Cambridge University Press
978-1-107-10999-5 — Computational Modeling of Cognition and Behavior
Simon Farrell , Stephan Lewandowsky
Excerpt
More Information

1 Introduction

This introductory chapter pursues three principal goals. First, we show that computational modeling is essential to ensure progress in cognitive science. Second, we provide an introduction to the abstract idea of modeling and its many and varied applications. Third, we survey some of the issues involved in the interpretation of model output, including in particular how models can help constrain scientists' own thinking.

1.1 Models and Theories in Science

Cognitive scientists seek to understand how the mind works. That is, we want to *describe* and *predict* people's behavior, and we ultimately wish to *explain* it, in the same way that physicists predict the motion of an apple that is dislodged from its tree (and can accurately describe its downward path) and explain its trajectory (by appealing to gravity). For example, if you forget someone's name when you are distracted seconds after being introduced to her, we would like to know what cognitive process is responsible for this failure. Was it lack of attention? Forgetting over time? Can we know ahead of time whether or not you will remember that person's name?

The central thesis of this book is that to answer questions such as these, cognitive scientists must rely on quantitative mathematical models, just like physicists who research gravity. We suggest that to expand our knowledge of the human mind, consideration of the data and verbal theorizing are insufficient on their own.

This thesis is best illustrated by considering something that is (just a little) simpler and more readily understood than the mind. Have a look at the data shown in Figure 1.1, which represent the position of planets in the night sky over time.

How might one describe this peculiar pattern of motion? How would you explain it? The strange loops in the otherwise consistently curvilinear paths describe the famous "retrograde motion" of the planets – that is, their propensity to suddenly reverse direction (viewed against the fixed background of stars) for some time before resuming their initial path. What explains retrograde motion? It took more than a thousand years for a satisfactory answer to that question to become available, when Copernicus replaced the geocentric Ptolemaic system with a heliocentric model. Today, we know that retrograde motion arises from the fact that the planets travel at different speeds along their orbits; hence, as Earth "overtakes" Mars, for example, the red planet appears to reverse direction as it falls behind the speeding Earth.

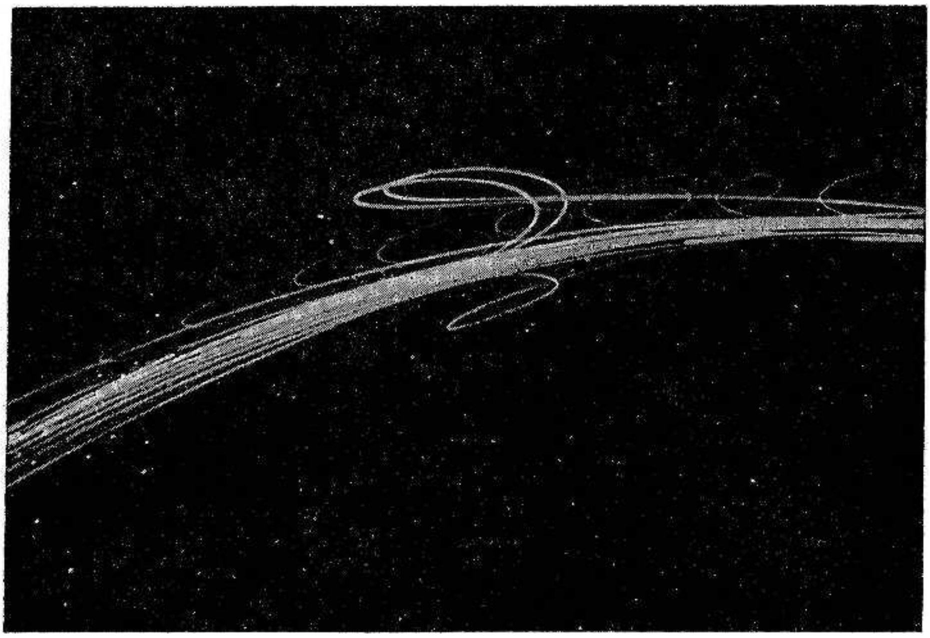

Figure 1.1 An example of data that defy easy description and explanation without a quantitative model. Figure taken from Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, Vol. 336, No. 1604, A Symposium on Planetary Science in Celebration of the Quincentenary of Nicolaus Copernicus 1473–1543. (Jan. 15, 1974), pp. 105–114. Reprinted with permission.

This example permits several conclusions that will be relevant throughout the remainder of this book. First, the pattern of data shown in Figure 1.1 defies description and explanation unless one has a *model* of the underlying process. It is only with the aid of a model that one can describe and explain planetary motion, even at a verbal level (readers who doubt this conclusion may wish to invite friends or colleagues to make sense of the data without knowing their source).

Second, any model that explains the data is itself unobservable. That is, although the Copernican model is readily communicated and represented (so readily, in fact, that we decided to omit the standard figure showing a set of concentric circles), it cannot be directly observed. Instead, the model is an abstract explanatory device that "exists" primarily in the minds of the people who use it to describe, predict, and explain the data.

Third, there nearly always are *several* possible models that can explain a given data set. This point is worth exploring in a bit more detail. The overwhelming success of the heliocentric model often obscures the fact that, at the time of Copernicus' discovery, there existed a fairly successful alternative, namely the geocentric model of Ptolemy shown in Figure 1.2. The model explained retrograde motion by postulating that while orbiting around the Earth, the planets also circle around a point along their orbit. On the additional assumption that the Earth is slightly offset from the center of the planets' orbit, this model provides a reasonable account of the data, limiting the positional

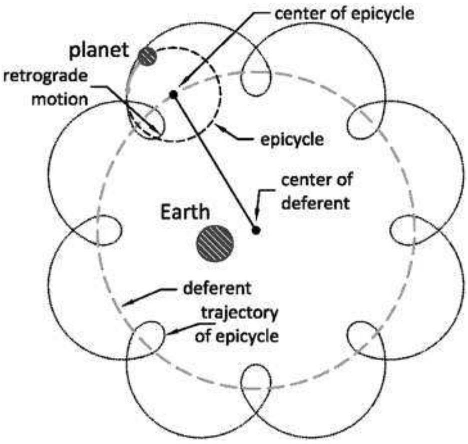

Figure 1.2 The geocentric model of the solar system developed by Ptolemy. It was the predominant model for some 1,300 years.

discrepancies between predicted and actual locations of, say, Mars to about 1° (Hoyle, 1974). Why, then, did the heliocentric model so rapidly and thoroughly replace the Ptolemaic system?[1]

The answer to this question is quite fascinating and requires that we move toward a *quantitative* level of modeling.

Conventional wisdom holds that the Copernican model replaced geocentric notions of the solar system because it provided a better account of the data. But what does "better" mean? Surely it means that the Copernican system predicted the motion of planets with less quantitative error – that is, less than the 1° error for Mars just mentioned – than its Ptolemaic counterpart? Intriguingly, this conventional wisdom is only partially correct. Yes, the Copernican model predicted the planets' motion in latitude better than the Ptolemaic theory, but this difference was slight compared to the overall success of both models in predicting motion in longitude (Hoyle, 1974). What gave Copernicus the edge, then, was not "goodness-of-fit" alone[2] but also the intrinsic elegance and simplicity of his model: compare the Copernican account by a set of concentric circles with the complexity of Figure 1.2, which only describes the motion of a single planet.

There is an important lesson to be drawn from this fact: The choice among competing models – and remember, there are always several to choose from – inevitably involves an *intellectual judgment* in addition to quantitative examination. Of course, the quantitative performance of a model is at least as important as are its intellectual attributes. Copernicus would not be commemorated today had the predictions of his model been *inferior* to those of Ptolemy; it was only because the two competing models were on an essentially

[1] Lest one think that the heliocentric and geocentric models exhaust all possible views of the solar system, it is worth clarifying that there is an infinite number of equivalent models that can adequately capture planetary motion because relative motion can be described with respect to *any* possible vantage point.

[2] "Goodness-of-fit" is a term for the degree of quantitative error between a model's predictions and the data; this important term and many others are discussed in detail in Chapter 2.

equal quantitative footing that other intellectual judgments, such as a preference for simplicity over complexity, came into play.

If the Ptolemaic and Copernican models were quantitatively comparable, why do we use them to illustrate our central thesis that a purely verbal level of explanation for natural phenomena is insufficient and that all sciences must seek explanations at a quantitative level? The answer is contained in the crucial modification to the heliocentric model offered by Johannes Kepler nearly a century later. Kepler replaced the circular orbits in the Copernican model by ellipses with differing eccentricities (or "egg-shapedness") for the various planets. By this straightforward mathematical modification, Kepler achieved a virtually perfect fit of the heliocentric model, with near-zero quantitative error. There no longer was any appreciable quantitative discrepancy between the model's predictions and the observed paths of planets. Kepler's model has remained in force essentially unchanged for more than four centuries.

The acceptance of Kepler's model permits two related conclusions, one that is obvious and one that is equally important but perhaps less obvious. First, if two models are equally simple and elegant (or nearly so), the one that provides the better quantitative account will be preferred. Second, the predictions of the Copernican and Keplerian models cannot be differentiated by verbal interpretation alone. Both models explain retrograde motion by the fact that Earth "overtakes" some planets during its orbit, and the differentiating feature of the two models – whether orbits are presumed to be circular or elliptical – does not entail any differences in predictions that can be appreciated by purely verbal analysis. That is, although one can talk about circles and ellipses (e.g. "one is round, the other one egg-shaped"), those verbalizations cannot be turned into testable predictions. Remember, Kepler reduced the error for Mars from 1° to virtually zero, and we challenge you to achieve this by verbal means alone.

Let us summarize the points we have made so far:

1. Data never speak for themselves but require a model to be understood and to be explained.
2. Verbal theorizing alone ultimately cannot substitute for quantitative analysis.
3. There are always several alternative models that vie for explanation of data and we must select among them.
4. Model selection rests on both quantitative evaluation and intellectual and scholarly judgment.

All of these points will be explored in the remainder of this book. We next turn our attention from the night sky to the inner workings of our mind.

1.2 Quantitative Modeling in Cognition

1.2.1 Models and Data

Let's try this again: Have a look at the data in Figure 1.3. Does it remind you of planetary motion? Probably not, but it should be at least equally challenging to discern

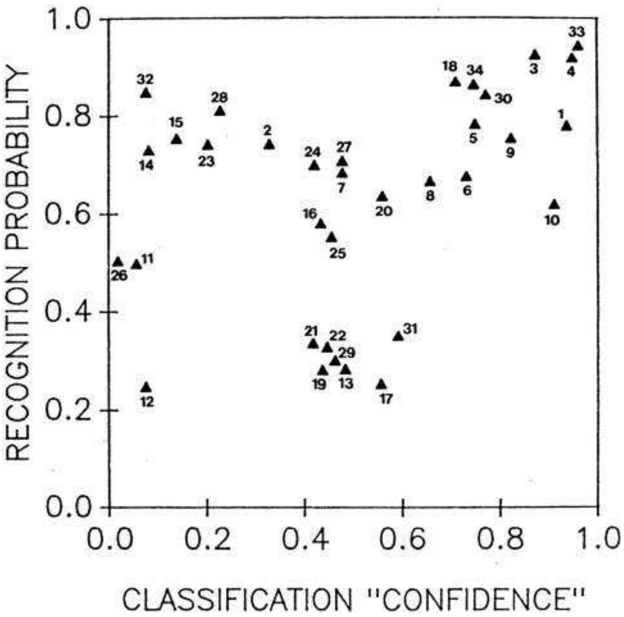

Figure 1.3 Observed recognition scores as a function of observed classification confidence for the same stimuli (each number identifies a unique stimulus). See text for details. Figure reprinted from Nosofsky, R. M., Tests of an exemplar mode for relating perceptual classification and recognition memory, *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 3–27, 1991, published by the American Psychological Association, reprinted with permission.

a meaningful pattern in this case at it was in the example from astronomy. Perhaps the pattern will become recognizable if we tell you about the experiment conducted by Nosofsky (1991) from which these data are taken. In that experiment, people were trained to classify a small set of cartoon faces into two arbitrary categories. We might call the two categories the Campbells and the MacDonalds, and their members might differ on a set of facial features such as length of nose and eye separation.

On a subsequent transfer test, people were presented with a larger set of faces, including those used at training plus a number of new ones. For each face, people had to make two decisions. The first decision was which category the face belonged to and the confidence of that decision (called "classification" in the figure, shown on the X-axis). The second decision was whether or not the face had been shown during training ("recognition" on the Y-axis). Each data point in the figure, then, represents those two responses, averaged across participants, for a given face (identified by ID number, which can be safely ignored). The correlation between those two measures was found to be $r = 0.36$.

Before we move on, see if you can draw some conclusions from the pattern in Figure 1.3. Do you think that the two tasks have much to do with each other? Or would you think that classification and recognition are largely unrelated and that knowledge of one response would tell you very little about what response to expect on the other

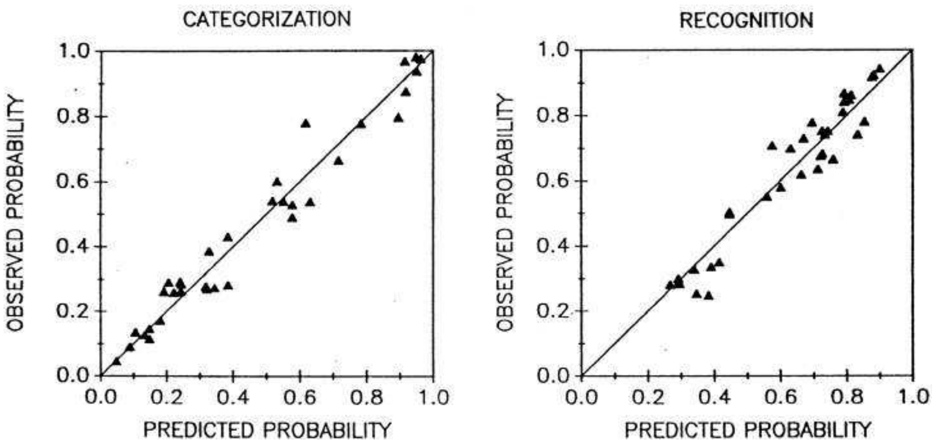

Figure 1.4 Observed and predicted classification (left panel) and recognition (right panel). Predictions are provided by the GCM; see text for details. Perfect prediction is represented by the diagonal lines. Figure reprinted from Nosofsky, R. M., Tests of an exemplar mode for relating perceptual classification and recognition memory, *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 3–27, 1991, published by the American Psychological Association, reprinted with permission.

task? After all, if $r = 0.36$, then knowledge of one response reduces uncertainty about the other one by only 13%, leaving a full 87% unexplained, right?

Wrong. There is at least one quantitative cognitive model (called the GCM and described a little later), which can relate those two types of responses with considerable certainty. This is shown in Figure 1.4, which separates classification and recognition judgments into two separate panels, each showing the relationship between observed responses (on the Y-axis) and the predictions of the GCM (X-axis). To clarify, each point in Figure 1.3 is shown twice in Figure 1.4, once in each panel, and in each instance it is plotted as a function of the *predicted* response obtained from the model.

The precision of predictions in each panel is remarkable: If the model's predictions were 100% perfect, then all points would fall on the diagonal. They do not, but they come close (accounting for 96% and 92% of the variance in classification and recognition, respectively). The fact that these accurate predictions were provided by the same model tells us that classification and recognition can be understood and related to each other within a common psychological theory. Thus, notwithstanding the low correlation between the two measures, there is an underlying model that explains how both tasks are related and permits accurate prediction of one response from knowledge of the other. This model will be presented in detail later in this chapter (Section 1.2.3); for now, it suffices to acknowledge that the model relies on the comparison between each test stimulus and all previously encountered exemplars in memory.

The two figures enforce a compelling conclusion: "The initial scatterplot . . . revealed little relation between classification and recognition performance. At that limited level of analysis, one might have concluded that there was little in common between the fundamental processes of classification and recognition. Under the guidance of the formal model, however, a unified account of these processes is achieved" (Nosofsky, 1991, p. 9). Exactly paralleling the developments in 16*th*-century astronomy, data in contemporary psychology are ultimately only fully interpretable with the aid of a quantitative model. We can thus reiterate our first two conclusions from above and confirm that they apply to cognitive psychology as well, namely that *data never speak for themselves, but require a model to be understood and to be explained*, and that *verbal theorizing alone cannot substitute for quantitative analysis*. But what about the remaining earlier conclusions concerning model selection?

Nosofsky's (1991) modeling included a comparison between his favored exemplar model, whose predictions are shown in Figure 1.4, and an alternative "prototype" model. The details of the two models are not relevant here; it suffices to note that the prototype model compares a test stimulus to the *average* of all previously encountered exemplars, whereas the exemplar model performs the comparison one-by-one between the test stimulus and each exemplar and sums the result.[3] Nosofsky found that the prototype model provided a less satisfactory account of the data, explaining only 92% and 87% of the classification and recognition variance, respectively, or about 5% less than the exemplar model. Hence, the earlier conclusions about model selection apply in this instance as well: There were several alternative models, and the choice between them was based on clear quantitative criteria.

Thus far, we initiated our discussion with the data and we then – poof! – revealed a quantitative model that spectacularly turned an empirical mystery or mess into theoretical currency. In many circumstances, this is what modelers might do: they are confronted with new data but have an existing model at hand, and they wish to examine how well the model can handle the data. In other circumstances, however, researchers might invert this process and begin with an idea "from scratch." That is, you might believe that some psychological process is worthy of exploration and empirical test. The next chapter provides an in-depth example of how one might proceed under those circumstances. Before we get into those details, however, we briefly describe how the large number of models and mode applications can be differentiated into two broad categories, namely models that simply describe data vs. models that explain the underlying cognitive processes.

1.2.2 Data Description

Knowingly or not, we have all used models to describe or summarize data, and at first glance this appears quite straightforward. For example, we probably would not hesitate to describe the salaries of all 150 members of the Australian House of Representatives by their average because in this case there is little doubt that the mean is the proper "model" of the data (notwithstanding the extra allowances bestowed upon Ministers). Why would we want to "model" the data in this way? Because we are replacing the

[3] Astute readers may wonder how the two could possibly differ. The answer lies in the fact that the similarity rule involved in the comparisons by the exemplar model is non-linear; hence, the summed individual similarities differ from that involving the average. This non-linearity turns out to be crucial to the model's overall power. The fact that subtle matters of arithmetic can have such drastic consequences further reinforces the notion that purely verbal theorizing is of limited value.

data points ($N = 150$ in this instance) with a single estimated "parameter."[4] In this instance, the parameter is the sample mean, and reducing 150 points into one facilitates understanding and efficient communication of the data.

However, we must not become complacent in light of the apparent ease with which we can model data by their average. As a case in point, consider U.S. President Bush's 2003 statement in promotion of his tax cut, that "under this plan, 92 million Americans receive an average tax cut of $1,083." Although this number, strictly speaking, was not incorrect, it arguably did not represent the best model of the proposed tax cut, given that 80% of taxpayers would receive less than this cut, and nearly half (i.e. some 45 million people) would receive less than $100 (Verzani, 2004). The distribution of tax cuts was so skewed (bottom 20% of income earners slated to receive $6 compared to $30,127 for the top 1%) that the median or a trimmed mean would have been the preferable model of the proposed legislation in this instance.

Controversies about the proper model with which to describe data also arise in cognitive science, although fortunately with more transparency than in the political arena. In fact, data description, by itself, can have considerable psychological impact. As a case in point, consider the debate on whether learning of a new skill is best understood as following a "Power Law" or is better described by an exponential improvement (Heathcote et al., 2000). There is no doubt that the benefits from practice accrue in a non-linear fashion: The first time you try your hands at a new skill (for example, creating an Ikebana arrangement), things take seemingly forever (and the output may not be worth writing home about). The second and third time round, you will notice vast improvements, but eventually, after some dozens of trials, chances are that further improvements will be small indeed.

What is the exact functional form of this pervasive empirical regularity? For several decades, the prevailing opinion had been that the effect of practice is best captured by a "Power law" – that is, by the function (shown here in its simplest possible form),

$$RT = N^{-\beta}, \tag{1.1}$$

where RT represents the time to perform the task, N represents the number of learning trials to date, and β is the learning rate. Parameters of models are often represented by Greek letters, and Appendix A lists these in full; in this case, β is the Greek letter Beta. Figure 1.5 shows sample data, taken from Palmeri (1997)'s Experiment 3, with the appropriate best-fitting power function superimposed as a dashed line. Participants judged the numerosity of random dot patterns that contained between 6 and 11 dots. Training extended over several days and each pattern was presented numerous times. The figure shows the training data for one participant and one particular pattern.

Heathcote et al. (2000) argued that the data are better described by an exponential function given by (again in its simplest possible form),

$$RT = e^{-\alpha N}, \tag{1.2}$$

[4] We will provide a detailed definition of what a parameter is in Chapter 2. For now, it suffices to think of a parameter as a number that carries important information and that determines the behavior of the model.

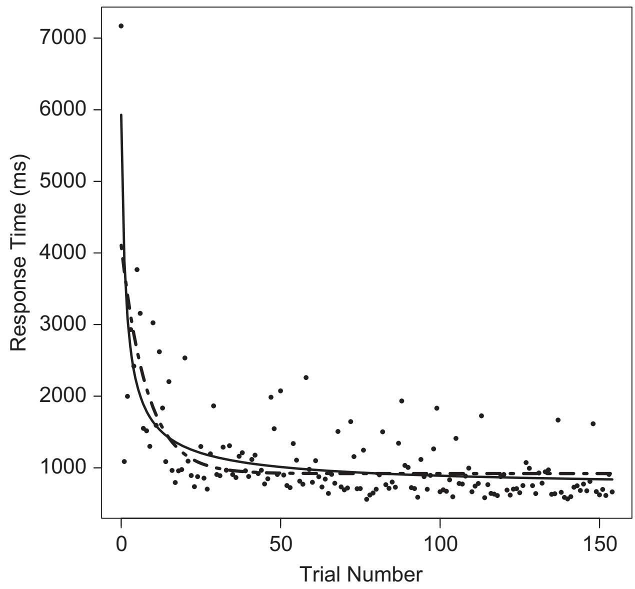

Figure 1.5 Sample power law learning function (solid line) and alternative exponential function (dashed line) fitted to the same data. Data are represented by dots and are taken from Palmeri (1997)'s Experiment 3 (Subject 3, Pattern 13). To fit the data, the power and exponential functions were a bit more complex than described in Equations 1.1 and 1.2 because they additionally contained an asymptote (A) and a multiplier (B). Hence the power function took the form $RT = A_P + B_P \times (N+1)^{-\beta}$ and the exponential function was $RT = A_E + B_E \times e^{-\alpha N}$.

where N is as before and α the learning rate. The best-fitting exponential function is shown by the dashed line in Figure 1.5; you will note that the two competing descriptions or models do not appear to differ much.[5] The power function captures the data well, but so does the exponential function, and there is not much to tell between them: The residual mean-squared deviation (RMSD), which represents the average deviation of the data points from the predicted function, was 482.4 for the Power function compared to 526.9 for the exponential. Thus, in this instance the Power function fits "better" (by providing some 50 ms less error in its predictions than the exponential), but given that RT's range from somewhere less than 1,000 ms to 7 seconds, this difference may not be considered particularly striking.

So, why would this issue be of any import? Granted, we wish to describe the data by the appropriate model, but surely neither of the models in Figure 1.5 misrepresents essential features of the data anywhere near as much as U.S. President Bush did by reporting only the average implication of his proposed tax cut. The answer is that the choice of the correct descriptive model, in this instance, carries important implications about the psychological nature of learning. As shown in detail by Heathcote et al. (2000), the mathematical form of the exponential function necessarily implies that the

[5] For now, we just present those "best-fitting" functions without explaining how they were obtained. We begin the discussion of how to fit models to data in Chapter 3.

learning rate, relative to what remains to be learned, is constant throughout practice. That is, no matter how much practice you have had, learning continues by enhancing your performance by a constant fraction. By contrast, the mathematics of the power function imply that the relative learning rate is slowing down as practice increases. That is, although you continue to show improvements throughout, the rate of learning *decreases* with increasing practice. It follows that the proper characterization of skill acquisition data by a descriptive model, in and of itself, has considerable psychological implications (we do not explore those implications here; see Heathcote et al., 2000, for pointers to the background).

Just to wrap up this example, Heathcote et al. (2000) concluded after re-analyzing a large body of existing data that the exponential function provided a better description of skill acquisition than the hitherto presumed “Power law.” For our purposes, their analysis permits the following conclusions. First, quantitative description of data, by itself, can have considerable psychological implications because it prescribes crucial features of the learning process. Second, the example underscores the importance of model selection that we alluded to earlier; in this instance, one model was chosen over another on the basis of strict quantitative criteria. We revisit this issue in Chapter 10. Third, the fact that Heathcote et al.’s model selection considered the data of individual subjects, rather than the average across participants, identifies a new issue – namely the most appropriate way in which to apply a model to the data from more than one individual – that we consider in Chapter 5.

The selection among competing functions is not limited to the effects of practice. Debates about the correct descriptive function have also figured prominently in the study of forgetting. Does the rate of forgetting differ with the extent of learning? Is the rate of information loss constant over time? Although the complete pattern of results is fairly complex, two conclusions appear warranted (Wixted, 2004a). First, the degree of learning does not affect the rate of forgetting. Hence, irrespective of how much you cram for an exam, you will lose the information at the same rate – but of course this is not an argument against dedicated study; if you learn more, you will also retain more, irrespective of the fact that the rate of loss per unit time remains the same. Second, the rate of forgetting *decelerates* over time. That is, whereas you might lose some 30% of the information on the first day, on the second day the loss may be down to 20%, then 10%, and so on. Again, as in the case of practice, two conclusions are relevant here. First, quantitative comparison among competing descriptive models was required to choose the appropriate function (it is a Power function, or something very close to it). Second, although the shape of the “correct” function has considerable theoretical import because it may imply that memories are “consolidated” over time *after* study (see Wixted, 2004a; 2004b, for a detailed consideration, and see Brown and Lewandowsky, 2010, for a contrary view), the function itself has no psychological content.

The mere description of data can also have psychological implications when the behavior it describes is contrasted to *normative* expectations (Luce, 1995). Normative behavior refers to how people would behave if they conformed to the rules of logic or probability. For example, consider the following syllogism involving two premises (P) and a conclusion (C). P1: All polar bears are animals. P2: Some animals are white.