# Understanding Similarity in Choice Behavior: A Connectionist Model

**Frank Y. Guo (fyguo@ucla.edu)**
UCLA, Department of Psychology, 405 Hilgard Ave.
Los Angeles, CA 90095-1563, USA

**Keith J. Holyoak (holyoak@psych.ucla.edu)**
UCLA, Department of Psychology, 405 Hilgard Ave.
Los Angeles, CA 90095-1563, USA

**Abstract**

Classical choice theories assume choice behavior is based on value maximization computed over the entire choice set. However, empirical evidence has revealed violations of axioms of rational choice that cannot be explained by value maximization. We argue that choice behavior can be reconceptualized as value maximization constrained by categorization processes, and describe a neural network model developed to account for key empirical findings. The model simulates two important phenomena that have been construed as irrational choice behavior, namely, the similarity effect and the attraction effect. We argue that there are important commonalities among choice behavior, categorization and perception.

## Introduction

Many axiomatic theories of choice behavior are based on the assumption that decision making is based on a process of value maximization performed over all attributes (c. f., Tversky & Simonson, 1993). However, empirical evidence has demonstrated that axioms of rational decision making are often violated in choice behavior, and value maximization alone is unable to explain these violations. Recently, an alternative perspective that is concerned with the relations between similarity processes and decision processes has been proposed to conceptualize choice behavior and to understand violations of rational decision making (Medin, Goldstone, & Markman, 1995). That view has been embodied in a comprehensive computational model of choice behavior (Roe, Busemeyer, & Townsend, 2001).

In the spirit of this alternative perspective, we have developed a connectionist model to account for two key violations of rational choice, namely, the similarity effect and the attraction effect. Both of these phenomena involve adding a third alternative (decoy) to a choice set of two options, thereby leading to inconsistency of choice. If the decoy is similar and competitive (two alternatives are competitive when their additive utilities are almost identical to each other) to one of the original options, then the addition of the decoy decreases the choice probability of that option. This phenomenon is called the similarity effect (Tversky, 1972). If the decoy is similar to and dominated by one of the two original alternatives but not the other, then the addition of the decoy increases the choice probability of the dominant option more than the other alternative. This phenomenon is referred to as the attraction effect (Huber, Payne, & Puto, 1982). Both phenomena can potentially lead to violations of rational choice. Few theories were able to provide an integrated explanation of both phenomena prior to the model proposed by Roe et al. (2001), which is a neural network instantiation of the decision field theory (Busemeyer & Townsend, 1993). That model explains the two effects (in addition to several other important choice phenomena) by taking into consideration similarity relations among options and the dynamic nature of decision processes. The model described here is similar to that of Roe et al. in that it also takes into account similarity among alternatives; however, the manner in which similarity is represented and processed differs between the two models. We will briefly discuss the relationship between the two models after we present our proposal.

Neural network models have been one of the major modeling tools in cognitive science (Rumelhart, McClelland, & PDP Research Group, 1986). However, such models have had only limited applications to decision behavior (Holyoak & Simon, 1999; Roe et al., 2001; Thagard & Millgram, 1995). The model we describe here, like that of Roe et al. (2001), uses a neural network approach to provide an account of the similarity and attraction effects.

## Operation of the Model

### Decision Scenario and Model Architecture

The decision scenario used here is adapted from that used by Roe et al. (2001). The decision maker has to choose one car from a set of two or three alternatives by evaluating their ratings on two attributes: gas mileage and performance (see Figure 2). A simple neural network is constructed for this scenario. Figure 2 shows the architecture of the model, adapted from ECHO (Thagard, 1989), a neural network model of how people achieve coherence in making explanations. Two nodes represent the attributes, gas mileage and performance, and three others represent the three alternatives. One special node, labeled as External Driver in Figure 2, represents the motivational and attentional sources that drive the decision process. The lines between nodes represent node connections. Each attribute or alternative is thus represented by one node in the network, with relations among attributes and alternatives represented by connection weights.

Bidirectional excitatory links (represented by dark arrowheads in Figure 2) connect attribute nodes to their respective alternatives. The alternative nodes send out inhibitory influences (represented by empty arrowheads in Figure 2) to one another. Node activation ranges from 0.0 to 1.0. The special node, which drives the decision-making process, always feeds excitatory influence to the attribute nodes, thereby initiating and maintaining activation throughout the entire network. The special node has a constant activation of 1.0, and the weight of its connections to the attribute nodes is 0.05 (there are no reciprocal connections to the special node from the attribute nodes, as the former is intended to be the source of activation). Because the three alternative nodes compete via inhibitory connections with one another, one winning node generally achieves a much higher activation than the rest.

## Setting Connection Weights and Initial Activations

Initially, the connection weight between an attribute and an alternative node (called *attribute-alternative weight* from now on) is set to the rating of the alternative on the corresponding attribute. For example, in Figure 2, the option Target is rated 8 and 2 on performance and gas mileage, respectively, so its initial weights are set to 8.0 and 2.0 for the performance-target and gas-mileage-target connections, respectively.

Next, each initial weight is normalized:

$$w_{ij} = \eta + \frac{(w_{ij} - \min(w)).(\kappa - \eta)}{\max(w) - \min(w)}. \qquad (1)$$

Here, $w_{ij}$ is the weight of the connection to node *i* from *j*. Weights are normalized according to their range; $\kappa$ and $\eta$ are maximum (set to 0.8) and minimum (set to 0.2) values for that range, respectively. Accordingly, the normalized weight

Figure 1. A summary of the phenomena simulated. The letters S and A stand for where the decoy is positioned: Decoy S yields the similarity effect; decoy A yields the attraction effect. The numbers in parentheses are the attribute ratings of the nearby alternative: The first number is the rating of that alternative on gas mileage and the second number is its rating on performance.

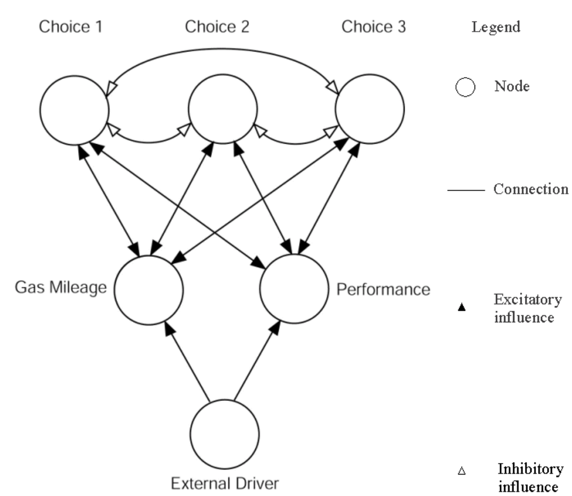


Figure 2: The architecture of the model. Choice 1, Choice 2 and Choice 3 are the alternatives, and Gas Mileage and Performance are the attributes. External Driver represents the motivational and attentional sources that drive the decision process.

should always be within the range of 0.2 to 0.8. The choice of this range is arbitrary, but it reflects the assumption that the perception of an attribute value should never actually reach 0, which can be viewed as reflecting no value at all, nor should it reach 1, which can be viewed as reflecting sublime satisfaction. The range of actual attribute value is computed by max($w$) - min($w$), where max($w$) and min($w$) are the largest and smallest attribute values obtained for all attributes.

The attribute-alternative weights as defined in Equation 1 are linearly related to the actual attribute ratings. In choice behavior, we are concerned with a *subjective* measure of utility in which the impact of a given increase in rating declines with the absolute magnitude of the rating. Accordingly, attribute-alternative weights are transformed:

$$w_{ij} = \frac{w_{ij} l}{w_{ij} + \lambda}. \qquad (2)$$

Here, both $l$ and $\lambda$ are constants. After exploration of the parameter space, $l$ was set to 1.4 and $\lambda$ was set to 0.5 to achieve good simulation results. Equation 2 describes a basic psychophysical function in which sensitivity to an increase of stimulus strength declines as the stimulus strength increases. Finally, weights undergo a linear transformation specified by

$$w_{ij} = w_{ij}\tau / 10.0. \qquad (3)$$

Here, $\tau$ (set to 4.0) is a parameter intended to amplify the attribute-alternative weights so that the same difference between attribute values now has a larger impact on node activations (see Equations 4 and 5). Finally, these weights are divided by 10.0 so that they are kept reasonably small in relation to node activations. Although the model has several parameters, and specific values for them were selected after extensive search of parameter space, the choices of parameter values do not affect the underlying conceptual framework of the model. Moreover, it is very likely that other sets of parameter values exist that would allow the model to exhibit desired behavior.

The inhibitory connections among the alternative nodes are all set to -0.60. The initial activations are set to 1.0 for the special node and 0.5 for all other nodes (0.5 is the middle point of the activation range, 0.0 - 1.0). To increase psychological realism, some randomness is introduced: The initial activation of an alternative node is a random number within the range of 0.5 $\pm$ 0.01. The generation of random numbers conforms to a uniform distribution. There is no randomness for the activations of the special node and the attribute nodes.

## Running the Model

The model runs in an iterative fashion. In each iteration the activation of a node is updated by a commonly-used activation function,

$$a_i(t+1) = \qquad (4)$$
$$\begin{cases} input_i(MAX - a_i(t))\gamma + a_i(t)(1-\theta) & \text{if } input_i > 0 \\ input_i(a_i(t) - MIN)\gamma + a_i(t)(1-\theta) & otherwise \end{cases}$$

$a_i(t+1)$ is the activation of node $i$ at iteration $t$ + 1; it is a function of $a_i(t)$, the activation of the same node at the previous iteration. *MAX* and *MIN* are the upper (1.0) and lower (0.0) limits of node activation. $\theta$ (set to 0.015) is a decay parameter specifying how much the activation decays in each iteration, and $\gamma$ (set to 0.12) is a growth rate specifying the increment of activation as a function of the input. The parameter $input_i$ is the total influence received by node $i$ from other nodes connected to it, specified by

$$input_i(t) = \sum_j w_{ij} a_j(t). \qquad (5)$$

The model runs iteratively according to Equations 4 and 5 until the activation of each node no longer changes from the previous iteration by more than a settling criterion (set to 0.001 here). According to Equation 4, a major determinant of node activation is the total input a node receives from other nodes; and according to Equation 5, this input depends on the attribute-alternative weights. It follows that an alternative with a high additive attribute rating tends to have a higher node activation than those with low additive attribute ratings; this is an instantiation of the value maximization principle, which implies that the winning choice should have the highest additive utility summed across all attributes.

The choice probability of an alternative depends on the activation of the corresponding node. Luce's (1959) choice model is used to convert the activation into choice probability for alternative $i$:

$$probability(i) = \frac{activation(i)}{\sum_j activation(j)} \qquad (6)$$

# Simulations and Results

The two phenomena simulated are schematized in Figure 1. For each phenomenon, 100 simulations were run and the results were averaged for each attribute and alternative. The averaged results are presented both as node activations, which are the final activation values of the nodes (see Table 1), and choice probabilities, which are converted from activations using Equation 6 (see Table 2).

Figure 3: Decision process of binary choice. The activation of alternative nodes is plotted as a function of number of iterations.

## Binary Choice

The original choice set contains two alternatives, one of which is arbitrarily selected as the target, and the other the competitor (see Figure 1). Both cars receive ratings on a 10-point (1 - 10) scale for gas mileage and performance. To simplify the choice scenario, the two options are made equal in terms of additive attribute rating: The competitor is rated 8 on gas mileage and 2 on performance, whereas the target is rated 2 on gas mileage and 8 on performance. It is a trivial prediction that (assuming the two attributes are equally important) the two alternatives should be equally likely to be chosen. The model makes this prediction: when these two alternatives are equally attractive, both have a 50% chance of being chosen (see Table 2).

## Similarity Effect

If the decoy is similar and competitive compared to one of the two original choices, the target, the introduction of the decoy reduces the probability of the target being chosen relative to that of the other choice in the original set, the competitor. This similarity effect (Tversky, 1972) can lead to a violation of an axiom of rational choice, independence of irrelevant alternatives, which implies that adding an alternative to a choice set will not alter the rank order of the original options. To produce a similarity effect, the decoy should be roughly as good as the target in terms of additive attribute rating. In the simulation, the decoy is chosen to have attribute values of 2.5 and 6.5 for gas mileage and performance, respectively (see Figure 1).

To model the similarity effect, we first run the model on a choice set that includes only the target and the decoy. After the network settles for that comparison, we run it on the entire set of three alternatives. The psychological rationale is that because the target and the decoy are similar to each other, they are grouped together in a manner similar to a perceptual grouping (e.g., in visual perception, when two shapes are close to each other, they are perceived as belonging to the same cluster). Our assumption is that the two similar alternatives are perceived as belonging to the same category, and therefore are compared to each other before all three alternatives are compared.

The simulation was thus divided into two stages: a binary comparison in which only the target and the decoy were compared, and a trinary comparison in which all three alternatives were compared. The activations are carried over from the first to the second stage; accordingly, any activation differences from the first stage will have an effect on the second stage. At the end of the binary-comparison stage, the target has an activation lower than 0.5, the baseline activation, due to its competition with the decoy. This low activation is carried over to the trinary-comparison stage, where the competitor joins the comparison with the default initial activation of 0.5. Thus in the trinary-comparison stage the target starts with a lower activation as compared to the competitor; as a result, the target attains a lower activation and choice probability as compared to the

Table 1: Simulation results as node activations.

| Choice scenarios | average node activations | | | | |
|---|---|---|---|---|---|
| | gas mileage | performance | competitor | target | decoy |
| Binary choice | 0.647 | 0.647 | 0.398 | 0.398 | ----- |
| Similarity effect | 0.695 | 0.729 | 0.424 | 0.343 | 0.317 |
| Attraction effect | 0.708 | 0.741 | 0.465 | 0.627 | 0.019 |

*Note.* Each node activation displayed here is the average of activations for the corresponding node calculated over 100 simulation runs.

competitor at the end of simulation. The dynamic process of the two-stage comparison is shown in Figure 4, where the sudden change in activation indicates the transition from the first to the second stage. The final choice probabilities of the target and the competitor are 0.317 and 0.391 respectively (see Table 2), indicating that the competitor ranks higher in terms of preference. Since in the binary choice the choice probabilities of the two alternatives are equal, the altered rank order is a violation of the principle of independence of irrelevant alternatives.

Table 2 Simulation results as choice probabilities.

| Choice scenarios | average choice probabilities | | |
|---|---|---|---|
| | competitor | target | decoy |
| Binary choice | 0.500 | 0.500 | ----- |
| Similarity effect | 0.391 | 0.317 | 0.292 |
| Attraction effect | 0.419 | 0.564 | 0.017 |

*Note.* Each choice probability displayed here is the average of choice probabilities for the corresponding node calculated over 100 simulation runs.

## Attraction Effect

Huber et al. (1982) showed that when the additional alternative (a dominated decoy) is similar to and obviously inferior to one of the alternatives (the target) of the original choice set, the introduction of this decoy will increase the probability of the target being chosen more than that of the competitor. This effect can potentially increase the probability that the target is chosen, thereby leading to violation of an axiom of rational choice, the regularity principle, which states that adding additional alternatives into the choice set would not increase the choice probabilities of options in the original choice set (cf. Huber et al., 1982). The violation of the regularity principle is a stronger form of preference reversal than the violation of independence of irrelevant alternatives.

The same two-stage comparison is employed to model the attraction effect, because the target and the decoy are similar to each other and therefore form a natural grouping. At the end of the binary comparison, the target has an activation higher than 0.5, the baseline activation, due to its superiority as compared to the decoy. This advantage in activation is carried over to the trinary comparison, and as a result the target has a relatively high activation and choice probability at the end of the simulation run. The dynamic process of the two-stage comparison is shown in Figure 5, where the sudden change in activation indicates the transition between the two stages of comparison. The final choice probability of the target is 0.564 (see Table 2). In the original binary choice set, the target has a choice probability of 0.5 (see Table 2); thus adding the decoy leads to a violation of regularity principle.

Figure 4: Decision process of similarity effect. Axes are the same as Figure 3. The vertical dashed line indicates the transition from binary comparison to trinary comparison.

Figure 5: Decision process of attraction effect. Axes are the same as Figure 3. The vertical dashed line indicates the transition from binary comparison to trinary comparison.

In simulating both effects, the model still computes a form of value maximization; however, the computation is carried out in a local instead of global manner during the first stage of comparison, due to the categorization process in which two similar alternatives are grouped and processed together independently of the third alternative.

## Conclusions

The connectionist model presented here explains two perplexing empirical findings in choice behavior using a straightforward neural network algorithm and simple psychological principles. It has been argued that the principle of value maximization underlying rational choice is in conflict with some apparently irrational choice behaviors (Simonson & Tversky, 1992). However, the present model shows that choice behavior can be viewed as value maximization constrained by categorization processes.

Roe et al. (2001) also used similarity relations to account for the similarity and attraction effects. In their neural network model, lateral inhibition among alternatives is set in such a way that the more similar two options are, the stronger is the lateral inhibition between them. This differential inhibition provides a foundation for modeling similarity-related findings. In contrast, in the present model similarity is assumed to lead to a grouping effect, which in turn leads to the two-stage comparison process. Thus while both models emphasize the role of similarity in choice behavior, Roe et al.'s algorithm models the impact of similarity by variations in a continuous parameter for inhibition; whereas the present algorithm hold inhibition constant and instead assumes that similarity alters the grouping of options, leading to a multi-stage comparison process. Further empirical investigations will be required to distinguish between these two possible mechanisms by which similarity may modulate choice behavior.

The present model has several limitations that will need to be addressed in future work. For example, the choice scenario is constructed in a highly schematic way, and more complex and realistic choice scenarios need to be used in future studies. Also, the way the connection weights are set by explicit equations is rather artificial; future efforts need to address how the weights may be acquired using a connectionist learning mechanism. Perhaps most importantly, the critical assumption that similar choices are grouped together and therefore processed together in choice behavior requires further empirical investigation.

The present model may have implications for applied work. Expert systems based on the current model can be developed to analyze and predict choice behavior. In contrast to more traditional axiom-based systems, such systems may make it possible to analyze apparently irrational choice and decision processes, thereby leading to more accurate predictions of human decisions.

## Acknowledgments

This research was supported by NSF Grant SES-0080375. We are grateful for comments from Jerry R. Busemeyer, Patricia W. Cheng, Aimee Drolet and Shi Zhang.

## References

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*, 432-459.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General, 128*, 3-31.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9*, 90-98.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

Medin, D. L., Goldstone, R. L., & Markman, A. B. (1995). Comparison and choice: Relations between similarity processes and decision processes. *Psychonomic Bulletin & Review, 2*, 1-19.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review, 108*, 370-392.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel distributed processing : explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research, 29*, 281-295.

Thagard, P. (1989). Explanatory coherence. *Behavioral & Brain Sciences, 12*, 435-502.

Thagard, P., & Millgram, E. (1995). Inference to the best plan: A coherence theory of decision. In A. Ram & D. B. Leake (Eds.), *Goal-driven learning* (pp. 439-454). Cambridge, MA: MIT Press.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review, 79*, 281-299.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science, 39*, 1179-1189.