is risky: Discounting of delayed and uncertain outcomes. *Behavioral Processes, 94*, 9-18. http://dx.doi.org/10.1016/j.beproc.2012.11.005

Brown, C. (1995). *Chaos and Catastrophe Theories.* Sage.

Burgos, J. (2001). A neural-network interpretation of selection in learning and behavior. *Behavioral and Brain Sciences, 24*(3), 531-533. https://doi.org/10.1017/S0140525X01254168

Byrne, D. (1997). Chaotic places or complex places: Cities in a post-industrial era. In: S. Westwood, & J. Williams (Eds.), *Imagining Cities* (pp.50-72). Routledge.

Catania, A. C. (2005). The operant reserve: A computer simulation in (accelerated) real time. *Behavioural Processes, 69*, 257-278. https://doi.org/10.1016/j.beproc.2005.02.009

Catania, A. C. (2013). *Learning: Fifth Edition*. The Cambridge Center.

Chaitin, G. J. (1992). *Information, Randomness, and Incompleteness.* World Scientific.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

Collier, G., Johnson, D. F., & Morgan, C. (1992). The magnitude-of-reinforcement function in closed and open economies. *Journal of the Experimental Analysis of Behavior, 57*(1), 81-89. https://doi.org/10.1901/jeab.1992.57-81

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied Behavior Analysis* (3rd ed.). Pearson Education, Inc.

Correig, A. M., Urquizu, M., & Vila, J. (1997). Aftershock series of event February 18, 1996: An interpretation in terms of self-organized criticality. *Journal of Geophysical Research: Solid Earth, 102*(B12), 27407-27420. https://doi.org/10.1029/97JB02487

Cox, D. J. (2019). The many functions of quantitative modeling. *Computational Brain & Behavior, 2,* 166-169. https://doi.org/10.1007/s42113-019-00048-9

Cox, D. J., & Brodhead, M. T. (2021). A proof-of-concept analysis of decision-making with time-series data. *The Psychological Record, 71*, 349-366. https://doi.org/10.1007/s40732-020-00451-w

Cox, D. J. & Dallery, J. (2016). Effects of delay and probability combinations on discounting in humans. *Behavioral Processes, 131,* 15-23. http://dx.doi.org/10.1016/j.beproc.2016.08.002

Cox, D. J., & Dallery, J. (2018). Influence of second outcome on monetary discounting. *Behavioral Processes, 153,* 84-91. https://doi.org/10.1016/j.beproc.2018.05.012

Cox, D. J., Javed, A., Sosine, J., Cordeiro, C., & Sotomayor, J. (2022). Data recording and analysis. In: J. K. Luiselli (Ed.), *Applied Behavior Analysis Advanced Guidebook: A Manual for Professional Practice* (2nd ed., pp. 217-239). Academic Press.

Critchfield, T. S., & Shue, E. Z. H. (2018). The dead man test: A preliminary experimental analysis. *Behavior Analysis in Practice, 11*(4), 381-384. https://doi.org/10.1007%2Fs40617-018-0239-7

Dallery, J., Kurti, A., & Erb, P. (2015). A New Frontier: Integrating Behavioral and Digital Technology to Promote Health Behavior. *The Behavior Analyst, 38*(1), 19-49. https://doi.org/10.1007/s40614-014-0017-y

Dallery, J., & Soto, P. L. (2013). Quantitative description of environment-behavior relations. In G. J. Madden (Ed.), *APA handbook of behavior analysis: Vol. I. Methods and principles* (pp. 219-249). American Psychological Association. https://doi.org/10.1037/13937-010

Dalton, J. (1808). *A new system of chemical philosophy: Part I*. Bickerstaff.

Davison, M., & McCarthy, D. (1988). *The matching law: A research review*. Erlbaum.

Diller, J. W., Barry, R. J., & Gelino, B. W. (2016). Visual analysis of data in a multielement design. *Journal of Applied Behavior Analysis, 49*, 980-985. https://doi.org/10.1002/jaba.325

Donahoe, J. W., Burgos, J. E., & Palmer, D. C. (1993). A selectionist approach to reinforcement. *Journal of the Experimental Analysis of Behavior, 60*(1), 17-40. https://doi.org/10.1901/jeab.1993.60-17

Estle, S. J., Green, L., Myerson, J., & Holt, D. D. (2006). Differential effects of amount on temporal and probability discounting of gains and losses. *Memory & Cognition, 34*(4), 914-928. https://doi.org/10.3758/BF03193437

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Appleton-Century-Crofts. https://doi.org/10.1037/10627-000

Fisher, W. P., Jr. (2011). Singular values and the transformation of learning history. *Constructivist Foundations*, *6*(2), 241-252.

Fleisch, D. A. (2011). *A Student's Guide to Vectors and Tensors*. Cambridge University Press. https://doi.org/10.1017/CBO9781139031035

Foster, T. M., Blakman, K. A., & Temple, W. (1997). Open versus closed economies: Performance of domestic hens under fixed-ratio schedules. *Journal of the Experimental Analysis of Behavior, 67*(1), 67-89. https://doi.org/10.1901/jeab.1997.67-67

Geiger, K. B., Carr, J. E., & LeBlanc, L. A. (2010). Function-based treatment for escape-maintained problem behavior: A treatment-selection model for practicing behavior analysts. *Behavior Analysis in Practice, 3*(1), 22-32.

Golub, G. H., & Van Loan, C. F. (1996). *Matrix Computations* (3rd ed.). Johns Hopkins University Press.

Green, L., Myerson, J., Oliveira, L., & Chang, S. E. (2013). Delay discounting of monetary rewards over a wide range of amounts. *Journal of the Experimental Analysis of Behavior, 100*(3), 269-281. https://doi.org/10.1002/jeab.45

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research 3*, 1157-1182. https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf

Hall, G. A., & Lattal, K. A. (1990). Variable-interval schedule performance in open and closed economies. *Journal of the Experimental Analysis of Behavior, 54*(1), 13-22. https://doi.org/10.1901/jeab.1990.54-13

Heacox, W. D. (2015). *The expanding universe*. Cambridge University Press.

Herrnstein, R. J. (1970). On the law of effect. *Journal of The Experimental Analysis of Behavior, 13*(2), 243-266. https://doi.org/10.1901/jeab.1970.13-243

Hidalgo, B., & Goodman, M. (2013). Multivariate or multivariable regression? *American Journal of Public Health, 103*(1), 39-40. https://doi.org/10.2105/AJPH.2012.300897

Holland, J. H. (1992). Complex adaptive systems. *Daedalus, 121*(1), 17-30. https://www.jstor.org/stable/20025416

Horn, R. A., & Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511810817

Huang, X., Wu, L., & Ye, Y. (2019). A review on dimensionality reduction techniques. *International Journal of Pattern Recognition and Artificial Intelligence, 33*(10), 1950017. https://doi.org/10.1142/S0218001419500174

Hursh, S. R., & Fantino, E. (1973). Relative delay of reinforcement and choice. *Journal of the Experimental Analysis of Behavior, 19*(3), 437-450. https://doi.org/10.1901/jeab.1973.19-437

Hursh, S. R., & Silberberg, A. (2008). Economic demand and essential value. *Psychological Review, 115*(1), 186-198. http://dx.doi.org/10.1037/0033-295X.115.1.186

Kahng, S. W., Chung, K. M., Gutshall, K., Pitts, S. C., Kao, J., & Girolami, K. (2010). Consistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis, 43*, 35-45. https://doi.org/10.1901/jaba.2010.43-35

Kessel, R., & Lucke, R. L. (2008). Ana analytic form for the interresponse time analysis of Shull, Gaynor, and Grimes with applications and extensions. *Journal of the Experimental Analysis of Behavior, 90*(3), 363-386. https://doi.org/10.1901/jeab.2008.90-363

Killeen, P. R. (1994). Mathematical principles of reinforcement. *Behavioral and Brain Sciences, 17*(1), 105-135. https://doi.org/10.1017/S0140525X00033628

Killeen, P. R. (2009). An additive-utility model of delay discounting. *Psychological Review, 116*, 602-619. https://doi.apa.org/doi/10.1037/a0016414

Killeen, P. R. (2019). Predict, control, and replicate to understand: How statistics can foster the fundamental goals of science. *Perspectives on Behavior Science, 42*, 109-132. https://doi.org/10.1007/s40614-018-0171-8

Kirby, M., & Kennedy, C. H. (2013). Variable-interval reinforcement schedule value influences responding following REM sleep deprivation. *Journal of the Experimental Analysis of Behavior, 80*(2), 253-260. https://doi.org/10.1901/jeab.2003.80-253

Kurti, A. N., & Dallery, J. (2014). Integrating technological advancements in behavioral interventions to promote health: Unprecedented opportunities for behavior analysts. *Revista Mexicana de Análisis de la Conducta, 40*(2), 106-126. https://doi.org/10.5514/rmac.v40.i2.63670

Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied Linear Regression Models*. McGraw-Hill.

LeBlanc, L. A., Ratez, P. B., Sellers, T. P., & Carr, J. E. (2016). A proposed model for selecting measurement procedures for the assessment and treatment of problem behavior. *Behavior Analysis in Practice, 9*, 77-83.

Lowe, J. (1985). *The Dynamics of Collapse: A Systems Simulation of the Classic Maya Collapse*. University of New Mexico Press.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* (NeurIPS 2017).

Manson, S. M. (2001). Simplifying complexity: a review of complexity theory. *Geoforum, 32*(3), 405-414. https://doi.org/10.1016/S0016-7185(00)00035-X

Marr, M. J. (2015). Reprint of “Mathematics as verbal behavior”. *Behavioral Processes, 114*, 34-40. https://doi.org/10.1016/j.beproc.2015.03.008

Mayr, E. (1996). The modern evolutionary theory. *Journal of Mammology, 77*(1), 1-7.

McDowell, J. J. (2005). On the classic and modern theories of matching. *Journal of the Experimental Analysis of Behavior, 84*(1), 111-127. https://doi.org/10.1901%2Fjeab.2005.59-04

McDowell, J. J. (2013a). Calculi of complexity: How phenomena emerge from rules. *Journal of the Experimental Analysis of Behavior, 99*(2), 234-244. https://doi.org/10.1002/jeab.16

McDowell, J. J (2013b). A quantitative evolutionary theory of adaptive behavior dynamics. *Psychological Review, 120*, 731-750. https://doi.org/10.1037/a0034244

McDowell, J. J. (2019). On the current status of the evolutionary theory of behavior dynamics. *Journal of the Experimental Analysis of Behavior, 111*(1), 130-145. https://doi.org/10.1002/jeab.495

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* https://doi.org/10.48550/arXiv.1802.03426

Merriam-Webster (2022). Model. Retrieved from: https://www.merriam-webster.com/dictionary/model

Michael, J., Palmer, D. C., & Sundberg, M. L. (2011). The multiple control of verbal behavior. *The Analysis of Verbal Behavior, 27*(1), 3-22. https://dx.doi.org/10.1007%2FBF03393089

Motulsky, H. (2014). *Intuitive biostatistics: A nonmathematical guide to statistical thinking*. Oxford University Press.

Moxley, R. (1982). Graphics for three-term contingencies. *The Behavior Analyst, 5*(1), 45-51. https://doi.org//10.1007/BF03393139

Nevin, J. A., Mandell, C., & Atak, J. R. (1983). The analysis of behavioral momentum. *Journal of the Experimental Analysis of Behavior, 39*(1), 49-59. https://doi.org/10.1901%2Fjeab.1983.39-49

Nijkamp, P., & Reggiani, A. (1990). Logit models and chaotic behavior: a new perspective. *Environment and Planning A: Economy and Space 22*, 1455-1467. https://doi.org/10.1068%2Fa221455

Ninci, J., Vannest, K. J., Willson, V., & Zhang, N. (2015). Interrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification, 39*, 510-541. https://doi.org/10.1177/0145445515581327

Norton, B. G., & Ulanowicz, R. E. (1992). Scale and biodiversity policy: a hierarchical approach. *Economics of Biodiversity, 21* (3), 244-249. https://www.jstor.org/stable/4313934

Patten, B. C. (1985). Energy cycling in the ecosystem. *Ecological Modelling, 28*(1-2), 1-71. https://doi.org/10.1016/0304-3800(85)90013-4

Pearson, J. M., & Platt, M. L. (2013). Change detection, multiple controllers, and dynamic environments: insights from the brain. *Journal of the Experimental Analysis of Behavior, 99*(1), 74-84. https://doi.org/10.1002/jeab.5

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O… & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research, 12*, 2825-2830. https://dl.acm.org/doi/pdf/10.5555/1953048.2078195

Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.) *Advances in Large Margin Classifiers* (pp. 61-74). MIT Press.

Rachlin, H. (2006). Notes on discounting. *Journal of the Experimental Analysis of Behavior, 85*(3), 425-435. https://doi.org/10.1901%2Fjeab.2006.85-05

Rachlin, H., Arfer, K. B., Safin, V., & Yen, M. (2015). The amount effect and marginal value. *Journal of the Experimental Analysis of Behavior, 104*(1), 1-6. https://doi.org/10.1002/jeab.158

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001

Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of dimensionality reduction techniques on big data. *IEEE Access, 8,* 54776-54788. https://doi.org/10.1109/ACCESS.2020.2980942

Renfrew, C., & Cooke, K. L. (1979). *Transformations: Mathematical Approaches to Cultural Change.* Academic Press.

Saini, V., Betz, A. M., Gregory, M. K., Leon, Y., & Fernandez, N. (2017). A survey-based method to evaluate optimal treatment selection for escape-maintained problem behavior. *Behavior Analysis in Practice, 10*(3), 214-227.

Sagan, C. (1994). *Pale blue dot: A vision of the human future in space*. Ballantine Books.

Santangelo, J. S., Johnson, M. T. J., & Ness, R. W. (2018). Modern spandrels: The roles of genetic drift, gene flow and natural selection in the evolution of parallel clines. *Proceedings of the Royal Society B: Biological Science, 285*, 20180230. https://doi.org/10.1098/rspb.2018.0230

Schrödinger, E. (1926). An undulatory theory of the mechanics of atoms and molecules. *The Physical Review, 28*(6), 1049-1070. http://people.isy.liu.se/jalar/kurser/QF/references/Schrodinger1926c.pdf

Shahan, T. A., Browning, K. O., & Nall, R. W. (2020). Resurgence as choice in context: Treatment duration and on/off alternative reinforcement. *Journal of the Experimental Analysis of Behavior, 113*(1), 57-76. https://doi.org/10.1002/jeab.563

Shawler, L. A., Miguel, C. F., Cengher, M., Zhelezoglo, K. N., & Luoma, S. M. (2022). Establishing derived reinforcers via stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 117*(2), 180-200. https://doi.org/10.1002/jeab.739

Skinner, B. F. (1935). Two types of conditioned reflex and a pseudotype. *Journal of General Psychology, 12*, 66-77.

Skinner, B. F. (1937). Two types of conditioned reflex: A reply to Konorski and Miller. *The Journal of General Psychology, 16*(1), 272-279. https://doi.org/10.1080/00221309.1937.9917951

Skinner, B. F. (1938). *The behavior of organisms*. Appleton-Century-Crofts.

Skinner, B. F. (1953). *Science and human behavior*. The B.F. Skinner Foundation.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*, 199-222. https://link.springer.com/article/10.1023/B:STCO.0000035301.49549.88

Strang, G. (2003). *Introduction to Linear Algebra* (3rd ed.). Wellesley-Cambridge Press.

Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science, 290*(5500), 2319-2323. https://doi.org/10.1126/science.290.5500.2319

Tiger, J. H., Hanley, G. P., & Bruzek, J. (2008). Functional communication training: a review and practical guide. *Behavior analysis in practice, 1*(1), 16-23. https://doi.org/10.1007/BF03391716

Trefethen, L. N., & Bau, D. (1997). *Numerical linear algebra.* SIAM: Society for Industrial and Applied Mathematics.

USDHHS (2020). Computational modeling. Retrieved from: https://www.nibib.nih.gov/science-education/science-topics/computational-modeling

Vanderveldt, A., Green, L., & Myerson, J. (2015). Discounting of monetary rewards that are both delayed and probabilistic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 148-162. http://dx.doi.org/10.1037/xlm0000029

Wolfe, K., & Slocum, T. A. (2015). A comparison of two approaches to training visual analysis of AB graphs. *Journal of Applied Behavior Analysis, 48*(2), 472-477. https://doi.org/10.1002/jaba.212

Yi, R., de la Piedad, X., & Bickel, W. K. (2006). The combined effects of delay and probability in discounting. *Behavioral Processes, 73*, 149-155. http://dx.doi.org/10.1016/j.beproc.2006.05.001

Zeiler, M. D. (2013). Reversed schedule effects in closed and open economies. *Journal of the Experimental Analysis of Behavior, 71*(2), 171-186. https://doi.org/10.1901/jeab.1999.71-171