Fitting the ANCHOR Model to Individual Data: A Case Study in Bayesian Methodology

Alexander A. Petrov (apetrov@andrew.cmu.edu)
Department of Psychology; Carnegie Mellon University
Pittsburgh, PA 15213 USA

Abstract

This paper presents a memory-based model of direct psychophysical scaling. The model is based on an extension of the cognitive architecture ACT-R and uses *anchors* that serve as prototypes for the stimuli classified within each response category. Using the ANCHOR model as a specific example, a general Bayesian framework is introduced. It provides principled methods for making data-based inferences about models of this kind. The internal representations in the model are analyzed as hidden variables that are constructed from the stimuli according to probabilistic representation rules. In turn, the hidden representations produce overt responses via probabilistic performance rules. Incremental learning rules transform the model into a dynamic system. A parameter-fitting algorithm is formulated and tested on experimental data.

Introduction

Parameter search is usually relegated to an appendix in the modeling literature, if it is discussed at all. And this is understandable: the emphasis usually is, as it should be, on describing the particular model and discussing its implications for the phenomena under study. Moreover, the search for suitable parameter values is often done by trial and error or by brute-force methods. This, however, need not necessarily be the case. There are principled ways of making data-based inferences about quantities used in a model, including parameter values. The immediate benefit of such methods is to speed up the model design cycle and expand the power, scope, and detail of the simulations. In addition, they can improve the conceptual understanding of the model by exposing relationships that would otherwise remain buried inside the black box.

This paper outlines a general Bayesian framework for probabilistic models involving internal representations modifiable by learning mechanisms. One such model—ANCHOR (Petrov & Anderson, 2000)—serves as a concrete example. First, the model is described in its own terms and is related to empirical data. It is then analyzed from a Bayesian perspective and a method for making data-based inferences is formulated. Next, the terms are generalized so that the resulting framework can be applied to a wide class of models. Finally, the algorithms are illustrated on experimental data.

The ANCHOR Model

ANCHOR is a memory-based model of psychophysical scaling tasks. It is based on the cognitive architecture ACT-R (Anderson & Lebière, 1998) extended with mechanisms for dealing with continuous (or analog) internal representations called *magnitudes*.

Empirical Phenomena

The model deals with unidimensional sensory continua such as loudness of tones or length of lines. The participants in direct scaling studies are shown stimuli of various physical intensities and are asked to report the corresponding subjective magnitudes. The reports are verbal and typically use some numerical scale. There are several variations of this basic paradigm, only one of which is chosen for the purposes of this paper. The responses in *category rating* are restricted to a fixed set of categories—for example, the numbers from 1 to 9. The participants are instructed to rate the least intense stimuli with 1, the most intense ones with 9, and to space the remaining categories evenly in between.

A number of robust empirical phenomena are well documented in the direct-scaling literature (e.g., see Gescheider, 1988, for review). Only the most salient ones are mentioned here—enough to provide some psychological substance to the abstract considerations in the following sections. Human performance is stochastic and a whole distribution of responses is observed for each stimulus level. The means of these response distributions are described by the so-called *Stevens' law* and their standard deviations by *Ekman's law*. The former states that the mean ratings tend to be a power function of the stimulus: $R=kS^n$ (Stevens, 1957). The latter states that the standard deviation of each response distribution is approximately proportional to its mean (Ekman, 1959).

The immediate stimulus on each trial is the major determinant of the corresponding response but other factors exert robust and measurable effects as well. These include the overall distribution of the stimuli used in the experiment (*context effects*, e.g., Parducci & Wedell, 1986), previous stimuli and responses (*sequential effects*, e.g., Jesteadt, Luce, & Green 1977), and various memory-related factors (*memory effects* and *practice effects*, e.g., Siegel, 1972). The ANCHOR model offers a unified explanation of these phenomena.

ANCHOR: Mechanisms and Equations

The ANCHOR model stands at the intersection of two broad theories—Thurstonian psychophysics (Thurstone, 1927) and the memory theory incorporated in ACT-R (Anderson & Lebière, 1998). From the former it adopts the notion of a subjective continuum of magnitudes. A stimulus with physical intensity S is represented by a magnitude M. Due to the intrinsic stochasticity of the perceptual mechanisms, a whole distribution of magnitudes is associated with each stimulus. It is assumed that these distributions are Gaussian, with mean and variance dependent on S. In keeping with Stevens law ANCHOR postulates that the mean of each distribution is a power function of the corresponding stimulus. Furthermore, in keeping with Ekman's law it assumes that each standard deviation is proportional to the corresponding mean. These are not the only choices that are theoretically possible and consistent with the available empirical data (Norwich & Wong, 1997). Still they are the most straightforward pair and are adopted for this reason. Note also that they imply Weber's law.

The data set used in this paper deals with a particularly convenient sensory modality—physical length. The Stevens exponent for length is very nearly one (Stevens, 1957; Petrov & Anderson, 2000) and hence the power function becomes a simple linear function. This leads to Equation 1, which defines the perceptual transformation of the ANCHOR model. The magnitude M is roughly proportional to the stimulus S but there is also some multiplicative noise. This "perceptual noise" ε_p is a normally distributed random variable with zero mean and standard deviation s_p, which is a free parameter of the model and does not depend on S.

$$M = k\,S\,(1 + \varepsilon_p) \qquad\qquad [1]$$

The task of the participant is to translate the subjective sensation M into an overt response. Three classes of factors constrain the mapping between magnitudes and responses. First, there is an arbitrary component chosen by the experimenter and established by the instructions. This includes the number of response categories and their particular labels. Second, there is a constraint for homomorphism—a pressure to align the intrinsic ordering of the magnitude continuum with the ordering of the response scale. For instance, if M_1 is more intense than M_2, the corresponding responses should also be ordered appropriately. Finally, there is a tacit constraint for consistency—a pressure to use the same response for repeated presentations of the same stimulus throughout the experimental session.

The first and third of these constraints strongly suggest that memory plays an important role in the direct scaling tasks. The same hypothesis is supported by many empirical phenomena and in particular the sequential, memory, and practice effects mentioned earlier (Siegel, 1972; Petrov & Anderson, 2000). This leads to the centerpiece of the model—the *anchor*.

An anchor is an association between a magnitude and a response. There is one anchor for each response category. The anchor magnitude represents the prototypical member of this category. The collection of all anchors thus defines a mapping between the magnitude continuum and the response scale. When a stimulus is presented for rating, its corresponding magnitude M serves as a memory cue. The anchors then compete to match that target and one of them is retrieved. An alternative but equivalent conceptualization is that a pattern completion process fills in the response label given the magnitude. Either way, the final outcome is described by Equations 2 and 3.

$$G_i = -\left| M - \overline{A}_i \right| + h\,B_i \qquad\qquad [2]$$

$$P(A_i) = \frac{\exp(G_i / T)}{\sum_j \exp(G_j / T)} \qquad\qquad [3]$$

Anchor retrieval is a stochastic process, just as any act of memory retrieval in the ACT-R architecture. The probability $P(A_i)$ of selecting an anchor depends on two factors: (*i*) the similarity between the anchor magnitude $\overline{A}_i$ and the target magnitude M and (*ii*) the current availability of the anchor measured by its *base-level activation* B_i. The two factors are combined into a "goodness score" G_i (Eq. 2; h is a scaling parameter). The anchors then compete on the basis of their scores through the "softmax" Equation 3. The "temperature" T is a free parameter that controls the degree of non-determinism of the selection process.

Memory retrieval is noisy and is prone to biases B_i. Therefore it is not guaranteed to provide on each trial the anchor that best matches the target magnitude. The verbal protocols of human participants suggest that they are aware of the unreliability of their memory and seem to adopt a correction strategy. A typical report is, "This length looks like a *7*… No, it's too short for a *7*. I'll give it a *6*."

The ANCHOR model implements these corrections in the following way. The magnitude A_i of the anchor retrieved from memory is compared against the target magnitude M. If the discrepancy $D = (M - A_i)$ is less than some cutoff value c (a free parameter of the model), the response associated with the anchor is chosen as the final response on the trial. Otherwise the anchor response is corrected by $+/-1$ or occasionally even $+/-2$ depending on the algebraic difference D (cutoffs $+/-c$ and $+/-3c$, respectively). The anchor magnitude is degraded by "memory noise" ε_a (Eq. 4) analogous to the "perceptual noise" in the target magnitude (Eq. 1). Due to this noise in the inputs, the correction mechanism is stochastic and error-prone too. Nevertheless, it plays a useful and important role in the model. Among other things, it promotes the homomorphism between magnitudes and responses.

$$A_i = \overline{A}_i(1 + e_a) \qquad [4]$$

So, the stimulus has been encoded, matched against anchors, and a response has been produced. Is this the end of the trial? According to the ANCHOR model and the broader memory theory, the answer is no. The cognitive system is plastic (within limits) and each experience seems to leave a mark on it. There are obligatory *learning mechanisms* that update the anchor magnitudes (Eq. 5) and their base-level activations (Eq. 6). Consider the magnitudes first. On each trial, only one of the anchors is updated—the one that corresponds to the actual response. The new magnitude $\overline{A}^{(t+1)}$ of this anchor is pulled slightly toward the target magnitude M (Eq. 5). The parameter a controls the learning rate. In the long run, each anchor magnitude becomes a weighted average of the magnitudes of all stimuli classified in the corresponding response category. Thus the anchors are true prototypes. However, recent stimuli weigh more than earlier ones. This introduces various sequential, transfer, and context effects (Petrov & Anderson, 2000).

$$\overline{A}^{(t+1)} = (1 - a)\,\overline{A}^{(t)} + a\,M^{(t)} \qquad [5]$$

In contrast to the selective update of magnitudes, the base-level learning Equation 6 affects the availability of all anchors. The formula is not transparent and need be discussed only briefly here. It is an approximate and parameter-free version of the base-level learning equation in ACT-R (Anderson & Lebière, 1998, p. 124). The availability B of each anchor reflects the frequency and recency of use of the corresponding response. The formula disregards the detailed history of the anchor; it retains only three critical pieces of information: the lag since the most recent use t_{last}, the total time since the beginning of the experiment t_{life}, and the total number of uses n.

$$B = \ln\left[t_{last}^{-0.5} + \frac{2(n-1)}{\sqrt{t_{last}} + \sqrt{t_{life}}} \right] \qquad [6]$$

Qualitatively, Equation 6 captures three important aspects of memory dynamics: sharp transient boost immediately after use, gradual buildup of strength with frequent use, and gradual decay in the absence of use. In terms of observable behavior, the rapid transient manifests itself in sequential assimilation and the gradual strengthening/decay in non-uniform response distributions (Petrov & Anderson, 2000).

General Framework

ANCHOR is just a specific instance of a whole class of models. It is useful to introduce a general framework for analyzing such models. Such broader view sharpens the understanding of the particular model and makes the discussion relevant to other modeling efforts as well.

It is assumed that time progresses in discrete steps or "trials". The *stimuli* $S^{(t)}$ and *responses* $R^{(t)}$ are the only observable quantities. The model builds and manipulates some *internal representations* $H^{(t)}$ to mediate between the two: $S \otimes H \otimes R$. These representations are hidden variables (or "hypotheses" in Bayesian terms).

The inner structure of the model specifies two kinds of rules: *representation rules* governing the transition $S \otimes H$ and *performance rules* governing the transition $H \otimes R$. Both kinds of rules are probabilistic. Thus on each trial the representation rules define a whole probability distribution over the space of possible representations conditional on the current stimulus $S^{(t)}$. Some sequence of chance events selects one specific realization $H^{(t)}$, which then drives the performance. The performance rules specify a probability distribution over the possible responses. Another chance event determines the final response $R^{(t)}$. Then the next stimulus is presented and the whole cycle repeats.

Each step of the process described so far depends on two kinds of parameters. The *global parameters* q are relatively few in number, apply throughout the model, and remain fixed during the whole session. There are also *local parameters* $I^{(t)}$ associated with the individual representational and/or processing units. Importantly, these local parameters are not static but may be updated on each step according to *learning rules*. This paper deals with the class of models whose learning rules are deterministic and incremental. That is, the updated values $I^{(t)}$ depend only on their previous values $I^{(t-1)}$, the current representations $H^{(t)}$, and on global parameters such as learning rates. In symbols:

$$I^{(t)} = F_L(I^{(t-1)}, H^{(t)}, q) \qquad [7]$$

The learning mechanisms make the model a dynamic system. The performance on a given trial depends not only on the current stimulus but also on the whole history of the computation. In the class of models discussed here, this history is condensed solely in the most recent update of the local parameters $I^{(t-1)}$. In other words the following *Markov property* is assumed: the operation at time t depends only on $S^{(t)}, H^{(t)}, q$ and the most recent values $I^{(t-1)}$. The representation rules can be written as Eq. 8 and the performance rules as Eq. 9. (Eq. 8 can be extended to include residual representations $H^{(t-1)}$ but this possibility is not pursued here.)

$$P(H^{(t)}) = F_R(S^{(t)}, I^{(t-1)}, q) \qquad [8]$$

$$P(R^{(t)}) = F_P(H^{(t)}, I^{(t-1)}, q) \qquad [9]$$

All these abstract notions can be illustrated on the ANCHOR model. It has six global parameters: perceptual noise s_p, weight h, softmax temperature T, anchor noise s_a, correction cutoff c, and magnitude learning rate a. This set of six values is collectively

denoted $\mathbf{q}$ There are two kinds[1] of internal representations: magnitudes M and anchors A. (In this case the symbol A denotes not an anchor magnitude but a discrete value indicating which of the several anchors is retrieved on the particular trial.) Thus, the hidden states (or hypotheses) in the model can be conceptualized as pairs $H=<M,A>$. They are constructed in two steps with probabilities given by two separate representation rules. First, Equation 1 produces a magnitude $M^{(t)}$ given the stimulus $S^{(t)}$. Based on this magnitude, Equations 2 and 3 then select an anchor $A^{(t)}$. This illustrates an important point about the formalism. The hidden variables $H^{(t)}$ (and, for that matter, $S^{(t)}$ and $R^{(t)}$) need not be unitary or constructed in a single step. Various intermediate stages can be involved. Still, it is always possible in principle to consider the Cartesian product of all the components and think in terms of a (joint) distribution conditional on the stimulus (Eq. 8). Two local parameters $\mathbf{I}^{(t-1)}$ are associated with each anchor: its magnitude $\overline{A_i}$ and availability[2] B_i. The corresponding update rules are given by Equations 5 and 6, respectively.

Model Tracing and Parameter Search

The stage is now set to confront the issue of estimating unobservable quantities from experimental data. For concreteness, consider the problem of finding suitable values of the global parameters $\mathbf{q}$ One approach is to treat the model as a black box and use Monte Carlo simulations to evaluate its performance under different parameter settings. Based on these estimates, a search algorithm such as gradient descent can be used to optimize some goal function of interest. For example, if the empirical accuracy is 80%, the optimization goal would be to bring the model accuracy as close as possible to this target value. This approach requires little thought and is easy to implement but has serious drawbacks both on technical and conceptual grounds. Monte Carlo methods are computer-intensive and their estimates are blurred by sample fluctuations. This slows down the search and derails all optimization algorithms that depend on comparisons between close choices for $\mathbf{q}$ Moreover, there is danger of overfitting because the validity of the model is measured by the same criterion that guides the search (e.g., overall accuracy).

This paper advocates another approach. It is to open the black box and work out in detail the various probability distributions underlying the overt behavior. This approach requires careful analytical derivations and a separate computer implementation. The returns on this

intellectual investment, however, can be considerable. The main leverage comes from the following idea: the probabilities are calculated directly from the structural equations of the model instead of being estimated by sampling. This information can then be used for various purposes as discussed below.

The calculations are done on a trial-by-trial basis. Assume the process has been carried up to trial $t–1$ and estimates of the local parameters $\mathbf{I}^{(t-1)}$ are available. The whole calculation is conditional on some global parameter set $\mathbf{q}$ Under these circumstances, when the next stimulus $S^{(t)}$ arrives the representation rules (Eq. 8) allow to calculate the conditional distribution:

$$P(H^{(t)} \mid S^{(t)}, \mathbf{I}^{(t-1)}, \mathbf{q}) \qquad [10]$$

In Bayesian terms, this is the prior distribution over the hypotheses. To implement that, one follows the logic of the model but at any point in which the original program would do a probabilistic choice and commit to a particular alternative, the modified version keeps all branches open and maintains a table of their respective probabilities.

To illustrate, the first step in the "primary" ANCHOR implementation is to generate a random number $\mathbf{e}_i$ and use it to compute the internal magnitude $M^{(t)}$ according to Equation 1. In contrast, the model-tracing version has to represent the whole distribution of potential magnitudes. The implementation (which is written in MATLAB) approximates the Gaussian distribution with an array of 15 discrete bins centered on $kS^{(t)}$ (Eq. 1) and spread out to cover 99% of the density. Thus, the modified program considers 15 alternative values in parallel instead of the single commitment $M^{(t)}$. Each of these values can occur in conjunction with any of the nine possible anchors. This creates a space of 135 hypotheses $H_{ij}=<M_i,A_j>$. The program represents it as a matrix and calculates the probability of each pair from Equations 2 and 3. Note that this requires knowledge of the current anchor magnitudes $\overline{A}_j$ and base-level activations B_j (that is, the local parameters $\mathbf{I}^{(t-1)}$).

Back to general terms, once the prior probabilities of all hypotheses have been calculated (Eq. 10), a straightforward application of the performance rules (Eq. 9) gives the joint distribution of hypotheses and responses:

$$P(R^{(t)} \times H^{(t)} \mid S^{(t)}, \mathbf{I}^{(t-1)}, \mathbf{q}) = \qquad [11]$$
$$= P(R^{(t)} \mid H^{(t)}, \mathbf{I}^{(t-1)}, \mathbf{q}) \cdot P(H^{(t)} \mid S^{(t)}, \mathbf{I}^{(t-1)}, \mathbf{q})$$

This distribution contains a wealth of information. From that point on, the hidden variables introduced at earlier stages begin to be averaged out. This elimination can be engineered in various ways depending on the goals of the modeler. The following subsections outline some of the potential applications.

---

[1] Strictly speaking, the final response must also be represented internally. Due to the one-to-one correspondence with the overt response, however, this representation need not be considered among the hidden variables.

[2] Actually, three values must be maintained for each B_i but they stay hidden behind Equation 6.

Predicting the Responses and Fitting Parameters

The obvious way to proceed from Equation 11 is to average out the unknown internal representations $H^{(t)}$. This produces the (marginal) probabilities of all possible responses $R^{(t)}$:

$$P(R^{(t)} \mid S^{(t)}, \boldsymbol{I}^{(t-1)}, \boldsymbol{q}) \qquad [12]$$

If the model is a good approximation to the actual process generating the observations, Equation 12 should predict the actual response $R^{(t)} = r_k$ with high probability on most trials. The log-likelihood L (Eq. 13) measures the overall goodness of fit on all available data D (the full sequence of stimuli and responses). This quantity can be used to guide the parameter search. Notice that L is a deterministic function of the parameters $\boldsymbol{q}$ and $\boldsymbol{I}^{(0)}$ (given that the data are fixed). Thus, unlike in Monte Carlo methods, gradient algorithms can track the optimum much faster and with high precision.

$$L = \log P(D \mid \boldsymbol{q}, \boldsymbol{I}^{(0)}) = \sum_{t=1}^{N} \log P(R^{(t)} \mid S^{(t)}, \boldsymbol{I}^{(t-1)}, \boldsymbol{q}) \quad [13]$$

It is worth pointing out that under this scheme the parameters are fitted using one criterion (L) while the final evaluation of the model is done by separate criteria (e.g., accuracy, response variability, sequential dependencies, and so forth) This reduces overfitting and discourages "fishing" for results.

Model Tracing. The internal representations that people use are of particular interest in applications such as intelligent tutoring (Corbett, Anderson, & O'Brien, 1995) or cognitive assessment (Martin & VanLehn, 1995). When a cognitive model is available for some task domain, the Bayesian framework advocated here can be used to trace the probable path that a student is following. In such applications, it is of little utility to eliminate the hidden variables in order to predict the next response. On the contrary, one wants to treat the observed response as an additional piece of information and obtain from Equations 11 and 12 posterior estimates about $H^{(t)}$ via Bayes rule:

$$P(H^{(t)} \mid R^{(t)}, S^{(t)}, \boldsymbol{I}^{(t-1)}, \boldsymbol{q}) \qquad [14]$$

Updating the Local Parameters. In addition to being useful in their own right, these posterior estimates (Eq. 14) are necessary to close the tracing cycle. The local parameters $\boldsymbol{I}^{(t-1)}$ must be updated at the end of trial t, so that the new values $\boldsymbol{I}^{(t)}$ are available on trial $t+1$. The learning rules of the model (Eq. 7) define this transition but they involve the specific internal representation H_i. It is impossible to know which of the many potential structures $H^{(t)}$ has actually been realized on that particular trial. The most that can be calculated are the posterior probabilities (Eq. 14). The cleanest option at this point would be to consider all possible updates $\boldsymbol{I}_i$

and propagate them in parallel through all subsequent trials keeping track of the respective probabilities. Unfortunately, this scheme is computationally intractable. Some approximation must be used instead.

There are several possibilities. One is to sample the posterior—to single out one value H_i at random according to Equation 14 and use it in $\boldsymbol{I}^{(t)}$. This approach, however, introduces randomness into the estimation process. The log-likelihoods calculated on this basis (Eq. 13) are no longer deterministic.

Another possibility is to single out the maximum *a posteriori* (MAP) hypothesis and use it in the update rule (Eq. 7). This approach restores the determinism of the estimation process and has certain intuitive appeal. It, however, systematically overestimates the coupling between the internal representations $H^{(t)}$ and the overt responses $R^{(t)}$ (cf. Eq. 9).

There is an important and widespread special case in which it is meaningful to combine alternative parameter estimates $\boldsymbol{I}_i$ into weighted linear combinations. In such cases, a "mean field" approximation is available. The idea is to take all alternative H_i, calculate the corresponding updates $\boldsymbol{I}_i$, and combine them into a single aggregate $\boldsymbol{I}^{(t)}$ that is passed on to the next trial. The posterior probabilities (Eq. 14) serve as weighting coefficients. (A related strategy using the prior probabilities (Eq. 10) can be preferable in some cases. The priors factor in the stimulus $S^{(t)}$ but not the response $R^{(t)}$.)

The ANCHOR model has two learning rules and thus illustrates the problem. The update of the base-level activations B presents little difficulty because no hidden variables are involved (Eq. 6). It is clear which anchor is strengthened on each trial—the one corresponding to the overt response. In contrast, the update of the anchor magnitudes is more problematic because Equation 5 involves the unobservable target magnitude M. As the magnitudes are additive, the mean field approximation is applicable. Recall that the hypotheses in the ANCHOR case are pairs $H_{ij} = \langle M_i, A_j \rangle$. To produce the posterior distribution of the magnitudes M_i alone, the anchor indicator variables must be averaged out from the joint posterior (Eq. 14). This leaves 15 probabilities $P(M_i)$ corresponding to the 15 target magnitude bins. The mean-field version of Equation 5 thus becomes:

$$\overline{A}^{(t+1)} = (1 - \boldsymbol{a})\, \overline{A}^{(t)} + \boldsymbol{a} \sum M_i^{(t)} P(M_i^{(t)}) \qquad [15]$$

Empirical Test

This final section applies the global parameter search algorithm to data from a category rating experiment (Petrov & Anderson, 2000). The original paper reports group-level fits of the ANCHOR model. The present framework allows us to extend the analysis to the level of individual participants.

The stimuli for the experiment were pairs of dots presented on a monitor. The task of the 24 participants

was to rate the distance between the two dots using the numbers from 1 to 9. Each session involved 450 trials.

The results replicated all classical category-rating phenomena that fell within the scope of the experiment, including sequential effects of various kinds and the linearity of the psychophysical scale for length. A battery of statistical measures designed to quantify these effects were calculated from the data of each participant. The ANCHOR model was then run on the same 24 stimulus sequences and the same statistics were calculated. The fits were very good on all dimensions (Petrov & Anderson, 2000, Table 1). The mean of each statistic calculated from the model data fell very close to the corresponding empirical mean. The variability in the empirical sample, however, was somewhat greater (by 40% on average) than that of the model.

The latter result is not surprising. The empirical data contain both within- and between-subject variability. All ANCHOR runs, on the other hand, used the same set of global parameters. Thus, the model data reflected only the inherent stochasticity of the mechanisms.

A better way to confront the model with the empirical data is to have an individual parameter set q_p for each participant p. The algorithms described in this paper make this easy to do. Each stimulus-response sequence D_p generates a log-likelihood "landscape" (Eq. 13) over the space of global parameters q. A general-purpose optimizer (the *fmincon* function in MATLAB) can then be used to find the maximum-likelihood estimator q_p.

We are now ready to derive individualized predictions. For concreteness, suppose we are interested in the overall accuracy (measured by the stimulus-response correlation R^2) and the shape of the response distribution (measured by its standard deviation SD). We fix a parameter set q_p, run the model 10 times to generate 10 response sequences, and calculate R^2 and SD for each. After averaging over the 10 runs, we obtain a single prediction for R^2_p and one for SD_p.

This procedure yields a distribution of 24 predicted R^2 values—one for each participant—and similarly for the SDs. On the other hand, the corresponding empirical distributions are also available. The model is successful to the extent in which the predicted distributions are indistinguishable from their observed counterparts.
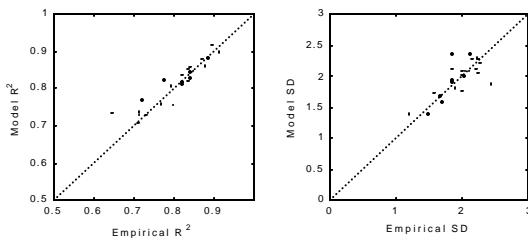

Figure 1: Correlations between model data and empirical data. Left panel: stimulus-response correlations. Right panel: response standard deviations.

Figure 1 illustrates the distribution overlap for our data. The correlation between the 24 empirical R^2 and the corresponding predicted R^2 is 0.93. The correlation among the standard deviations is 0.74. These results clearly indicate that the ANCHOR model is able to reproduce the data well and that the Bayesian algorithms are able to find appropriate parameter sets.

Acknowledgments

References

Anderson, J. R. & Lebière, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Corbett, A., Anderson, J. R., & O'Brien, A. (1995). Student modeling in the ACT programming tutor. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ekman, G. (1959). Weber's law and related functions. *The Journal of Psychology, 47*, 343-352.

Gescheider, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology, 39*, 169-200.

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3* (1), 92-104.

Martin, J. & VanLehn, K. (1995). A Bayesian approach to cognitive assessment. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Norwich, K. H. & Wong, W. (1997). Unification of psychophysical phenomena: The complete form of Fechner's law. *Perception & Psychophysics, 59* (6), 929-940.

Parducci, A. & Wedell, D. H. (1986). Category effects with rating scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception and Performance, 12* (4), 496-516.

Petrov, A. & Anderson, J. R. (2000). ANCHOR: A memory-based model of category rating. *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 369-374). Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegel, W. (1972). Memory effects in the method of absolute judgment. *Journal of Experimental Psychology, 94* (2), 121-131.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64* (3), 153-181.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273-286.