Psychological Review
2005, Vol. 112, No. 2, 383–416

0033-295X/05/$12.00 DOI: 10.1037/0033-295X.112.2.383

The Dynamics of Scaling: A Memory-Based Anchor Model of Category Rating and Absolute Identification

Alexander A. Petrov
University of California, Irvine

John R. Anderson
Carnegie Mellon University

A memory-based scaling model—ANCHOR—is proposed and tested. The perceived magnitude of the target stimulus is compared with a set of *anchors* in memory. Anchor selection is probabilistic and sensitive to similarity, base-level strength, and recency. The winning anchor provides a reference point near the target and thereby converts the global scaling problem into a local comparison. An explicit correction strategy determines the final response. Two incremental learning mechanisms update the locations and base-level activations of the anchors. This gives rise to sequential, context, transfer, practice, and other dynamic effects. The scale unfolds as an adaptive map. A hierarchy of models is tested on a battery of quantitative measures from 2 experiments in absolute identification and category rating.

Category rating is a widely used method of data collection in experimental psychology. Ratings come in a wide variety of guises: psychophysical scales, similarity judgments, typicality judgments, confidence ratings, attitude questionnaires, health self-reports, and many others. The participants in all these tasks are asked to rate things using an ordered set of categories such as 1, . . . , 7 or *strongly agree, . . . , strongly disagree.* Most people can do this without effort, and most psychologists tend to take their responses at face value. Ratings are among our primary dependent measures.

Yet the field lacks a comprehensive theory of how these responses are produced. There are many partial theories, each within the confines of a particular discipline, and the practitioners in each discipline make conscious efforts to minimize the impact of factors outside their scope. Thus, psychophysicists treat nonperceptual factors such as sequential dependencies and guessing strategies as nuisance biases that must be minimized through randomization, counterbalancing, and averaging (Stevens, 1957). Cognitive psychologists, conversely, tend to ignore low-level perceptual features in their stimuli, experiments, and theories.

The problem with this divide-and-conquer approach is that each theory is incomplete. Both peripheral and central mechanisms are engaged in most real-life situations. The overall behavior stems from the joint action and interaction of these various components. Thus, careful study of the isolated components is valuable and necessary, but it must be followed by systemic efforts to put the pieces back together.

The pursuit of integration is hardly new. Miller's (1956) classic article devotes equal amounts of space to short-term memory and absolute identification, a close relative of category rating. Both are subject to capacity limitations, and Miller cast them both in information-processing terms. The research that sprouted from this influential early effort, however, splintered into increasingly divergent traditions. Forty years later, a review of Miller's impact concludes that "the dominant theories and paradigmatic approaches in these two domains have gone their separate ways, with little, if any, cross talk" (Shiffrin & Nosofsky, 1994, p. 360).

Integration was premature in 1956, but perhaps in 2005 the time is ripe. The psychophysics and memory traditions have amassed a rich collection of empirical data and theoretical insights. By pooling knowledge from all these sources, one can hope to achieve unified understanding.

We propose a theory bridging the gap between psychophysics and memory. It is based on the ACT–R architecture (Anderson & Lebière, 1998), which is deeply rooted in the memory tradition (Anderson, 1983; Anderson & Bower, 1973). At the same time, the theory addresses two quintessentially psychophysical tasks—absolute identification and category rating. We conduct two behavioral experiments and compile a list of empirical constraints for the theory. We operationalize most of them with quantitative targets for modeling. The list includes effects that to our knowledge are novel contributions to the empirical literature. The response distributions are noticeably nonstationary and nonuniform even when the stimulus distributions are stationary and uniform. Also, the context effect induced by skewed stimulus distributions apparently reverses direction depending on the presence or absence of feedback. These and various other dynamic effects are successfully accounted for by a process model called ANCHOR. It formalizes the theoretical principles in mathematical equations and implements them in a computer program. A set of memory-based *anchors* compete to match the perceived magnitude of the target stimulus. An explicit strategy corrects most (but not all) memory fluctuations. Incremental competitive learning updates the locations of the anchors, and activation learning updates their availability. The response scale unfolds as an adaptive map from a single arbitrarily placed anchor. The correction strategy generates

Alexander A. Petrov, Department of Cognitive Sciences, University of California, Irvine; John R. Anderson, Department of Psychology, Carnegie Mellon University.

This research was supported in part by Air Force Office of Scientific Research Grant F49620-99-10086. We thank Boicho Kokinov, Duncan Luce, Encho Gerganov, Lynne Reder, Mark Steyvers, Scott Brown, and Stefan Mateeff for discussions and comments on a draft.

Correspondence concerning this article should be addressed to Alexander A. Petrov, Department of Cognitive Sciences, University of California, 3151 Social Science Plaza, Irvine, CA 92697. E-mail: apetrov@uci.edu

novel responses and enforces the local consistency of the stimulus–response mapping, whereas competitive learning consolidates the local consistency into a global homomorphism. As the model reinforces its own responses during category rating without feedback, its dynamic stability depends vitally on the correction mechanism. Under skewed stimulus distributions, activation learning induces assimilation, whereas competitive learning induces compensation. The direction of the overall context effect depends on the relative strength of these competing tendencies. As competitive learning is silenced by external feedback, the context effect reverses direction during absolute identification exactly as in the behavioral data. Extensive simulations reveal the value added by each ANCHOR mechanism on a battery of operational measures. Finally, the limitations of the model are discussed, and it is compared with alternative proposals.

Main Principles of the Theory

The theory sets out to characterize the information-processing mechanisms engaged in unidimensional scaling tasks. At the most generic level, they all consist of establishing and maintaining a systematic correspondence between stimuli and responses. The absolute identification task requires one-to-one mapping, typically defined and constantly reinforced by external feedback. The entire stimulus set consists of a relatively small number of distinct stimuli: for instance, a set of nine lines of different lengths. The participant is asked to identify each stimulus by its corresponding label. The category rating task is similar, but the number of stimuli is greater than the number of responses, and hence a many-to-one mapping is required. Feedback is seldom provided because the purpose of the procedure is to estimate the subjective magnitude of perceived stimulus intensity.

ANCHOR stands at the intersection of two broad theories—Thurstonian psychophysics (Green & Swets, 1966; Thurstone, 1927; Torgerson, 1958) and the theory of memory incorporated in the ACT–R architecture (Anderson, 1983; Anderson & Lebière, 1998; Anderson & Milson, 1989). The link between the two theories is the construct of internal *magnitude* (see Figure 1). It is assumed that some sensory processes, collectively referred to as the *perceptual subsystem,* construct an internal magnitude M that encodes the intensity of the physical stimulus S. This magnitude is then processed within the *central subsystem* to determine the overt response R. A central claim of the ANCHOR theory is that the latter transition is memory based.

Figure 1 is provided for expository purposes only and is deliberately simplified. In particular, the open loop suggested by the diagram obscures one very important feature of ANCHOR: The central subsystem maintains an internal state that evolves from trial to trial. Thus the response R depends not only on the immediate stimulus S but also, at least in principle, on all previous stimuli and responses.

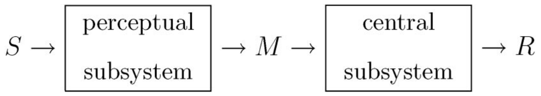

Figure 1. ANCHOR has two subsystems communicating via internal magnitudes M. A central claim of the theory is that the $M \rightarrow R$ transition is memory based. S = stimulus; R = response.

Following the lead of many psychophysical theories and models (e.g., Baird, 1997; Braida et al., 1984; Green & Swets, 1966; Nosofsky, 1997; Treisman & Williams, 1984), ANCHOR assumes that the perceptual subsystem operates independently of the central one and that the internal magnitude M is the only piece of information exchanged between them. We believe that this independence holds to a good approximation in a wide range of circumstances, including those typical in scaling studies.

The theory deals exclusively with unidimensional continua. Such continua are a subclass of the multidimensional psychological spaces for which several well articulated theories exist (see Ashby, 1992; Ashby & Maddox, 1998, for reviews), including some excellent memory-based theories (e.g., Kruschke, 1992; Nosofsky, 1986, 1991, 1997). Unidimensional continua, however, have a special property that sets them apart: They are ordered. A magnitude M_1 is either greater or less than another magnitude M_2. This ordering relation, and the concomitant homomorphism between magnitudes and responses, is fundamental to the whole notion of scaling (Luce, 1959).

At the most general level, our theory rests on four main principles: internal magnitude continuum, content-addressable memory, explicit correction strategies, and obligatory learning. Each of them is widely accepted in its respective scientific community.

Internal Magnitude Continuum

The first principle postulates an internal continuum of magnitudes. On each trial, the external stimulation generates a magnitude M. It is this internalized quantity that can be committed to memory and compared against other magnitudes. More generally, magnitudes are a form of analog representations: Relative positions and distances on the internal continuum correspond systematically to relative intensities and similarities among the physical stimuli.

The intrinsic stochasticity of the perceptual subsystem entails some magnitude variability even when the stimulus remains fixed across multiple presentations. Thus, although a single magnitude is realized on a trial, a whole distribution of magnitudes must be considered for each stimulus level (Thurstone, 1927).

Content-Addressable Memory

The second principle postulates content-addressable memory involving these magnitudes. In particular, it is possible to establish associations between a magnitude and the label of a response category. Such associations are called *anchors.* They substantiate the mapping between magnitudes (and hence the stimuli represented by them) and responses. When a new target magnitude is produced by the perceptual subsystem, the memory fills in the corresponding response label. This completion process is stochastic and depends on two factors: (a) the location of the target magnitude with respect to the various anchors in memory and (b) the frequency and recency of use of each response category. The latter factor is captured by the *base-level activations* (or *biases*) of the anchors. These activations play an important role in the theory and make direct contact with many memory-related phenomena.

Explicit Correction Strategies

Because the memory system is noisy and prone to biases, it is not guaranteed to provide on each trial the anchor that best matches the target magnitude. The verbal protocols of human observers suggest that they are aware of the unreliability of their first guesses and adopt explicit correction strategies. Consequently, the third main principle of the ANCHOR theory provides for such explicit corrections. Phenomenologically, an introspective report of a trial might go like this, “I see the stimulus. It looks like a *7.* No, it’s too short for a *7;* I’ll give it a *6.*” Such increments and decrements have far-reaching implications and are vital for the stability of the overall system, especially in the absence of feedback.

There is strong evidence that people rely on such anchor-plus-adjustment heuristics in uncertain situations (Tversky & Kahneman, 1974). Anchoring effects have been found in probability estimation, risk estimation, utility assessment, stock market investment, and various social cognition phenomena. The effects are robust and persist even when people are forewarned and motivated to avoid them (Wilson, Houston, Etling, & Brekke, 1996).

Obligatory Learning

So, the stimulus has been encoded and matched against anchors, and a response has been produced. Is this the end of the trial? According to the fourth principle of the theory, the answer is no. The cognitive system is plastic (within limits) and each experience seems to leave a mark on it. ANCHOR, and the ACT–R architecture in general, postulates obligatory learning mechanisms that incrementally update the internal state of the system at the end of each trial. There are two learning mechanisms in the current version of the model: One updates the base-level activations of the anchors, and the other updates their locations. This makes ANCHOR an adaptive dynamic system and gives rise to various sequential, context, and transfer effects.

There is ample evidence for such dynamic effects in the experimental literature. Sequential effects, in particular, have been found in virtually every scaling experiment in which the issue has been examined and reported, regardless of the experimental procedure or perceptual modality (Baird, Green, & Luce, 1980; DeCarlo & Cross, 1990; Holland & Lockhead, 1968; Jesteadt, Luce, & Green, 1977; King & Lockhead, 1981; Luce, Nosofsky, Green, & Smith, 1982; Mori & Ward, 1995; Petzold, 1981; Purks, Callahan, Braida, & Durlach, 1980; Schifferstein & Frijters, 1992; Ward, 1979; Ward & Lockhead, 1970, 1971). They indicate that some kind of internal state persists across trials, blocks, and even days and influences subsequent processing. Memory seems the most natural candidate to perform this function. ANCHOR explores a specific, prototype-based variant of this general idea.

The experimental literature on psychophysical scaling, despite all its admirable riches, typically reports only very coarse statistics—aggregates over many observers and thousands of trials. These data do establish a wealth of empirical regularities in qualitative terms but do not constrain adequately a mechanistic model on a trial-by-trial basis. Moreover, the various phenomena are documented under widely different experimental settings. Such diversity is crucial for the generalizability of the findings but obstructs the construction of an integrated model. To remedy these shortcomings, we decided to collect some detailed, unified data ourselves.

Experimental Design

We performed two experiments: one on absolute identification and the other on category rating. Both studies use exactly the same stimulus material, response categories, presentation sequence, and participant population. The experimental design seeks to replicate as many scaling phenomena as possible within this unified framework. The goal is to collect a single data set that consolidates the disparate reports scattered throughout different journals over many years. We compile a battery of quantitative measures for over a dozen phenomena (summarized in Tables 1 and 3 below). In addition to being worthwhile in itself, such consolidation imposes much tighter constraints on the model as all effects must be produced under unified parameter values.

Perceptual Modality

We have chosen a particularly fundamental and convenient perceptual modality: length of lines in the frontal plane. Several considerations recommend this choice. The subjective perception of length seems linear to an excellent approximation—Stevens’s exponent is very close to 1.0 (see Wiest & Bell, 1985, for a meta-analysis of 70 studies). This permits the use of convenient analytic tools such as linear regression. It has been argued (Krantz, 1972) that subjective length is the paradigmatic example of a ratio scale and the gold standard in cross-modality matching.

The stimuli in our experiments are not lines but pairs of bright dots against a dark background. The task of the observers is to judge the distance between them. This is a concrete, physically tangible rendition of some of the abstractions implicit in multidimensional scaling (Schiffman, Reynolds, & Young, 1981) and distance-based similarity metrics (Nosofsky, 1992).

Stimulus Presentation Schedule

From the standpoint of the memory hypothesis, we are particularly interested in the dynamic aspects of scaling. Our design relies on nonuniform stimulus distributions to induce context effects and manipulates them within subjects to induce transfer effects (see Figure 2). The main experimental manipulation involves the presentation frequency of the different stimuli. Each session is divided in five blocks in an alternating schedule sche-

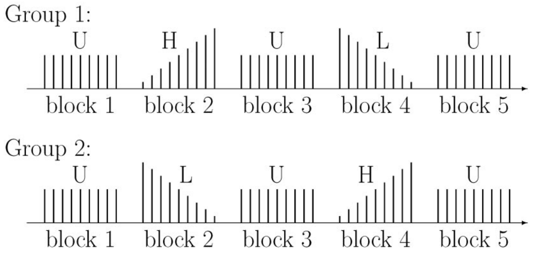

Figure 2. The presentation schedules alternate uniform (U), low (L), and high (H) blocks. Bar heights on the diagram depict stimulus presentation frequencies within each block.

matized in Figure 2. Exactly the same stimuli are used throughout; only their respective frequencies are changing.

Three kinds of frequency distributions are used: uniform (U), positively skewed (low, L), and negatively skewed (high, H). Blocks 1, 3, and 5 are always uniform—all stimuli have equal presentation probabilities. Blocks 2 and 4 have triangular distributions as illustrated in Figure 2. A low block presents shorter stimuli with progressively higher frequencies than longer ones; in a high block the situation is reversed. The skew direction is counterbalanced within subjects, and the overall schedule is counterbalanced between subjects: UHULU for Group 1 and ULUHU for Group 2.

Skewed presentation distributions are interesting because they tend to induce context effects. The response to any given stimulus depends not only on the stimulus itself but also on the distribution of the other stimuli in the block (Chase, Bugnacki, Braida, & Durlach, 1983; Marks, 1993; Parducci, 1965, 1974; Parducci, Knobel, & Thomas, 1976; Parducci & Perrett, 1971; Parducci & Wedell, 1986; Schifferstein & Frijters, 1992). Given that this distributional information can only be accumulated over time, context effects are a valuable tool for studying the internal state maintained across trials.

Two kinds of context effects are possible: assimilative and compensatory. For concreteness, suppose the block is dominated by long stimuli. By definition, if the stimuli tend to be systematically overestimated under such conditions, there is an *assimilatory context effect*—the responses are attracted toward the densely populated end of the scale. If the stimuli tend to be systematically underestimated instead, there is a *compensatory context effect.* Metaphorically speaking, assimilation makes the rich even richer: The response distribution is even more skewed than the stimulus distribution that drives the process. Compensation is egalitarian: The response distribution is a compromise between the skewness in the stimuli and a uniform ideal.

The experimental literature abounds with reports of compensatory context effects, obtained almost invariably under between-subjects designs (e.g., Parducci & Wedell, 1986; Schifferstein & Frijters, 1992). Assimilative effects are occasionally reported too (Chase et al., 1983). Marks (1993), in particular, reported consistent assimilation in nine studies with shifting stimulus ranges. Thus the direction of the context effect is an open research question, especially under within-subject transfer manipulations.

The alternating schedules in Figure 2 are designed to induce context effects within the skewed blocks and transfer effects from the skewed blocks to the uniform ones and vice versa. Such dynamic transfer is especially informative from the standpoint of the memory hypothesis.

Earlier transfer studies shift either the stimulus range (e.g., Haubensak, 1990, 1992; Marks, 1993) or the stimulus spacing (Wedell, 1984). Both manipulations introduce new stimuli at the transition from one block to the next. This may alert the participants to the shift and induce conscious changes in their response strategies. Our frequency manipulation is far less conspicuous as it uses the same stimuli throughout the experiment.

We are not aware of any previous investigation of the transfer of frequency-induced context effects. Thus the present studies are an opportunity to try a novel variation of the experimental paradigm and contribute to the empirical literature. In addition, they collect a consolidated data set as a target for modeling.

Experiment 1: Absolute Identification

Method

Participants. Twenty-four undergraduate students enrolled in an introductory psychology course at Carnegie Mellon University participated in the experiment to satisfy a course requirement. Twelve were randomly assigned to Group 1 and 12 to Group 2.

Stimuli and apparatus. The stimuli were pairs of white dots presented against a uniformly black background on a 17-in. AppleVision monitor. The viewing distance was approximately 600 mm. The independent variable was the distance between the centers of the two dots. The stimulus set consisted of nine dot pairs with the following distances: 275, 325, 375, . . . , 675 pixels (275 pixels ≈ 88 mm ≈ 8.4 degree visual angle [dva]; 675 pixels ≈ 216 mm ≈ 20 dva). The full width of the monitor was 1,000 pixels (320 mm, 32 dva). The imaginary segment formed by the dots was always horizontal and was randomized with respect to its absolute horizontal and vertical position on the screen. The stimulus set for each participant was generated and randomized separately. Each dot was roughly circular in shape with a diameter of 16 pixels (5 mm, 0.5 dva).

Procedure. The participants were instructed that there were nine stimuli and nine responses and that their task was to identify each stimulus with a number from *1* to *9.* The stimuli were presented on the monitor one at a time according to the schedules described below. Each trial began with a 500-ms alert sound followed by a 3,300-ms stimulus display. The participants entered their responses on the numeric keypad of the computer keyboard. As soon as the observer pressed a key the dots were cleared from the screen and a big white digit indicating the correct identification appeared. The feedback stayed for 1,300 ms or until the end of the 3,300 ms presentation window, whichever lasted longer. The computer recorded the response and the latency (measured from the onset of the stimulus). After a 200-ms intertrial interval, the next trial began.

There were 17 demonstration and 450 experimental trials divided into 10 periods with short breaks after Trials 56, 112, 157, 202, 247, 292, 337, and 393. The demonstration introduced the stimuli first in increasing and then in decreasing order, with feedback *1, 2, 3, . . . , 8, 9, 8, . . . , 1,* respectively. The participants were encouraged to practice pressing the corresponding keys. The whole session lasted about 40 min. After completing the computer-administered procedure, each observer was asked to give an informal retrospective account of “what was going on in your head as you were doing the task.”

Presentation schedules. The presentation schedule was UHULU in Group 1 and ULUHU in Group 2 (see Figure 2). From the standpoint of logical design, the sequence of 450 experimental trials consisted of 5 blocks of 90 trials each. The uniform (U) blocks presented the nine stimuli 10 times. The low (positively skewed) blocks comprised 18, 16, 14, . . . , 2 presentations of Stimuli *1, 2, 3, . . . , 9,* respectively. The presentation frequencies in the high (H) blocks were skewed in the opposite direction.

The order of presentation within each block was randomized. The break periods never coincided with a boundary between logical blocks. From the observer’s point of view it all looked like a long homogenous sequence of randomized trials. Note that each stimulus is presented to each observer exactly 50 times overall.

Results and Discussion

All data were analyzed individually for each observer. We focus almost exclusively on the identification responses, mentioning the response times and the retrospective protocols only in passing. The data set consists of 10,636 valid cases (10,800 trials total, 164

Table 1
Empirical Constraints Derived From the Identification Experiment

Phenomenon	Brief description	Empirical	Model
Capacity limitation	Absolute identification is error prone and the amount of transmitted information is limited.	T = 1.68 (0.21)	T = 1.57 (0.20)
Nonuniform response distribution	The response distribution has a peak in the middle even when the stimulus distribution is uniform.	s = 2.40 (0.09)	s = 2.50 (0.04)
Edge (bow) effects	Both accuracy (% correct) and discriminability (d') are higher at the edges of the stimulus range relative to the interior.	bow = +.14 (.40)	bow = −.31 (.33)
Sequential effect	The current response R_t is positively correlated with the previous response R_{t-1}.	Figure 4	Figure 10
Similarity effect	The magnitude of the sequential effect depends on the similarity between the consecutive stimuli S_{t-1} and S_t.	Figure 4	Figure 10
Repetition effect	When the same stimulus is repeated on two consecutive presentations, the identification accuracy on the second trial is greater than average.	rep = .11 (.09)	rep = .04 (.08)
Assimilative context effect	Under skewed stimulus distributions, identification responses shift in the direction of the skew.	d = +.14 (.24)	d = +.11 (.22)
Practice effect	The identification accuracy improves over time.	pr = .06 (.12)	pr = .02 (.07)

Note. The measures of the various effects are defined in the text. Means and standard deviations (in parentheses) are reported for the sample of 24 observers and for 9,600 model runs. See Table 3 for a complementary list of category rating phenomena and Table 4 for additional model fits.

responses missing). The complete data set and the software that administered the experiment can be downloaded from the ACT–R web server or obtained from Alexander A. Petrov on request.[1]

We used the following methodology for quantifying the empirical results and evaluating the model performance. We define a numerical measure for each phenomenon of interest and develop software estimating it from the stimulus–response sequence. Each individual data set thus yields a battery of measures. Table 1 summarizes nine absolute identification phenomena discussed in turn below and reports the mean and standard deviation of each measure over the sample of 24 participants. Statistics for 9,600 model runs are also listed for comparison.

Descriptive statistics. The first measure in our battery characterizes the overall performance. Ever since Miller's (1956) classic 7 ± 2 article, absolute identification capacity is traditionally quantified by the amount of transmitted information T (Equation 1). To the extent a (stochastic) relationship between stimuli and responses exists, knowing the stimulus on a given trial reduces the uncertainty (or entropy H) of the response. The magnitude of this reduction thus measures the strength of the stimulus–response relationship:

$$T = H(R) - H(R|S) \tag{1}$$

$$= -\sum_r P(r)\log_2 P(r) + \sum_s P(s) \sum_r P(r|s)\log_2 P(r|s).$$

In our sample, the transmitted information ranges from 1.36 to 2.12 bits, with mean T = 1.68 and a standard deviation of 0.21. This corresponds to perfect identification of three to four items and replicates the classic capacity limitation results (Baird, Romer, & Stein, 1970; Braida & Durlach, 1972; Luce, Green, & Weber, 1976; Miller, 1956). The probability of giving a correct response varies from .41 to .74 (M = .54, SD = .08). That is, the observers make mistakes on almost half the trials.

The overwhelming majority of the errors, however, are only one unit away from the correct response (see Table 2). Hence, although the probability of being absolutely correct is relatively low, the stimulus–response correlation is very high (mean R^2 = .90, SD = .034). Thus the identification performance is not nearly as bad as the information transmission values may seem to imply.

The response distributions tend to have a peak in the middle and even clearer depressions at the extremes.[2] Quantitatively, the standard deviation of the individual response distributions has mean s = 2.40 in the sample (SD = 0.09). This is below the value 2.58 that corresponds to perfect uniformity.

Practice effect. The identification accuracy seems to increase in the course of the experiment. The overall proportion of correct responses is 0.49, 0.52, 0.54, 0.58, and 0.55 during Blocks 1 through 5, respectively. (Recall that Blocks 2 and 4 involve triangular distributions.) We define a new statistic, $pr = P_5 - P_1$, to measure the increase in accuracy of Block 5 relative to Block 1. This statistic has a mean of .06 and a standard deviation of .12 in our sample. The increase is statistically significant, matched-sample $t(23) = 2.59$, $p < .01$, and replicates earlier reports of practice effects in absolute identification (Hartman, 1954; Rouder, 2001; D. L. Weber, Green, & Luce, 1977). Though frequently neglected, this practice effect is of obvious importance from a memory perspective.

Edge effects. Given that the experiment involves only nine stimuli, it is straightforward to characterize the performance separately for each category. Figure 3 plots the profiles of four different measures. They all point to the same conclusion: Stimuli S_1 and S_9 (and to a lesser extent S_2 and S_8) elicit superior perfor-

[1] The software is written in Macintosh Common Lisp. The MATLAB implementation of various analytic algorithms is available at http://www.socsci.uci.edu/~apetrov/ and http://act-r.psy.cmu.edu/.

[2] The overall response histogram is 860, 1,175, 1,211, 1,279, 1,397, 1,381, 1,277, 1,180, and 876.

Table 2
Frequency of Various Kinds of Identification Errors for Each Stimulus, Pooled Across All 24 Observers

	Stimulus length (pixels)										
Response error	275	325	375	425	475	525	575	625	675	Total	%
3	7	9	11	4	3	1				35	0.3
2	32	58	72	55	27	20	13			277	2.6
1	350	365	371	360	288	242	188	151		2,315	21.8
0	804	704	613	576	618	607	568	584	710	5,784	54.4
−1		52	111	177	221	277	340	357	385	1,920	18.1
−2			2	8	19	31	46	68	80	254	2.4
−3				1	2	3	13	18	14	51	0.5
Total	1,193	1,188	1,180	1,181	1,178	1,181	1,168	1,178	1,189	10,636	100.0

mance than those in the interior of the range. This replicates earlier reports of *edge* (or *bow*) effects (Braida & Durlach, 1972; Lacouture, 1997; Luce et al., 1982; Mori & Ward, 1995; W. Siegel, 1972; D. L. Weber et al., 1977).

It is useful to distinguish between *accuracy edge effect* and *resolution edge effect* (Treisman, 1985). The former refers to the increased proportion of correct responses near the edges and could stem from the simple fact that there are fewer possibilities for mistake there (cf. Table 2). This cutoff undoubtedly contributes to the bows in the two left panels in Figure 3, but it cannot account for the bow in the interstimulus discriminability profile in the top right panel.

One can treat each boundary $i|(i + 1)$ as a binary discrimination and compute the associated $d'_{i,i+1}$ by the method of Luce et al. (1982). Whenever stimulus S_{i+1} is presented, all responses greater than or equal to $i + 1$ count as hits, and those less than or equal to i count as misses. Across the boundary, on trials with S_i, responses greater than or equal to $i + 1$ count as false alarms, and those less than or equal to i count as correct rejections. The d' is then calculated in the usual way, separately for each participant. The top right panel in Figure 3 plots the group average.

The d' profile is asymmetrical, indicating that short distances are generally more discriminable than long ones. There is a hint of a slight rise at the right-hand side, however, which cannot be attributed to Weber's law. To test whether this rise is significant, we define a new variable *bow* according to Equation 2. This statistic has a mean of 0.14 and a standard deviation of 0.40 across the 24 observers. The difference from zero is significant, $t(23) =$

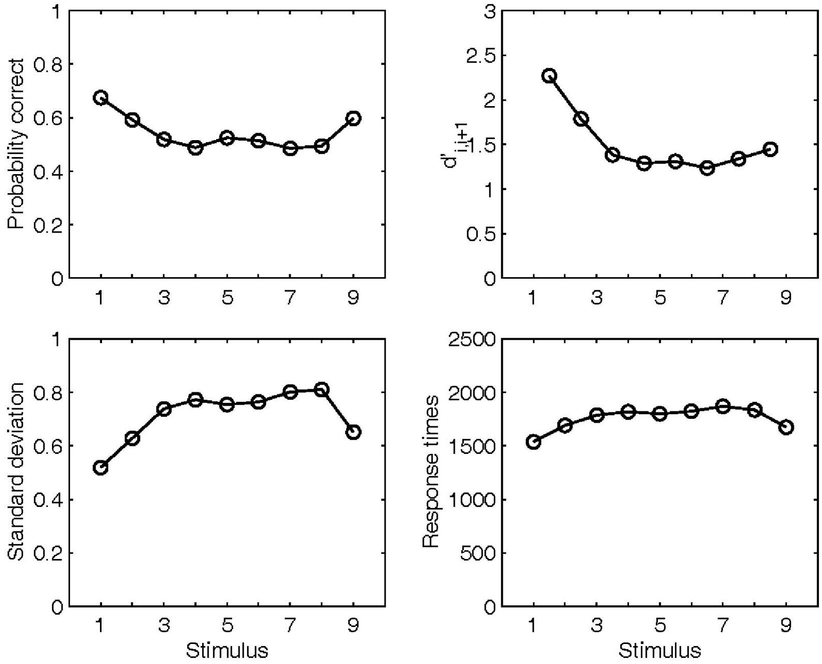

Figure 3. Bow effects in absolute identification. Top left: Overall accuracy for each stimulus. Bottom left: Standard deviation of responses. Top right: d' for each interstimulus boundary. Bottom right: Mean latency in milliseconds.

1.72, one-tail $p < .05$, which is evidence for a resolution edge effect strong enough to overcome the monotonic decrease in discriminability.

$$\text{bow} = d'_{8,9} - \frac{1}{4}(d'_{3,4} + d'_{4,5} + d'_{5,6} + d'_{6,7}). \quad (2)$$

Sequential effects. There is clear evidence for an assimilative sequential effect. Figure 4 plots the conditional probabilities of various kinds of error. Following Luce et al. (1982), we track “overshoot” and “undershoot” errors separately, conditioning on the difference $\Delta S = S_{t-1} - S_t$ between the stimuli on consecutive trials $t - 1$ and t. Consider the line labeled “Overshoot by 1” in the top panel, for example. It describes the probability of giving a response R_t that is one unit greater than the correct label of the stimulus S_t presented on the current trial. The likelihood of such overshoot error is between .10 and .15 when S_{t-1} is less than S_t. This is followed by a sharp steplike increase when S_{t-1} becomes greater than S_t—the probability quickly rises to .35 and then remains there. The undershoot curve shows the opposite pattern. Mistakes by two or more units are rare, and the inflection points of their respective profiles appear to be around $\Delta S = \pm 3$.

The profiles in the top panel of Figure 4 and especially the cumulative probability plot in the bottom panel show a clear assimilative tendency. This is one more replication of this robust and ubiquitous finding (Garner, 1953; Holland & Lockhead, 1968; Luce et al., 1982; Petzold, 1981; Purks et al., 1980; Ward & Lockhead, 1970). Our data do not discriminate whether this assimilation is driven by the previous stimulus S_{t-1}, previous feedback (which is perfectly confounded with S_{t-1}), or the previous response R_{t-1}. When R_{t-1} is used in place of S_{t-1}, the resulting plot is virtually identical to Figure 4. It seems well established, however, that whatever the critical factors may be, they act in relation to the position of the current stimulus (similarity effect).

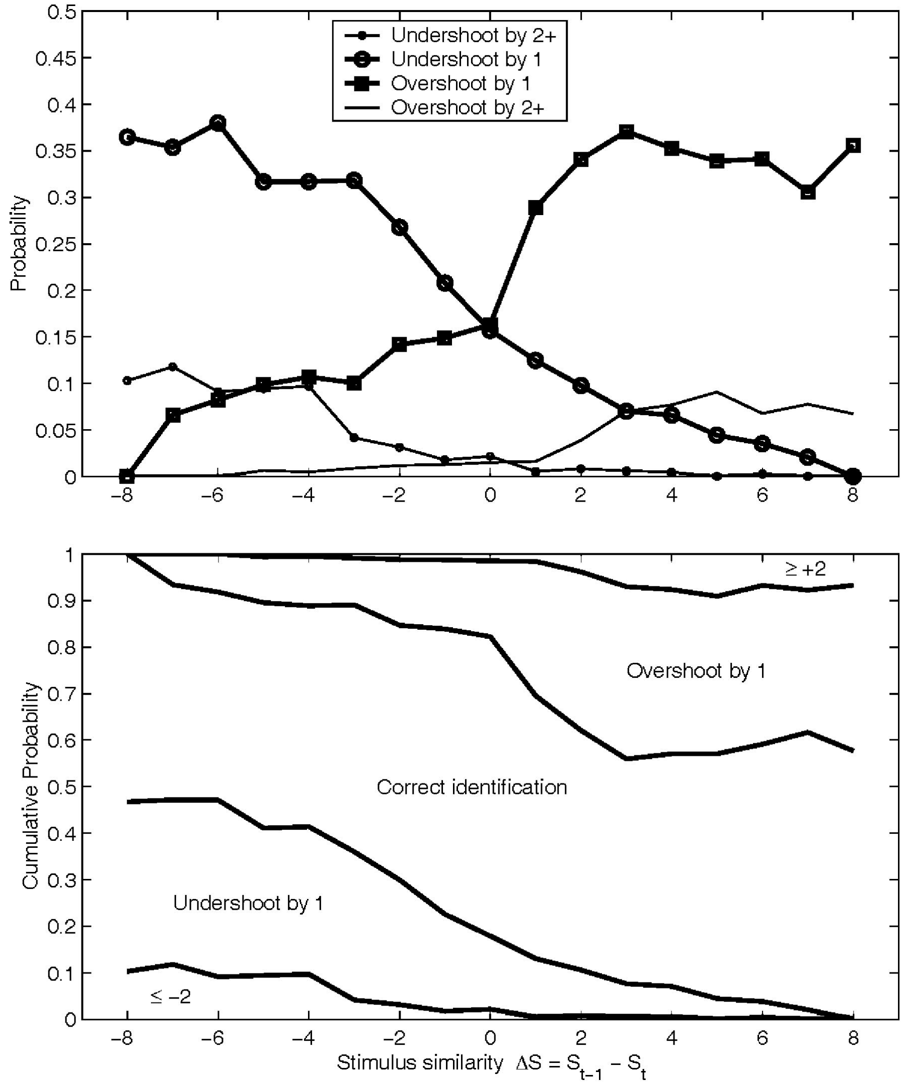

Figure 4. Top: Conditional probabilities for various identification errors as a function of the similarity between consecutive stimuli. Bottom: Cumulative-probability plot of the same data.

Repetition effect. Repetition of the same stimulus on two consecutive trials brings noticeable improvement in the identification performance (see Figure 5). To check whether this repetition effect extends farther back in time, a new variable N_t is constructed from the trial sequence by counting back to the most recent repetition of the stimulus S_t (J. A. Siegel & Siegel, 1972; W. Siegel, 1972). For instance, $N_t = 1$ if and only if $S_t = S_{t-2}$ and $S_t \neq S_{t-1}$. Because of the randomization, approximately 1⁄9 of all trials have $N_t = 0$, 1⁄9 of the rest have $N_t = 1$, and so forth. Figure 5 plots the identification accuracy and the mean response times for the first few levels of *N*.

There is a clear difference between repetition ($N = 0$) and nonrepetition trials both in terms of accuracy (about 11% benefit) and latency (about 160 ms on average). Our data show little further influence of the number of intervening items, with the exception perhaps of a marginal (40 ms) effect on the response times after many intervening trials. This replicates the results of W. Siegel (1972), who reported consistent and clear-cut accuracy increases for $N = 0$ in several experiments and occasional marginal effects for $N = 1$.

We define a new variable rep to quantify the repetition benefit for each observer:

$$\text{rep} = P_r - P_n, \tag{3}$$

where P_r and P_n denote the probabilities of correct identification on trials with $N = 0$ and $N \geq 1$, respectively. This difference is positive for 21 of the 24 participants ($M = .11$, $SD = .09$).

Context effect. We introduce here a general method for calculating the average response levels of individual observers during various time periods. It is used to quantify the context and transfer effects in all experiments and simulations in this article.

Suppose the response policy of a given participant during a particular time period is characterized by some known Stevens function $\bar{R} = F(S)$ mapping stimuli to their average responses. Suppose further that it was possible to freeze the observer's state of mind and probe the response policy with a long sequence of uniformly distributed stimuli. We call the expected mean response under such conditions the *average response level* (ARL). By definition, ARL equals the area under the Stevens curve divided by the stimulus range. When the stimulus range is fixed (as in our experiment) this quantity depends only on the function *F*. Assuming a power law $\bar{R} = R_0 + aS^n$, the ARL is defined by Equation 4, where $S_{\min}$ and $S_{\max}$ denote the minimal and maximal stimulus, respectively. When $n = 1$, this expression simplifies to Equation 5; the ARL for linear scales equals the expected response to the stimulus in the middle of the range. In our setting, this is the stimulus with a length of 475 pixels.

$$\text{ARL} = \frac{1}{(S_{\max} - S_{\min})} \int_{S_{\min}}^{S_{\max}} (R_0 + aS^n)dS \tag{4}$$

$$= R_0 + a\frac{(S_{\max}^{n+1} - S_{\min}^{n+1})}{(n+1)(S_{\max} - S_{\min})};$$

$$\text{ARL} = R_0 + a(S_{\min} + S_{\max})/2. \tag{5}$$

Thus if we know the Stevens function, we can calculate the ARL. In practice the Stevens function is not known and must be estimated from the observed stimulus–response pairs. For the linear response scale in our experiment this amounts to estimating a slope *a* and intercept R_0 by linear regression of the responses *R* on the stimuli *S*. ARL then equals $R_0 + 475a$.

Note that the ARL calculated according to this procedure is not the same as the arithmetic mean of the observed responses. The latter average reflects not only the response policy but also the stimulus distribution, which is not necessarily uniform. ARL is closely related but not identical to two other measures in the literature: the area under the Stevens curve (e.g., Parducci & Wedell, 1986) and the adaptation level (Helson, 1964). ARL is always proportional to the area in our experiment because the stimulus range is held constant. It, however, has the advantage of being expressed in the units of the response scale. Helson (1964) defined the adaptation level as the stimulus that corresponds to the

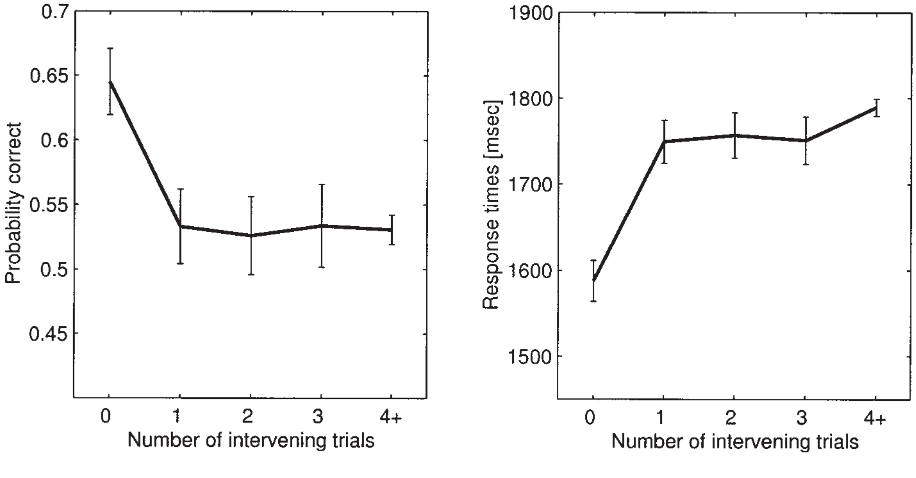

Figure 5. Benefit of repeating a stimulus on two consecutive trials. Left: Identification accuracy (plus or minus 95% confidence interval). Right: Mean response times (plus or minus 95% confidence interval within subjects).

average response. ARL is the complement of this—the response that corresponds to the average stimulus (when the Stevens function is linear).

The response policy changes over time—both endogenously and because of external factors such as context manipulations. Indeed, we want to estimate the ARL precisely to track such changes. There is a tradeoff. On one hand we need a sufficient number of observations to get a reasonable estimate of the Stevens function *F*. On the other hand if we average over too many trials, the dynamics of the ARL would be smoothed away. As a compromise, we have chosen to segment the data into periods of 45 trials each.

In summary, we calculate the ARL profile of each observer by segmenting the data into 10 nonoverlapping periods, fitting a regression line in each period, and applying Equation 5. This procedure transforms the sequence of 450 stimulus–response pairs into a profile of 10 ARLs. Final averaging across participants produces the two group profiles plotted in Figure 6.

The ARLs in Figure 6 stay close to 5.0, which is consistent with the stable feedback level. The limited variability that remains around this overall baseline, however, appears to be context dependent. The two profiles coincide during the first uniform block (90 trials, two points on each profile). Then they seem to diverge slightly over the 90 trials with skewed presentation frequencies. The third block, with uniform distribution, apparently brings the two response levels back together. They diverge again during the subsequent nonuniform Block 4 and stay apart until the end of the session.

The context effect is assimilative: The ARL tends to increase when high stimuli dominate the presentation schedule and decrease when low stimuli dominate. To quantify this effect and to test its significance, we single out two points on the profile of each observer. Point *h* is the ARL during the second half[3] of the high block; point *l* is the ARL during the second half of the corresponding low block. The group averages are $h = 5.15$ and $l = 4.91$ in Group 1, and $h = 5.12$ and $l = 5.07$ in Group 2.

$$d = h - l. \tag{6}$$

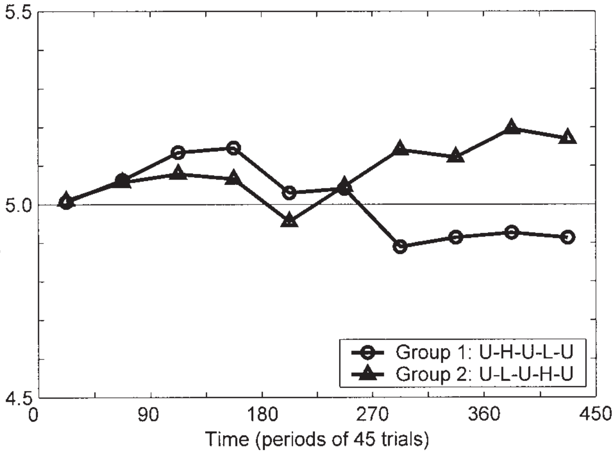

Figure 6. Assimilative context effect in absolute identification. The average response levels tend to increase in negatively skewed (H) and decrease in positively skewed (L) blocks. U = uniform.

We adopt the difference *d* in Equation 6 as an overall measure of the context effect. Positive values indicate assimilation, and negative values indicate compensation. The context effect in our case is clearly assimilatory (mean $d = +0.14$), matched-sample $t(23) = 2.90$, $p < .01$. This finding is at odds with the compensatory effects typically found in most context studies (e.g., Parducci & Wedell, 1986) and our own category rating data in Experiment 2. We attribute the unusual direction of the present context effect to the explicit feedback in the identification task, for reasons discussed at length later.

Verbal protocols. Nineteen informal retrospective reports are available. Six of them contain explicit and spontaneous statements that the previous stimulus S_{t-1} is used as a reference (cf. Laming, 1984). Five people mentioned "stable images of *1* and *9*" anchoring the rest of the scale. Four people singled out the midpoint. Some observers referred to multiple strategies depending on the circumstances. For instance, "I compared it [the stimulus] with the previous one if it was in the same ballpark, otherwise I worked from *1* or *9*."

In summary, Experiment 1 replicates all phenomena falling within its scope: limited information transmission, practice, sequential, and edge effects of various kinds. It also provides evidence for an assimilative context effect. Quantitative measures of most effects are available for each individual observer as summarized in Table 1.

Mechanisms of the ANCHOR Model

Armed with these empirical data we are now in a position to discuss the computational mechanisms of the model in detail. Figure 7 shows a schematic diagram of the main variables involved and the dependencies among them.

The model instantiates and concretizes the four principles outlined in the introduction (cf. Figure 1). On each trial, a stimulus with intensity *S* is presented to the model. The perceptual subsystem builds, stochastically, an internal representation of this stimulus. This target magnitude *M* then acts as a memory cue, and the anchors compete to match it. The anchor selection mechanism determines, stochastically, a single anchor as winner. The magnitude of this anchor is denoted by *A* in Figure 7. The correction mechanism then compares the target magnitude *M* with the anchor magnitude *A* to determine whether a correction *I* is needed. If the increment *I* in Figure 7 is zero, the response associated with the anchor is adopted as the final response *R* on this trial. Otherwise the increment *I,* which can be either positive or negative, is added to the anchor response to determine *R.* Finally, two learning mechanisms update various anchor parameters, thereby changing the internal state of the model. The updated internal state controls the behavior on the next trial, and the whole cycle repeats.

There are five computational mechanisms in ANCHOR: perception, anchor selection, base-level activation, correction, and competitive learning. To demonstrate the value added by each mechanism, a hierarchy of models introduces them one by one. Each successive model incorporates one new feature relative to its

[3] Trials 136–180 for Group 1 and 316–360 for Group 2, see Figure 2. The first half of each block is left out in an effort to minimize transfer-related contamination and for consistency with the analysis of Experiment 2.

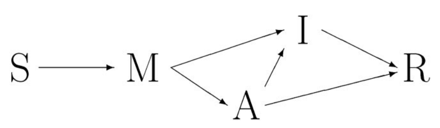

Figure 7. Main ANCHOR variables and the dependencies among them: stimulus intensity *S,* internal magnitude *M,* anchor *A,* corrective increment *I,* and response *R.* Compare with Figure 1.

predecessor. Models $\mathcal{M}_{\mathrm{P}}$ and $\mathcal{M}_{\mathrm{S}}$ are just the perceptual or the anchor-selection mechanisms working in isolation, respectively. $\mathcal{M}_{\mathrm{PS}}$ puts these two components together, and $\mathcal{M}_{\mathrm{PSA}}$ adds base-level activations to the anchors. Next, model $\mathcal{M}_{\mathrm{PSAC}}$ introduces the correction mechanism. Finally, the competitive learning (or running average) mechanism yields $\mathcal{M}_{\mathrm{PSACR}}$, which is synonymous with the full-blown ANCHOR model.

The Perceptual Subsystem

The present article does not seek to establish a sophisticated theory of sensory transduction. Indeed, we are content with a minimalist generic formulation of the perceptual subsystem that still takes into account the fundamental empirical constraints expressed by Weber's and Stevens's laws. This allows us to focus on the central subsystem.

The whole perceptual subsystem is reduced to a single equation that describes the distribution of magnitudes as a function of the stimulus. It abstracts away factors such as contrast, attention, habituation, Gestalt, and so forth. The current version of the model assumes that the internal magnitude *M* depends solely on the intensity of one particular attribute of the stimulus. The effects of the other attributes, the surrounding context, and the previous stimuli are assumed negligible. Thus, as a first approximation, the stimulus intensity *S* is the only independent variable in the perceptual equation of the model. Additional factors can be included in the future without disrupting the rest of the theory.

Weber's law and Stevens's law. One empirical constraint that cannot be neglected by any credible scaling system is that the intensity difference threshold ΔS tends to be proportional to *S* over much of the dynamic range of the stimulus (Fechner, 1860/1966; E. H. Weber, 1849). Thus the ratio of the two—the Weber fraction—is approximately constant for a given perceptual modality:

$$\frac{\Delta S}{S} = k = \text{constant}. \tag{7}$$

Another major empirical regularity comes from a vast array of magnitude estimation and category rating studies (Marks, 1974; Stevens, 1957, 1975; Stevens & Galanter, 1957). For intensive (or *prothetic*) continua the average rating $\bar{R}$ varies approximately as a power function of the stimulus intensity *S:*

$$\bar{R} = aS^n. \tag{8}$$

Both Weber's and Stevens's laws are subject to qualifications, and various alternative formulations and corrections have been proposed (e.g., Ekman, 1959; Jesteadt, Wier, & Green, 1977; Krueger, 1989; Laming, 1986; Marks & Stevens, 1968; Norwich & Wong, 1997). Most of them deal with slight deviations near the low absolute threshold and can be put aside for our present purposes because their impact on direct scaling performance is probably negligible.

Perceptual equation in ANCHOR. The internal ANCHOR representation of a stimulus *S* is a magnitude *M* (see Figure 7). Following Thurstone (1927), *M* is a Gaussian random variable whose mean and variance depend on the stimulus intensity *S.* The standard interpretation of Weber's law is that the mean of each magnitude distribution is proportional to the logarithm of the corresponding stimulus. Assuming equal variance (Fechner's postulate), this explains the progressively poorer discriminability at higher intensity levels.

There is an alternative proposal, however, that is equally consistent with the data. It is possible, indeed likely,[4] that the standard deviation of each magnitude distribution grows in proportion to its mean (Ekman's law; Ekman, 1959). With such multiplicative perceptual noise, the spacing among the means can be less compressive and still produce poorer discriminability at higher intensities. In fact, power-law compression is sufficient (see proof in Appendix A). This leads to Equation 9, in which *n* is the exponent from Stevens's power law (Equation 8), *a* is an arbitrary scaling factor, and k_p is a dimensionless coefficient of proportionality. The noise ε_p has zero mean and unit variance.

$$M = aS^n(1 + k_p\varepsilon_p). \tag{9}$$

Various other forms have also been proposed, and in general the psychophysical function cannot be inferred unambiguously from the available data (Norwich & Wong, 1997; Treisman, 1964, 1985). Most theoretical alternatives, however, differ in rather subtle features and have comparable qualitative shapes over finite stimulus ranges away from the absolute threshold (for instance, a logarithm versus a power function with an exponent less than 1). Moreover, the indeterminacy is contained within the perceptual subsystem and has little or no implications for the memory-based central processing of main interest here. Simulations with the ANCHOR model have shown that it can fit the empirical profiles in Figure 3 equally well using either additive noise and strong (logarithmic) compression or multiplicative noise and moderate (power-law) compression (Petrov, 2003). Thus one can investigate the sequential, context, and other dynamic effects independently of the details of the perceptual subsystem.

We adopt the multiplicative-noise Equation 9 for our present purposes. It takes a particularly convenient linear form for our perceptual modality. The Stevens exponent is set to $n = 1.0$ in agreement with the general consensus that the scale for physical length is very nearly linear (e.g., Stevens, 1957; Stevens & Galanter, 1957; Wiest & Bell, 1985). Our rating Experiment 2 also corroborates this choice. It is convenient to set the scaling factor to $a = {}^{1}\!/_{1{,}000}$, yielding magnitudes *M* in the [0, 1] range:

$$M = S(1 + k_p\varepsilon_p)/1{,}000. \tag{10}$$

The only thing left to specify is the confusability parameter k_p that scales the variance of the magnitude distribution in proportion to its mean. Appendix A derives an upper bound on this parameter in

[4] The equal-variance assumption contradicts the frequent finding of normalized receiver operating characteristic curves with slopes less than unity (Swets, 1986).

terms of the empirically observable exponent n (Equation 8) and the Weber fraction[5] k (Equation 7):

$$k_p \leq nk. \tag{11}$$

As $n = 1.0$ in our case the relationship becomes particularly simple: $k_p \approx k$. An estimate of the Weber fraction for the dot pairs used in our scaling experiments would therefore eliminate one free parameter from the model. Laming and Scheiwiller (1985) reported two discrimination experiments with stimuli similar to ours. The details of this study are outlined in Appendix A, and a Weber fraction $k = .04$ is obtained. The same value is listed in two secondary sources (Baird & Norma, 1978; Laming, 1986).

On the basis of this information we can fix $k_p = .04$ in Equation 10. This provides a parameter-free specification of the perceptual subsystem for all subsequent simulations.

Anchors

All processing in the model involves anchors in one way or another. There is one anchor for each response category—nine in our case. Taken together, the anchors map the internal magnitude continuum onto the overt response scale. They contain succinct information about what stimuli were labeled with each particular response in the past. The whole internal state of the model is encapsulated in the current locations and base-level activations of the anchors.

The location L_i of each anchor i represents the current estimate of the magnitude of the prototypical member of the corresponding response category. For instance, suppose we are dealing with Anchor 7 and its current location is $L_7 = .612$. Under the calibration of the model, an internal magnitude of .612 corresponds to physical length of 612 pixels. This stimulus is estimated to be the prototypical *7* at the moment.

What the system actually responds to, however, is the anchor magnitude A_i, which is a noisy version of its location L_i. Suppose that halfway through a category rating experiment we ask the participant, "Please close your eyes and visualize a typical *7*." ANCHOR assumes that memory access is noisy and introduces fluctuations and distortions. The specific proposal is formalized in Equation 12, which is directly analogous to the perceptual Equation 10. The memory noise ε_m is a Gaussian variable with zero mean and unit variance.

$$A_i = L_i(1 + k_m\varepsilon_m) \quad \text{for each anchor } i. \tag{12}$$

The anchor magnitude A_i is a random variable centered on the current anchor location L_i, just as the target magnitude M is a random variable centered on the stimulus intensity S. We assume multiplicative noise in memory, as in perception. The standard deviation of each magnitude distribution is proportional to its mean. The coefficient of proportionality k_m (Equation 12) is a dimensionless parameter analogous to k_p (Equation 10). It is the same for all anchors. The value $k_m = .07$ yields good fits to our experimental data.

Note that four different variables have the prefix *anchor* in their names. The first of them is the anchor label i that identifies the particular anchor. In our setting the labels are the numerals from *1* to *9*, but in general they can be any verbal labels such as *similar, very similar,* and so on. The second quantity is the anchor location L_i. It is updated dynamically by the competitive learning mechanism described later. The anchor magnitude A_i is produced by injecting multiplicative noise to L_i on a trial-by-trial basis (Equation 12). Finally, each anchor has a base-level activation B_i, also described later.

Anchor Selection Mechanism

The anchors compete on each trial to match the target magnitude M. The outcome of this competition is described by two equations in the ANCHOR model. Equation 13 produces goodness scores G_i, and the *softmax* Equation 14 converts them into selection probabilities P_i. A random number generator then draws the specific winner i^* on that trial.

$$G_i = -|M - A_i| + HB_i; \tag{13}$$

$$P_i = \frac{\exp(G_i/T)}{\sum_j \exp(G_j/T)}. \tag{14}$$

Each goodness score G_i is a sum of two terms: similarity $-|M - A_i|$ and history HB_i. The first is simply the negation of the mismatch between the target magnitude M and the anchor magnitude A_i. The second term does not depend on the current target M (and hence the stimulus S) at all: It introduces a priori bias B_i multiplied by a fixed weight H. In this section we concentrate on the similarity factor and temporarily assume $H = 0$.

The anchor selection is a stochastic process. This stochasticity comes from two sources. First, the anchor magnitudes A_i are noisy (Equation 12), and this noise propagates into the goodness scores G_i. Moreover, it is not guaranteed that the anchor with the best score wins; it only has the highest probability of winning (Equation 14). The temperature parameter T controls the degree of randomness: Values close to zero produce nearly deterministic choice, whereas large values result in nearly uniform selection probabilities P_i. The temperature is measured in the units of the magnitude scale and scales the standard deviation of the selection noise implicit in Equation 14. A typical value is $T = .050$, which corresponds to 50 pixels on the stimulus continuum.

Equation 14 plays a pivotal role in a remarkably diverse array of memory models. In the ACT–R architecture, it governs the retrieval of chunks from declarative memory and the conflict resolution in procedural memory (Anderson & Lebière, 1998). It is motivated in terms of a stochastic competition among many alternatives, the goodness score of each being perturbed by selection noise. An equivalent equation (with different notation) is standard in instance-based models (e.g., Nosofsky, 1986, 1988; Nosofsky & Palmeri, 1997), where it is motivated as biased choice (Luce, 1963) with exponentially scaled similarities (Shepard, 1957, 1987). The same Boltzmann–Gibbs distribution describes the thermodynamic equilibrium of autoassociative memory networks (e.g., Hinton & Sejnowski, 1986; Hopfield, 1982).

It is easier to illustrate the anchor selection mechanism in a model in which all other mechanisms are switched off. We denote this model $\mathcal{M}_S$, with parameters $k_m = .07$ and $T = .050$ (the default

[5] Throughout this article, Weber fractions are defined on the basis of difference thresholds ΔS that elicit 75% correct responses in two-alternative forced-choice comparisons.

values for the full model). The nine anchors are fixed at evenly spaced locations $L_1 = .275$, $L_2 = .325, \ldots, L_5 = .475, \ldots, L_9 = .675$. Model $\mathcal{M}_S$ is run 100,000 times with a target $M = .475$ to estimate the anchor selection probabilities. The resulting estimates are .010, .029, .080, .207, .322, .209, .092, .037, and .014. Thus, exact matches occur on only about 32% of the trials. The immediate neighbors in either direction are not far behind, and more distant anchors have their chances too. Note the slight asymmetry of the probability distribution. It stands above the Monte Carlo fluctuations and stems from the multiplicative nature of the memory noise.

One can probe the model with target magnitudes corresponding to the nine experimental stimuli and collect a confusion matrix of anchor selection probabilities. This matrix can then be converted to a d' profile and compared with the corresponding empirical profile in the top right panel of Figure 3. We extend our simulations to include four different models: $\mathcal{M}_P$, $\mathcal{M}_{S1}$, $\mathcal{M}_{S2}$, and $\mathcal{M}_{PS}$. In model $\mathcal{M}_P$ the perceptual noise parameter k_p is set to .04, and the anchor selection mechanism is noise free ($k_m = 0$, $T = 0$). Hence all confusions in this model arise in the perceptual subsystem (Equation 10). This is equivalent to a Thurstonian system with fixed criteria (Torgerson, 1958). In contrast, $\mathcal{M}_{S1}$ and $\mathcal{M}_{S2}$ experiment with noise-free perception ($k_p = 0$) but introduce randomness in the goodness scores ($\mathcal{M}_{S1}$: $k_m = .07$, $T = 0$) or softmax selection ($\mathcal{M}_{S2}$: $k_m = 0$, $T = .050$). Finally, all three sources of variability are present in $\mathcal{M}_{PS}$ ($k_p = .04$, $k_m = .07$, $T = .050$). Each model is probed 100,000 times with each of the nine targets.

Figure 8 shows the d' profiles for the four models. They should be interpreted in qualitative terms; no effort for quantitative fit to the empirical profile has been made (see Petrov, 2003, for such effort). The sliding profiles of $\mathcal{M}_P$ and $\mathcal{M}_{S1}$ reflect the multiplicative noises in Equations 10 and 12, respectively. This feature is reminiscent of Weber's law and is also present in the human data. The softmax-only model ($\mathcal{M}_{S2}$, Equation 14) treats all anchors equally, and its profile is correspondingly flat. When the softmax rule operates on noisy magnitudes (model $\mathcal{M}_{PS}$), the resulting profile accommodates the flatness of $\mathcal{M}_{S2}$ with the steep slopes of $\mathcal{M}_P$ and $\mathcal{M}_{S1}$. Also, the overall performance level drops considerably as there are now multiple sources of confusion. Note that the $\mathcal{M}_{PS}$ profile, which is generated with default ANCHOR parameters, lies quite below the empirical values (average $d' = 1.51$; see Figure 3). It is the responsibility of the explicit correction strategy to bring it back up in the final ANCHOR model. Finally, none of the synthetic profiles show the upward bow noticeable in the human data.

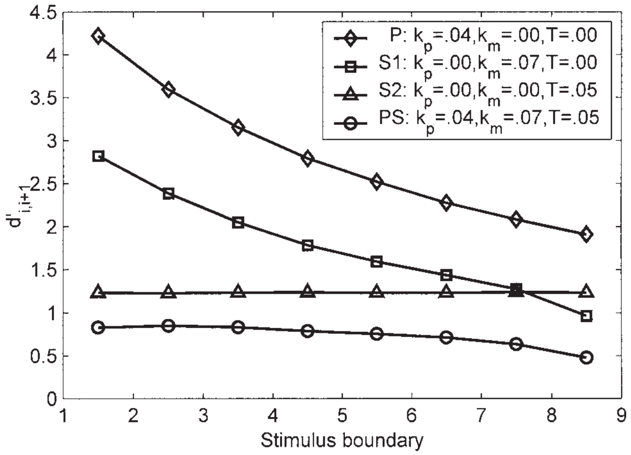

Figure 8. Discriminability profiles for four partial models (P, S1, S2, and PS). Perceptual noise (k_p), memory noise (k_m), and softmax temperature (T) parameters are described in the legend. Compare with Figure 3, top right.

There are three noise sources in ANCHOR and three corresponding parameters: k_p, k_m, and T. The two multiplicative sources have essentially identical signatures, and it is tempting to subsume k_p and k_m into a single parameter k. Instead, we eliminate one degree of freedom using an independent data set ($k_p = .04$; see Appendix A). Keeping the perceptual and memory noises separate improves the realism and modularity of the model.

Base-Level Activation

ANCHOR has two history-dependent, incremental learning mechanisms. The first of them is based on the rational analysis of memory (Anderson, 1990; Anderson & Milson, 1989) and plays a key role in the ACT–R architecture.

Activation defined. Each anchor has a base-level activation (or bias) B_i that controls its overall availability. Whenever a response category is used on a given trial the activation of the corresponding anchor increases, thereby making it more available on subsequent trials. At the same time, the activations of all unused anchors decay away. Figure 9 plots two typical base-level curves. The dashed line belongs to an anchor that is used only 6 times (marked by the open triangles) as compared with 82 uses for the solid line.

Note the three distinctive features of the activation dynamics: sharp transient peak immediately after each use, decay in the absence of use, and gradual buildup of strength with frequent use. These features are consistent with a huge body of experimental evidence and are widely used in memory models (e.g., Anderson & Milson, 1989; Conway, 1997). They also match the statistical structure of the environment (Anderson & Milson, 1989; Anderson & Schooler, 1991).

The base-level activation of each anchor is a concise summary of the history of its use. Equation 15, taken verbatim from ACT–R (Anderson & Lebière, 1998, p. 124), is the conceptual backbone of our approach. It is a logarithm of a sum of powers with decay rate d. Each new use of the anchor adds another term to this sum, which then decays independently. The total count so far is denoted by n, and t_l are the individual time lags from the present. It is easy to see that this equation captures the three desired properties. Moreover, it has been applied successfully in many ACT–R models, and the value $d = .5$ of the decay parameter has proven to work well in a wide range of circumstances (Anderson & Lebière, 1998). Therefore, we can eliminate this degree of freedom and fix the ANCHOR decay rate to the ACT–R default $d = .5$.

$$B = \ln\left[\sum_{l=1}^{n} t_l^{-d}\right]; \qquad (15)$$

$$B \approx \ln\left[t_{\text{last}}^{-.5} + \frac{2(n-1)}{\sqrt{t_{\text{life}}} + \sqrt{t_{\text{last}}}}\right]. \qquad (16)$$

Equation 16 is an excellent approximation to the computationally expensive Equation 15. It retains only three critical pieces of

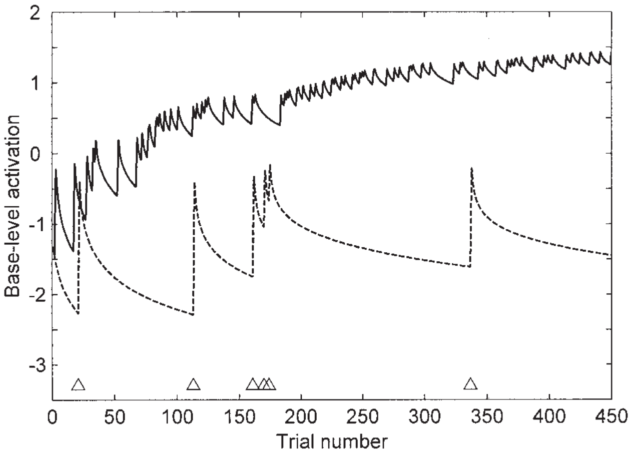

Figure 9. Typical activation dynamics. The dashed line tracks an anchor that is used 6 times (marked by the open triangles) as compared with 82 uses for the solid line.

information about the anchor: the time since its creation t_{life}, the time since its most recent use t_{last}, and the total number of uses n. The parameter-free Equation 16 is used in Figure 9 and all simulations in this article.

Activations in ANCHOR. To fully specify the activation mechanism in ANCHOR, we must define which anchor is considered "used" on a trial. When there is explicit feedback, this is the anchor corresponding to the correct response. When there is no feedback, the system's own response is taken as the best available estimate. Note that it is possible, for instance, to select Anchor 3 from memory, make a correction and respond "4," and finally receive feedback "5." Only Anchor 5 is strengthened on such a trial; all other anchors, 3 and 4 included, suffer decay.

Recall from Equation 13 that each goodness score G_i is a weighted sum of two terms: similarity $-|M - A_i|$ and base level HB_i. They correspond directly to the context and history factors in Anderson and Milson's (1989) rational analysis of memory. The history weight parameter H varies across individuals and experimental conditions. Typical values are $H = .080$ for absolute identification and $H = .100$ for category rating.

Because of the normalization in Equation 14, all anchor selection probabilities depend only on differences among the competing goodness scores. Consequently, all activation levels are invariant up to an additive constant. This implies in turn that one can change the unit of time in Equations 15 and 16 without affecting the behavior of the model or disrupting any estimated parameters. For compatibility with our empirical studies, all lags in Equation 16 are measured in seconds, and the duration of each trial is 4 s.

Observable manifestations of the activation dynamics. The introduction of base-level activations has profound implications for the model. It is no longer a static system and is now capable of (and liable to) sequential, context, transfer, repetition, and practice effects.

Model $\mathcal{M}_{\text{PSA}}$ explores the value added by the activation mechanism relative to the feedforward model $\mathcal{M}_{\text{PS}}$. It has four parameters: $k_p = .04$, $k_m = .07$, $T = .050$, and $H = .080$. The simulations replicate the identification experiment: The model is run on sequences of 450 trials with feedback. Each sequence comprises five blocks with uniform (U), low (L), or high (H) stimulus distributions. There are 1,000 runs with schedule UHULU (Group 1) and another 1,000 with schedule ULUHU (Group 2; see Figure 2).

The simulation shows straightforward repetition effects. The overall accuracy is 34%, which is comparable to that of model $\mathcal{M}_{\text{PS}}$. Unlike its predecessor, however, $\mathcal{M}_{\text{PSA}}$ can be primed by prior events. In particular, it is more accurate on repetition than on nonrepetition trials: rep $= .13 = .46 - .33$.

The repetition effect is easy to explain. Whenever a stimulus is repeated on two consecutive trials, the feedback on the first trial reinforces the corresponding anchor. The resulting spike in that anchor's activation improves its goodness score on the subsequent trial and thus makes it more likely that the same anchor is selected again. On repetition trials, this happens to favor the correct response. The activation spike is transient and decays away after a few intervening trials (see Figure 9), though not as quickly as the empirical data in Figure 5 suggest.

The activation-based priming gives rise to sequential assimilation as well. For example, suppose the stimulus on trial $t - 1$ is $S_{t-1} = 5$, and hence Anchor 5 is strengthened. If the same stimulus were repeated on the next trial, facilitation would result. Suppose, however, that the new stimulus is somewhat different instead, but not too different. For concreteness, let $S_t = 4$. The residual activation still favors Anchor 5, and consequently, the response is likely to be "5" again. Under the circumstances, such a response constitutes an assimilative error. Finally, if the new stimulus is very different from the old (e.g., $S_t = 1$), the mismatch penalty of the selection mechanism effectively eliminates Anchor 5 from the competition regardless of its elevated activation level.

Figure 10 illustrates this process. It plots the conditional probabilities of overshoot and undershoot errors as a function of the stimulus difference $\Delta S = S_{t-1} - S_t$, just as Figure 4 does for the empirical data. Two nested models are explored: $\mathcal{M}_{\text{PSA}}$ (top) and $\mathcal{M}_{\text{PSAC}}$ (bottom). The latter introduces the correction mechanism and is described in detail in the next section.

The assimilative effect is clearly evident. Moreover, it is controlled by the interstimulus similarity, especially in the top panel. The probability to err by ± 1 or ± 2 rises sharply for $\Delta S = \pm 1$ or ± 2, respectively.[6] The correction mechanism smooths out the curves by converting most of the large errors to smaller ones and some of the small errors to correct responses. The corrections, however, are incomplete and cannot eliminate the assimilation in full. Model $\mathcal{M}_{\text{PSAC}}$ thus makes few errors of magnitude ± 2 or more, but the near misses persist and retain their assimilative bias, just as they do in the experimental data (see Table 2).

The simulated butterfly profile in the bottom panel of Figure 10 is strikingly similar to the empirical one (see Figure 4). The steplike discontinuities around $\Delta S = 0$ are clearly replicated. Even the slight asymmetry between the two wings is reproduced, driven by the asymmetry in the multiplicative noise sources in Equations 10 and 12.

The pattern of sequential effects depends on the availability of feedback. ANCHOR predicts sequential assimilation toward the previous correct response when feedback is available and toward

[6] The far flanks of the curves reflect edge-related artifacts rather than assimilation. For example, $\Delta S = 8$ implies $S_t = 1$, the undershoot probability is forced to 0, and some of it spills to the overshoot curves.

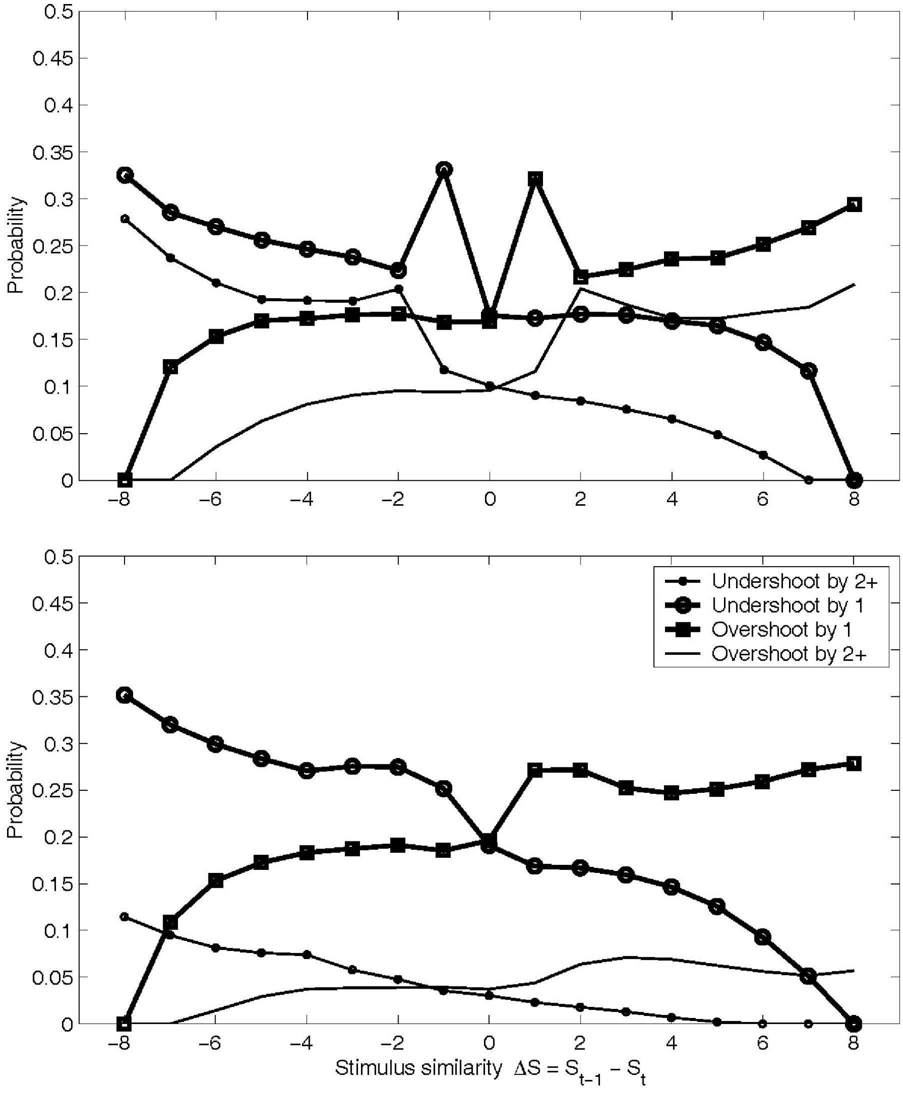

Figure 10. Sequential effects arise from the base-level activation mechanism (model $\mathcal{M}_{\text{PSA}}$; top) and are modulated by the correction mechanism ($\mathcal{M}_{\text{PSAC}}$; bottom). Compare with Figure 4.

the previous own response when feedback is not available. This is exactly what is reported by Mori and Ward (1995), who alternated feedback and no-feedback blocks within subjects and evaluated the relative strength of the time-lagged predictors. In three identification experiments, R_t was more dependent on S_{t-1} rather than R_{t-1} in the feedback blocks and more dependent on R_{t-1} rather than S_{t-1} in the no-feedback blocks. Given that the feedback was in one-to-one correspondence with the stimulus S_{t-1}, this finding confirms the ANCHOR prediction.

The activation mechanism produces long-term counterparts to the short-term sequential effects as well. The gradual buildup of strength with frequent use gives rise to assimilative context and transfer effects. Figure 11 shows the average response levels of models $\mathcal{M}_{\text{PSA}}$ and $\mathcal{M}_{\text{PSAC}}$ under the two presentation schedules. The ARL profiles are calculated from the simulation data in exactly the same way as in the empirical study: segmenting each response sequence into 10 nonoverlapping periods, fitting a regression line in each period, and applying Equation 5.

Model $\mathcal{M}_{\text{PSA}}$ exhibits strong assimilative context and transfer effects. The two thin lines in Figure 11 begin at ARL = 5.0—the middle of the response scale—during the first uniform block (90 trials, two points on each profile). Then they diverge during the second block as the stimulus presentation frequencies become skewed in opposite directions. The diagram is symmetrical; consider the profile for Group 1, schedule UHULU. This group gets disproportionately many long stimuli on Trials 91–180. As a result, the base-level activations of the "long" anchors in the model grow progressively stronger. The model in effect learns that a higher response is a better bet than a lower one, everything else being equal. The observable consequence of this bias is a gradual increase in the ARL. This assimilation is in qualitative agreement with the empirical ARL profiles from the absolute-identification experiment (see Figure 6).

The presentation schedule switches back to uniform during the third block (Trials 181–270) in Figure 11. The two ARL profiles begin to move closer to the baseline and to each other. Notice,

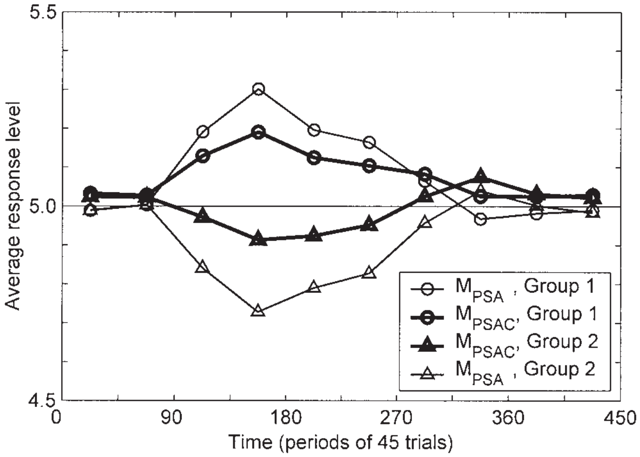

Figure 11. Assimilative context effect produced by models $\mathcal{M}_{\mathrm{PSA}}$ (thin lines) and $\mathcal{M}_{\mathrm{PSAC}}$ (thick lines) with feedback. The average response levels are calculated as in Figure 6. See text for details.

however, that there is a clear transfer effect: The two groups continue to differ even though the environmental conditions are now identical. The two profiles cross during the second skewed block (Trials 270–360) and finally return to the baseline by the end of the session.

The correction mechanism (model $\mathcal{M}_{\mathrm{PSAC}}$) attenuates these effects without altering their fundamental character. Recall from Equation 6 that we quantify the context effect by the difference $d = h - l$ between the ARLs in high and low blocks. This statistic is 0.14 in the psychophysical data, 0.32 for model $\mathcal{M}_{\mathrm{PSA}}$ with default parameters, and 0.16 for $\mathcal{M}_{\mathrm{PSAC}}$.

Correction Mechanism

Successful performance in any scaling task depends on a body of prior knowledge. It includes facts such as "5 is one unit greater than 4" and at least tacit awareness of the principles of homomorphism. Human observers apparently adopt a variety of strategies that engage this prior knowledge. ANCHOR implements one such strategy in its correction mechanism. Although it hardly exhausts the possibilities, this particular strategy is very powerful and has far reaching consequences.

The correction mechanism combines five pieces of information: (a) the target magnitude M, (b) the anchor magnitude A, (c) the response associated with this anchor R_A, (d) knowledge about the approximate category width, and (e) knowledge about the ordering of the response scale. The target and anchor magnitudes are compared to estimate the discrepancy D (Equation 17). If it is not too large, the anchor label is adopted as the final response R. Otherwise the target is judged too different to belong to the category represented by the anchor, and the response is incremented or decremented accordingly.

$$D = M - A. \tag{17}$$

Specifically, ANCHOR adds an increment I to the anchor response R_A (Equation 18). It can be zero, positive, or negative depending on the discrepancy D. Five increments are currently implemented: $I \in \{-2, -1, 0, 1, 2\}$. The decision rule is based on four correction thresholds c_i. For example, an increment of 1 is made when $c_1 < D \leq c_2$. The corrected response R is clipped at 1 or 9 if necessary:

$$R = R_A + I \quad \text{clipped between } R_{\min} \text{ and } R_{\max}. \tag{18}$$

Note that all adjustments are made locally. Corrections by ±2 units are rare, and larger ones are impossible. Thus the model relies on a two-tiered approach. First a global, similarity-driven process narrows the field and provides a reference point in the vicinity of the target. This transforms the absolute judgment task into a relative one. A local comparison then fine-tunes the response.

The correction mechanism is stochastic because of perceptual and memory fluctuations in the discrepancy D. As M and A are already specified (Equations 10 and 12), it is fully parameterized by its four thresholds. For parsimony, they are set to fixed multiples of two free parameters: c^+ controls the ease of upward corrections, and c^- controls the ease of downward corrections. To illustrate, consider the arrangement with $c^+ = c^- = .025$ and thresholds at $\{-3c^-, -c^-, c^+, 3c^+\}$. It centers the no-correction interval ($I = 0$) around $D = 0$ and makes the width of each interval equal to one category width (.050 magnitude units, 50 pixels). On the basis of this a priori analysis, we fix the outer thresholds to 3 times the corresponding inner thresholds in all simulations.

The upward and downward corrections are not necessarily symmetric. This allows ANCHOR to produce slow systematic drifts of the ARLs—a dynamic feature detectable in the human data as well. In particular, when increments are easier to make than decrements ($c^+ < c^-$), the correction mechanism tends to drive the responses upward. Such slow upward drift is evident in the data from our category rating Experiment 2 (see below). Motivated by this finding and a desire for parsimony, we decided to fix the ratio c^+/c^- to 0.9 and express all thresholds in terms of a single parameter: $\{-3c, -c, 0.9c, 2.7c\}$. This arrangement is informed by the qualitative direction of the upward drift in the data, but no quantitative optimizations have been made.[7] The only parameter optimized to fit the data is the overall cutoff c.

The cutoff c controls the ease of correction. It is estimated separately for each observer and typically falls in the range .030–.050. This indicates a conservative correction strategy: Substantial discrepancy D is required to trigger any changes. Several factors contribute to this result. It is possible that the participants do not know the true category width, especially on the early trials. Also, it is hardly obvious to them that a discrepancy of half category width warrants a full correction point, just as the number 0.51 rounds up to 1.00. Finally, it seems very likely that the correction strategy is not applied on every trial.

As the correction mechanism makes only insufficient adjustments, it matters which anchor is used as reference; the response is assimilated toward it. The insufficiency of adjustment is a recurring theme in the diverse literature on anchoring effects (e.g., Tversky & Kahneman, 1974; Wilson et al., 1996).

Model $\mathcal{M}_{\mathrm{PSAC}}$ explores the value added by the correction mechanism. The cutoff is $c = .040$, and all nonthreshold parameters are the same as in model $\mathcal{M}_{\mathrm{PSA}}$. Figures 10 and 11 illustrate the

[7] In retrospect, a ratio less than .9 probably accounts for the empirical drifts better.

sequential and context effects in the two models. Keeping the variability of all noise sources fixed, the introduction of corrections raises the overall percentage correct from 34% ($\mathcal{M}_{\text{PSA}}$) to 50% ($\mathcal{M}_{\text{PSAC}}$) and the amount of transmitted information from $T = 0.94$ to $T = 1.49$.

The contributions of the correction mechanism may seem insignificant at first, but it truly comes into its own in the absence of feedback. It binds the anchors into a coherent ordered set, which is crucial for category rating and contributes enormously to the long-term stability and robustness of the system. The category rating task is more challenging than absolute identification because no feedback is available to guide the responses. Rating data thus impose strong additional constraints on the model mechanisms.

Experiment 2: Category Rating

Experiment 2 extends our investigation to the more demanding task of category rating. It replicates the design of Experiment 1 with two modifications. First and foremost, there is no feedback. Second, more stimulus levels are involved—hundreds of physically distinct lengths in Experiment 2 versus only nine in Experiment 1. This is a relatively minor difference because perceptual variability makes the internalized magnitude distribution continuous and smooth even when the external stimuli are discrete and spaced apart. The presentation schedule alternates uniform and nonuniform blocks of opposite skewness as in Experiment 1 (see Figure 2).

Method

Participants. Forty undergraduate students participated in Experiment 2. None of them were involved in Experiment 1. Twenty were randomly assigned to Group 1, and 20 to Group 2.

Stimuli and apparatus. The apparatus was the same as in Experiment 1. The stimuli were dot pairs as before but with 451 distinct lengths covering the range from 250 pixels (80 mm, 7.7 dva) to 700 pixels (224 mm, 22 dva).

Procedure. The participants were asked to "rate the distance between the two dots" on a scale from *1* to *9*. The instructions stated explicitly that there were more "possible distances" than responses and, therefore, one "has to give the same response to slightly different stimuli." There was no feedback; the screen turned blank when the participant responded and stayed blank until the end of the presentation window. In all other respects the procedure was the same as in Experiment 1: 500-ms alert sound plus 3,300-ms stimulus presentation plus 200-ms intertrial interval.

There were 17 demonstration trials presenting Stimuli 275, 325, 375, . . . , 625, 675, 625, . . . , 275, in that order. The participants were encouraged to practice pressing the keys *1, 2, 3, . . . , 8, 9, 8, . . . , 1*. The main sequence of 450 experimental trials followed, with break periods as in Experiment 1.

Presentation schedules. The presentation schedule was UHULU in Group 1 and ULUHU in Group 2 (see Figure 2). The stimuli in each block were sampled with replacement from the corresponding distribution: uniform (U), triangular (L) ascribing 451 times greater probability to $S = 250$ than to $S = 700$, or triangular of opposite skewness (H). Each sequence was generated and randomized individually.

Results and Discussion

The data set consists of 17,743 valid stimulus–response pairs (18,000 trials total, 257 responses missing). We follow the same analytic methodology as in Experiment 1: calculating a battery of quantitative measures from each stimulus–response sequence. Table 3 lists the means and standard deviations of these measures over the sample of 40 participants and a comparison sample of 9,600 model runs.

Linearity and overall accuracy. All 40 data sets reveal a substantial degree of stochasticity, and a whole distribution of responses is obtained for each stimulus level. The corresponding conditional mean ratings $\bar{R}(S) = \mathbf{E}(R|S)$ constitute the Stevens function for these data. It is linear to an excellent approximation for our perceptual modality: The Stevens's exponent n in Equation

Table 3
Empirical Constraints Derived From the Category Rating Experiment

Phenomenon	Brief description	Empirical	Model
Stevens's law	The mean rating $\bar{R}$ is approximately a power function of the stimulus intensity: $\bar{R} = aS^n$. For line length, $n \approx 1.0$.	$R^2 = .77$ (.08)	$R^2 = .77$ (.09)
Nonuniform response distribution	The response distribution has a peak in the middle even when the stimulus distribution is uniform.	$s = 1.77$ (0.24)	$s = 1.93$ (.55)
Nonstationary response distribution	The response distribution becomes progressively less uniform over time.	$\Delta s = .55$ (.35)	$\Delta s = .21$ (.37)
Gradual trend	There is spontaneous gradual drift of the average response levels (ARLs).	$\Delta\text{ARL} = .49$ (.58)	$\Delta\text{ARL} = .27$ (.64)
Compensatory context effect	Under skewed stimulus distributions, category ratings shift in the direction that attenuates the skew.	$d = -.21$ (.43)	$d = -.53$ (.51)
Transfer effect	When the context changes, the old response levels persist (temporarily) under the new circumstances.	Figure 13	Accounted for
Sequential effect	The current response R_t is positively correlated with the previous response R_{t-1}.	$r = .34$ (.12)	$r = .17$ (.11)
Similarity effect	The magnitude of the sequential effect depends on the similarity between the consecutive stimuli S_{t-1} and S_t.	Figure 14	Accounted for

Note. The measures of the various effects are defined in the text. Means and standard deviations (in parentheses) are reported for the sample of 40 observers and for 9,600 model runs. See Table 1 for a complementary list of identification phenomena and Table 4 for additional model fits.

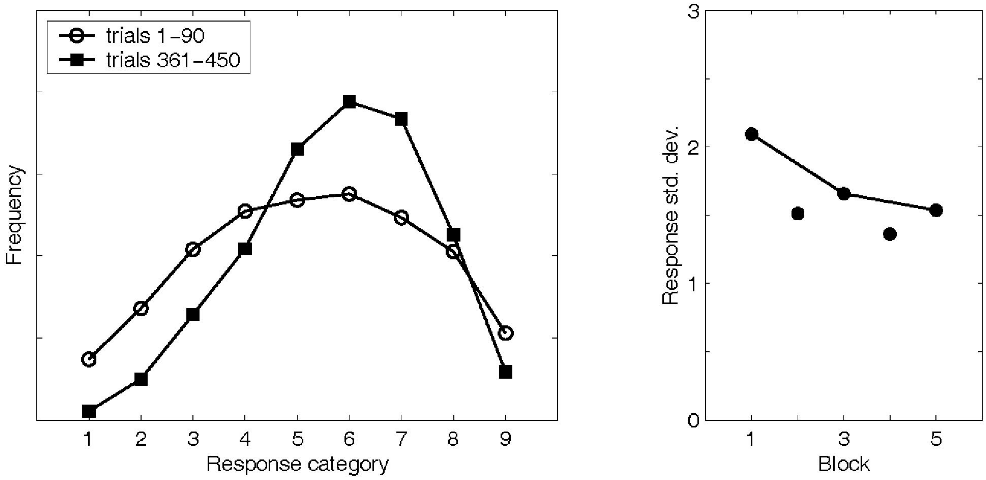

Figure 12. The response distribution becomes progressively less uniform over time. Left: Response histograms during Blocks 1 and 5. Right: Response standard deviations (std. dev.), averaged across participants.

8 is empirically indistinguishable from unity.[8] This replicates the results of a pilot study (Petrov & Anderson, 2000) and numerous other sources (see Wiest & Bell, 1985, for a meta-analysis of 70 studies).

This linearity justifies the linear perceptual Equation 10 in the model and the use of linear (rather than geometric) averages and regressions in the data analysis. In particular, the accuracy of each observer is conveniently defined as the squared correlation between stimuli and responses. This statistic ranges from .57 to .90 over the sample of 40 participants (mean $R^2 = .77$, $SD = .082$).

Response distributions. One salient feature of the data is that the response distributions are distinctly nonuniform. The overall response histogram is 276, 775, 1,624, 2,396, 3,220, 3,501, 3,088, 2,146, and 717. All 40 observers used the middle of the scale more often than the extremes. This replicates our earlier results (Petrov & Anderson, 2000).

As in Experiment 1, we use the response standard deviation as a measure of this nonuniformity. It ranges from 1.10 to 2.31 in the sample (mean $s = 1.77$, $SD = 0.24$). For comparison, if the 450 responses were evenly distributed in 9 categories, s would be 2.58. Values below 2.32 indicate significant departure from uniformity ($p < .01$).

The nonuniformity of response distributions appears to increase with time. To track the dynamics, we partitioned each sequence into five 90-trial blocks and pooled the responses separately. The left panel in Figure 12 plots the histograms for Blocks 1 and 5. The late distribution is markedly more peaked than the early one. This is quantified in the right panel by the response standard deviation for each block, calculated individually and then averaged across participants. There is a clear decreasing trend. (Recall that Blocks 1, 3, and 5, but not 2 and 4, have uniform presentation frequencies.)

This trend is evident at the level of individual participants as well. Let s_1 and s_5 denote the standard deviations of responses during Trials 1–90 and 361–450, respectively. The ratio s_1/s_5 is greater than unity for 38 of the 40 observers and reaches statistical significance for 29 of them ($p < .05$). This strongly suggests that the response distributions tend to become progressively less uniform over time. This nonstationarity is an important manifestation of the dynamism of the system. To our knowledge such analyses have not been reported in the past, and therefore the robustness of this phenomenon is a matter for further investigation.

As a target for modeling, the difference $\Delta s = s_1 - s_5$ quantifies the decrease of response variability for a given observer. This statistic has a mean of 0.55 and a standard deviation of 0.35, matched-sample $t(39) = 9.94$, $p < .0001$. We add it to the battery of measures in Table 3.

Context, transfer, and primacy effects. The response policy changes over time, both endogenously and in response to external factors such as context manipulations. The *average response level* (ARL) defined in Equation 5 is the tool we use to track these changes. It is calculated in exactly the same way as in Experiment 1. The profiles of the individual participants are quite noisy, but with averaging a pattern begins to emerge (see Figure 13).

Perhaps the most salient feature of the ARL profiles, both at the individual and group levels, is a general trend upward. The ARLs increase by one-half category unit (on average) by the end of the session. The thin line in Figure 13 (top) accounts for over 6% of the variance of the individual profiles, $F(2, 397) = 15.0, p < 10^{-6}$. This highly significant, if unexpected, result provides additional evidence that the response distributions are not stationary.

Equation 19 quantifies the magnitude of this effect as a target for modeling. See Table 3 for details.

$$\Delta\text{ARL} = (\text{ARL}_9 + \text{ARL}_{10} - \text{ARL}_1 - \text{ARL}_2)/2. \quad (19)$$

The effects of the experimental manipulations are superimposed over the general trend and obscured by it. In particular, the trend counteracts the local attempts to deflect the ARL downward. To aid interpretation and comparison with the identification data, we subtracted the trend from all ARLs to produce the detrended profiles in the bottom panel of Figure 13. Consider Group 1, presentation schedule UHULU. The first uniform block (90 trials;

[8] The correlations between the functions $S^{0.95}$, $S^{1.00}$, and $S^{1.05}$ are greater than .99 in the domain [250; 700].

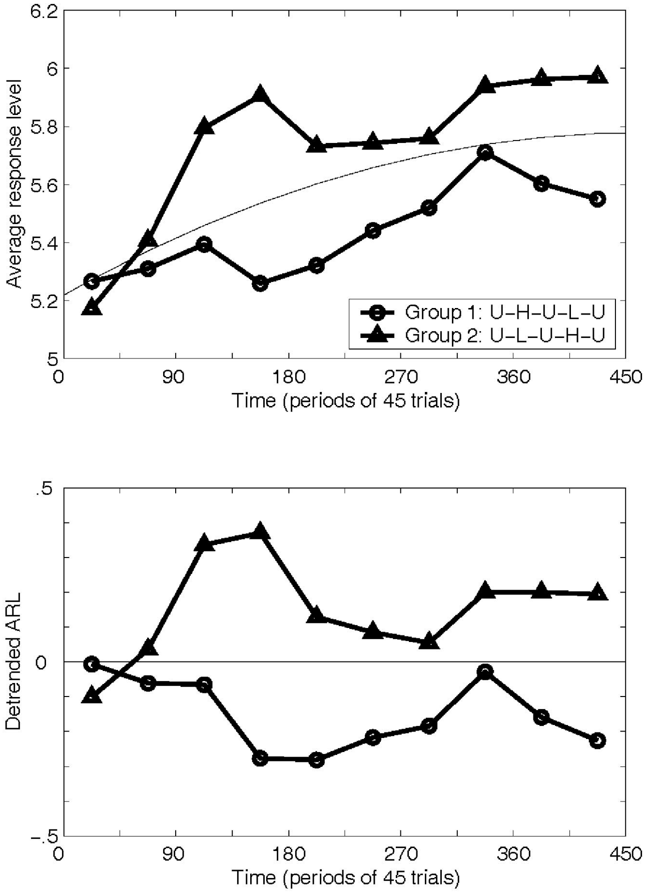

Figure 13. Compensatory context effect in category rating. The average response levels tend to decrease in negatively skewed (H) and increase in positively skewed (L) blocks. There is also a general trend upward (top). It is subtracted away (bottom) to aid comparison with Figure 6. U = uniform; ARL = average response level.

two points on the profile) establishes a baseline ARL. The high (negatively skewed) block of the next 90 trials gradually deflects the ARL downward. This suggests a compensatory effect: Under a preponderance of long stimuli, the responses slide toward the short end of the scale in agreement with the classic results (e.g., Parducci, 1965; Parducci & Wedell, 1986). Note that this is the exact opposite of the assimilatory effect observed in Experiment 1 (see Figure 6).

The third block reverts to uniform presentation frequencies, and the response level in Group 1 turns upward again. It fails to reach the baseline established by the first block, however. Thus we have two blocks with identical stimulus distributions and nonidentical detrended response levels. This suggests transfer from Block 2. The fourth block is low (positively skewed) and deflects the ARL upward. Finally, the last uniform block washes away the effect of Block 4. All transitions are gradual, providing further evidence for transfer effects. The profile for Group 2 shows a complementary pattern.

The approximate, though admittedly imperfect, symmetry of the two profiles increases our confidence that the variance in Figure 13 is not due to sample fluctuations. A repeated measures analysis of variance points to the same conclusion. Each participant contributes five data points to this analysis: u_1, l, u_2, h, and u_3. Each point is the detrended ARL during the second half of the corresponding block. In particular, point h is calculated over Trials 136–180 for Group 1 and 316–360 for Group 2; point l is calculated over Trials 316–360 for Group 1 and 136–180 for Group 2. The first half of each block is left out to minimize transfer-related contamination.

The results show a highly significant compensatory context effect, $F(2, 156) = 6.61$, $p < .002$, within subjects. The mean detrended ARLs are $+0.17$, -0.03, and -0.04 under distributions L, U, and H, respectively. The difference $d = h - l$ in Equation 6 has a mean d of -0.21 and a standard deviation of 0.43, matched-sample $t(39) = -3.06$, $p < .002$.

The between-subjects Group factor is significant too, $F(1, 38) = 4.38$, $p < .05$. Under our counterbalanced design, this can be attributed only to the relative order of the nonuniform blocks. More concretely, Figure 13 suggests that the early manipulation in Block 2 produces a deflection in the ARL that is not fully undone by the subsequent manipulation in the opposite direction. Such primacy effect has been observed before (Haubensak, 1990, 1992) and is consistent with the memory-based view advocated here.

Sequential effects. The rating data contain two kinds of sequential dependencies: short term, extending one trial back, and long term, extending some tens of trials back. The exact pattern and relative importance of these dependencies are analyzed in a hierarchy of autoregression models detailed in Appendix B and Figure B1. Briefly, the response R_t on trial t depends on three components as described in Equation 20. The overall Stevens function, averaged across the whole session without recourse to any time-dependent variables, is represented by the static component aS_t. A rapid transient component replicates the classic sequential effects: assimilation toward the previous response R_{t-1} and contrast with the previous stimulus S_{t-1} (e.g., DeCarlo & Cross, 1990; Jesteadt, Luce, & Green, 1977; Lockhead & King, 1983). Note that these two variables do not act in isolation but through the residual-like term $(R_{t-1} - aS_{t-1})$ in Equation 20. Finally, there is also a slow component $c\text{ARL}_t$ that accounts for the gradual drift of the average response levels. The variable ARL_t is calculated over a roving temporal window extending 30 trials back in time. To our knowledge, the latter component has not been documented before, and therefore its robustness and generality are only tentatively established.

$$\hat{R}_t = aS_t + b(R_{t-1} - aS_{t-1}) + c\text{ARL}_t. \qquad (20)$$

The sequential structure in the data thus appears fully compatible with the two-tiered processing in ANCHOR (see Figure 1). The perceptual transformation is completely atemporal (Equation 10). The central subsystem then introduces sequential effects at time scales reminiscent of the twofold activation dynamics in Figure 9.

Moreover, the sequential effects are modulated by interstimulus similarity. This similarity effect echoes the butterfly pattern in the identification data (see Figure 4) and supports the anchor selection mechanism in the model, which is similarity driven too (Equation 13). To quantify the magnitude of the sequential effects in the rating data, we calculate the autocorrelation coefficient $r = \text{corr}(\text{res}_t, \text{res}_{t-1})$ of the residual time series $\text{res}_t = R_t - aS_t$ and plot it as a function of the interstimulus

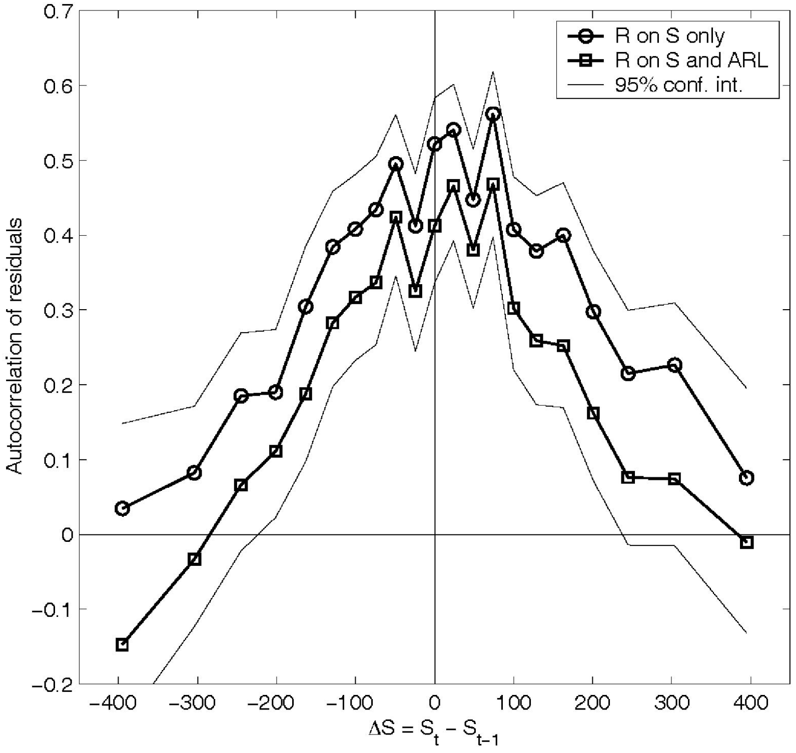

Figure 14. The magnitude of the sequential effect depends on the similarity between the consecutive stimuli S_{t-1} and S_t. The two curves plot the autocorrelation of the residuals from Equations B1 and B3, respectively. See Appendix B for details. Compare with Figure 4. R = response; ARL = average response level; conf. int. = confidence interval.

difference $\Delta S = S_t - S_{t-1}$ (see Appendix B for details). The clear triangular pattern in Figure 14 replicates earlier reports of the similarity effect (DeCarlo, 2003; DeCarlo & Cross, 1990; Jesteadt, Luce, & Green, 1977; Ward, 1979).

As a target for modeling, we summarize the sequential effects with a single number per observer. The autocorrelation coefficient of all residuals is an easily computed, reliable overall measure. Its mean in our sample is $r = .34$ ($SD = .12$).

Table 3 summarizes the yield of the category rating experiment. It replicates all phenomena falling within its scope: linearity of the scale for length, context, transfer, primacy, and sequential effects of various kinds. A novel effect—nonstationarity of the response distribution—is discovered, furnishing further evidence for the dynamic nature of scaling.

Building a Scale Through Competitive Learning

Four of the five ANCHOR mechanisms—perception, anchor selection, base-level activation, and correction—account successfully for a broad range of identification phenomena (see Table 1). The anchors in all simulations so far, however, are set manually in advance. This leaves the model wide open to the criticism that it fails to address the real essence of the scaling problem. The competitive learning mechanism imbues ANCHOR with the ability to learn the locations of its anchors.

Updating the Anchor Locations

The competitive learning rule is straightforward and well known (Kohonen, 1995; Rumelhart & Zipser, 1985). Whenever a stimulus is classified under a particular category, the corresponding anchor location is updated to accommodate the new member. The new anchor location $L_{i*}^{(t+1)}$ is simply a weighted sum of the old location $L_{i*}^{(t)}$ and the target magnitude $M^{(t)}$ on trial t (Equation 21). The learning rate α is a free parameter that we fix to .3 in all simulations. Exactly one anchor, with index $i*$, is updated on each trial. If there is feedback, this is it; otherwise the system's own response designates the anchor for update.

$$L_{i*}^{(t+1)} = (1 - \alpha)L_{i*}^{(t)} + \alpha M^{(t)}. \tag{21}$$

This learning mechanism has far-reaching consequences. First and foremost, the location of each anchor is a running average of the magnitudes of all stimuli classified under the associated response category. Therefore the anchors represent true prototypes. Second, the relative weight of each training exemplar decreases exponentially as new exemplars are averaged in. The model can thus quickly readjust its anchors if the environmental conditions change. Third, in the absence of feedback the model reinforces its own policy, thereby promoting the consistency of the stimulus–response mapping. More concretely, the location L_{i*} moves closer to the target M, and should the same target appear again, the match between the two will be better than before, which in turn improves the chances of responding R_{i*} again (cf. Equations 13 and 17).

The fourth implication of the learning rule is that the anchors tend to spread out and cover all available regions of the stimulus probability distribution. This property emerges from the interaction with the softmax selection mechanism (Equation 14) and is a characteristic feature of all competitive learning systems. Consider an extreme example with only two anchors, A and B, planted in the exact middle of the magnitude continuum. On the first trial one of them will move as the first exemplar is averaged in. For concreteness, suppose that Anchor A moves upward. It now has advantage in future competitions for long targets, whereas B has advantage for short ones. If a short stimulus now comes, it will probably be averaged into Anchor B, thereby pulling it downward. As this process continues, the two anchors partition the magnitude space among themselves.

The Importance of Corrections

The correction mechanism makes a crucial contribution at this point. The ordering it establishes among the anchors binds them into a self-organizing map with linear topology (Kohonen, 1995). The anchors not only partition the space but do so in a systematic manner: Anchor 1 takes the lowest "estate," Anchor 2 the second lowest, and so forth. This happens very reliably regardless of the initial anchor locations.

The following simulation demonstrates this in the extreme case of reverse initialization. The full ANCHOR model, with all five mechanisms in place, is presented with a sequence of stimuli generated at random in the range of 250 to 700 pixels. The model runs without feedback under its default parameters. However, the initial anchor locations are "upside down": $L_1 = .675$, $L_2 = .625, \ldots, L_9 = .275$.

Figure 15 plots the evolution of three anchor locations. Anchor 1 traverses most of the magnitude continuum during the first 250 trials; Anchor 9 undertakes an equally resolute march in the opposite direction. To illustrate this process, consider the first trial after the initialization. Suppose the first target is $M = .500$, and the softmax mechanism happens to select Anchor 7, location $L_7 = .375$. The large discrepancy triggers a two-point correction and the model responds "9" (Equation 18). Now, Anchor 9 is updated to location $L_9 = .343 = .7 \times .275 + .3 \times .500$ (Equation 21). In other words, Anchor 9 makes a large step upward driven by the pressure that L_9 should be greater than L_7. Before long the anchors rearrange themselves in agreement with the ordering relations imposed by the correction mechanism. A stimulus–response homomorphism settles in.

Of course, this homomorphism does not come ex nihilo. It is grounded in prior knowledge about numbers, on one hand, and in the systematic perceptual transformation (Equation 10), on the other. This is precisely the kind of knowledge that human observers, aided by the instructions, bring to the task. The contribution of the

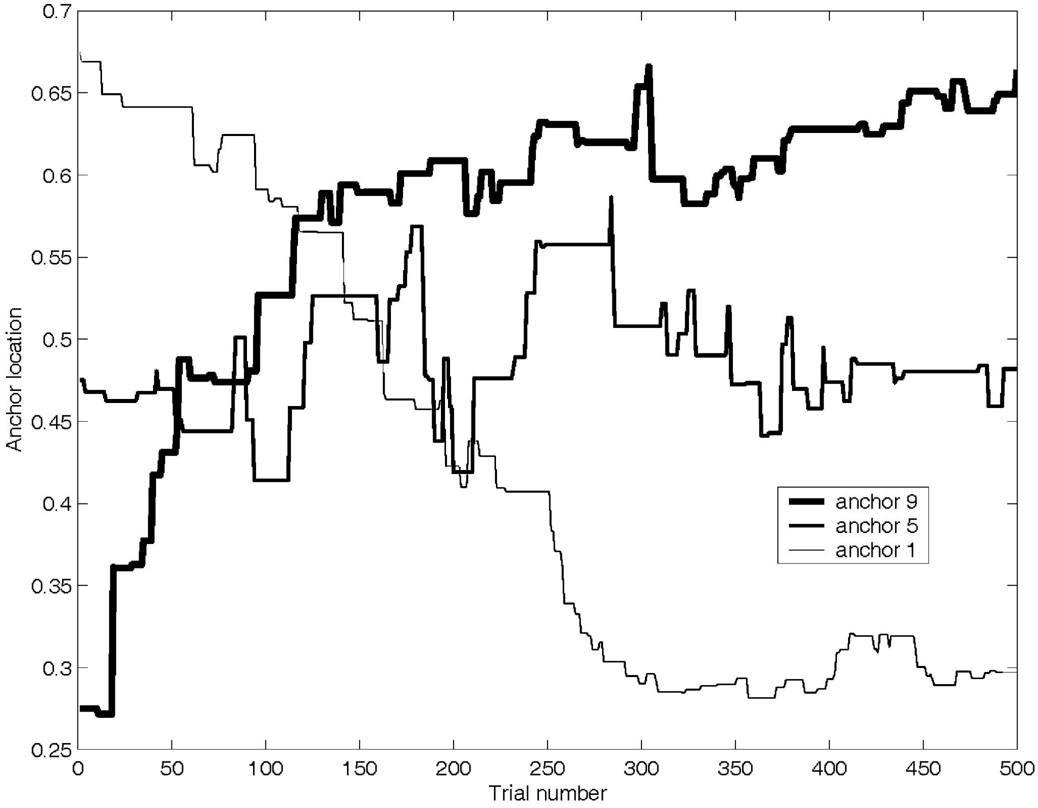

Figure 15. The competitive learning mechanism sets the anchor locations in agreement with the ordering imposed by the correction mechanism. No feedback is required. See text for details.

correction mechanism is to align, locally, the ordering of magnitudes with the ordering of responses. Competitive learning then consolidates this local alignment into a global homomorphism.

With time, the initial anchor configuration is washed out and the system enters a steady state (see the late trials in Figure 15). Each anchor finds its preferred location on the continuum and all remaining variability is just random walk around this equilibrium level. As long as the stimulus distribution remains stationary, the preferred locations do not change. Our simulations indicate that the initial anchor configuration has no effect on the equilibrium levels, because of the exponential discounting implicit in Equation 21.

Where Do All These Anchors Come From?

We have established so far that once a set of anchors is present in memory, the model is able to adjust the anchors' locations properly. This still leaves the deeper question of the genesis of these anchors in the first place. How are all these anchors created in memory? And when are they created?

At the level of analysis pursued in this article, the ability to form anchors is postulated as an architectural primitive. That is, it is taken for granted that the human brain is able to establish magnitude-label associations. The growing field of memory psychophysics provides abundant evidence that such associations can be formed in human memory and maintained over extended periods of time (Algom, 1992; Kerst & Howard, 1978; Moyer, Bradley, Sorensen, Whiting, & Mansfield, 1978; Wiest & Bell, 1985). A typical mnemophysical experiment comprises two sessions, usually on consecutive days. On Day 1 the participants are trained to associate verbal codes (such as consonant–vowel–consonant syllables) with the stimuli. On Day 2 they are asked to imagine each stimulus and estimate its magnitude from memory, cued only by its verbal code. A very robust finding is that people give systematic power-law ratings comparable with those given to immediately perceived stimuli (see Algom, 1992, for a review).

In the ACT–R architecture, anchors are instances of its fundamental declarative memory primitive: the *chunk* (Anderson & Lebière, 1998). The typical ACT–R chunk combines two or more discrete (symbolic) pieces of information. They are created either through perception or in the action side of production rules. ACT–R avoids the creation of duplicate chunks. Rather, the new exemplar is merged with the old, boosting its base-level activation (Equation 15).

ANCHOR introduces continuous (analog) magnitudes into this general framework. Anchors are heterogeneous associations: One of their elements is continuous; the other is discrete. The two kinds have different properties with respect to anchor creation. Anchor magnitudes are averaged along the continuum according to Equation 21. Anchor labels, alternatively, are discrete and cannot be mixed. They establish the distinctive identity of each anchor. Whenever a new label is encountered, a new anchor is created.

In scaling situations, these labels are categories on the response scale. Thus, the generation of new anchors ultimately reduces to generation of novel response labels. There are two sources of this novelty. The most straightforward one is external demonstration and/or feedback whereby the response associated with a given stimulus is simply presented to the system. This is sufficient for absolute identification. Category rating, however, requires an internal generative source. This is yet another important function of the correction mechanism. On the basis of explicit prior knowledge, and within bounds established by the instructions, it can construct responses that have never appeared before.

This generative property allows the model to unfold the whole scale from a single arbitrarily placed anchor. To illustrate, suppose there is only one anchor in memory, labeled *5* and located in the middle of the magnitude continuum. Suppose further that the stimulus presented on the first experimental trial is rather short. The anchor is selected for lack of competition, but it doesn't quite match this target (Equation 17). Therefore, the anchor response is decremented by one category unit. This novel response triggers the creation of a new anchor with label *4.* Its location is set to the current target magnitude. The first stimulus classified under the new category thus becomes the prototype for it. Still more anchors will be created on subsequent trials until there is an anchor for each response within the range specified by the instructions. (Recall that Equation 18 clips the responses between $R_{\min}$ and $R_{\max}$.) The competitive learning mechanism soon adjusts the new set of anchors to the appropriate locations. The scaling problem is solved.

The Dynamics of Scaling

ANCHOR is a dynamic adaptive system. It maintains an internal state that determines its responses to external stimuli. It constantly adjusts this internal state: *Obligatory learning* is one of its foundational principles. Two learning mechanisms incrementally update the base-level activations and locations of the anchors (Equations 16 and 21). Presumably, the cognitive architecture has evolved machinery to track those statistics of the stimulus distribution that tend to optimize performance in an ever-changing world (Anderson, 1991). The unfolding of the rating scale is an excellent example of this adaptability.

The same mechanisms that are so instrumental for the system's overall success, however, have some unintended consequences that lead to suboptimal behavior in the carefully counterbalanced environment of a scaling experiment. Many of the phenomena listed in Tables 1 and 3 are of this nature. They are very important theoretically because they reveal the underlying architecture of cognition. The sequential and transfer effects are the most obvious manifestation of the incremental learning mechanisms (cf. Figures 10 and 11). The dynamic origin of the other effects is harder to discern because, as many unintended consequences do, they emerge from subtle interactions in the system.

Nonuniform Response Distributions

The base-level learning mechanism is prone to unstable dynamics in the absence of feedback. Because the model reinforces its own responses, a single runaway anchor could strengthen its own activation and take over the selection process. The correction mechanism eliminates this danger by providing a means for redistribution of strength among the anchors. For instance, suppose that for some reason the base-level activation of Anchor 4 completely dominates all others. This anchor is selected on every trial, but its magnitude does not match every stimulus. Hence large discrepancies occur and trigger frequent corrections. This generates a fair number of 2s, 3s, 5s, and 6s among the final responses. The corresponding anchors are thus strengthened, and their competitiveness increases. Soon the dominance of Anchor 4 is broken. The redistribution effect then propagates to Anchors 1, 7, and so forth.

As an emergent consequence of this redistribution, ANCHOR's response histograms tend to be smooth. If a response category occurs with a certain frequency, chances are that the neighboring categories occur with comparable frequencies. This constrains the shape of the response distributions; they cannot have gaps or too many modes.

This does not mean that the response distributions must be flat. Quite the contrary, the self-reinforcing dynamics of the activation learning mechanism is still in place. It tends to amplify any deviations from uniformity. Flat distributions are therefore unstable; they are repelling points in the phase portrait of the system. Even under strictly uniform stimulus frequencies, the inevitable symmetry-breaking fluctuations grow to macroscopic proportions. The stabilizing role of the correction mechanism is to smooth out this process and prevent it from getting out of hand. Some anchors gain strength at the expense of others but are soon forced to redistribute most of it back. The interaction of these opposing mechanisms typically results in smooth, nonuniform distributions that grow progressively peaked over time until they reach a steady state. This is exactly the kind that is reliably produced by humans as well (see Figure 12). The response histogram buckles up rather than down because of restricted correction opportunities and hence weaker redistribution of strength toward the edges.

Context Effects: Push and Pull

What happens when the stimulus distribution itself is not uniform but skewed in one direction or another? Of course, the strong stimulus–response correlation immediately skews the response distribution in the same direction. This obvious consequence can be partialled out mathematically to address the more interesting question of whether the response policy changes. The *average response level* (ARL) is defined to track these changes in a commensurable way (Equation 5).

Apart from global parameters such as correction thresholds that remain fixed at all times, the response policy of the model depends entirely on the locations and base-level activations of its anchors. Both are systematically affected by the stimulus distribution. ANCHOR thus predicts context effects on a principled basis.

Figure 16 illustrates the main factors involved. A configuration with uniformly located, equally active anchors serves as baseline (top). The cone around each anchor represents the corresponding goodness score as a function of the target magnitude (Equation 13). The shaded areas delineate the resulting partitions on the magnitude continuum. (The stochasticity in Equation 14 is ignored for simplicity.) Each anchor is characterized by two independent quantities: location L_i and base-level activation B_i plotted on the horizontal and vertical axes in Figure 16, respectively. Each quantity is updated by a separate mechanism and affects the observable responses in specific ways. Notably, the two mechanisms push in opposite directions, and the overall context effect depends on the relative strengths and interactions between these opposing forces.

Activation learning leads to assimilation. Whenever the activation B_i of an anchor i is strengthened, its chances for selection on subsequent trials increase. The "sphere of influence" of the reinforced anchor thus expands on both sides (see Figure 16, middle). Some targets that used to be labeled *1* or *3* in the baseline configuration are now labeled *2.*

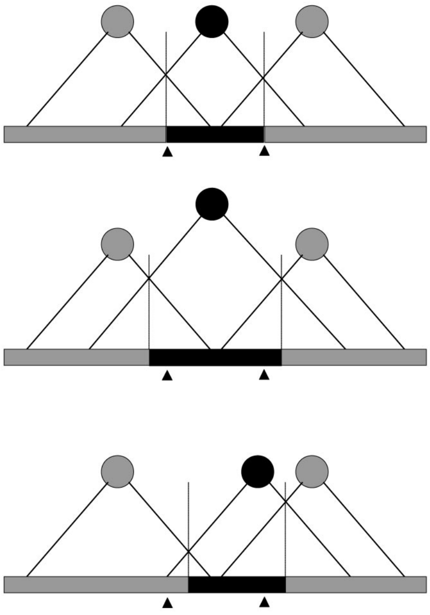

Figure 16. Schematic comparison of ANCHOR's learning mechanisms. The horizontal axis plots the location of each anchor, and the vertical axis plots its base-level activation. Top: Baseline configuration. Middle: Assimilation in activation learning. Bottom: Inversion in location learning—whenever an anchor shifts to the right, the responses shift to the left. See text for details.

Under a skewed stimulus distribution the base-level activations quickly form an ascending ladder (not shown in Figure 16). This introduces a progressively stronger bias toward the more frequent responses. The observable consequence is an assimilatory shift of the average response levels (see Figure 11). As earlier simulations demonstrated, the correction mechanism attenuates this assimilation but does not alter its direction. The process is very similar to the self-reinforcing response nonuniformity discussed in the previous section. The only difference is that in skewed contexts the nonuniformity is imposed from the outside and the dynamic equilibrium is shifted toward the corresponding end of the scale rather than relaxing in its middle. Again, redistribution of strength by the correction mechanism prevents runaway activations in the absence of feedback.

Competitive learning leads to compensation. The effect of the location learning mechanism (Equation 21) is less transparent, even somewhat counterintuitive. Consider the bottom panel in Figure 16. The location of Anchor 2 has been shifted to the right, presumably by averaging in a long stimulus. The zone dominated by this anchor shifts accordingly without growing in size. Comparison with the baseline configuration (marked by the little triangles) reveals that a region formerly labeled *2* is now labeled *1* and a region formerly labeled *3* is now labeled *2.* Thus, there is a systematic decrement of the overt responses.

It is convenient to formulate a descriptive rule of thumb to refer to this effect. According to this *inversion rule,* whenever the location of any anchor increases, the average response level decreases, and vice versa.

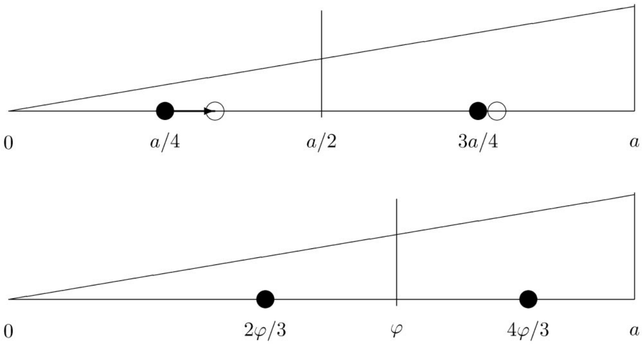

Figure 17. Schematic illustration of the location shifts in skewed contexts. Top: Initial configuration with two uniformly spaced anchors (solid circles). The open circles indicate the expected values (or barycenters) of the resulting partitions of the magnitude continuum. The competitive learning mechanism tends to move the anchors toward these barycenters. The system converges to a steady state (bottom) with boundary at $\varphi = (\sqrt{5} - 1)a/2$. See the main text and Appendix C for details.

The running averaging Equation 21 tracks the probability density of the magnitude distribution. In skewed contexts, this shifts the anchor locations toward the heavier end of the continuum. Figure 17 illustrates the qualitative situation. Let us assume for simplicity that there are only two anchors and their initial locations are indicated by the solid circles in the top panel. The resulting boundary bisects the magnitude range (assuming deterministic anchor selection). The open circles indicate the expected values (or barycenters) of the two partitions. Notice that they are systematically displaced to the right of the anchors. Thus, averaging new targets in tends to shift the anchors rightward. This in turn redefines the partition boundary and the barycenter locations, setting the stage for further shifts. The process converges to the steady state illustrated in the bottom panel of Figure 17. Each anchor is located at the exact barycenter of its corresponding partition and the system is in stable equilibrium (see proof in Appendix C). The steady state is unique and all initial anchor configurations converge to it.

Our simulations indicate that the dynamics of the full ANCHOR model is qualitatively the same. In skewed contexts, the anchors converge to a steady state shifted toward the frequently sampled end of the continuum. By the inversion rule, this deflects the ARLs in the opposite direction—a compensatory context effect.

This compensation counteracts the assimilation driven by the activation mechanism. Hence the two learning mechanisms in ANCHOR tend to push the response levels in opposite directions. This opposition dampens any big fluctuations in either direction and aids the correction mechanism in preserving the balance of the system as a whole.

The overall context effect is determined by the relative strength of these opposing forces. The exact outcome depends on the model parameters and in particular the weight H in Equation 13 and the learning rate α in Equation 21. Our simulations indicate that the compensatory influence is much stronger than the assimilatory one across most of the parameter space, including the defaults. This is probably due to the logarithmic compression in the base-level Equation 16. Thus ANCHOR accounts for the compensatory context effect in category rating (see Figure 13).

The role of feedback. The competitive learning mechanism is switched off by external feedback. More precisely, the explicit feedback in absolute identification fixes the anchor locations to the internal images of the corresponding stimuli regardless of their presentation frequencies. The only location variability comes from random fluctuations in the perceptual subsystem (Equation 10). Thus, no systematic shift of the ARL is introduced by the competitive learning mechanism during absolute identification.

The base-level activations of the anchors, conversely, are qualitatively the same regardless of feedback. Skewed stimulus distributions always lead to progressively ascending or descending activation levels, reflecting the presentation frequencies either directly through feedback or indirectly through the system's own responses, which are highly correlated with the stimuli. Thus, the base-level learning mechanism consistently promotes assimilation.

In summary, ANCHOR makes the following predictions about the context effects in different tasks. In absolute identification, the system must exhibit assimilatory context effects because the compensatory tendency is switched off by the explicit feedback. This strong prediction is in excellent agreement with the results of Experiment 1. The context effects in category rating are not so unequivocally determined. Two opposing tendencies operate in the absence of feedback: activation-driven assimilation and location-driven compensation via the inversion rule. The overall outcome is parameter dependent and hence can vary across participants and experimental conditions. The compensatory tendency dominates most of the time: The system usually exhibits compensation but is also capable of assimilation in rare cases. This explains the prevailing pattern in the literature and the results of Experiment 2.

Simulation Experiments

ANCHOR accounts for a wide range of phenomena in absolute identification and category rating. The simulations so far, however, have been mostly qualitative and focused on isolated effects. But can the model satisfy these diverse constraints all at once with a unified parameter setting? The existence of opposing tendencies and trade-offs in the system makes this question far from trivial. We need to test the model on the battery of quantitative measures in Tables 1 and 3.

The main unit of analysis is the stimulus–response sequence for each run, just as the behavioral data are analyzed individually for each observer. The statistics are computed from each model sequence by the same software that analyzed the behavioral data: six measures for absolute identification (see Table 1) and six more for category rating (see Table 3). This methodology imposes the strongest possible empirical constraints on the model: unified fits to a comprehensive data set collected with identical stimuli.

To increase the informativeness of the simulations even further, we test a whole suite of partial models alongside the full ANCHOR model. This systematic approach reveals the value added by each computational mechanism. It also validates the statistical measures. For example, a static model that cannot learn should score zero on all dynamic effects.

Method

The simulation method is designed to mimic the behavioral experiments as closely as possible. The performance of each model is assessed both in absolute identification with feedback and in category rating without feedback.

Models and parameters. A hierarchy of seven models cover the five computational mechanisms postulated by the theory. Each new model introduces one new mechanism to its predecessor(s): $\mathcal{M}_{\mathrm{P}}$, $\mathcal{M}_{\mathrm{PS}}$, $\mathcal{M}_{\mathrm{PSC}}$, $\mathcal{M}_{\mathrm{PSA}}$, $\mathcal{M}_{\mathrm{PSAC}}$, $\mathcal{M}_{\mathrm{PSCR}}$, and $\mathcal{M}_{\mathrm{PSACR}}$.

Model $\mathcal{M}_{\mathrm{P}}$ consists of the perceptual mechanism (Equation 10) with deterministic nearest-neighbor anchor selection. All other mechanisms are silenced by setting their controlling parameters to zero. For absolute identification $\mathcal{M}_{\mathrm{P}}$ has only one free parameter: the perceptual confusability coefficient k_p in Equation 10. The nine anchors are fixed at the internal images of the stimuli. For category rating, a second free parameter L_1 allows the location of Anchor 1 to vary. Anchor 9 is fixed at $L_9 = .675$, and the other anchors are spaced evenly in between.

Model $\mathcal{M}_{\mathrm{PS}}$ introduces memory noise (Equation 12) and nondeterministic anchor selection (Equations 13 and 14). In this and all subsequent models, the perceptual confusability parameter is fixed to the value $k_p = .04$ derived from the Weber fraction (Equation 11). The overall accuracy of model $\mathcal{M}_{\mathrm{PS}}$ is thus controlled by the memory noise k_m and the softmax temperature T. There are no corrections and no learning ($H = 0$, $\alpha = 0$). The anchors are fixed evenly within their ranges as before: from $L_1 = .275$ to $L_9 = .675$ for absolute identification and from an adjustable L_1 to $L_9 = .675$ for category rating.

Model $\mathcal{M}_{\mathrm{PSC}}$ introduces the correction mechanism (Equations 17 and 18). It has one additional free parameter relative to $\mathcal{M}_{\mathrm{PS}}$: the cutoff c. The four thresholds for the five correction increments are set to fixed multiples of this parameter: $\{-3c, -c, 0.9c, 2.7c\}$. The slight asymmetry in favor of upward corrections is motivated by the upward trend in the empirical response levels (see Figure 13).

Model $\mathcal{M}_{\mathrm{PSA}}$ introduces the base-level activation mechanism to model $\mathcal{M}_{\mathrm{PS}}$. This is the first model in the hierarchy that learns during the run and is thus capable of dynamic effects. Without the correction mechanism, however, the activations are prone to self-reinforcing collapse in the absence of feedback. Hence $\mathcal{M}_{\mathrm{PSA}}$ can perform only the identification task. Its free parameters are k_m and T, as in $\mathcal{M}_{\mathrm{PS}}$, plus the history weight H in Equation 13. The activation Equation 16 itself is parameter free.

Model $\mathcal{M}_{\mathrm{PSAC}}$ combines the activation and correction mechanisms of $\mathcal{M}_{\mathrm{PSC}}$ and $\mathcal{M}_{\mathrm{PSA}}$. The only missing component relative to the full ANCHOR model is the competitive learning mechanism; it is switched off by setting its learning rate to zero. Four parameters are adjusted for absolute identification: k_m, T, H, and c. Category rating requires an additional free parameter L_1 as $\mathcal{M}_{\mathrm{PSAC}}$ still lacks a more principled means for adjusting its anchor locations.

Model $\mathcal{M}_{\mathrm{PSCR}}$, by contrast, has competitive learning but lacks activation learning. (The subscript R stands for the running averaging in Equation 21.) For the first time the anchor locations are not fixed in advance but are induced from the stimuli and change over time. In emphasis of the generative potential of this model, only two anchors are available at the beginning of each run: Anchor 1 located at $L_1 = .275$ and Anchor 9 at $L_9 = .675$. This initialization favors the extreme responses on the scale and thereby biases the model against the peaked empirical distribution that it has to fit. If $\mathcal{M}_{\mathrm{PSCR}}$ develops a preference for the middle of the scale despite that bias, this would be very strong evidence for the robustness of this preference. The learning rate in Equation 21 is poorly determined by the data and is therefore fixed to $\alpha = .3$ for all runs. This leaves only three parameters free: k_m, T, and c. The history weight H is set to zero to silence the activation mechanism.

Finally, model $\mathcal{M}_{\mathrm{PSACR}}$ includes all five mechanisms postulated by the theory and hence is synonymous with the full ANCHOR model. Only Anchors 1 and 9 are available initially, just as in $\mathcal{M}_{\mathrm{PSCR}}$. Two parameters are fixed for all runs ($k_p = .04$ and $\alpha = .3$), and four parameters are free (k_m, T, H, and c).

To factor in the individual differences in the human population, we estimated a separate parameter set for each observer. See Appendix D and Petrov (2001) for details.

Procedure. The first part of the simulation emulates the absolute identification Experiment 1. A three-step procedure is applied for each of the seven models. First, 24 parameter sets are estimated from the 24 empirical stimulus–response sequences, obeying the constraints of the particular model. Appendix D outlines the parameter search algorithm. Second, the model is run on 9,600 fresh stimulus sequences—400 for each parameter set, with feedback. The stimulus presentation schedule is exactly the same as in Experiment 1 and is counterbalanced between runs and within parameter sets: 200 runs under schedule UHULU and 200 runs under ULUHU. Third, a battery of six measures is calculated for each run just as for the behavioral data (see Table 1). The results are then aggregated by averaging each statistic and are reported in the top half of Table 4. To facilitate comparison across models, we presented the same 9,600 stimulus sequences to each model.

The second part of the simulation emulates the category rating Experiment 2. An analogous three-step procedure is applied for each of the seven models. Only this time 40 parameter sets are estimated as there are 40 human observers. Each parameter set is tested on 120 fresh UHULU and 120 ULUHU sequences without feedback. The resulting set of 9,600 sequences is presented to each model. A battery of six measures is calculated for each run, then averaged and reported in the bottom half of Table 4. Part of the measures (e.g., the response standard deviation s) are the same as in absolute identification; the rest are unique for category rating (cf. Table 3).

Results and Discussion

Table 4 reports the mean of each statistic over the 9,600 runs with each model. The rightmost column provides the empirical summaries for comparison. The second, third, and fourth columns quantify the performance of the static models $\mathcal{M}_{\mathrm{P}}$, $\mathcal{M}_{\mathrm{PS}}$, and $\mathcal{M}_{\mathrm{PSC}}$, respectively. Overall, these models fit the gross accuracy levels and little else. As they do not update their internal state from trial to trial, they cannot exhibit any dynamic effects. The corresponding entries in the table are all zeros (or very close thereto), which attests to the validity of the measures.

The introduction of the base-level activation mechanism changes this situation dramatically. Model $\mathcal{M}_{\mathrm{PSA}}$ exhibits strong repetition, context, and practice effects. The correction mechanism in model $\mathcal{M}_{\mathrm{PSAC}}$ attenuates them without altering their qualitative character. It also supplies the checks and balances necessary to stabilize the system during category rating. The rapid transient component of the activation dynamics (see Figure 9) gives rise to an assimilative sequential effect, detected by the residual autocorrelation r. The gradual buildup of strength with frequent use gives rise to an assimilative context effect in both tasks. This is reflected in the positive difference d between the average response levels in high and low presentation blocks (Equation 6). The response distribution becomes markedly nonuniform in the no-feedback condition as $\mathcal{M}_{\mathrm{PSAC}}$ reinforces its own responses, but the redistribution of strength keeps the activations under control. Without corrections, runaway activation leads to absurd nonuniformity and model $\mathcal{M}_{\mathrm{PSA}}$ typically ends up giving the same hyperactive response on all trials.

It is very instructive to compare models $\mathcal{M}_{\mathrm{PSCR}}$ and $\mathcal{M}_{\mathrm{PSAC}}$. Each features a learning mechanism and thereby exhibits dynamic effects. The different mechanisms, however, have different signatures. Most notably, competitive learning gives rise to compensatory context effect in category rating, in sharp contrast to the activation-driven assimilation. The mobility of the anchor locations makes the response distribution nonstationary: $\mathcal{M}_{\mathrm{PSCR}}$ captures both the gradual trend in the average response level (ΔARL)

Table 4
Performance of a Hierarchy of Models on a Battery of Measures of Various Phenomena

	Model							
Phenomenon and statistic	$\mathcal{M}_{\mathrm{P}}$	$\mathcal{M}_{\mathrm{PS}}$	$\mathcal{M}_{\mathrm{PSC}}$	$\mathcal{M}_{\mathrm{PSA}}$	$\mathcal{M}_{\mathrm{PSAC}}$	$\mathcal{M}_{\mathrm{PSCR}}$	$\mathcal{M}_{\mathrm{PSACR}}$	Empirical
Absolute identification								
Transmitted information (*T*)	1.79	1.53	1.57	1.50	1.66	1.57	1.57	1.68
Nonuniform response distribution (*s*)	2.59	2.55	2.52	2.55	2.50	2.50	2.50	2.40
Edge effect (*bow*)	−.42	−.47	−.30	−.38	−.29	−.27	−.31	+.14
Repetition effect (*rep*)	.01	.01	.01	.15	.05	.01	.04	.11
Assimilative context effect (*d*)	−.00	+.00	+.00	+.14	+.11	+.00	+.11	+.14
Practice effect (*pr*)	.00	.00	.00	.06	.01	.01	.02	.06
Bayesian information criterion (BIC)	504	496	491	522	493	504	504	—
Category rating								
Overall accuracy (R^2)	.82	.68	.80	n. a.	.82	.73	.77	.77
Nonuniform response distribution (*s*)	2.08	2.06	2.05	n. a.	1.91	2.56	1.93	1.77
Nonstationary distribution (Δs)	.00	.00	.00	n. a.	−.01	.12	.21	.55
Sequential effect (*r*)	.00	.00	.00	n. a.	.04	.08	.17	.34
Gradual trend (ΔARL)	.00	.00	.00	n. a.	−.01	.08	.27	.49
Compensatory context effect (*d*)	+.00	+.01	+.01	n. a.	+.26	−.62	−.53	−.21
Bayesian information criterion (BIC)	615	595	614	n. a.	537	552	472	—

Note. Each cell reports the mean of 9,600 model runs. Model $\mathcal{M}_{\mathrm{P}}$ has perceptual variability only. Models $\mathcal{M}_{\mathrm{PS}}$ and $\mathcal{M}_{\mathrm{PSC}}$ add anchor selection and response correction. Models $\mathcal{M}_{\mathrm{PSA}}$, $\mathcal{M}_{\mathrm{PSAC}}$, and $\mathcal{M}_{\mathrm{PSCR}}$ introduce activation learning and running averaging of anchor locations. Model $\mathcal{M}_{\mathrm{PSACR}}$ is synonymous with the full ANCHOR model. Compare with the empirical data in Tables 1 and 3. Dashes indicate cells in which BIC is not defined. n. a. = not applicable; ARL = average response level.

and the sharpening of the response variance (Δs). These slow trends also masquerade as sequential effects.

Note that $\mathcal{M}_{\mathrm{PSCR}}$ exhibits no dynamic effects in absolute identification. In fact it is then virtually equivalent to $\mathcal{M}_{\mathrm{PSC}}$ because the external feedback keeps the anchor locations fixed to the images of the stimuli at all times (Equation 21).

Finally, $\mathcal{M}_{\mathrm{PSACR}}$ puts all mechanisms together and completes the model hierarchy. The two learning mechanisms act synergistically with respect to some effects and antagonistically with respect to others. Thus, the sequential effect and the drift of the ARLs are much stronger under the full model than under any of the partial ones. The magnitude of the no-feedback context effect is a compromise between the assimilation in $\mathcal{M}_{\mathrm{PSAC}}$ and the compensation in $\mathcal{M}_{\mathrm{PSCR}}$. $\mathcal{M}_{\mathrm{PSACR}}$ is the only model that can account for the opposite directions of the context effects in the two tasks.

The full model $\mathcal{M}_{\mathrm{PSACR}}$ thus accounts for the qualitative pattern in the data better than any partial model. A quantitative analysis corroborates this conclusion. The *Bayesian information criteria* (BICs; Schwartz, 1978; Wassermann, 2000) for each model are given in Table 4. This is a log-likelihood metric for model selection that includes a penalty term for the number of free parameters (see Appendix D for details). It should be interpreted with caution as many of the technical assumptions justifying its use (e.g., independence) are violated by our learning models. Moreover, finding the optimal log-likelihood is a challenging computational problem, and the values in Table 4 are only approximate solutions. With these caveats, the full model emerges as the winner in the category rating corpus. The identification results seem inconclusive: the BICs of all models are identical within the margin of estimation error. Apparently, the (approximate) BIC is not sensitive enough to resolve the subtle dynamic effects when the task is dominated by feedback.

In conclusion, ANCHOR ($\mathcal{M}_{\mathrm{PSACR}}$) seems to fit the empirical data very well. Qualitatively, it predicts the direction of over a dozen interlocking behavioral effects at once, with the sole exception of the edge effect. Quantitatively, the model reproduces closely the means of the statistical measures in Table 4. It reproduces the range and standard deviations around these means as well, as reported in Tables 1 and 3. This variability is due in part to between-subjects differences captured by the individualized parameter sets and in part to within-subject fluctuations driven by the stochasticity of the mechanisms.

General Discussion

Explanation of the Phenomena

To recapitulate, ANCHOR offers an integrated, principled explanation of a wide range of scaling phenomena:

Capacity limitations. There are random fluctuations and inefficiencies throughout the system. Most of them are nonperceptual and hence persist even for perfectly discriminable stimuli. Various sequential and context effects also erode the accuracy. Furthermore, the decay of the base-level activations limits the total number of anchors that can maintain availability.

Gradual trend. The drift of the average response levels without feedback is a direct consequence of the obligatory learning in the system. As it continuously and incrementally tracks the statistics of the environment, slow oscillations are inevitable. The specific direction of the trend depends on the initial anchor set, stimulus material, various explicit strategies, and other factors specific to each particular setting. The fundamental and robust fact, however, is that the responses are dynamic and nonstationary.

Nonuniform response distribution. The central peak in the response distribution emerges from the self-reinforcing activation dynamics, tempered by the redistribution of strength by the correction mechanism. Even when the stimulus distribution is uniform, the small symmetry-breaking fluctuations grow to macroscopic proportions.

Nonstationary response distribution. The decrease of the response standard deviation over time is a corollary of the drift of the average response levels.[9]

Sequential effects. The positive correlation between consecutive responses is a signature property of activation-mediated priming. The negative correlation between the current response R_t and the previous stimulus S_{t-1} seems mostly an artifact of the regression analyses typically performed in the literature (see Figure B1). To the extent that this "perceptual contrast" is real, it can be attributed to the emergent inversion within the competitive learning mechanism. As illustrated in the bottom panel of Figure 16, whenever an anchor location shifts up the responses tend to shift down and vice versa. Given that the location shifts toward the previous stimulus S_{t-1}, a negative correlation with R_t results.

Similarity effect. The magnitude of the sequential effect depends on the similarity between the consecutive stimuli S_{t-1} and S_t as a direct consequence of the similarity term in the anchor-selection Equation 13.

Repetition effect. When the same stimulus is repeated on two successive presentations, the identification accuracy on the second trial is greater than average because the correct anchor has just been reinforced.

Assimilative context effect in absolute identification. Under skewed stimulus distributions with feedback, the response levels shift in the direction of the skew because a base-level bias accumulates in favor of the frequently used anchors.

Compensatory context effect in category rating. Under skewed stimulus distributions without feedback, the response levels shift in a compensatory direction because of the inversion effect of the competitive learning mechanism. The anchor locations move into the densely sampled stimulus region, which in turn drives the responses in the opposite direction. The assimilatory tendency of the activation mechanism is usually too weak to reverse this effect.

Transfer effect. When the context changes, the old response levels persist under the new circumstances because of the incremental nature of the learning mechanisms. As both Equations 16 and 21 discount the distant past, the transfer effect eventually decays away.

Practice effect. The identification accuracy improves over time because the system works with suboptimal anchors at the beginning of the session until the learning mechanisms fine-tune them to the statistics of the environment.

Resolution Edge Effect: An Open Issue

There is one phenomenon that ANCHOR currently does not account for: the elevated discriminability at the edges of the stimulus range evident in the d' profile in the top right panel of Figure 3. Other kinds of edge effects such as the elevated percentage correct and the reduced frequencies of extreme responses have straightforward explanations. The resolution edge effect, however, is a challenging theoretical problem. It has been replicated many times and various interpretations are proposed in the literature.

Four general ideas appear in one form or another. One approach is to postulate that the variance of the stimulus representations is reduced in proportion to the distance to the nearest edge of the range (e.g., Nosofsky, 1997). This postulate is interpreted in various psychological terms, but its main contact with empirical data is precisely the resolution edge effect. The variance profiles along the magnitude continuum are parameterized in advance and must be reset manually whenever the stimulus range changes. In our opinion this amounts to little more than a redescription of the d' data.

A related idea attributes the resolution edge effect to criterion variability rather than perceptual variability (Treisman, 1985). Careful examination of this proposal, however, reveals some flaws as discussed in the next section.

The third idea explains the edge effect in terms of two perceptual anchors located at the ends of the stimulus range (Braida & Durlach, 1972; Braida et al., 1984) or in terms of a rehearsed frame (Marley & Cook, 1984, 1986). According to these interesting proposals, the interior stimuli are judged on the basis of their distance to these anchors, in noisy multiples of a category unit. In addition to the bow in the d' profile, this explains the slower response times for intermediate stimuli and the overall capacity limitation. A related hypothesis attributes the edge effects to attention bands located near the extreme intensities (Luce et al., 1976, 1982; D. L. Weber et al., 1977).

The fourth idea comes from a connectionist framework. Lacouture and Marley (1991, 1995) proposed a feedforward neural network in which the stimulus intensity is represented by the activation level of a single hidden unit. There are N linear output units, one per category, and the unit with maximal activation determines the response. The solution of this encoder problem assigns greater weights to the peripheral output units (Lacouture & Marley, 1995). The activation functions of these units thus have steeper slopes and, given the constant noise across the output layer, achieve better discriminability near the edges. The whole scheme depends strongly on the assumption of linearity, and the weights are set in advance according to parametric formulas depending on N. An earlier, trainable version of the model "does not usually yield the end anchor effect for all set sizes" (Lacouture & Marley, 1991, p. 427).

All in all, we are not aware of any account in which the resolution edge effect emerges from the learning mechanisms in the system rather than being designed into it in advance. See Stewart, Brown, and Chater (2004) for a thorough review and an interesting new proposal. The elemental perceptual units in their relative judgment model (RJM) are not absolute magnitudes but rather representations of the differences between consecutive stimuli S_{t-1} and S_t. Stripping away details we cannot consider here, the scale in RJM is based on the proportionality $(R_t - F_{t-1})/(\ln S_t - \ln S_{t-1})$ = constant, where R_t is the mean of a random variable $\mathbf{R}_t$ that determines the response on trial t, and F_{t-1} is the correct response on trial $t - 1$. The standard deviation of $\mathbf{R}_t$ varies from trial to trial in proportion to the range ρ of possible responses between the previous feedback and the relevant edge of the scale. This follows from the assumption that people can rule out the responses on the side of F_{t-1} that is inconsistent with the sign of the stimulus difference. For example, if F_{t-1} is *4* and S_t is less than S_{t-1}, then only responses *1* through *3* are represented within a constant limited capacity (i.e., $\rho = 3$). The exact psychological mechanism of the adjustable, range-dependent variability in $\mathbf{R}_t$ is not

[9] The self-reinforcing activation dynamics also unfolds over time, but simulations with model $\mathcal{M}_{\text{PSAC}}$ suggest that this process is more or less complete by the end of the first experimental block.

specified. This variability is largest when S_t is in the middle of the stimulus domain because then the range ρ, averaged over all possible S_{t-1}, is largest. RJM thus accounts for the bow effect, at least qualitatively, although the simulated d' profiles cannot be made deep enough to fit the data without additional assumptions (N. Stewart, personal communication, January 17, 2005).

Many of these ideas can be incorporated in ANCHOR without disrupting any of the successful predictions of the theory. We could easily introduce an edge-dependent term into the perceptual Equation 10, for instance. The compatibility between the perceptual anchor model of Braida and colleagues (1984) and ANCHOR's correction mechanism is obvious. Thus, the negative bow effect in Table 4 does not point to a structural defect in the model. This is an area for future extensions that seem unrelated to the dynamic focus of the present article. Instead of adding machinery whose sole raison d'être is to fit this one effect, it seems more productive to leave the question open and invite more research. The verdict of Luce, Nosofsky, Green, and Smith is no less true today than it was in 1982: "Concerning the bow in absolute identification data, we remain unsure of its source" (p. 406).

Range Effects and the Magical Number Seven

The identification experiment in this article emphasized the dynamic aspects of scaling and did not manipulate the range or number of the stimuli. A limited information capacity ($T = 1.68$) was observed, replicating the classic results (Braida & Durlach, 1972; Miller, 1956). ANCHOR reproduced easily both the group mean of this statistic and the individual differences in the sample (see Tables 1 and 4). As parameters were varied freely, however, these fits do not address the more stringent capacity limitation implied by the *range effect.* For roughly equally spaced stimuli, increasing the stimulus spacing leads, at best, to only modest improvements in the absolute identification performance (Braida & Durlach, 1972; Luce et al., 1976). A related phenomenon is that the discriminability between two fixed stimuli appears to decrease as the overall range of the other stimuli increases (Gravetter & Lockhead, 1973). Also, the introduction of new stimuli can drastically impair the identification of a previously error-free stimulus set (Miller, 1956; Pollack, 1953; Shiffrin & Nosofsky, 1994).

Can ANCHOR account for these important and challenging phenomena? Three factors jointly restrict the identification capacity of the model. First, both perceptual noise (Equation 9) and memory noise (Equation 12) are multiplicative and thus scale up when stimulus intensities increase. This strongly attenuates the advantage of spacing out the stimuli but does not always eliminate it fully, particularly when the logarithm of the range increases. Second, activation decay (Equation 16) and the interference inherent in stochastic anchor selection (Equation 14) restrict the number of anchors that can be maintained strong enough (cf. Taatgen, 2001; see also Cowan, 2001, for an extensive discussion of the analogous short-term memory limitation).

The third capacity-limiting factor is the imperfect correction mechanism. The importance of this factor was illustrated in an earlier simulation ($T = 1.49$ with corrections vs. $T = 0.94$ without, all else being equal). Now, wider stimulus ranges entail larger category sizes and hence inflated correction thresholds and fewer corrections (Equations 17 and 18). If the intensities of two particular stimuli remain fixed while the overall range expands, the local discriminability is indeed expected to decrease as reported by Gravetter and Lockhead (1973). Finally, it seems plausible that when forced to operate across larger distances, the correction mechanism becomes progressively inefficient and inaccurate, although this is not implemented in the current version of the model.

In summary, ANCHOR's potential, as currently defined, to account for the range effects is best characterized as promising but untested. Given the complexity of the model, proper testing requires systematic analyses and simulations beyond the scope of this article. Should it fail, ANCHOR may have to be revised. One possibility is to incorporate some form of normalization or gain control over and above that already implicit in Equation 9 (e.g., Parker, Murphy, & Schneider, 2002). Another possibility, which accounts for the edge effect as well, is to introduce a strategy anchored at the edges and counting inward on some trials (Braida et al., 1984). None of these extensions seem to jeopardize the successful predictions of the theory.

Comparison With Related Models

Various mechanistic accounts of absolute identification and/or category rating are discussed in the literature (e.g., Baird, 1997; Haubensak, 1992; Kokinov, Hristova, & Petkov, 2004; Laming, 1984; Rouder, 2001; J. A. Siegel & Siegel, 1972). Most models maintain some kind of internal state to account for the ubiquitous sequential effects in the data. Perhaps the most widespread idea, implied by all autoregression models, is to keep the previous stimulus S_{t-1} and response R_{t-1} (DeCarlo, 2003; DeCarlo & Cross, 1990; Jesteadt, Luce, & Green, 1977; Lockhead & King, 1983). The relative judgment model (Stewart & Brown, 2004; Stewart et al., 2004) dispenses with long-term representations altogether and bases the categorization decision entirely on comparison with the previous exemplar. The range-frequency model (Parducci & Wedell, 1986) maintains a search set of the 12 most recent presentations. Other proposals include a shifting adaptation level (Helson, 1964) and a roving attention band (Luce et al., 1976). See Marks and Algom (1998) for a recent review.

The theory of criterion setting (TCS) is among the most comprehensive proposals (Treisman, 1985; Treisman & Williams, 1984). Following Thurstone (1927) and Torgerson (1958), the internal magnitude continuum is partitioned into response regions by $N - 1$ criteria. Two dynamic mechanisms adjust these criteria relative to a set of static reference values. A tracking mechanism pushes the criteria away from the previous response R_{t-1}, and a stabilization mechanism pulls them toward the previous stimulus S_{t-1}. The size of each adjustment, or indicator trace, is inversely proportional to the distance to the corresponding criterion and decays over time. The effective position of each criterion on a given trial is a linear combination of its fixed reference value and two families of transient indicator traces. This gives rise to sequential assimilation toward the previous response, contrast with the previous stimulus, and various other effects.

The fact that (nearly) all criteria are adjusted on every trial has an important impact on criterion variability. A criterion near the edge is pushed predominantly in one direction, as most stimuli and responses fall on one side of it. The variability of this criterion is therefore relatively low. In contrast, a criterion in the interior of the range is pushed irregularly up and down and thus has greater variability. The resulting dome-shaped variance profile has the potential to account for the observed dip in the *d′* profiles (Treisman, 1985).

Despite the appeal of this idea, several technical obstacles apparently undermine its potential. First, the *d′* bows cannot be made deep enough to match the data unless the criterion variance exceeds the stimulus variance by an order of magnitude (see Figure 10 in Treisman, 1985). A direct estimation of the two sources of variability via an independent technique suggests that their relative strengths are in fact reversed (Nosofsky, 1983). The simulated *d′* bows also seem very noisy and parameter dependent (Treisman, 1985). The high learning rates necessary to induce such substantial criterion variability probably induce exaggerated sequential effects as well. This casts doubt on the model's ability to meet all empirical constraints simultaneously on the same run.

Technical detail notwithstanding, the criterion-based and anchor-based systems are functionally equivalent in most respects. In fact, there is a complementary duality: The former emphasizes the boundaries between response regions, whereas the latter emphasizes the centroid of each region. TCS updates all criteria on each trial; ANCHOR updates a single anchor. Tracking and stabilization in TCS correspond to activation and location learning in ANCHOR, respectively, as illustrated in Figure 16.

The biggest difference between the two theories is that TCS has no counterpart of ANCHOR's correction mechanism and consequently cannot promote the homomorphism between stimuli and responses autonomously. In the final analysis, the stimulus–response correspondence in TCS is largely imposed by the criterion reference values that are fixed a priori, essentially like free parameters. The theory gives only a vague informal description of the reference system responsible for this crucial part of the scaling problem. The reference values must be set in the beginning of the session when the relevant statistics are largely unknown. All subsequent adjustments fine-tune the criteria around these fixed home positions. This stands in sharp contrast with ANCHOR's ability to unfold the scale from a single arbitrarily placed anchor and to track the stimulus density dynamically. Repeated ANCHOR runs with fixed parameters can produce quite different scaling solutions reflecting idiosyncratic frozen accidents (Gell-Mann, 1994) from the early trials that become entrenched later on. For instance, the model may respond *1* frequently on a given run and hardly ever on another, shifting the rest of the scale up or down accordingly.

ANCHOR builds on the memory-based models of categorization (e.g., Kruschke, 1992; Nosofsky, 1986). In particular, Nosofsky (1997) sketched how the exemplar-based random-walk model of speeded classification (EBRW; Nosofsky & Palmeri, 1997) can be applied to the absolute identification task. The specific proposal is similar in many respects to the ANCHOR subset labeled model $\mathcal{M}_{\mathrm{PS}}$ in Table 4. Like in $\mathcal{M}_{\mathrm{PS}}$, a noisy representative of each response category competes to match the target. The winner determines the response; there are no corrections. Unlike in $\mathcal{M}_{\mathrm{PS}}$, the competition is resolved in a prolonged random walk rather than in a single step. EBRW thus predicts reaction times in addition to choice probabilities—a capacity that ANCHOR currently lacks but can inherit from the ACT–R architecture. If the memory strengths of the individual exemplars were allowed to vary dynamically (Nosofsky, 1988, 1991), EBRW would exhibit sequential, repetition, and assimilative context effects similar to those of model $\mathcal{M}_{\mathrm{PSA}}$ in Table 4. Like $\mathcal{M}_{\mathrm{PSA}}$, however, it would also become unstable without external feedback.

ANCHOR uses a prototype-based representation of the response categories, continuously updated by the competitive learning mechanism. An alternative method is to store the individual exemplars themselves and rely on retrieval-time processing (similarity, random walk) to consolidate and smooth out the category representations (Nosofsky, 1986). This instance-based method has greater representational power (Ashby & Alfonso-Reese, 1995). Unidimensional scaling, however, does not need that extra power as categories are always convex and there are no exceptions. The prototype representation is fully adequate in this case, and much more economical. It would be very interesting to compare the two kinds of representations within a category-rating framework, adding to the debate in multidimensional categorization where the notion of homomorphism plays no role (e.g., Minda & Smith, 2002; Nosofsky & Zaki, 2002; Smith & Minda, 1998).

In our opinion, the key ANCHOR innovation relative to other memory-based models is the introduction of the correction mechanism. The idea that people often adjust their responses is of course introspectively familiar to everyone. However, the profound impact that even occasional corrections can have on the dynamical stability of a memory-based system has not been sufficiently appreciated. The correction mechanism introduces ordinal relations and enforces the stimulus–response homomorphism that is so essential for scaling. It is qualitatively different from the random-walk mechanism in other models (e.g., Nosofsky & Palmeri, 1997). Both mechanisms smooth out memory fluctuations and improve the overall accuracy by combining several pieces of evidence. However, whereas the random walk simply resamples the same memory pool, the correction mechanism introduces qualitatively different, relational knowledge—ordinal comparisons between magnitudes (Equation 17) and responses (Equation 18)—as well as knowledge that these two ordered structures should be aligned. The random walk cannot generate any novel responses; it is confined to whatever already exists in memory. The correction mechanism, in contrast, can and does generate novel responses. Finally and very importantly, it redistributes strength among the anchors and thereby prevents runaway activation dynamics in the absence of feedback. This brings the category rating task within the scope of the memory-based paradigm.

The emphasis on learning and its attendant dynamic manifestations is another distinguishing feature of our approach. Competitive learning in particular unfolds the scale starting from a single arbitrarily placed anchor. The stimulus–response correspondence is not set a priori but emerges from the dynamic equilibrium of various interacting forces. This level of adaptability and emergence has no parallel in existing scaling models. To our knowledge, no other model can account for the feedback-dependent context effects and the nonstationary response distributions that ANCHOR predicts so naturally.

Finally, ANCHOR is an integrated model. Its computational mechanisms mesh seamlessly with the huge corpus of memory-related theory and data. ANCHOR can thus serve as a building block for integrated models of even greater scope. In effect, it brings rating-based measures within the scope of the ACT–R architecture and thereby connects psychophysical scaling to the numerous domains in which this architecture has been successfully applied (Anderson, 1983; Anderson & Lebière, 1998).

Three classes of factors collectively shape ANCHOR's behavior. First, the statistics of the environment are paramount as the system is built to adapt to the prevailing conditions. Second, the cognitive architecture both predisposes the model to capitalize on certain particularly diagnostic information and limits the maximal discriminability that it can achieve. Finally, an explicit strategy compensates for some inherent limitations and interacts with the environmental factors in an attempt to maximize performance. ANCHOR thus illustrates a pervasive characteristic of cognition—the delicate and sometimes surprising interplay of environment, architecture, and strategy.

References

Algom, D. (1992). Memory psychophysics: An examination of its perceptual and cognitive prospects. In D. Algom (Ed.), *Psychophysical approaches to cognition* (pp. 441–513). Amsterdam: Elsevier.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* Mahwah, NJ: Erlbaum.

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

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96,* 703–719.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396–408.

Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449–483). Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology, 39,* 216–233.

Ashby, F. G., & Maddox, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making* (pp. 251–301). San Diego, CA: Academic Press.

Baird, J. C. (1970). A cognitive theory of psychophysics: II. Fechner's law and Stevens' law. *Scandinavian Journal of Psychology, 11,* 89–102.

Baird, J. C. (1997). *Sensation and judgment: Complementarity theory of psychophysics.* Mahwah, NJ: Erlbaum.

Baird, J. C., Green, D. M., & Luce, R. D. (1980). Variability and sequential effects in cross modality matching of area and loudness. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 277–289.

Baird, J. C., & Norma, E. (1978). *Fundamentals of scaling and psychophysics.* New York: Wiley.

Baird, J. C., Romer, D., & Stein, T. (1970). Test of a cognitive theory of psychophysics: Size discrimination. *Perceptual and Motor Skills, 30,* 495–501.

Braida, L. D., & Durlach, N. I. (1972). Intensity perception. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America, 51,* 483–502.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., & Purks, S. R. (1984). Intensity perception. XIII. Perceptual anchor model of context-coding. *Journal of the Acoustical Society of America, 76,* 722–731.

Chase, S., Bugnacki, L. D., Braida, L. D., & Durlach, N. I. (1983). Intensity perception. XII. Effect of presentation probability on absolute identification. *Journal of the Acoustical Society of America, 73,* 279–284.

Conway, M. A. (Ed.). (1997). *Cognitive models of memory.* Cambridge, MA: MIT Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

DeCarlo, L. T. (2003). An application of a dynamic model of judgment to magnitude production. *Perception & Psychophysics, 65,* 152–162.

DeCarlo, L. T., & Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General, 119,* 375–396.

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

Fechner, G. T. (1966). *Elements of psychophysics* (Vol. 1; H. E. Adler, Trans.). New York: Holt, Rinehart & Winston. (Original work published 1860)

Garner, W. R. (1953). An information analysis of absolute judgments of loudness. *Journal of Experimental Psychology, 46,* 373–380.

Gell-Mann, M. (1994). *The quark and the jaguar: Adventures in the simple and the complex.* New York: Freeman.

Gravetter, F., & Lockhead, G. R. (1973). Criterial range as a frame of reference for stimulus judgment. *Psychological Review, 80,* 203–216.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hartman, E. B. (1954). The influence of practice and pitch-distance between tones on the absolute identification of pitch. *American Journal of Psychology, 67,* 1–14.

Haubensak, G. (1990). Primacy effects in absolute judgments. In H.-G. Geissler (Ed.), *Psychophysical explorations of mental structures* (pp. 104–113). Toronto, Ontario, Canada: Hogrefe & Huber.

Haubensak, G. (1992). The consistency model: A process model for absolute judgments. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 303–309.

Helson, H. (1964). *Adaptation-level theory.* New York: Harper & Row.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Exploration in the microstructure of cognition. Vol. I. Foundations* (pp. 282–317). Cambridge, MA: MIT Press.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics, 3,* 409–414.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA, 79,* 2554–2558.

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

Jesteadt, W., Wier, C. C., & Green, D. M. (1977). Intensity discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America, 61,* 169–178.

Kerst, S. M., & Howard, J. H., Jr. (1978). Memory psychophysics for visual area and length. *Memory & Cognition, 6,* 327–335.

King, M. C., & Lockhead, G. R. (1981). Response scale and sequential effects in judgment. *Perception & Psychophysics, 30,* 599–603.

Kohonen, T. (1995). *Self-organizing maps* (3rd ed.). Berlin, Germany: Springer-Verlag.

Kokinov, B., Hristova, P., & Petkov, G. (2004). Does irrelevant information play a role in judgment? In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 720–725). Mahwah, NJ: Erlbaum.

Krantz, D. H. A. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology, 9,* 168–199.

Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral and Brain Sciences, 12,* 251–320.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Lacouture, Y. (1997). Bow, range, and sequential effects in absolute identification: A response-time analysis. *Psychological Research, 60,* 121–133.

Lacouture, Y., & Marley, A. A. J. (1991). A connectionist model of choice and reaction time in absolute identification. *Connection Science, 3,* 401–433.

Lacouture, Y., & Marley, A. A. J. (1995). A mapping model of bow effects in absolute identification. *Journal of Mathematical Psychology, 39,* 383–395.

Laming, D. (1984). The relativity of "absolute" judgments. *British Journal of Mathematical and Social Psychology, 37,* 52–183.

Laming, D. (1986). *Sensory analysis.* London: Academic Press.

Laming, D., & Scheiwiller, P. (1985). Retention in perceptual memory: A review of models and data. *Perception & Psychophysics, 37,* 189–197.

Link, S. W. (1992). *The wave theory of difference and similarity.* Hillsdale, NJ: Erlbaum.

Lockhead, G. R., & King, M. C. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 461–473.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review, 66,* 81–95.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103–189). New York: Wiley.

Luce, R. D., Green, D. M., & Weber, D. L. (1976). Attention bands in absolute identification. *Perception & Psychophysics, 20,* 49–54.

Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Perception & Psychophysics, 32,* 397–408.

Marks, L. E. (1974). *Sensory processes.* New York: Academic Press.

Marks, L. E. (1993). Contextual processing of multidimensional and unidimensional auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 227–249.

Marks, L. E., & Algom, D. (1998). Psychophysical scaling. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making* (pp. 81–178). San Diego, CA: Academic Press.

Marks, L. E., & Stevens, J. C. (1968). The form of the psychophysical function near threshold. *Perception & Psychophysics, 4,* 315–318.

Marley, A. A. J., & Cook, V. T. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical and Statistical Psychology, 37,* 136–151.

Marley, A. A. J., & Cook, V. T. (1986). A limited capacity rehearsal model for psychophysical judgements applied to magnitude estimation. *Journal of Mathematical Psychology, 30,* 339–390.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 275–292.

Mori, S., & Ward, L. M. (1995). Pure feedback effects in absolute identification. *Perception & Psychophysics, 57,* 1065–1079.

Moyer, R. S., Bradley, D. R., Sorensen, M. H., Whiting, J. C., & Mansfield, D. P. (1978, April 21). Psychophysical functions for perceived and remembered size. *Science, 200,* 330–332.

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

Nosofsky, R. M. (1983). Information integration and the identification of stimulus noise and criterial noise in absolute judgment. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 299–309.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 3–27.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43,* 25–53.

Nosofsky, R. M. (1997). An exemplar-based random-walk model of speeded categorization and absolute judgment. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 347–365). Mahwah, NJ: Erlbaum.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 924–940.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,* 407–418.

Parducci, A. (1974). Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception: Vol. II. Psychophysical judgment and measurement* (pp. 127–141). New York: Academic Press.

Parducci, A., Knobel, S., & Thomas, C. (1976). Independent contexts for category ratings: A range-frequency analysis. *Perception & Psychophysics, 20,* 360–366.

Parducci, A., & Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency. *Journal of Experimental Psychology Monographs, 89,* 427–452.

Parducci, A., & Wedell, D. (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,* 496–516.

Parker, S., Murphy, D. R., & Schneider, B. A. (2002). Top-down gain control in the auditory system: Evidence from identification and discrimination experiments. *Perception & Psychophysics, 64,* 598–615.

Petrov, A. A. (2001). Fitting the ANCHOR model to individual data: A case study in Bayesian methodology. In E. M. Altmann, A. Cleeremans, C. D. Schunn, & W. D. Gray (Eds.), *Proceedings of the 2001 Fourth International Conference on Cognitive Modeling* (pp. 175–180). Mahwah, NJ: Erlbaum.

Petrov, A. A. (2003). Additive or multiplicative perceptual noise? Two equivalent forms of the ANCHOR model. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 922–927). Hillsdale, NJ: Erlbaum.

Petrov, A. A., & Anderson, J. R. (2000). ANCHOR: A memory-based model of category rating. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 369–374). Hillsdale, NJ: Erlbaum.

Petzold, P. (1981). Distance effects on sequential dependencies in categorical judgments. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1371–1385.

Pollack, I. (1953). The information of elementary auditory displays. II. *Journal of the Acoustical Society of America, 25,* 765–769.

Purks, S. R., Callahan, D. J., Braida, L. D., & Durlach, N. I. (1980). Intensity perception. X. Effects of preceding stimulus on identification performance. *Journal of the Acoustical Society of America, 67,* 634–637.

Rouder, J. N. (2001). Absolute identification with simple and complex stimuli. *Psychological Science, 12,* 318–322.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9,* 75–112.

Schifferstein, H., & Frijters, J. (1992). Contextual and sequential effects on judgments of sweetness intensity. *Perception & Psychophysics, 52,* 243–255.

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to* *multidimensional scaling: Theory, methods, and applications.* New York: Academic Press.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22,* 325–345.

Shepard, R. N. (1987, September 11). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Shiffrin, R., & Nosofsky, R. (1994). Seven, plus or minus two: A commentary on capacity limitations. *Psychological Review, 101,* 357–361.

Siegel, J. A., & Siegel, W. (1972). Absolute judgment and paired-associate learning: Kissing cousins or identical twins? *Psychological Review, 79,* 300–316.

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

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1411–1436.

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

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects.* New York: Wiley.

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology, 54,* 377–411.

Stewart, N., & Brown, G. D. A. (2004). Sequence effects in the categorization of tones varying in frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 416–430.

Stewart, N., Brown, G. D. A., & Chater, N. (2004). *Absolute identification by relative judgment.* Manuscript submitted for publication.

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos.* Reading, MA: Addison-Wesley.

Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99,* 181–198.

Taatgen, N. A. (2001). Dispelling the magic: Towards memory without capacity. *Behavioral and Brain Sciences, 24,* 147–148.

Teghtsoonian, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review, 78,* 71–80.

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

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Treisman, M. (1964). Sensory scaling and the psychophysical law. *Quarterly Journal of Experimental Psychology, 16,* 11–22.

Treisman, M. (1985). The magical number seven and some other features of category scaling: Properties of a model for absolute judgment. *Journal of Mathematical Psychology, 29,* 175–230.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91,* 68–111.

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Ward, L. M. (1979). Stimulus information and sequential dependencies in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 444–459.

Ward, L. M., & Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology, 84,* 27–34.

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics, 9,* 73–78.

Wassermann, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44,* 92–107.

Weber, D. L., Green, D. M., & Luce, R. D. (1977). Effects of practice and distribution of auditory signals on absolute identification. *Perception & Psychophysics, 22,* 223–231.

Weber, E. H. (1849). Der Tastsinn und das Gemeingefühl [The sense of touch and general sensation]. In R. Wagner (Ed.), *Handworterbuch der Physiologie* (Vol. 3, pp. 481–588). Braunschweig, Germany: Vieweg.

Wedell, D. H. (1984). A process model for psychophysical judgment. *Dissertation Abstracts International, 45,* 3102-B. (UMI No. 8428589)

Wiest, W. M., & Bell, B. (1985). Stevens's exponent for psychophysical scaling of perceived, remembered, and inferred distance. *Psychological Bulletin, 98,* 457–470.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General, 125,* 387–402.

Appendix A

Proof That Equation 9 Is Consistent With Weber's Law

Let us consider two stimuli with intensities $S_1 = S$ and $S_2 = (1 + k)S$. According to Equation 9, the means and standard deviations of the respective magnitude distributions are as follows: $\mu_1 = aS^n$, $\sigma_1 = k_p aS^n$; $\mu_2 = aS^n(1 + k)^n$, $\sigma_2 = k_p aS^n(1 + k)^n$. Assuming that all variability in the system comes from perceptual sources and is accurately summarized by σ_1 and σ_2, the probability of correct response in a two-alternative forced-choice (2AFC) comparison is equal to the normal integral of the following quantity:

$$z = \frac{\mu_2 - \mu_1}{\sqrt{\sigma_2^2 + \sigma_1^2}} = \frac{aS^n[(1 + k)^n - 1^n]}{k_p aS^n \sqrt{(1 + k)^{2n} + 1^{2n}}}. \qquad \text{(A1)}$$

Notice that the aS^n terms cancel out and hence z does not depend on the absolute intensity levels S_1 and S_2. Keeping in mind that $k \ll 1$, we can approximate

$$z = \frac{(1 + k)^n - 1}{k_p \sqrt{(1 + k)^{2n} + 1}} \approx \frac{(1 + nk) - 1}{k_p \sqrt{(1 + 2nk) + 1}} \approx \frac{nk}{k_p \sqrt{2}}. \qquad \text{(A2)}$$

Given that k_p and n are fixed for each sensory modality, the probability of correct response varies only as a function of k—that is, of the intensity ratio $S_2/S_1 = (1 + k)/1$. This is exactly the relation implied by Equation 7. Hence the multiplicative-noise Equation 9 is consistent with Weber's law within the margin of measurement error.

Moreover, we can use the empirical Weber fraction k to inform our choice of the parameter k_p. For concreteness, we define k as the relative increment ΔS needed to yield 75% correct 2AFC performance. Restated in terms of Equation A2, k is adjusted experimentally until its corresponding z value equals $z_{75} = \Phi^{-1}(0.75)$, where Φ^{-1} is the inverse of the normal cumulative distribution function. Solving for k_p, we get

(*Appendixes continue*)

$$k_p \approx \frac{nk}{z_{75}\sqrt{2}} \approx \frac{nk}{0.674 \times 1.41} \approx 1.05nk. \qquad \text{(A3)}$$

This estimate is based on the assumption that all the uncertainty of the responses is due to perceptual sources as described in Equation 9. There are other sources of variability and errors in actual psychophysical experiments. These include the need to maintain in memory the magnitude of the first stimulus during successive presentations. Extraneous factors such as mistaken keypresses and lapses of attention also inflate the Weber fraction. Consequently, the observable k imposes an upper bound on k_p but does not determine it completely. Thus we arrive at the rule-of-thumb inequality $k_p \leq nk$ (Equation 11).

Incidentally, Baird (1970), Teghtsoonian (1971), and Link (1992) independently compiled lists of exponents and Weber fractions for numerous modalities and found strong correlations between n and $1/k$. The same inverse relation follows from Equation 11 if one assumes that k_p reflects some invariant property of the neural substrate and therefore varies little across perceptual modalities.

Laming and Scheiwiller (1985) reported two discrimination experiments with stimuli very similar to ours—lines presented successively at randomized locations on a video display. Their Experiment 2 seems a good source for the estimate we need. On each trial, a standard was chosen at random among seven different lengths (3–12 dva) and presented for 0.5 s. It was followed by a comparison line that was either shorter or longer than the standard. There were seven interstimulus intervals ranging from 0 to 4 s. The trial sequence randomly interleaved 500 trials per condition. The 1-s interstimulus interval is representative for all conditions and will suffice for our purposes.

Two motivated and experienced observers (the authors themselves; Laming & Scheiwiller, 1985) rated on a 6-point scale their confidence that the second stimulus was shorter or longer than the first. The relative difference Δ between the two lines was fixed at 4% for observer D.L. and 8% for P.S. The following d' values were obtained: $d' = 1.52$ for observer P.S. and $d' = 1.39$ for D.L. We can invert these values to produce Weber fractions. Under an equal-variance normal model and 2AFC task, 75% correct corresponds to $d' = z_{75}\sqrt{2} \approx 0.954$. Although this relationship is theory dependent, it suffices to justify an estimate. Hence,

$$k \approx 0.954 \frac{\Delta}{d'}. \qquad \text{(A4)}$$

This yields Weber fractions $k = .03$ for observer D.L. and $k = .05$ for P.S. These estimates are in good agreement with the value $k = .04$ listed in two secondary sources (Baird & Norma, 1978; Laming, 1986) as a typical Weber fraction for line length. They also agree with the thresholds obtained in Laming and Scheiwiller's (1985) Experiment 1.

Appendix B

Sequential Effects in Category Rating

The sequential effects in the data from Experiment 2 are analyzed in a hierarchy of autoregression models. Some preliminary remarks are in order. First, the logarithmic transformation $\log R = \log a + n \log S$ common in the literature seems inappropriate here. The relationship between stimuli and responses is already linear in our data ($n \approx 1$), there is homogeneity of variance, and the response distributions are (nearly) symmetrical. A transformation would disrupt these conditions. Second, all blocks with nonuniform presentation frequencies are excluded from the analyses. This facilitates the interpretation by eliminating the autocorrelational structure among the stimuli. The earliest 20 trials and the cases with missing values are also excluded. This leaves 250 stimulus–response pairs per observer, Trials 21–90, 181–270, and 361–450; there were 9,696 valid and 304 missing cases overall. Third, outliers are corrected relative to the residual standard deviation of a preliminary regression: Responses deviating by more than $\pm 3\sigma$ are replaced by exactly $\pm 3\sigma$, separately for each participant. Less than 1% of the raw responses need correction according to this criterion. Finally, all variables are standardized to zero mean and unit variance. This eliminates the need for constant terms in the equations and makes the regression coefficients comparable.

Our analytic strategy is to fit separate regressions for each observer and then aggregate the individual R^2. Figure B1 summarizes eight different regression models. The mean R^2 in each box averages 40 squared multiple correlation coefficients.

Equation B1 is the simplest model in the hierarchy. The estimated mean response $\hat{R}_t$ on trial t is proportional to the corresponding stimulus S_t in a direct application of Stevens's law (Equation 8). This simple equation accounts for 78% of the response variance on average, with individual R^2 values ranging from .55 to .92.

$$\hat{R}_t = aS_t. \qquad \text{(B1)}$$

Equation B2 is the standard tool for sequential analysis in the literature. It accounts for 81% of the response variance on average (see Figure B1). Thus, the time-lagged variables S_{t-1} and R_{t-1} increase the fit by $\Delta R^2 = .037$, with individual increases as large as .15 for some observers. Thirty-nine of the 40 R^2 increments are significant at the .05 level.

$$\hat{R}_t = aS_t + bR_{t-1} + dS_{t-1}. \qquad \text{(B2)}$$

The regression coefficient for R_{t-1} is consistently positive, and the one for S_{t-1} is negative for all individuals (mean $a = .88$, $SD = .05$; mean $b = .33$, $SD = .14$; mean $d = -.25$, $SD = .11$). This is a clear replication of the classic sequential effects: assimilation to the previous response and contrast with the previous stimulus (e.g., DeCarlo & Cross, 1990; Jesteadt, Luce, & Green, 1977; Lockhead & King, 1983).

But what do these correlational results mean? One possible interpretation is that they reflect the slow drift of the response scale. We know from Figure 13 that the early response levels tend to be one-half category unit (on average) smaller than the ones at the end of the session. This tendency counts as error in Equation B1, but it can be accounted for by Equation B2 because the pair $\langle S_{t-1}, R_{t-1}\rangle$ contains implicit information about the current drift position. For instance, if R_{t-1} falls below the value predicted by S_{t-1}, then trial $t - 1$ probably occurs early in the session, and hence R_t too is likely to fall below its expected value. The residuals $(R_{t-1} - aS_{t-1})$ and $(R_t - aS_t)$ tend to have the same sign.

To capture the effect of the trend, we include the explicit average response level in the model. We construct a new variable ARL_t for each trial t by applying Equation 5 to the data from trials $t - 31$ to $t - 2$. Note that for some values of t this roving window extends into nonuniform blocks that are otherwise excluded from the analysis. The

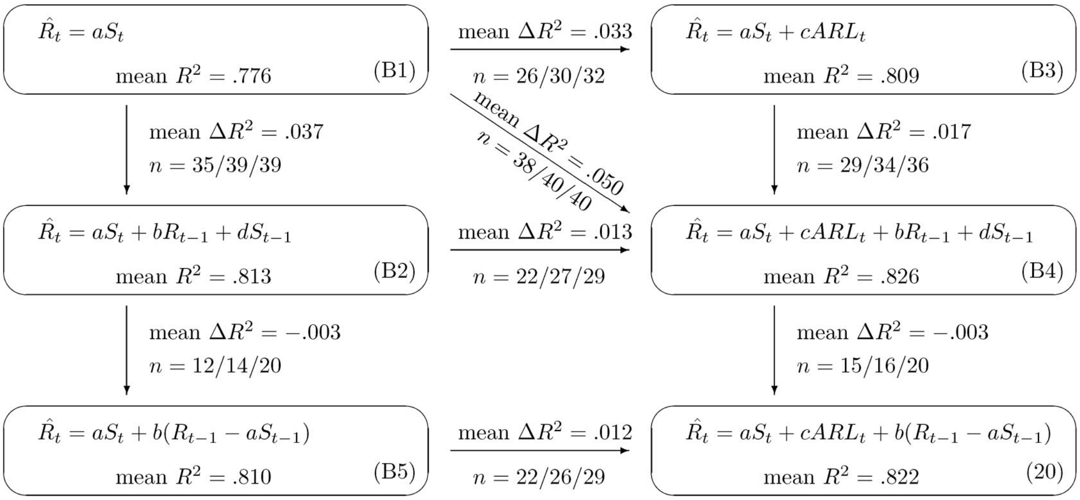

Figure B1. A hierarchy of autoregression models. Equation numbers are the same as in Appendix B. Each mean R^2 is based on 40 individual squared correlations. The notation $n = \cdot/\cdot/\cdot$ denotes the number of observers for whom the corresponding transition is significant at $\alpha = .01$, .05, and .10, respectively.

sliding ARLs are calculated separately for each observer and standardized to zero mean and unit variance. The regression model then becomes

$$\hat{R}_t = aS_t + c\text{ARL}_t. \tag{B3}$$

Equation B3 represents the hypothesis that all systematic variance in the responses is attributable to two sources: the current stimulus S_t and a gradual trend ARL (which may incorporate transfer and context effects). This model accounts for 81% of the response variance on average. The mean increase relative to Equation B1 is $\Delta R^2 = .033$; 30 individual increases are significant at the .05 level. This is comparable to the fit of Equation B2. Thus, a single variable extending 30 trials in the past accounts for nearly as much variance as R_{t-1} and S_{t-1} do together.

We have, therefore, two alternative interpretations of the sequential effect: short-term carryover across neighboring trials (Equation B2) and long-term gradual trend (Equation B3). Either one of them alone accounts for about 3.5% of the response variance on average.

These two mechanisms are not incompatible and can be combined. The combined Equation B4 accounts for 83% of the response variance on average—an increase of $\Delta R^2 = .050$ relative to Equation B1 and a statistically significant improvement over both Equations B2 and B3. The data therefore seem to contain two kinds of sequential dependencies: short term, extending one trial back, and long term, extending some tens of trials back. The top four panels in Figure B1 suggest that on average, 1.7% of the response variance is uniquely attributable to short-term effects and 1.3% to long-term drift, and an additional 2.0% are shared between the two. In summary, 78% are attributable to the immediate stimulus S_t and 5% to various sequential effects.

$$\hat{R}_t = aS_t + c\text{ARL}_t + bR_{t-1} + dS_{t-1}. \tag{B4}$$

The short-term component is most relevant to the analysis of sequential effects. Hence it is useful to partial out the nontransient components and consider the residuals. We calculate two sets of residuals $\text{res}_t = R_t - \hat{R}_t$, where $\hat{R}_t$ is defined by Equation B1 or B3, respectively. The autocorrelation coefficients for the first set (Equation B1) have a mean of .34 over the sample (SD = .12). This is the measure used in Tables 3 and 4 in the main text. It overestimates the magnitude of transient sequential effects because it is inflated by the slow component. When ARL is also partialled out (Equation B3), the mean residual autocorrelation drops to .23 (SD = .09). It underestimates the magnitude of transient effects because it suppresses the nonsystematic slow fluctuations that occur naturally in autocorrelated time series. Thus, the "true" value probably lies between .23 and .34.

The magnitude of the sequential effects has been shown to depend on the interstimulus similarity (DeCarlo, 2003; DeCarlo & Cross, 1990; Jesteadt, Luce, & Green, 1977; Ward, 1979). The notion of a fixed "true" value of the short-range autocorrelation is therefore inaccurate; it is better to measure it as a function of the difference $\Delta S = S_t - S_{t-1}$. To that end, the data from all observers are pooled together and then grouped into 21 bins according to ΔS. There are about 470 observations per bin. The correlation between res_t and res_{t-1} is then calculated within each bin. The upper solid curve in Figure 14 is based on residuals from Equation B1 and the lower curve on Equation B3. The characteristic triangular pattern is clearly replicated.

One can revise the regression models to capitalize on the autocorrelation structure of residuals and reduce the number of free parameters. Equation B5 formalizes the hypothesis that the two time-lagged variables R_{t-1} and S_{t-1} in Equation B2 do not act independently but as a residual-like unit of the form $(R_{t-1} - aS_{t-1})$.

$$\hat{R}_t = aS_t + b(R_{t-1} - aS_{t-1}). \tag{B5}$$

This in effect says that Equation B2 is really Equation B5 in disguise. This hypothesis can be tested in two ways. First, it predicts that the coefficient d in front of S_{t-1} in Equation B2 should equal the product of the other two coefficients (DeCarlo & Cross, 1990). In symbols, $d = -ab$. This relationship is easy to test as we have independent estimates of a, b, and d for each observer. We construct a new variable $d' = -ab$ and check if it predicts d. The correlation between the two is .92. Thus, as DeCarlo and Cross (1990) also pointed out, the negative coefficient in front of S_{t-1} in Equation B2 may appear for reasons very different from the usual "perceptual contrast" interpretation.

A second test is to compare the empirical fits of Equations B2 and B5. Equation B5 accounts for 81% of the response variance on aver-

(*Appendixes continue*)

age.[B1] The mean decrease relative to Equation B2 is only $\Delta R^2 = -.003$. The individual decreases are not significant for at least half of the participants even under the high statistical power of the analysis. This evidence leads us to prefer the more parsimonious Equation B5.

Equation B4 can be modified in an analogous way with similarly negligible decrease in R^2, leading to Equation 20 in the main text. See Figure B1 for details.

[B1] Equation B5 is not linear in the coefficients and hence the standard regression procedures no longer apply. We used a general-purpose optimizer (MATLAB's lsqnonlin; see Footnote 1) to find coefficients *a* and *b* that minimize the sum of squares $\Sigma\ (R_t - \hat{R}_t)^2$.

Appendix C

Derivation of the Steady State in Figure 17

The dynamics of the simplified example in Figure 17 can be solved analytically. Let L_t and R_t denote the locations of the two anchors. The competitive learning mechanism can be approximated by a deterministic iterative process in which L_t and R_t jointly determine the boundary x_t, and the barycenters of the resulting partitions determine the next anchor locations. For the triangular distribution in Figure 17 the partitions have expected values $L_{t+1} = 2x_t/3$ and $R_{t+1} = 2x_t/3 + 2a^2/3(x_t + a)$. Resetting the boundary halfway in between produces the following iterated map:

$$x_{t+1} = f(x_t) = \frac{2}{3}\,x_t + \frac{a^2}{3(x_t + a)}\,. \tag{C1}$$

The steady state must satisfy $x^* = f(x^*)$. This equation has a unique solution in the range $[0, a]$—the "golden section" $\varphi = a(\sqrt{5} - 1)/2$. The stability of this fixed point is governed by the eigenvalue $\lambda = f'(x^*)$ (Strogatz, 1994). Because $0 < f'(x) < 1$ for all $x \in [0, a]$, φ is a global attractor with monotonic (nonoscillatory) convergence.

Appendix D

Parameter Optimization and Model Selection

To factor in the individual differences in the human population, we estimated a separate parameter set for each observer in Experiments 1 and 2. The optimization algorithm is quite technical and is presented at length elsewhere (Petrov, 2001). Briefly, it is based on a Bayesian framework that treats the internal representations in the model as hidden variables. The conditional probability distributions relating these variables to one another and to the observable stimuli and responses are calculated from the structural equations of the model. A special model-tracing version of the ANCHOR software calculates the probabilities of producing each of the nine possible responses on a given trial. The global goodness of fit to a given stimulus–response sequence can then be quantified by the summary log-likelihood. A general-purpose optimizer (MATLAB's fmincon; see Footnote 1) searches the parameter space to maximize the fit.

In the interest of space we report only the parameters for the full model $\mathcal{M}_{\mathrm{PSACR}}$. The means and standard deviations for the 24 absolute identification sequences are memory noise $k_m = .058$ ($SD = .024$), softmax temperature $T = .050$ ($SD = .009$), history weight $H = .071$ ($SD = .036$), cutoff $c = .36$ ($SD = .09$), and log-likelihood $L = 492$ ($SD = 66$). For the 40 category rating sequences, they are $k_m = .076$ ($SD = .016$), $T = .050$ ($SD = .008$), $H = .108$ ($SD = .036$), $c = .43$ ($SD = .08$), and $L = 460$ ($SD = 83$). The temperature variability is constrained by tight search bounds (.040–.060); the bounds used for the other parameters are quite liberal. The optimal parameters for the partial models are very similar, with lower noise levels compensating when the correction mechanism is disabled. Details are available online at http://www.socsci.uci.edu/~apetrov/.

Equation D1 defines the Bayesian information criterion (BIC) for a given model $\mathcal{M}$ and a given stimulus–response sequence D (Schwartz, 1978; Wassermann, 2000):

$$\mathrm{BIC}(\mathcal{M}) = -\log P(D|M, \hat{\theta}) + p_M \log(n)/2, \tag{D1}$$

where $\hat{\theta}$ is the parameter set that maximizes log P for that sequence, p_{M} is the number of free parameters in the model, and n is the number of trials. For $n = 450$, each parameter incurs a penalty of about 3 points. The values reported in Table 4 are averaged across observers, which is equivalent to a master BIC calculated for all available data.

Note that several factors limit the reliability of BIC for our data. The log-likelihood is computationally intractable unless certain approximations and simplifications are used (Petrov, 2001). The approximate log P guides the parameter search well, but its value as a model-selection criterion is less certain. Furthermore, many of the technical assumptions behind Equation D1 are violated: The trials are neither independent nor identically distributed, the models are not strictly nested, and none of them is the "true" model because none can account for the bow effect. BIC, therefore, should be interpreted with caution. We tend to put much greater emphasis on the qualitative analysis of the phenomena predicted by the models.

Received February 28, 2004
Revision received June 16, 2004
Accepted June 28, 2004 ■