Nomological Machines, Data Science and Dynamics

Rick Dale
Department of Communication, UCLA
https://co-mind.org/rick

In this newsletter contribution, my primary aim is to introduce equation discovery using surprisingly simple methods from statistics and data science. I illustrate that under certain conditions, a relatively simple regression method can precisely recover a dynamical system's governing equations from observed data (Brunton, Proctor, & Kutz, 2016; Dale & Bhat, 2018). I illustrate this in a very simple system familiar, I assume, to readers here: the logistic map. Under empirical research circumstances, in which we use natural and noisy human time series data, it may not be possible to recover easy-to-interpret equations. I summarize how the equation discovery method may be useful in these situations, too. In general, this methodology figures into the emergence of data science, in which we wrangle noisy human data through machine learning and related techniques. Psychologists interested in nonlinear dynamics may find fruitful new tools here.

Rather than jumping right into this demonstration, I will begin instead with some broader theoretical reflections. From a methodological vantage point, what I showcase below is simply an application of well-known regression methods. But I believe the method has broader relevance. Equation discovery raises intriguing theoretical discussion about psychological dynamics and scientific procedures. Indeed, some of my own interest in equation discovery is rooted in training I received in the philosophy of science long ago while a student. This training continues to sneak up on me from time to time. And so, at the risk of overdoing it, I begin with a shamelessly broad backstory.

1. Shamelessly broad backstory

The great philosopher of science Nancy Cartwright challenges the traditional notion of universal scientific laws. Her book *The Dappled World* is one that impacted me while a student (Cartwright, 1999). In it, she defends what I still see as a compelling position, common among the "scientific pluralists" of philosophy (e.g., Kellert, Longino, & Waters, 2006). The position, put simply, is that a scientific law may be true, but that does not imply that the law is universal. To us psychologists, this can seem comically intuitive. Most of us, I would assume, admit that laws of psychology govern psychological beings. For example, we cannot ascribe "introversion" to subatomic particles, not even neutrinos.

But consider basic physical theory instead. Many of us speak as if its laws are the supreme example of universality that psychology should achieve.[1] Cartwright (1999) argues that we should interpret the success of physics quite differently, not as a demonstration of its universality, but rather as a demonstration of how sharp the boundaries are around its explanations: "…the special kinds of circumstances that fit the models of a single theory turn out to be hard to find and difficult to construct." (p. 10) In other words, even the best models have sharp boundaries around them. The most famous example of this can also seem comical. It is also relevant to this newsletter's Society: Add one body to two, chaos.[2]

In her great book, Cartwright tries to dismantle the intuitions we have about physical theories being universal. She orients the reader around a different framework for science: Scientific practices are systematic methods that identify *the many regularities* of the universe. These practices are what she calls

[1] Obviously there are other properties of physical law we wish to achieve, such as its amazing predictive precision, its sophisticated network of relationships, such as with chemistry, and so on. It is obvious that these are things worth achieving in our own theories (Muthukrishna & Henrich, 2019). Psychophysics does a pretty good job! In any case, here I speak only of ontological commitments, such as universality.
[2] I recognize the three-body problem does not break physics. It was just tricky for theory at the time, or so I've heard. It is easy to consult various online resources summarizing this issue, along with the exciting contributions of Poincaré, an important precursor to the theoretical underpinnings of this newsletter's Society (e.g., bit.ly/2DTSJvQ).

nomological machines—law-making machines—contexts and procedures that reliably generate some observable behavior in the universe. Experimental psychology works this way, of course. We generate regularities in behavior by controlling stimulus order, stimulus properties, task context, and so on. Where these laws breakdown, too, we find interesting potential for expanding our understanding (Chater & Brown, 2008). The upshot of all this, to Cartwright, is that we live in a dappled world, one beautifully pockmarked with regularities under different conditions, at different levels of analysis and through particular manipulations. These regularities are found through the systematic and bounded procedures of our nomological machines.

2. Nomological machines of psychology

Cartwright's "nomological machine" is a concept that psychologists can easily recognize. Psychology has long used two prominent kinds of nomological machine. The first of these is *experimental control*: We setup experiments and observation equipment in just a given way. This generates regularities and, on occasion, laws. But break the conditions, and the laws may no longer be useful.

A second kind of nomological machine is *statistical control* (useful discussion in Darlington & Hayes, 2016, Chap. 1). Many readers of the Society's newsletter are no doubt experts on this kind of "machine," as well. Psychologists who study complex clinical matters, personality, emotion dynamics, and so on, do not always have the benefits of rigid experimental control, especially if their interests are in natural circumstances. Instead, the researcher collects data of various kinds, often under noisy conditions. The researcher then subjects these data to extensive statistical modeling to render stable regularities and relationships when these variables are held constant relative to each other.

In this sense, perhaps one of the most prominent nomological machines is the structural equation modeling (SEM) approach that is now widely used in many areas of psychology. The variables in an SEM model—often latent variables inferred from covariation among response items—are specified as theoretical constructs themselves. Their relationships to other key variables are then established through statistical control. So the practice of collecting data systematically, but in situations that are varied and difficult to control, can be plugged into the statistical framework to reveal the regularities that Cartwright refers to.

3. What about dynamics?

Many frameworks of statistical control, including SEM, are not great at capturing psychological dynamics, especially if these dynamics violate certain assumptions of these models (see chapters in Riley & Van Orden, 2005). SEM, for example, can be equipped with some useful extensions to handle time (for an elegant example see Keith, 2005, Chap. 20). But in general, as a framework for studying the rich dynamic character of psychological processes, traditional statistical frameworks have important limitations.[3]

The most common alternative quantitative framework for describing and modeling dynamic behaviors are ordinary differential equations, and basic difference equations. I will not offend this readership with a definition, but will only remind with a recognizable example. Consider our old friend the logistic map, a discrete difference equation specifying the time course of a single state variable:

$$x_{t+1} = ax_t(1 - x_t)$$

This little equation, as we all know, showcases many of the behaviors that interest researchers of dynamics. Under increasing values for control parameter a, it bifurcates repeatedly across limit cycles into chaos, through a period-doubling cascade.

This equation defines the system's evolution, and precisely expresses the relationships between its state variable and a control parameter. Some see such models as the *sine qua non* of scientific

---

[3] There are plenty of other options in the linear modeling framework. Going into these is outside the scope of this brief newsletter article. For example, growth-curve multilevel modeling is a promising way to look at temporal change (Mirman, 2014).

discovery—when we have these equations in hand, we have made great headway in understanding a phenomenon by encasing it in the language of mathematics (e.g., Gottman, Murray, Swanson, Tyson, & Swanson, 2005). For example, in my own area of cognitive science, an intriguing debate once erupted about this dynamic formalism as an explanatory construct (Bechtel, 1998; Eliasmith, 1996; Kaplan & Bechtel, 2011; Van Gelder, 1995). My favorite reflection on these dynamic models comes from one of the original cognitive scientists:

> "For systems that change through time, explanation takes the form of laws acting on the current state of the system to produce a new state – endlessly. Such explanations can be formalized with differential or difference equations. A properly programmed computer can be used to explain the behavior of the dynamic system that it simulates. Theories can be stated as computer programs." (Simon, 1992, p. 160)

Simon saw the relationship between early cognitive models, even classic computational ones, and dynamic equations. Dynamical systems are central to understanding cognition and behavior.

4. A new nomological machine? Computational model discovery

Finding differential equations for your phenomenon is difficult. Many cognitive scientists are interested in studying scientific practice as an interesting cognitive phenomenon itself, such as discovering or inventing explanatory laws (Klahr & Simon, 1999). Wide swaths of cognitive modeling, especially of the probabilistic sort, shed light on how we make generalizations from data (Chater, Tenenbaum, & Yuille, 2006; Tenenbaum, Kemp, Griffiths, & Goodman, 2011).

In the past few decades, some researchers have come up with data-driven models of scientific discovery. For example, Langley and colleagues (Langley, 1981; Langley, Sanchez, Todorovski, & Dzeroski, 2002) have developed probabilistic models that simulate aspects of human scientific discovery, specifically inducing general principles that underlie observed data. Crutchfield, Shalizi and others have developed an entire computational framework that operates over raw time series data to generate a kind of "abstract" theory for these data (Crutchfield, 1994, 2011; Shalizi & Crutchfield, 2001). A recent review by Sozou and colleagues (2017; also see our review in Dale & Bhat, 2018) offers an impressive bird's-eye view of these frameworks, from rule-based learning systems to evolutionary computation. From their review, you get the sense that interest in this approach is accelerating, perhaps alongside a still rapidly growing enthusiasm with data science and machine learning.

Let's consider a simple example of one of these frameworks. This example is from the framework that a collaborator and I have recently extended (Dale & Bhat, 2018). We also offer free tools that can be downloaded, with detailed examples.[4] This example and the framework it is built on are designed specifically to identify differential and difference equations. It overcomes some limiting assumptions of many prior frameworks. For example, in some of the early scientific discovery work, researchers had to specify the form of the underlying law (e.g., constraints on the equations). A recent approach by Brunton and colleagues (2016) shows that methods for such computational model discovery need not impose strict restrictions. They develop a framework they call *sparse identification of nonlinear dynamics* (SINDy). The approach is surprisingly simple. Here's the sequence of steps to run SINDy on your own data:

1) Obtain time series measurements from a behavior or behaviors of interest ($\mathbf{X}$).
2) Numerically differentiate these measurements to obtain an approximation of the first-order derivative of these variables ($d\mathbf{X}/dt$).
3) Find the regression coefficients ($\mathbf{B}$) that relate these derivatives ($d\mathbf{X}/dt$) to your original measurements and any transformations of them ($f(\mathbf{X})$).
4) Now, set to 0 any regression coefficient in $\mathbf{B}$ that does not achieve some threshold (set by the researcher), and redo your fit.

---

[4] https://github.com/racdale/sindyr

5) Once $\mathbf{B}$ stops losing coefficients to this thresholding, full stop and reconstruct the equations.

This is sparse regression *for dynamic equation discovery*—it finds equations that best fit the time-relative derivatives and a transformation of original measurements. SINDy can work even when you do a very large and unconstrained search through different transformations. For example, imagine having just one state variable (x) and fitting the following regression model through SINDy:

$$\frac{dx}{dt} = b_0 x + b_1 x^2 + \cdots + b_i \sin(x) + b_{i+1} sin^2(x) + \cdots$$

These x-derived terms are the "features" generated by transforming the observed data $\mathbf{X}$ (here, just one measurement, x). The coefficients are b_j, and populate a matrix $\mathbf{B}$, which is "sparsified" by only keeping the terms that have some level of impact on the derivative. In principle, we could have a very high-dimensional feature space, and SINDy will work to ferret out the most important variables. Once you've got lots of b_j set to 0, what you have left is a differential equation that describes your system. In the next section, I illustrate this with our very simple model, the logistic map.

5. Finding the equation for the logistic map

Let's save the logistic map's state variable over a series of values of its control parameter a. Such a time series is illustrated in Fig. 1 below. We take these values of x, and transform them so that we obtain a feature space, shown in the right panel in Fig. 1.

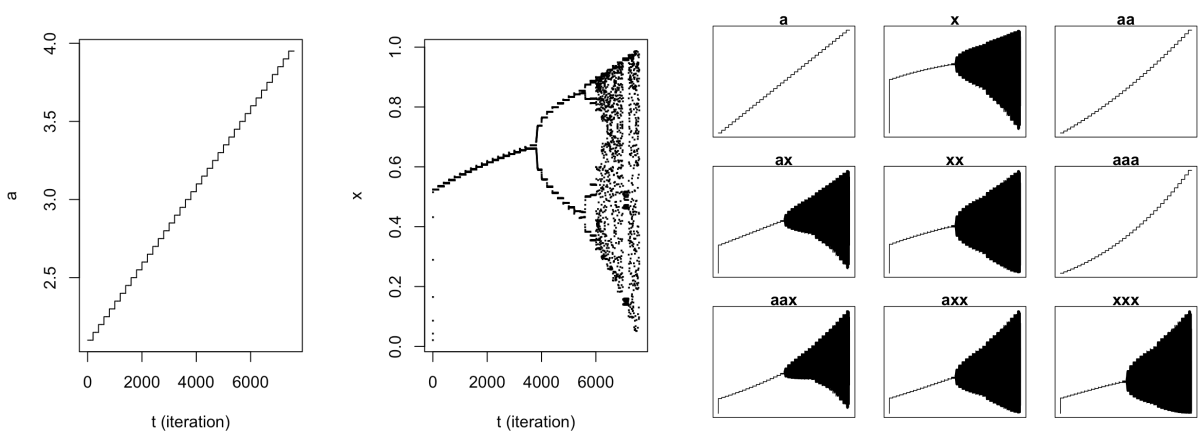

Figure 1: Data used as $\mathbf{X}$ for this demonstration, drawn from the familiar logistic map. Figure can be found in the worked up example on GitHub: bit.ly/2SYNJ2m. The leftmost panel shows values for the control parameter. The middle panel is the familiar bifurcation diagram for the state variable. On the right, these are the features that can be generated by these system measurements.

After carrying out the procedures set in Step 1-5 of the prior section, we obtain a sparsified matrix of coefficients $\mathbf{B}$ that relates the derivatives and variables. This is shown in Figure 2 below. When we extract the non-zero coefficients and specify the derived equation, we obtain the logistic map. As Brunton et al. (2016) show and we replicate (Dale & Bhat, 2018), this works even under some noise. In short, we have reconstructed a dynamic system, however simple, directly from observed data.

B

```
##                      a         x
## (Intercept) 0.000000  0.000000
## a           1.000077  0.000000
## x           0.000000  0.000000
## a:a         0.000000  0.000000
## a:x         0.000000  1.000136
## x:x         0.000000  0.000000
## a:a:a       0.000000  0.000000
## a:a:x       0.000000  0.000000
## a:x:x       0.000000 -1.000157
## x:x:x       0.000000  0.000000
```

$$\longrightarrow \quad a = 1.000a$$
$$x = 1.000ax - 1.000axx$$
$$\downarrow \textit{approx.}$$
$$x = ax(1 - x)$$

Figure 2: On the left, an illustration of the output of the algorithm. SINDy does not work perfectly, of course; its precision is to a number of decimal places. When we distil this output into the remaining coefficients, on the right, we can then factor and recover the logistic map from data.

6. Applications and limitations of SINDy

You may be asking, dear reader, "well that's a neat gizmo Rick, but big deal? What can you do with it?" In our recent paper exploring SINDy (Dale & Bhat, 2018), we showcase other examples of this reconstruction. We reconstruct the Lorenz equations, and a simplified variant of the Haken-Kelso-Bunz model for bistable decisions (Tuller, Case, Ding, & Kelso, 1994). We show how increasing noise does ultimately disrupt SINDy, though SINDy can reconstruct equations even in the presence of some noise. We also explore how SINDy may be useful for the study of social systems. If we obtain measurements from *two* people behaving, we can have terms in our equations that represent multiplicative interactions—mutual nonlinear influence between two systems. We showed that it is possible to estimate interactivity in a coupled logistic map model of communication introduced by Buder (1991). In our paper, we share extensive visualizations of all this, and if the reader is interested, the technical detail is also worked out thoroughly there.

In that same paper, we show some of SINDy's limitations. For example, when fitting the Buder (1991) model, the large number of features generated from $\mathbf{X}$ (e.g., higher-order polynomial and interactive terms) leads to classic issues with regression. Namely, with too many variables, a complicated model "soaks up" variance of the derivatives, and important variables can be lost to the thresholding procedure. In general though, this data-driven equation discovery approach finds the right *complexity* of the underlying system. We can pose such questions as: How nonlinearly interacting are a system's variables? How many terms underlie its control? How mutually constraining are subsets of variables?

7. Concluding notes: nomological machines and complexity

Let's return to the topic of nomological machines and the universality of discovered laws. In our paper on SINDy, we describe how the underlying equations describing psychological systems are unlikely to be stable. Contexts can change rapidly. Human behavior patterns may change sharply in response. In my own area of language and communication, we know that the properties of an interaction partner can have impacts on communicative behaviors. For example, speaking to someone new to your research area poses challenges distinct from speaking to a fellow expert. Interactive contexts, such as formality, may alter behavior sharply. Our choice of words may change, the manner of our speech, and so on.

Imagine collecting the dynamics of such varied behaviors, and deploying SINDy on these data. We would not expect there to be one single set of equations underneath these behaviors for the entire data set. Yet we could still use SINDy to say something potentially interesting. We could identify the discovered equations over a set of snapshots, and track how the equations describing the behavior are changing. This could be expressed in terms of the complexity of each model, a kind of "dynamic

Kolmogorov complexity." We could ask questions such as: How does the complexity of governing equations change in response to distinct contexts? What is the characteristic range of nonlinear interactivity among our variables, and how stable are these interactions as communication unfolds? These data-science inspired methods may therefore afford new metrics. These metrics could be expressed explicitly in mathematical formalisms that may link intriguingly to other dynamic models. Our simulations in Dale and Bhat (2018) show the promise of SINDy, the next step is wrestling with noisy natural time series.

In Cartwright's 1999 book, she notes that a great challenge we face is "to develop methodologies, not for life in the laboratory where conditions can be set as one likes, but methodologies for life in the messy world that we inevitably inhabit." (p. 18) The computational discovery approach I briefly reviewed here may be suited to help with this problem. The statistical control approach of SEM and so on is obviously well suited for it. But if we are concerned with dynamics, and characterizations of dynamic control for the complex systems we study, then tools like SINDy may be a handy new way of rendering regularities, as a new kind of nomological machine. My collaborator Harish Bhat has taken important next steps to improve this regression-based approach. He has devised a method that better integrates stochasticity in the discovery process, integrating terms for both deterministic control, but also probabilistic "fields" that involve drift. He has been finding that this improves recovery. Some recent summaries of his extended approach are available (see Bhat & Madushani, 2016; Bhat, Madushani, & Rawat, 2016).

In a prior newsletter article (January, 2015), David Schuldberg warns about "boosterism" on various matters, such as "systems boosterism," or "big data boosterism." The concern is instructive. I am not endorsing this computational approach as a future panacea of any kind; indeed, at the moment, the systems reconstructed have been rather minimal, of known systems. This concern reminds me of a prominent series of articles in *Wired* sometime ago, which serve as a related warning.[5] In 2008 the tech magazine published a series of articles on the "End of Theory," how "Big Data" are going to overwhelm theorists simply by generating new computational oracles that will tell us everything we need to know (even if we don't exactly know what's going on inside them). This series didn't age well, it would seem.

So big data needs big theory. Applied machine learning often involves careful intervention of a thoughtful researcher. In any case, one reason SINDy is enticing is that it is relatively transparent. You can see the equations. You can touch them, plug them in, see how they change. The method itself has the potential for creating an intriguing kind of dynamic with the investigator. It is a new probe, a new tool, to put on the shelf. I welcome readers to give it a try.

References

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science*, *22*(3), 295–318.

Bhat, H. S., & Madushani, R. (2016). Density tracking by quadrature for stochastic differential equations. *ArXiv Preprint ArXiv:1610.09572*.

Bhat, H. S., Madushani, R. W. M. A., & Rawat, S. (2016). Rdtq: Density Tracking by Quadrature (Version 0.1). Retrieved from https://CRAN.R-project.org/package=Rdtq

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932–3937.

Buder, E. H. (1991). A Nonlinear Dynamic Model of Social Interaction. *Communication Research*, *18*(2), 174–198. https://doi.org/10.1177/009365091018002003

Cartwright, N. (1999). *The dappled world*. Cambridge, UK: Cambridge University Press.

Chater, N., & Brown, G. D. (2008). From universal laws of cognition to specific cognitive models. *Cognitive Science*, *32*(1), 36–67.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287–291.

[5] https://bit.ly/2cBN4OE

Crutchfield, J. P. (1994). The calculi of emergence: computation, dynamics and induction. *Physica D: Nonlinear Phenomena*, *75*(1–3), 11–54.

Crutchfield, J. P. (2011). Between order and chaos. *Nature Physics*, *8*(1), 17–24.

Dale, R., & Bhat, H. S. (2018). Equations of mind: Data science for inferring nonlinear dynamics of socio-cognitive systems. *Cognitive Systems Research*, *52*, 275–290.

Darlington, R. B., & Hayes, A. F. (2016). *Regression Analysis and Linear Models: Concepts, Applications, and Implementation*. Guilford Publications.

Eliasmith, C. (1996). The third contender: A critical examination of the dynamicist theory of cognition. *Philosophical Psychology*, *9*(4), 441–463.

Gottman, J. M., Murray, J. D., Swanson, C., Tyson, R., & Swanson, K. R. (2005). *The mathematics of marriage: Dynamic nonlinear models*. The MIT Press.

Kaplan, D. M., & Bechtel, W. (2011). Dynamical models: An alternative or complement to mechanistic explanations? *Topics in Cognitive Science*, *3*(2), 438–444.

Keith, T. Z. (2005). *Multiple Regression and Beyond* (1st ed.). Allyn & Bacon.

Kellert, S. H., Longino, H. E., & Waters, C. K. (2006). *Scientific pluralism*. Minneapolis, MN: University of Minnesota Press.

Klahr, D., & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, *125*(5), 524–543.

Langley, P. (1981). Data-driven discovery of physical laws. *Cognitive Science*, *5*(1), 31–54.

Langley, P., Sanchez, J. N., Todorovski, L., & Dzeroski, S. (2002). Inducing process models from continuous data. *Proceedings of the Nineteenth International Conference on Machine Learning*, 347–354.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. CRC Press Boca Raton, FL.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, 1. https://doi.org/10.1038/s41562-018-0522-1

Riley, M. A., & Van Orden, G. C. (2005). *Tutorials in contemporary nonlinear methods for the behavioral sciences*. National Science Foundation.

Shalizi, C. R., & Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, *104*(3), 817–879.

Simon, H. A. (1992). What is an "explanation" of behavior? *Psychological Science*, *3*(3), 150–161.

Sozou, P. D., Lane, P. C., Addis, M., & Gobet, F. (2017). Computational scientific discovery. In *Springer Handbook of Model-Based Science* (pp. 719–734). Springer.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Tuller, B., Case, P., Ding, M., & Kelso, J. A. S. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 3–16.

Van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, *92*(7), 345–381.