Title
Does Irrelevant Information Play a Role in Judgment?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Authors
Kokinov, Boicho
Hristova, Penka
Petkov, Georgi

Publication Date
2004

Peer reviewed

Does Irrelevant Information Play a Role in Judgment?

Boicho Kokinov (bkokinov@nbu.bg)[12]
Penka Hristova (phristova@cogs.nbu.bg)[1]
Georgi Petkov (gpetkov@cogs.nbu.bg)[1]
[1]Central and East European Center for Cognitive Science, Department of Cognitive Science and Psychology,
New Bulgarian University, 21 Montevideo Street
Sofia 1635, Bulgaria

[2]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Street, bl.8
Sofia 1113, Bulgaria

Abstract

This paper presents an unusual prediction made by the DUAL-based model of judgment JUDGEMAP and its verification. The model is shortly presented as well as the simulation data obtained with it. These data predict that people will use the information on an irrelevant dimension when judging another dimension. This prediction is then tested in a psychological experiment and confirmed.

Introduction

Suppose that you are judging how tall a person is. Do you expect that the color of his or her eyes will play a role in that process? Or suppose you are judging the quantity of oil in the bottle you are buying, do you expect that the font used on its label will have an effect? Finally, suppose you are judging the length of a given line segment. Do you expect that the color of the line will make a difference?

Both our intuition and the theories of judgment would answer these questions negatively. Basically they would assume that when judging length we ignore all irrelevant features (including color) and only physical length plays a role. Of course, many other factors, like order of presentation and context, may play a role, but only the length of the lines will take part in the judgment.

This paper is challenging this assumption of standard theories of judgment and is trying to answer the above seemingly stupid and self-evident questions and surprisingly to show that all features (including the irrelevant ones) do matter or more precisely they may matter under certain circumstances.

Approaches to Judgment

There are a number of theories of judgment and a few running models. Most of the theories originate from psychophysics and are mathematical in their nature; they do not describe the process of judgment, but only characterize the end result. Since we are interested in describing the process of judgment we will briefly outline only the main approaches proposed so far in that direction.

Judgment as measuring similarity/dissimilarity with a standard. The classical ideal point approach proposed by Coombs (1964, Wedell & Pettibone, 1999) falls into this category. He believes individuals have their "ideal points" and therefore judging a stimulus can be described as comparing it to this standard and measuring the distance toward it. The Adaptation Level Theory (Helson, 1964) falls into the same category, however, here the standard (adaptation level) is changed depending on context. Finally, the Norm Theory (Kahneman & Miller, 1986) follows a similar approach, however, the standard here is called "norm" and what is more important is that this norm is constructed on the spot rather than retrieved from long-term memory. A comparison set is constructed in working memory consisting of known exemplars and its norm is computed. Thus all three theories can be described as relying on comparison of the target stimulus with a standard (Figure 1), but they differ in the degree to which they subscribe to the constructivist approach toward this standard.

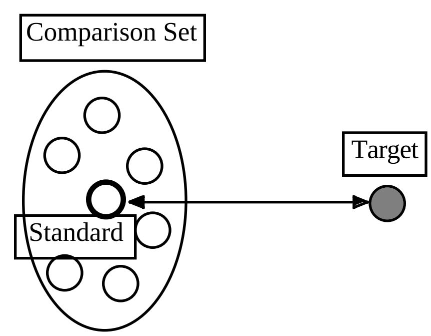

Figure 1. Judgment as comparison with a standard.

Judgment as classification task. Within this approach the comparison set is subdivided into subcategories each of them corresponding to a judgment label (or scale element) and the target stimulus is classified within one of these subcategories. The Range-Frequency Theory (Parduci, 1965, 1974) postulates the constraints which should be met by such category subdivision: the range of value variation within all subcategories should be about the same, and the number of examples in all subcategories should be about the same. The Theory of Criterion Setting (Treisman & Williams, 1984, Treisman, 1985) is a process model that explains how dynamically we change the boundaries of the subcategories. Finally, the ANCHOR model (Petrov &

Anderson, 2000, in press) describes the process of learning of these subcategories and solves the classification task by comparing the target stimulus to the prototypes of each subcategory, these prototypes are supposed to be hold in long-term memory and are called anchors (Figure 2). The comparison set represented by the set of anchors is dynamically formed.

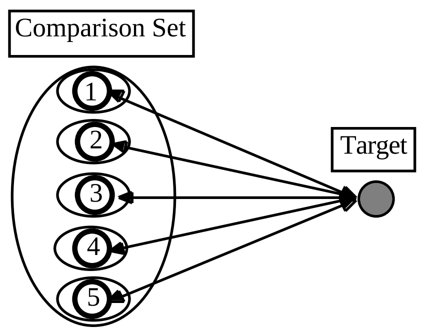

Figure 2. Judgment as classification task. Comparing the target to the standard of each of the subcategories.

Judgment as a mapping task. The DUAL-based model of judgment discussed in this paper follows a third approach: The target stimulus is not compared to the comparison set, but is rather included in it and then a mapping is established between the elements of the comparison set and the set of rating labels (or scale elements). This mapping should be as close as possible to a homomorphism, i.e. the relations among the elements of the comparison sets should be kept among their corresponding rating labels. Thus the process of judgment involves construction of the comparison set, joining the target to it, and mapping between the comparison set and the rating labels (Figure 3).

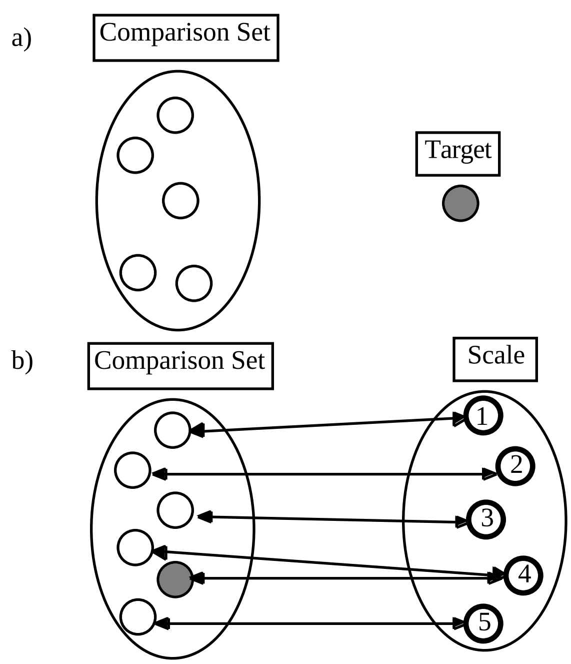

Figure 3. Judgment as mapping in the DUAL-based model.

DUAL-Based Model of Judgment

The current model – JUDGEMAP (Judgment as Mapping) – is based on a general cognitive architecture – DUAL (Kokinov, 1994b, 1994c). This architecture is a hybrid (symbolic/connectionist) one and is explicitly designed to model context-sensitivity of human cognition. It is based on decentralized representations of concepts, objects, and episodes and parallel emergent computations.

The AMBR1 (Kokinov, 1988, 1994a) and AMBR2 (Kokinov, 1998, Kokinov & Petrov, 2001) models are built on DUAL and integrate memory and analogy-making. Since the process of judgment, as described above, involves memory (construction of the comparison set in working memory) and mapping (which is a central mechanism in analogy-making) the JUDGEMAP model is most naturally integrated in DUAL and borrows many of the mechanisms developed for analogy-making in AMBR. Because of the lack of space the model is described only in broad strokes. Interested readers are invited to consult the literature on DUAL and AMBR for more details.

Construction of the comparison set. The comparison set is formed from perception (the target as well as potential context stimuli) and from long-term memory (familiar or recently presented exemplars as well as generalized prototypes, if such exist in LTM). The mechanism responsible for that construction is spreading activation. The sources of activation are the INPUT and GOAL nodes, i.e. the perceived target (and possibly context) stimuli and the goal to judge the stimuli on a scale predefined in the instruction (e.g. a scale from 1 to 7). Thus the representations of the target and the scale elements become sources of activation which is then spread through the network of micro-agents. Naturally, concepts related to the representation of the target become active, e.g. various features of the target – these include both relevant and irrelevant features (of course, relevant features receive more activation than irrelevant ones). The activation spreads further from the general concepts (like RED, GREEN, etc.) towards specific examples of the concepts (other red or green objects). However, there are only a few links from the general concepts to their exemplars – only to the most familiar (typical) exemplars or to recently experienced ones. Thus gradually a number of exemplars (and possibly prototypes) are activated and become part of working memory – all these form the comparison set (Figure 4).

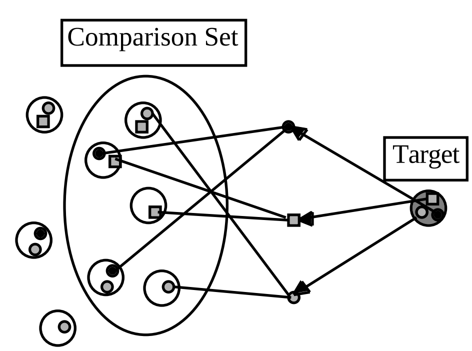

Figure 4. Formation of the comparison set in WM by the spreading activation mechanism of DUAL.

Mapping of the comparison set onto the scale elements. We can now consider the comparison set as a retrieved base and map it onto the scale elements which are the target. The mapping process should preserve the relations among the elements of the comparison set among their images on the scale. The mapping should also follow the range-frequency principle described in the previous section. How is the mapping achieved in JUDGEMAP? Similarly to AMBR, a constraint-satisfaction network is constructed by the marker-passing and structure-correspondence mechanisms. This network consists of temporal agent-hypotheses representing possible correspondences between members of the comparison set and elements of the scale. These initial hypotheses are formed according to the range principle. Excitatory and inhibitory links are constructed among the hypotheses and the spreading activation mechanism selects the winning hypotheses which form the mapping (Figure 5). The competition among the hypotheses implements the frequency principle. As result of this process not only the target stimulus but also each element of the comparison set receives a judgment. This does not mean that people would be aware of all these judgments – most or even all of them might remain unconscious.

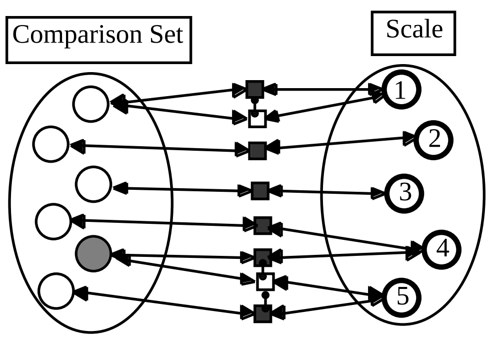

Figure 5. The process of mapping accomplished by the constraint satisfaction mechanism. The winning hypotheses are in black.

Speculative prediction. Since the activation spreads from the target stimulus (represented in a decentralized way by many agents), exemplars, similar in some respect to it (sharing some feature with the target), can be potentially activated and thus become members of the comparison set in working memory. This means that in addition to currently perceived stimuli, to recently activated exemplars, and to highly familiar (typical) exemplars, exemplars which are simply similar to the target will also participate in the comparison set. Moreover, these exemplars might be similar along the relevant (judged) dimension or along an irrelevant dimension.

Let us consider the following example. Suppose we are judging the length of line segments but the lines are colored. Let the target stimulus be a red line of certain length. In this case we may expect that there will be more red lines in the comparison set (Figure 6) – they will be activated through the RED concept which is shared with the target. On the other hand, if the target stimulus is a green line of the same length, more green lines will become part of the comparison set (Figure 7). Now, if it happens that the known red lines are longer than the known green lines, then the two target stimuli (differing only in color) will be included in different comparison sets and thus judged differently and there will be a shift in favor of the green target. Therefore the speculative prediction of JUDGEMAP will be that even such irrelevant feature of the line like its color will play a role in the judgment process. This prediction is in sharp contrast to all theories and models described in the first section, which assume that only the relevant features play a role.

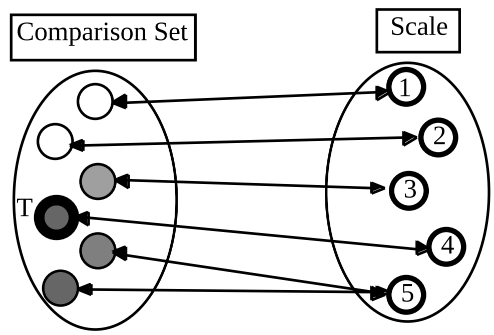

Figure 6. The target stimulus is red and therefore we expect more red exemplars in the comparison set. They happened to be larger in size and thus they compete for the upper part of the scale. In this case the target stimulus (of the same size as in Figure 7) will compete with them and will be mapped onto 4.

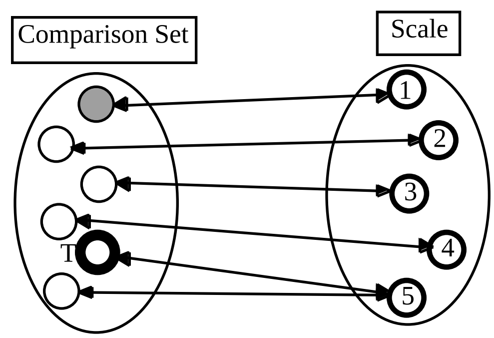

Figure 7. The target stimulus is green and therefore we expect more green exemplars in the comparison set. They happened to be smaller in size and thus they compete for the lower part of the scale. In this case the target stimulus (of the same size as in Figure 6) will compete with them and eventually will be mapped onto 5. In this way we receive an upward shift in the judgment.

Thus we will first describe a simulation experiment with JUDGEMAP that tests in practice this speculation and will also give us a rough estimation of the order of this color effect (if any). If we are successful, we will run a psychological experiment to text the model's prediction and thus verify the model.

Simulation Experiment

In this simulation experiment we use a stimulus set of 56 lines. They are all in the long-term memory of the model. The lines differ in length and color. There are 7 different sizes (from 10 units of length to 34 unit with increment of 4 units) and two different colors (red and green). Thus in each size group there are 8 lines. The frequency of the red (respectively green) lines varies across the size groups. In size group one (the shortest lines - length 10 units) there are 7 green and 1 red line, in the second shortest group (length 14 units) there are 6 green and 2 red lines, etc. In the largest group size (length of 34 units) there are 7 red lines and one green line. Thus we have positively skewed distribution of the green lines and negatively skewed distribution for the red lines.

Each line is represented by a coalition of 5 agents standing for the line itself, for its color, for its length, and for the two relations (color_of and length_of). In addition there are agents standing for the numbers from 0 to 8, but only the agents standing for 1 to 7 are instances of “scale element”.

On each run of the program we connect one of these lines to the input list thus simulating the perception of the target stimulus, and connecting the agent standing for “scale_from_1_to_7” to the goal node thus simulating the instruction for rating on a 7 point scale.

We have produced 42 variations of the knowledge base of the system thus simulating 42 different participants in the experiment. The knowledge bases differ mainly in the associative and instance links among the agents, thus although all our “artificial participants” will know the same lines and the same concepts, they will activate different instances in the comparison set.

For each of these knowledge bases we have run two judgment trials: one for a red line of size 22 and one for a green line of the same size.

Simulation Results

The results from the simulations are presented in Figure 8. As we can see the mean rating of the green lines are in most cases slightly higher than the mean rating of the red lines with the same length.

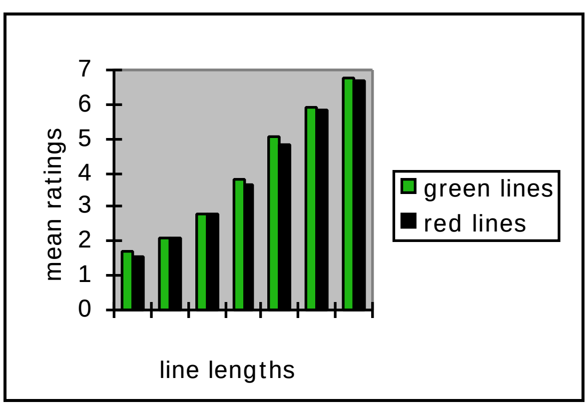

Figure 8. Simulation data. The mean rating of each line with a certain length (1-7) and color (green and red) obtained from all subjects.

Thus the mean of the mean ratings of all red categories is 4.012, while the mean of the mean ratings of all green categories is 4.065, which makes a difference of 0.053 which turns out to be almost significant tested with repeated measurements analysis (F(1,41)=3.917, p=0.055). The data show that the possible size of the color main effect is very small, but may still be significant. This prediction makes sense: on one hand it is small enough, so that we can ignore it in everyday life and this explains why our intuition says that irrelevant information does not play a role in judgment. On the other hand, the simulation predicts that the irrelevant information does play a role and shifts a bit the evaluation. This means that under specific circumstances this shift might be larger and become significant.

The experiment described below is designed to test this prediction of the model. Basically it replicates the simulation experiment with a larger number of lines.

Psychological Experiment

In this experiment human participants rate the length of red and green lines of various sizes. The interesting question is whether we will obtain a main effect of color, i.e. whether there will be a difference between the ratings of the red and green lines of the same size.

Method

Design

The experiment has a 14x2 within-subject factorial design. The independent variables are length (varying at 14 levels) and color (varying at 2 levels: green and red) of the lines. The dependent variable is the rating of the length of the lines on a 7-point-scale. The experimental question is whether there will be a main effect of color, which is supposedly an irrelevant factor in judging length.

Material

A set of 14 color lines has been presented horizontally against a gray background on a 17-inch monitor. The shortest line is 12 pixels, the longest one is 727 pixels and the increment is 55 pixels. Each particular line length has been shown eight times in red or green color. The short lines were predominantly green while the long ones were predominantly red. The color distribution within the set of all 112 lines (14 lengths x 8 times) is presented in Table 1. The frequency of the stimuli was calculated in order to receive a positively skewed distribution for the green color and a negatively skewed one for the red lines.

Table 1. Frequency of the presented stimulus lines (where 1 represents stimulus length 12 pixels, 2-67 pixels and so on).

lengths	***number of the green lines***	***number of the red lines***
1 & 2	7	1
3 & 4	6	2
5 & 6	5	3
7 & 8	4	4
9 & 10	3	5
11 & 12	2	6
13 & 14	1	7

Procedure
The participants were tested individually in front of a computer screen where all 112 stimuli were shown sequentially and in random order. They were instructed to judge the length of each line presented on the screen on a seven point scale: 1-“it is not long at all”, …, 7-“it is very long”. No feedback was provided to the participants and no time restrictions have been imposed on them. The whole experiment typically lasted about 15 minutes.

Participants
The participants were 18 undergraduate students (9 men and 9 women none of whom was color-blind) from the introductory classes in psychology at New Bulgarian University who participated in order to satisfy a course requirement.

Results and Discussion

We had 14x2=28 data points for each participant. The results averaged over subjects are shown in Figure 9. Each bar stands for the mean rating that a line of the corresponding size and color has received during the experiment. The repeated measurements analysis showed that the difference (0.046) between the mean judgment of the green lines (4.239) and the mean judgment of the red lines (4.193) is significant ($F(1, 17)=5.966$, $p=0.026$).

Surprisingly enough we obtained a difference (0.046) that is almost the same as the difference we obtained in the simulation (0.053). No tuning of the model was possible since we did not have the experimental data in advance.

Thus the prediction of the JUDGEMAP model has been experimentally confirmed.

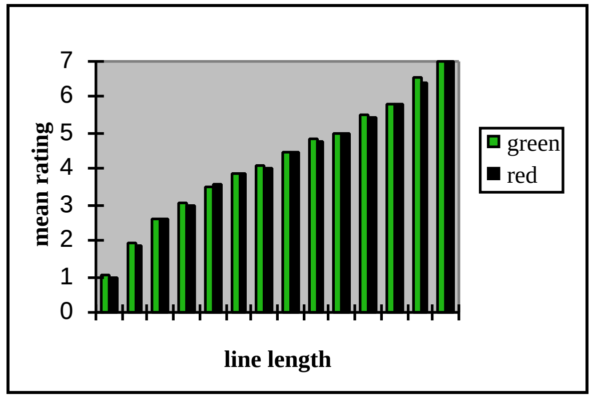

Figure 9. The mean rating of each line with a certain length (1-14) and color (green and red) obtained from all subjects.

Conclusions

The JUDGEMAP model of human judgment has been presented. This model is based on a general cognitive architecture (DUAL) and is thus integrated with the memory and analogy-making model AMBR. Moreover, this model inherits the underlying assumptions of DUAL and AMBR: human cognition is context-sensitive (Kokinov, 1994c), judgment included; human memory is constructive (Kokinov & Hirst, 2003), analogy-making is at the core of human cognition (Gentner, Holyoak & Kokinov, 2001) and its mapping mechanisms may be used in judgment.

The JUDGEMAP model is similar to the Norm theory and the ANCHOR model with respect to the constructive approach to the formation of the comparison set. However, unlike all the models described in the first section judgment in JUDGEMAP is not based on comparison of the target with some aspect of the comparison set, but rather the target stimulus is included in the comparison set and it receives a rating along with all other members of this set. This rating process is based on establishing a mapping between the comparison set and the set of scale elements which mapping preserves the order relations.

Unlike all other models JUDGEMAP does not ignore the irrelevant features of the to be judged targets, moreover these irrelevant features play a role in the construction of the comparison set (retrieving similar objects according to these irrelevant dimensions). The model makes a strange prediction that the color of the target line may play a role in the rating of its length and thus predicts a shift of the mean rating (although a small one) with the change of color. This prediction has been tested in a psychological experiment and has been confirmed.

The size of this color effect is very small, but the stimuli have been very simple and the features unremarkable. It is difficult to imagine that the green color reminds us of a particular green line. That is why we plan to repeat the experiment with more complex stimuli (human figures and clothes) and more memorable features (human faces). It is possible the size of the effect in this case to become larger.

Acknowledgments

We would like to thank the AMBR research team for their continuous support and stimulating environment as well as Stefan Mateeff and Maurice Grinberg for the valuable discussions.

References

Coombs, C. (1964). A Theory of Data. NY: Wiley.

Gentner, D., Holyoak, K. & Kokinov, B. (2001). The Analogical Mind. Cambridge, MA: MIT Press.

Helson, H. (1964) Adaptation-Level Theory: An Experimental and Systematic Approach to Behavior. NY: Harper and Row.

Kahneman, D., Miller, D. (1986). Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review*, vol. 93 (2), pp 136-153.

Kokinov, B. (1988). Associative memory-based reasoning: How to represent and retrieve cases. In T. O'Shea and V. Sgurev (Eds.), *Artificial intelligence III: Methodology, systems, applications*. Amsterdam: Elsevier.

Kokinov, B. (1994a). A hybrid model of reasoning by analogy. In K. Holyoak & J. Barnden (Eds.), *Advances in connectionist and neural computation theory: Vol. 2. Analogical connections* (pp. 247-318). Norwood, NJ: Ablex

Kokinov, B. (1994b). The DUAL cognitive architecture: A hybrid multi-agent approach. *Proceedings of the Eleventh European Conference of Artificial Intelligence*. London: John Wiley & Sons, Ltd.

Kokinov, B. (1994c). The context-sensitive cognitive architecture DUAL. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kokinov, B. (1998). Analogy is like cognition: Dynamic, emergent, and context-sensitive. In K. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research.* Sofia, Bulgaria: NBU Press.

Kokinov, B. & Hirst, W. (2003). Constructive Memory. Sofia: NBU Press.

Kokinov, B. & Petrov, A. (2001). Integration of Memory and Reasoning in Analogy-Making: The AMBR Model. In: Gentner, D., Holyoak, K., Kokinov, B. (eds.) *The Analogical Mind: Perspectives from Cognitive Science*, Cambridge, MA: MIT Press.

Luce, D. (1959). On the Possible Psychophysical Laws. *Psychological Review*, 66(2), 81-95.

Parducci, A. (1965). Category Judgment: A Range-Frequency Model. *Psychological Review*, vol. 72 (6), pp 407-418.

Parducci, A. (1974) Contextual Effects: A Range-Frequency Analysis. In: Carterette, E., Friedman, M. (eds) *Handbook of Perception.* vol. II. Psychophysical Judgment and Measurement. NY:Academic Press, pp 127-141.

Petrov, A. & Anderson, J. (2000). ANCHOR: A Memory-Based Model of Category Rating. In: L. Gleitman & A. Joshi (eds.) Proceedings of the 22nd Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Petrov, A. & Anderson, J. (submitted). The Dynamics of Scaling: A Memory-Based Anchor Model of Category Rating and Absolute Identification.

Sarris, V., Parducci, A. (1978) Multiple Anchoring of Category Rating Scales. *Perception and Psychophysics*, vol. 27 (1), pp 35-39.

Thurstone, L. (1927). A Law of Comparative Judgments. *Psychological Review*, 34, 273-286.

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. (1984). A Theory of Criterion Setting with an Application to Sequential Dependences. Psychological Review, 91(1), 68-111.

Wedell, D., Pettibone, J. (1999) *Preference and the Contextual Basis of Ideals in Judgment and Choice.* Journal of Experimental Psychology: General, vol. 128, pp 346-361.