Carnegie Mellon University
Research Showcase @ CMU

Department of Psychology | Dietrich College of Humanities and Social Sciences

8-2000

ANCHOR: A Memory-Based Model of Category Rating

Alexander A. Petrov
Carnegie Mellon University

John R. Anderson
Carnegie Mellon University

Follow this and additional works at: http://repository.cmu.edu/psychology

This Conference Proceeding is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Psychology by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

ANCHOR: A Memory-Based Model of Category Rating

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

John R. Anderson (ja+@cmu.edu)
Department of Psychology; Carnegie Mellon University
Pittsburgh, PA 15213 USA

Abstract

This paper attempts to draw a bridge between psychophysics and memory research by proposing a memory-based model of category rating. The model is based on the cognitive architecture ACT-R and uses *anchors* stored in memory that serve as prototypes for the stimuli classified within a response category. The anchors are retrieved by a partial matching mechanism and updated dynamically by an incremental learning mechanism. Anchors also have base-level activations that reflect the frequency and recency of the responses. These mechanisms give rise to sequential effects and nonuniform response distributions. A psychological experiment involving category rating of physical length is reported and the predictions of the model are compared against the empirical data. The psychophysical implications of the model are discussed.

Introduction

Category rating is a widely used method of data collection in experimental psychology. A category-rating situation arises whenever the participants are asked to assign each stimulus to one of several ordered categories such as *1, 2, …, 9* or *very dissimilar, …, very similar*. Procedures of this kind are common for many studies ranging from psychophysical scaling to similarity judgment to personality inventories. Therefore a detailed analysis of the cognitive mechanisms underlying this task is potentially relevant to a diverse set of situations.

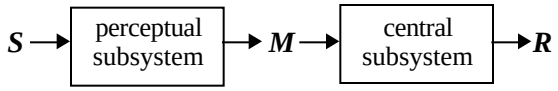

Figure 1: Simplified decomposition of the category-rating process. The external stimulus $\boldsymbol{S}$ maps to an internal magnitude $\boldsymbol{M}$ which in turn gives rise to the overt response $\boldsymbol{R}$.

A rough decomposition of the process of category rating is presented in Figure 1. (This diagram is by no means complete or accurate; it is provided for expository purposes only.) The *perceptual subsystem* maps the external stimulus $\boldsymbol{S}$ onto an internal representation $\boldsymbol{M}$ on a psychological continuum. In this paper the internal representation is called *magnitude*. The magnitude $\boldsymbol{M}$ then serves as a basis for generating an overt response $\boldsymbol{R}$ on the category scale. The latter transformation is the responsibility of the *central* (or *cognitive*) *subsystem*. Both subsystems are characterized with internal states that unfold in time and may differ from trial to trial. Thus each box in Figure 1 has underlying dynamics and the whole system is more complex than the open-loop pipeline suggested by the diagram.

The present paper focuses on the central subsystem and the computational mechanisms converting subjective magnitudes into external reports. While the perceptual aspects of the process are certainly important, they are not central to the research reported here. Therefore the research strategy has been to try to minimize the contribution of the perceptual subsystem so that the properties of the central one can show through. This dictated the choice of a modality for which the perceptual transformation is as simple as possible—physical length.

The empirical relation between stimulus intensities $\boldsymbol{S}$ and averaged category ratings $\boldsymbol{R}$ tends to follow a power function: $\boldsymbol{R}=\boldsymbol{k}.\boldsymbol{S}^{\boldsymbol{n}}$ (Stevens, 1957). The exponent $\boldsymbol{n}$ is characteristic of the perceptual modality. For physical length, this exponent is very close to 1.0 (Stevens, 1957). In other words, the scale is linear. Thus it seems reasonable to assume that the perceptual subsystem delivers veridical representations of physical length, with little if any systematic distortions (Krantz, 1972). Under this assumption, any patterns in the category-rating data for length are largely due to the central subsystem.

The psychophysical literature reports several phenomena related to category rating. The most basic finding is that the participants are able to perform this task without major difficulties and provide robust and regular data: the average rating values vary smoothly with stimulus intensity (Stevens, 1957). This is true whether or not feedback is provided (e.g. Ward & Lockhead, 1970). The second major finding is Stevens' power law stated above. In addition to these first-order results, there are several second-order effects as well.

The *sequential effects* are of special interest here because they shed light on the dynamics of the rating process. Numerous studies have indicated that the successive trials in a rating experiment are not independent (Ward & Lockhead, 1970; Jesteadt et al., 1977; Petzold, 1981; Schifferstein & Frijters, 1992). The responses, regarded as a time series, show autocorrelational structure. Typically the data are analyzed using multiple regression in which the stimulus $\boldsymbol{S}_{t-1}$ and the response $\boldsymbol{R}_{t-1}$ on the preceding trial enter as predictors after the contribution of the current stimulus $\boldsymbol{S}_t$ has

been partialed out. A robust finding is that current responses tend to be contrasted (i.e. negatively correlated) with previous stimuli and assimilated (positively correlated) toward previous responses. Moreover, there is an interaction between the two time-lagged variables $\boldsymbol{S}_{t-1}$ and $\boldsymbol{R}_{t-1}$. The assimilation towards the previous response seems to be modulated by the difference between the two consecutive stimuli $\boldsymbol{S}_{t-1}$ and $\boldsymbol{S}_{t}$ (Jesteadt et al., 1977; Petzold, 1981). The closer the stimuli, the stronger the assimilation.

Theoretical analysis of the task also invites the hypothesis that some form of memory is involved in the rating process. Consider a trial in a category-rating experiment. The presentation of the stimulus evokes some subjective percept in the participant. The participant is then faced with the problem of communicating this subjective percept using the particular response scale chosen by the experimenter. There is no a priori correspondence between the subjective magnitudes and the response categories. Such correspondence must be established at the beginning of the experiment and then applied consistently until the end. This is a role for memory.

This hypothesis is supported by a study of Ward and Lockhead (1970). The experiment involved 8 sessions on 8 consecutive days. Feedback was provided at the end of each trial. Unbeknown to the participants the feedback was manipulated so that the response categories were associated with different stimuli on different days. This caused systematic shifts in participants' responses.

The thesis of the present paper is that memory plays an important role in category rating and in particular in the transition from internal magnitudes to overt responses. Memory maintains the consistency of responses over periods of hours and even days. Moreover, the hypothesis is that failures to achieve perfect consistency—manifested as response drifts, sequential effects, and context effects—are due to the plasticity of the memory system and reflect the dynamics of its operation.

This paper reports the initial steps towards a memory-based theory of category rating. The theory is instantiated in a computational model called ANCHOR and the predictions of the model are compared with empirical data.

Psychological Experiment

The ANCHOR model makes detailed predictions on a trial-by-trial basis. To estimate the parameters of the model and evaluate its adequacy as a psychological theory one needs empirical data at the same level of granularity. The psychophysical literature cited in the introduction reports aggregate data only and hence falls short of this standard. Therefore, a psychological experiment was carried out. In addition to providing the necessary data, it replicates the sequential effects from the literature and tests the assumption of linearity of the scale of physical length.

Method

Stimulus Material. The stimuli were pairs of white dots presented against black background on a 17-inch AppleVision monitor. The only independent variable in the experiment was the distance between the two dots measured in pixels. The distance used on each trial was drawn independently from a uniform distribution ranging from 250 pixels (80 mm) to 700 pixels (224 mm). The viewing distance was approximately 500 mm. 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. The maximal distance representable on the monitor was 1000 pixels (320 mm). Each dot was roughly circular in shape with a diameter of 16 pixels (5 mm).

Participants: 24 students participated in the experiment to satisfy a course requirement.

Procedure. The participants were asked to rate the "distance between the dots" on a scale ranging from *1* to *9*. The participants entered their responses on the numeric keypad of the computer keyboard. Each trial began with a 500 ms beep followed by 3300 ms stimulus presentation followed by 200 ms inter-trial interval. There were 17 demonstration and 450 experimental trials divided into 10 blocks with short rest periods between the blocks. The demonstration presented stimuli of length 275, 325, 375, …,625, 675, 625, …, 275 pixels and the participants were encouraged to practice pressing the keys 1, 2, …, 8, 9, 8, …, 1. No feedback was given during the experimental trials. The whole procedure lasted about 40 minutes.

Results and Discussion

The data are analyzed at the level of individual participants.

Linearity of the Scale. To estimate the exponent of Stevens' power law, a function of the form $\boldsymbol{R} = \boldsymbol{a} + \boldsymbol{k}.\boldsymbol{S}^{\boldsymbol{n}}$ is fitted to the data of each individual participant. The exponents $\boldsymbol{n}$ range from 1.01 to 1.12 in the sample of 24 participants, with mean 1.06. Thus the exponent is empirically indistinguishable from unity for all participants. (The correlations between the functions $\boldsymbol{S}^{0.95}$, $\boldsymbol{S}^{1.00}$, and $\boldsymbol{S}^{1.10}$ are greater than 0.99 in the domain [250;700].) This suggests that the assumption of linearity of the scale is correct, at least within the precision of measurement.

Overall Accuracy. The linearity of the scale allows the data to be analyzed by simple linear regression of $\boldsymbol{R}$ on $\boldsymbol{S}$. The squared correlation coefficient R^2 is a measure of the accuracy of the respective participant. It ranges from 0.65 to 0.91 for the 24 participants, with mean 0.80 and std.dev. 0.070. In other words, the immediate stimulus accounts for full three quarters of the response variance, sometimes up to 90%.

Response Distributions. Even though the stimuli are uniformly distributed, the responses are not. Figure 2 shows the response distributions for two representative participants. A marked feature of these distributions is the predominance of responses in the middle of the scale at the expense of extreme ones. The response standard deviation ranges from 1.20 to 2.44, with mean 1.96 and s.d. 0.28. For comparison, if the 450 responses were evenly distributed in 9 categories, the standard deviation would be 2.58.

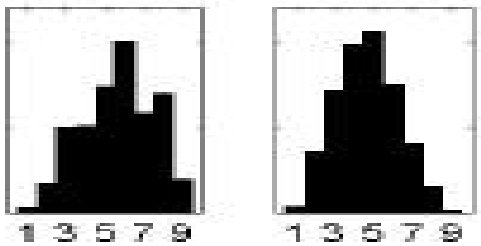

Figure 2: Response distributions for two representative participants.

It seems unlikely that the perceptual subsystem maps the uniform stimulus distribution onto a highly non-uniform distribution of internal magnitudes. Therefore the shape of the response distribution appears to be largely due to the cognitive subsystem. It is possible that the participants reserve the extreme responses for distances that are very short (close to zero) or very long (filling the width of the screen). Such extreme stimuli are not presented during the experiment and this may be one of the reasons for the non-uniformity of responses. However, this explanation does not address the peak in the middle of the scale. The memory-based theory of category rating offers an alternative explanation in terms of self-reinforcing buildup of strength for the frequent responses and corresponding loss of strength for the infrequent ones.

Sequential Effects. A multiple linear regression is performed with the following variables entering as predictors: the current stimulus $\boldsymbol{S}_t$, the previous stimulus $\boldsymbol{S}_{t-1}$, and the previous response $\boldsymbol{R}_{t-1}$. The signs of the regression coefficients of the time-lagged variables are of special interest. For the previous stimulus $\boldsymbol{S}_{t-1}$, the standardized coefficient β_S ranged from –0.53 to –0.08, with mean –0.25 and s.d. 0.10. Conversely, the standardized coefficient β_R for the previous response $\boldsymbol{R}_{t-1}$ ranged from +0.15 to +0.55, with mean +0.30 and s.d. 0.10. Thus all 24 participants without exception show evidence of stimulus-driven contrast and response-driven assimilation.

Additional regression analyses involving interaction terms replicate the finding of Jesteadt et al. (1977) that the assimilation towards $\boldsymbol{R}_{t-1}$ is modulated by the difference between the two consecutive stimuli $\boldsymbol{S}_{t-1}$ and $\boldsymbol{S}_t$. These analyses are not reported here because of lack of space.

Memory Based Model of Category Rating

As argued in the introduction, memory seems to play an important role in the category-rating process. The remainder of this paper outlines one particular proposal about the computational mechanisms that may carry out this process. The ANCHOR model proposed here is based on a general theory of memory incorporated in the ACT-R cognitive architecture (Anderson & Lebière, 1998). The ACT-R theory is consistent with a broad range of memory phenomena. Thus ANCHOR draws a bridge between psychophysics and memory research. The following two subsections describe the model first in general terms and then with details and equations.

Main Principles of the Model

The centerpiece of the ANCHOR model is the construct of an *anchor*. An anchor is an association between an internal magnitude and a category on the response scale. There is one anchor per category and it can be construed as an internal representation of the prototypical member of this category.

The collection of all anchors defines a mapping from the continuum of magnitudes to the discrete categories of the response scale. This mapping is partly constrained and partly arbitrary. The constraints come from the demand for homomorphism implied by the category-rating task. There is intrinsic ordering of the intensity of the physical stimuli and hence of the magnitudes on the subjective continuum. Also, there is ordering of the response categories. When reporting their subjective magnitudes, the participants try to align the ordering of the two domains.

Another constraint implied by the task is to maintain consistency over time. If, for whatever reason, a stimulus is labeled with a particular response on a given trial, there is pressure to label this stimulus with the same response on subsequent trials. This extends not only to the stimulus that happened to be presented but to other stimuli that evoke similar subjective magnitudes.

These constraints motivate the following mechanisms of the ANCHOR model. When a stimulus is presented and encoded as an internal magnitude, a *partial matching mechanism* activates an anchor whose magnitude is similar to the magnitude of the target stimulus. In so far as anchor magnitudes are relatively stable, categorization of the stimuli is consistent over time.

The partial matching is stochastic and depends on other factors besides similarity (viz. recency and frequency, discussed below). Therefore it is not guaranteed to retrieve on each trial the anchor that best matches the target magnitude. In the cases when there is large discrepancy between the target magnitude evoked by the stimulus and the anchor magnitude retrieved from memory, a *correction mechanism* may increment or decrement the response suggested by the anchor. The correction mechanism is stochastic and error-prone too but it does tend to enforce homomorphism between magnitudes and responses.

Phenomenologically, an introspective report of a category-rating trial might run like this, "I see the dots… The distance looks like a *7*… No, it's too short for a *7*. I'll give it a *6*."

So, the stimulus has been encoded, matched against anchors, and a response has been produced. Is this the end of the trial? According to the ANCHOR model and the broader ACT-R theory (Anderson & Lebière, 1998), the answer is no. The cognitive system is plastic (within limits) and each experience seems to leave a mark on it. It is impossible to step into the same river twice. The model postulates an obligatory *learning mechanism* that pulls the magnitude of the relevant anchor in the direction of the magnitude of the stimulus that has just been presented. Thus each trial results in a slight change of the magnitude of one of the anchors—namely the one that corresponds to the response given on that particular trial. The notion of obligatory learning is similar to the ideas of Logan (1988), although ANCHOR learns prototypes rather than individual instances.

The implications of this incremental learning mechanism are worth considering in detail. After a long sequence of trials, each anchor magnitude ends up being a weighted average of the magnitudes of all stimuli classified in the corresponding response category. Thus the anchors are true prototypes. However, recent stimuli weigh more heavily than earlier ones, introducing bias. The influence of the initial instructions and demonstrations gradually wash away.

More importantly, the performance of the system on each trial depends on the history of its performance on previous trials. This makes it a dynamic system capable and even forced to exhibit gradual shifts, sequential effects, and self-reinforcing preferences. Each run of the model becomes idiosyncratic in systematic ways apart from the random noise even when tested on the exact same sequence of stimuli.

One final aspect of the model remains to be introduced. There is abundant evidence that the human memory system is sensitive to the frequency and recency of the encoded material. These two factors enter the ACT-R theory and the ANCHOR model through a construct called *base-level activation (BLA)*. Each memory element, anchors included, has some base-level activation that goes up and down with time. The partial matching mechanism is sensitive not only to the similarity between the target magnitude and the anchor magnitudes but also to the activation levels of the anchors. Overall, anchors with high BLA are more likely to win in the matching process than anchors with low BLA, the target stimulus notwithstanding.

The form of the base-level learning equation (Eq. 6 below) entails that when a response is produced on a trial the BLA of the corresponding anchor receives a sharp transient boost followed by small residual increase. On the other hand, when some response is not used for a long time the activation of the corresponding anchor gradually decays away. In terms of observable behavior, the rapid transient manifests itself as sequential response assimilation and the long-term overall strength leads to rich-gets-richer differentiation of the response frequencies.

Details and Equations

Figure 3 shows a schematic diagram of the various quantities used in the model and the dependencies among them.

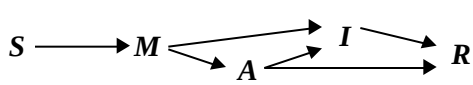

Figure 3: Schematic diagram of the quantities used in the model: physical intensity of the stimulus $\boldsymbol{S}$, target magnitude $\boldsymbol{M}$, anchor magnitude $\boldsymbol{A}$, increment $\boldsymbol{I}$, and overt response $\boldsymbol{R}$.

The perceptual subsystem (cf. Figure 1) is modeled by a single equation [1]. It transforms the physical intensity of the stimulus $\boldsymbol{S}$ into an internal magnitude $\boldsymbol{M}$. The transformation is linear, with some multiplicative noise. The magnitudes are arbitrarily scaled between 0.25 and 0.70, given that $\boldsymbol{S}$ varies between 250 and 700 pixels. The random variable ε is normally distributed with zero mean. Thus the term $(1+\varepsilon)$ is centered around 1.0. The standard deviation of the noise is a free parameter of the model. In the simulation experiments reported in the next section this parameter was set to 0.050. The multiplicative relationship between the scale value (i.e. the mean of the magnitude distribution induced by a given stimulus $\boldsymbol{S}$) and the noise term implements Ekman's law (Ekman, 1959).

$$\boldsymbol{M = S.(1+\varepsilon) \,/\, 1000} \qquad [1]$$

There are 9 anchors with magnitudes $\boldsymbol{A_1 \ldots A_9}$ respectively. The partial matching mechanism has to select one of them according to their similarity to the target magnitude $\boldsymbol{M}$ and their base-level activations $\boldsymbol{B_1 \ldots B_9}$. This process is governed by two equations. First, a *score* is produced for each anchor according to Eq. 2. Second, one anchor is chosen according to the *softmax* Equation 3.

$$\boldsymbol{Score_i = B_i - MP.|M - A_i|} \qquad [2]$$

The *mismatch* (or dissimilarity) between two magnitudes is simply the absolute difference between them. The mismatch is multiplied by a *mismatch penalty factor* $\boldsymbol{MP}$ and subtracted from the base-level activation of the anchor to produce the combined score for this anchor. $\boldsymbol{MP}$ is a free parameter of the model that scales the mismatches relative to the activation values. It was set to 7.0 in the simulations.

$$\boldsymbol{P_i = exp(Score_i \,/\, t) \,/\, \sum_j exp(Score_j \,/\, t)} \qquad [3]$$

Equation 3 converts scores into retrieval probabilities. $\boldsymbol{P_i}$ is the probability of retrieval of anchor $\boldsymbol{i}$ and $\boldsymbol{exp(\cdot)}$ denotes the exponentiation function. The *temperature* $\boldsymbol{t}$ is a free parameter of the model controlling the degree of non-determinism of the partial-matching process. It was set to 0.40 in the simulations.

Having retrieved an anchor, the model has to determine the correction $\boldsymbol{I}$ to produce the final response. Under the current settings of the model, the correction can be 0, +/–1, and occasionally +/–2. The correction depends, stochastically, on the discrepancy between the target magnitude $\boldsymbol{M}$ and the anchor magnitude $\boldsymbol{A}$. One free parameter of the model—$\boldsymbol{d}$—defines a set of five *discrepancy reference points* $\{\boldsymbol{-2d}, \boldsymbol{-d}, 0, \boldsymbol{d}, \boldsymbol{2d}\}$. They are compared with the algebraic difference $\boldsymbol{(M-A)}$ to produce *correction scores*:

$$\boldsymbol{CorrScore_k = \left| d_k - (M-A) \right|} \quad , k = -2 \ldots +2 \qquad [4]$$

The correction scores are converted to choice probabilities by an equation analogous to Eq. 3. The only differences are that the correction scores enter with negative signs, thus transforming the softmax rule into softmin, and that a separate temperature parameter is used. In the simulations this parameter was set to 0.040. The discrepancy reference parameter was $\boldsymbol{d}$=0.090. To illustrate these settings, suppose the anchor magnitude $\boldsymbol{A}$ is 0.050 below the target magnitude $\boldsymbol{M}$, which is roughly the width of one response category. Then there is 51% chance that the model will increment the anchor response by +1, 39% chance to leave it unchanged, and marginal chance to increment it by +2 or decrement it.

The final response $\boldsymbol{R}$ is the algebraic sum of the anchor label and the increment, clipped between 1 and 9 if needed.

At the end of the trial the learning mechanism updates the magnitude of the anchor corresponding to the response $\boldsymbol{R}$. (Note that this does not necessarily coincide with the anchor retrieved from memory.) The anchor magnitude $\boldsymbol{A}$ is updated according to Eq. 5, which is a form of competitive learning.

The learning rate α weighs the most recent trial relative to earlier ones. The simulation experiments used α=0.50.

$$new_A = \alpha.M + (1-\alpha).old_A \qquad [5]$$

The base-level learning equation is somewhat less transparent. The ACT-R theory postulates Equation 6a which contains an explicit term for each instant the anchor is updated (Anderson & Lebière, 1998, p.124). Suppose a particular response has been given at time lags $t_1,\ldots,t_n$ from the present trial. Then the base-level activation B of the corresponding anchor is the logarithm of a sum of powers [6a], where d is a decay parameter.

$$B = ln\left(\sum_i t_i^{-d}\right) \qquad [6a]$$

Because Equation 6a is computationally expensive, the model uses Eq. 6b which closely approximates the theoretical formula. The approximation disregards the detailed update history and retains only the time lag since the last usage t, the lag T since the beginning of the experiment, and the total number of times the corresponding response has been given up to the current trial. In the simulation experiments the decay parameter was set to d=0.5, which is a default value used in many ACT-R models. The duration of each trial was 4 sec, as in the psychological experiment.

$$B = ln\left[t^{-d} + n.(T^{1-d}-t^{1-d}) \;/\; [(1-d)(T-t)]\right] \qquad [6b]$$

Equations 2, 3, 4, and 6a are taken verbatim from the ACT-R architecture (Anderson & Lebière, 1998) and thus establish continuity between the ANCHOR model and a broad spectrum of memory-related models. Equation1 is ANCHOR's connection to Stevens' and Ekman's psychophysical laws.

Evaluation of the Model

Simulation Experiment

In order to test the model, its computer implementation was run on the 24 random sequences of stimuli used in the psychological experiment. To mimic the effect of the introductory demonstration, the magnitudes of the anchors were initialized as follows. Anchor *9* was set to 0.800—a compromise value between the longest stimulus presented on the demonstration (675 pixels) and the total width of the screen (1000 pixels). Anchor *1* was initialized to 0.150 and the remaining anchors were evenly spaced in between. The other parameters were set as reported in the previous section. The model generated 24 sequences of responses which were then analyzed in the same way as the psychological data.

Table 1: Comparison of the performance of the model and the psychological data. See text for details.

	Human data				**Model**			
Statistic	min	**mean**	max	s.d.	min	**mean**	max	s.d.
Accuracy (R^2)	.65	**.80**	.91	.07	.65	**.76**	.84	.05
Resp. std.dev.	1.20	**1.96**	2.44	.28	1.58	**1.81**	2.57	.21
Multiple R^2	.67	**.83**	.93	.07	.73	**.78**	.84	.03
Increase in R^2	.00	**.02**	.06	.01	.00	**.02**	.10	.02
β for S_t	.80	**.90**	.93	.04	.80	**.87**	.92	.03
β for S_{t-1}	–.53	**–.25**	–.08	.10	–.47	**–.23**	–.10	.09
β for R_{t-1}	+.15	**+.30**	+.55	.10	+.13	**+.25**	+.53	.10

Table 1 summarizes the outcome of these various analyses and compares the performance of the model with the human data. The overall accuracy of the model, operationalized as the squared correlation between stimuli and responses, ranges from 0.65 to 0.84 in the sample of 24 runs, with mean 0.76 and standard deviation 0.046. The mean R^2 for the psychological data is 0.80. The degree of non-uniformity of the response distribution is reflected in the standard deviations reported in the second row of Table 1.

The remainder of Table 1 summarizes the multiple regression analysis of the response R_t on the current stimulus S_t, previous stimulus S_{t-1}, and previous response R_{t-1}. The model shows the same pattern of sequential effects as the psychological data.

Overall, the results of the simulation experiment suggest that the ANCHOR model closely matches human category-rating behavior. The biggest discrepancy between the two data sets is that the model responses are less variable. The human data, however, includes both within-subject and between-subject variability whereas the parameter settings of the model were fixed for all 24 runs. Individual differences can be modeled by using different parameter settings for the different runs.

Explanation of the Empirical Phenomena

The fact that a model fits the data indicates that its computational mechanisms hang together and can be brought in line with the empirical observations. A much more acid test for the utility of the model, however, is the degree to which it contributes to the theoretical understanding of the psychological phenomena. This closing section discusses the empirical effects in light of the ANCHOR model.

Nonuniformity of the Response Distribution. The model shifts the level of theorizing from aggregate scale values to individual responses. At that level of granularity the entire response distribution becomes important. Two salient features of this distribution appear to be the predominance of responses in the middle of the scale and the relative infrequency of extreme responses (Figure 2). Several factors conspire to produce such distributions in the model. The base-level learning mechanism (Eq. 6a/b) tends to differentiate the response frequencies—more frequent anchors build up strength which in turn makes them more likely to be retrieved in the future. This makes flat distributions unstable—small differences tend to grow. This self-reinforcing dynamics cannot go out of hand, however, because of three stabilizing factors. First and foremost, the immediate stimulus controls about 75% of the response variance and hence the responses cannot stray too much from the stimuli. Second, the correction mechanism redistributes the strength among neighboring anchors. This inhibits the formation of isolated spikes or gaps in the distribution, making the smooth unimodal shape the most stable configuration. . The third stabilizing factor is related to the context effects discussed below.

Context Effect. If the stimuli control 75% of the response variance and the base-level learning tends to amplify inequalities, what happens when the stimuli are unevenly

distributed themselves? It may appear that the model would produce responses that are even more skewed. This would directly contradict the finding of several studies (Parducci, 1965; Parducci & Wedell, 1986; Schifferstein & Frijters, 1992). Empirically, the responses tend to be less skewed than the stimuli, not more so. However, simulation experiments with the ANCHOR model that are too long to be detailed here indicate that it produces context effects consistent with the empirical data. In a nutshell, this is due to the anchor adjustment Equation 5. Because the anchors are prototypes, they tend to cluster in those regions of the magnitude continuum that are densely populated with stimuli. In turn, this reduces the skewness of the response distribution.

Sequential Effects. The positive autocorrelation between responses on successive trials is a direct consequence of the recency component of base-level activations (Eq. 6a/b). When a particular response is given, the BLA of its corresponding anchor goes up, which in turn improves the probability of retrieving the same anchor on the next trial. This produces assimilation towards the previous response. However, the increase of the activation level matters only when the two successive stimuli are similar enough (cf. Eq. 2). If they are too far apart, the response on the first trial primes an anchor that is too remote from the target on the second trial to have any influence on the final outcome. The closer the two consecutive stimuli, the stronger the assimilation.

Another sequential effect is the negative correlation between the response $\boldsymbol{R}_t$ on a given trial and the stimulus $\boldsymbol{S}_{t-1}$ on the previous trial. Part of this effect is probably due to the perceptual subsystem and its tendency to enhance contrasts. The ANCHOR model, however, has a deliberately simplified front end that precludes any interaction between the stimuli at the perceptual level. Still, the model exhibits contrast effects due to the plasticity of anchor magnitudes (Eq. 5) and the discrepancy penalizing aspect of the partial matching mechanism (Eq. 2). The magnitude of the past stimulus $\boldsymbol{S}_{t-1}$ is averaged into the magnitude of one of the anchors, which then serves as a proxy of that stimulus on subsequent trials. The anchor magnitudes $\boldsymbol{A}_i$ are subtracted from the new target magnitude $\boldsymbol{M}$ during the partial matching process. In other words, one of the $\boldsymbol{A}_i$ terms in Eq. 2 is positively correlated with $\boldsymbol{S}_{t-1}$, $\boldsymbol{M}$ is positively correlated with $\boldsymbol{R}_t$,, and $\boldsymbol{A}_i$ and $\boldsymbol{M}$ are subtracted from each other. This creates negative relationship between the response $\boldsymbol{R}_t$ and the previous stimulus $\boldsymbol{S}_{t-1}$.

Memory-Related Effects. The anchors are stored in memory and decay only slowly with time. Therefore, the mapping from stimuli to responses implicit in these anchors can influence the performance hours and even days later.

This paper argues in favor of the hypothesis that category ratings are produced in a memory-based manner. A range of category-rating phenomena seem to arise naturally from a set of principles that are also consistent with a large body of memory research. In so far as the ANCHOR model is successfull, it illustrates the advantages of its integrative methodology and the utility of general architectures for cognitive modeling.

Acknowledgments

This research is supported in part by grant AFOSR F49620-99-10086 awarded to John Anderson. The authors thank Stefan Mateeff, Stephen Gotts, and two anonymous reviewers for their valuable comments on the paper. The contribution of Stefan Mateeff is especially gratefully acknowledged.

References

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

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

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

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

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95* (4), 492-527.

Parducci, A.(1965). Category judgment: A range-frequency model. *Psychological Review, 72* (6), 407-418.

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

Petzold, P. (1981). Distance effects on sequential dependencies in categorical judgments. *Journal of Experimental Psychology: Human Perception and Performance, 7* (6), 1371-1385.

Schifferstein, H. N. J. & Frijters, J. E. R. (1992). Contextual and sequential effects on judgments of sweetness intensity. *Perception & Psychophysics, 52* (3), 243-255.

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

Ward, L. M. & Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology, 84* (1), 27-34.