Psychological Review
2004, Vol. 111, No. 3, 757–769

0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.3.757

Loss Aversion and Inhibition in Dynamical Models of Multialternative Choice

Marius Usher
Birkbeck College, University of London

James L. McClelland
Carnegie Mellon University and the Center for the Neural Basis of Cognition

The roles of loss aversion and inhibition among alternatives are examined in models of the similarity, compromise, and attraction effects that arise in choices among 3 alternatives differing on 2 attributes. R. M. Roe, J. R. Busemeyer, and J. T. Townsend (2001) have proposed a linear model in which effects previously attributed to loss aversion (A. Tversky & D. Kahneman, 1991) arise from attention switching between attributes and similarity-dependent inhibitory interactions among alternatives. However, there are several reasons to maintain loss aversion in a theory of choice. In view of this, an alternative theory is proposed, integrating loss aversion and attention switching into a nonlinear model (M. Usher & J. L. McClelland, 2001) that relies on inhibition independent of similarity among alternatives. The model accounts for the 3 effects and makes testable predictions contrasting with those of the Roe et al. (2001) model.

Several interesting empirical discoveries have emerged from studies of how people choose between several objects that differ on two or more attributes. For example, someone might be given a choice among three automobiles, varying in performance quality and driving economy (Roe, Busemeyer, & Townsend, 2001). Experimental investigations of human decision making in such multialternative, multiattribute situations have revealed a series of effects that raise challenges for traditional theories of rational choice. These traditional theories are based on the normative principle of *independence of irrelevant alternatives* (Debreu, 1960), which states that the relative preference of any two alternatives is independent of all other alternatives. Despite the intuitive appeal of this principle, a number of violations have been discovered. Below, we describe three central effects that illustrate these violations, called *similarity*, *attraction*, and *compromise* effects (Roe et al., 2001). The situations all arise in choice among three alternatives defined by two attributes, as in the car example above. A graphical representation of the positions of the various alternatives within the two dimensional space defined by the position of each alternative with respect to its value or attractiveness on each of the attributes is provided in Figure 1.

Most cases we consider involve the alternatives *A* and *B* in Figure 1, which were preselected to be of equal binary preference, such that $P(A|A, B) = P(B|A, B)$: That is, the probability of choosing *A* equals the probability of choosing *B* when these are the only available choices. Each case also involves a single additional alternative. All points on the negative diagonal in the figure (the indifference line) are of equal binary preference; alternatives on this line, including S_e, S_c, and *C*, are *indifferent* alternatives with respect to *A* and *B*. All those below the line (such as *I*) are inferior alternatives of lesser binary preference relative to alternatives on the indifference line. An inferior alternative with a value less than another on one dimension and less than or equal to it on the other dimension is *dominated* by that alternative (e.g., *D*, *R*, and *F* are dominated by *A*).

1. *Similarity effect:* This effect arises when an indifferent competitive option *similar* to *A*, such as either S_e or S_c (the subscripts e and c correspond to extreme and compromise, respectively, relative to the options *A* and *B*) in Figure 1, is added to the set of alternatives. The added alternative results in a change of the choice probability between *A* and *B* in favor of the dissimilar alternative ($P(B|A, B, S_x) > P(A|A, B, S_x)$, x ∈ c, e; Tversky, 1972; Sjoberg, 1977). This effect violates the principle that an irrelevant alternative (S_x) should not influence the relative probability of choice between a specific pair of alternatives (*A* and *B*).

2. *Attraction effect:* When the third option (such as *D*, *R*, or *F*) is dominated by the similar alternative (*A*), its introduction can enhance the likelihood of choosing *A* over *B*, $P(A|A, B, D) > P(B|A, B, D)$ (Huber, Payne, & Puto, 1982; Simonson, 1989). In some cases, one even finds that adding the choice option to the response set increases the likelihood of choosing *A* compared with the situation in which *B* was the only alternative (e.g., $P(A|A, B, D) > P(A|A, B)$). Such a finding violates the so-called *regularity* principle, which is implied by a large class of random utility models (see, e.g., Marley, 1989). A further distinction is sometimes made between two types of dominated options, *R* and *F* (called *range* and *frequency* decoys, respectively). The attraction

Marius Usher, Department of Psychology, Birkbeck College, University of London, London; James L. McClelland, Department of Psychology, Carnegie Mellon University, and the Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania.

The work reported here was supported in part by National Institute of Mental Health Grant P50-MH64445. We also thank Jerry Busemeyer for very helpful comments on the article.

Correspondence concerning this article should be addressed to Marius Usher, School of Psychology, Birkbeck College, University of London, Malet Street, WC1E 7HX London, England. E-mail: m.usher@bbk.ac.uk

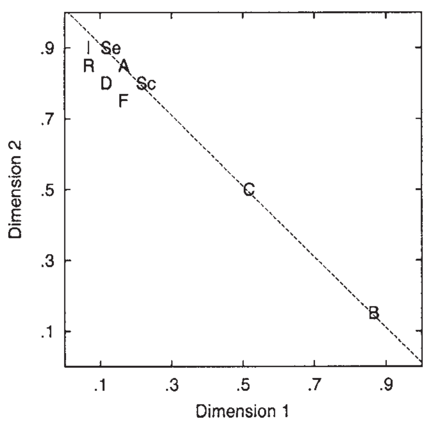

Figure 1. Illustration of the choice alternatives characterized by their values on two dimensions associated with the attributes that distinguish the alternatives. The alternatives might be automobiles, and the attributes might be performance quality and driving economy. Each alternative is defined as a point in this two dimensional space, consisting of a value or attractiveness with respect to each attribute within the normalized interval (0, 1), corresponding to the attractiveness of the alternative with respect to the attribute. The negative diagonal in the figure corresponds to the indifference line along which there is a perfect trade-off of attractiveness, so that when the choice is restricted to two alternatives on this line, each has a choice probability of .5.

effect has been shown to be larger with range decoys (Huber et al., 1982).

3. *Compromise effect:* This effect arises when the third alternative (such as *C* in Figure 1) is an indifferent one positioned halfway between *A* and *B*. In this condition, *C* tends to win over *A* and *B* ($P(C|A, B, C) > P(A|A, B, C) = P(B|A, B, C)$; Simonson, 1989; Tversky & Simonson, 1993).

An intensive effort to provide a causal explanation of choice behavior that meets these challenges was undertaken by Tversky and colleagues (Tversky, 1972; Tversky & Kahneman, 1991; Tversky & Simonson, 1993). Consistent with the theory of choice under uncertainty (Kahneman & Tversky, 1979, 1984), Tversky and colleagues' account of multialternative, multiattribute preference is based on two central tenets: (a) Options are evaluated relative to a reference frame, and (b) the value function for gains and losses (advantages and disadvantages relative to the reference) is S-shaped and asymmetric, with a higher slope in the loss domain that conforms to the principle of *loss aversion* (Tversky & Kahneman, 1991; Tversky & Simonson, 1993). Within this model, called the *context-dependent advantage* model, violations of the independence of irrelevant alternatives principle occur because decision makers treat all the alternatives in the option set as reference points in the evaluation of each option.

Although the context-dependent advantage model accounts for the attraction and compromise effects, Roe et al. (2001) have recently shown that it fails to account for the similarity effect. It is interesting to note that one of the first models of multiattribute, multialternative choice, the *elimination by aspects* (EBA) model of Tversky (1972), explains the similarity effect but not the attraction and compromise effects. In their recent article, Roe et al. have shown that a model for multialternative, multiattribute choice, based on their decision-field theory (DFT; Busemeyer & Townsend, 1993), accounts for all three effects. The model shares a property of the EBA model (Tversky, 1972) in assuming that decision makers attend to only one of the attributes (or aspects) that differentiate the alternatives at a time, sampling the aspects at random.

The model also shares with the context-dependent advantage model the principle that the input into the choice evaluation is driven by *valances*, which are computed contrasts between the options. However, the Roe et al. (2001) model differs from the context-dependent advantage model in that it does not rely on nonlinear and asymmetric value functions for gains and losses. As we discuss in detail below, it relies instead on the use of a distance-dependent lateral inhibition mechanism within a linear dynamical system in which alternatives with negative activations can boost the activation of similar competitors. This is perhaps the most interesting and provocative aspect of their model because it suggests that apparent loss aversion on the part of human subjects may be an emergent property of decision dynamics. For this reason, we refer to the Roe et al. model as the DFT model with distance-dependent inhibition ($\mathrm{DFT}_{\mathrm{DDI}}$).

In the present article, we offer an alternative to the $\mathrm{DFT}_{\mathrm{DDI}}$ model that also accounts simultaneously for the similarity, attraction, and compromise effects. Our approach, which arises within our leaky, competing accumulator (LCA) model of perceptual choice (Usher & McClelland, 2001), shares many of the same principles of the $\mathrm{DFT}_{\mathrm{DDI}}$ model but differs from it in its assumptions about competitive interactions among alternatives and loss aversion. Two considerations motivate our alternative treatment of these issues. First, we argue that the mechanism used by Roe et al. (2001) does not provide a unified account for all the situations in which loss aversion is found in decision making (Tversky & Kahneman, 1991). Second, the attraction effect arises in the $\mathrm{DFT}_{\mathrm{DDI}}$ model because it relies on a completely linear dynamical system in which activations below 0 are allowed to propagate to other units, thereby allowing a dominated alternative to boost the activation of a similar competitor via an "inhibitory" (i.e., negative-valued) connection. As we also review below, the propagation of activations below 0 is not allowed in the LCA model or in many other models of perceptual choice because it has undesired computational consequences. We address these considerations by showing that it is possible to account for the three effects in a version of an LCA model that incorporates stochastic attention switching and relies directly on loss aversion instead of distance-dependent inhibition and propagation of negated inhibition.

The model we propose addresses the point raised by Roe et al. (2001) that no single model incorporating loss aversion has heretofore simultaneously addressed the similarity effect together with the attraction and compromise effects. By retaining the core principles of the LCA model, it also builds a bridge between efforts to understand multialternative, multiattribute choice on the one hand and perceptual identification on the other. As we show below, the model also makes distinct predictions from the $\mathrm{DFT}_{\mathrm{DDI}}$ model, allowing for clear tests to be carried out that will discriminate between these alternative approaches. In what follows, we first review the $\mathrm{DFT}_{\mathrm{DDI}}$ model and then consider how it accounts for the three effects. We go on to present in more detail some reactions to this account, thereby motivating our alternative account based on the use of LCAs, attention switching, and loss aversion. We then proceed to demonstrate how the model accounts for the similarity, attraction, and compromise effects. Finally, we discuss the contrasting predictions made by the different models, indicating how they can be distinguished by future experiments.

The $\mathrm{DFT}_{\mathrm{DDI}}$ Model

The $\mathrm{DFT}_{\mathrm{DDI}}$ model was developed within DFT, which has been successfully applied to a wide range of decision-making situations (Busemeyer & Townsend, 1993; Diederich, 1997, 2003; Diederich & Busemeyer, 1999; Roe et al., 2001; for a survey, see Busemeyer & Diederich, 2002). As with some other DFT models, it relies on leaky integration of information. As Roe et al. (2001) noted, the model can be viewed as a neural network with four layers. The first layer corresponds to the input attribute values, which feed via weights into units at Level 2 that correspond to the two choice alternatives. An attentional mechanism stochastically selects between the attribute units, so that only one attribute (determined randomly) provides input to Level 2 at each time step. Level 3 computes valences for each by subtracting the average Level 2 activation of the two other alternatives from its own Level 2 activation. As the attention switches between the attributes, the valences vacillate from positive to negative values. Level 4 is the choice layer, which performs a leaky integration of the varying input from Level 3; the leaky integration is sometimes called an *Ornstein–Uhlenbeck diffusion process*[1] and is used in several other models including those of Busemeyer and Townsend (1993), Diederich (1997), and Usher and McClelland (2001). Competition between the options occurs at Level 4 and is mediated by bidirectional inhibitory connections with strengths that are assumed to be distance dependent. That is, the strengths of the inhibitory connections among units in the fourth layer decrease as the distance between them, in the attribute space shown in Figure 1, increases. In support of this, it is suggested that an inhibition function that decreases monotonically with distance (dissimilarity) is consistent with the computation of contrast in early sensory processing. Two different rules for selecting a response are considered: According to the external stopping rule, some event independent of the state of the decision-making process determines the time at which the choice is made, and the choice unit that has the highest activation at the given time is selected. According to the internal stopping rule, a unit is selected if it is the first to reach a response criterion. Roe et al. (2001) indicated that both approaches produce similar results. Because the external stopping rule is the basis of their main graphs, we focus on that case in our own simulations.

Central to understanding the way in which the $\mathrm{DFT}_{\mathrm{DDI}}$ model accounts for the similarity and compromise effects is the pattern of correlation among the activations of the response units. In cases in which three alternatives on the indifference line are used, the net input to each alternative always averages out to 0. By itself, this would lead to equal choice among all three alternatives, if it were not for the effects of correlations. The more two units are temporally correlated in their activation dynamics, the more they come to share the same opportunities for choice, leading to a choice advantage for alternatives that are relatively uncorrelated with other alternatives. In conjunction with the random alternation of attention between the attributes, this leads, as in the EBA model, to the similarity effect. Because of the attentional switching, the activations of similar options such as *A* and a similar alternative (S_e or S_c) tend to rise and fall together, as they are activated and deactivated together by the supporting attributes. When the stopping rule indicates it is time to make a choice, it will usually be the case that the correlated alternatives are both more active than the other, uncorrelated one or the correlated alternatives will both be less active. In the first case, they split between them the opportunity to be chosen, whereas in the second case, the uncorrelated alternative will be chosen, thereby producing the similarity effect. The compromise effect is also explained with regard to correlations between activations of choice units. This time, however, this correlation is induced not by the alternating attribute input but by the distance-dependent inhibition. The inhibition is higher between the middle or compromise option *C* and the two more extreme alternatives *A* and *B* than it is between the extreme alternatives *A* and *B* themselves. As a result, the activation of the compromise alternative becomes anticorrelated with the activation of the two extreme alternatives, thereby leading them to become correlated with each other. Once again, choice probability is split between the correlated alternatives, resulting in a compromise advantage.

The $\mathrm{DFT}_{\mathrm{DDI}}$ model gives a very different explanation for the attraction effect. What happens in the model is that the dominated alternative *D* has a negative valance input (averaged over the dimensions). As a result, the activation of the unit for this alternative quickly becomes negative. At this point, this alternative sends positive activation to the other alternatives. This is because a negative activation, when multiplied by a negative connection weight, produces a positive resulting activation. The effect of this negative activation on the similar alternative *A* is greater than the effect on *B* because of the distance dependence. Thus, *A* receives more boosting than *B*, accounting for the attraction effect.

Reactions to the $\mathrm{DFT}_{\mathrm{DDI}}$ Account

The $\mathrm{DFT}_{\mathrm{DDI}}$ model accounts for a set of phenomena that have previously been thought to require the use of a loss-averse value function. As such, it is intriguing and provocative. Loss-averse behavior arises in the model as an emergent consequence of a dynamic decision-making process based on principles Roe et al. (2001) attributed to the underlying neural computation. Because much of our previous work explored emergent consequences of neural networks (in place, e.g., of explicit rules, as in Rumelhart, McClelland, & the PDP Research Group, 1986), we find ourselves quite sympathetic to the approach. Yet, in this particular case, it may be worth considering further the merits of retaining loss aversion as an explicit contributing factor in explaining human choice behavior. We first summarize, following Tversky and Kahneman (1991), the need for loss aversion in multiattribute choice, pointing out that the “emergent” loss aversion produced by distant-dependent activation by negated inhibition does not cover all choice situations exhibiting this property. Although other features might be introduced into the DFT to address the remaining situations, thus far the coverage is incomplete, and the mechanisms that

[1] A diffusion process is a statistical process that tracks the evolution through time of a random variable subjected to perturbation by noise. The standard, or *Wiener*, diffusion process describes a perfect integration of the perturbations, such that they are simply summed over time to determine the state of the variable. In the Ornstein–Uhlenbeck diffusion, the random variable is also affected by a decay term that attracts the variable toward a baseline, with a force proportional to its deviation from it.

have been proposed are not without difficulties. We then examine the use of distance-dependent inhibition by negated activation itself, suggesting reasons why we see this mechanism as problematic. That background leads us to present our own model, similar to the DFT_{DDI} but incorporating loss aversion explicitly rather than relying on distant-dependent activation by negated inhibition. This is consistent with Tversky and Kahneman's (1991) position of loss aversion as a basic principle grounded in the fact that "the asymmetry of pain and pleasure is the ultimate justification of loss aversion in choice" (p. 1057).

Loss Aversion in Multiattribute Choice

A large number of studies indicate that loss aversion is a general principle underlying decision making in a wide range of contexts. We focus here on studies involving choice between options characterized by several attributes, ignoring studies of risk and monetary equivalents (but see, e.g., Kahneman & Tversky, 2000). Consider the following three situations.

The endowment/status-quo effect: Rather fight than switch? Three groups of participants are offered a choice between two objects of roughly equal value (a mug and a chocolate bar), labeled here as *A* and *B*. One group is first offered the *A* object and then the option to exchange it for the *B* object. The second group is offered the *B* object followed by the option to exchange it for *A*. The control group are simply offered a choice between the two objects. The results reported by Knetsch (1989) are striking. Whereas the control participants chose the two objects in roughly equal fractions (56% vs. 44%), 90% of the participants in either of the groups that were first offered one of the objects preferred to keep it rather than exchange it for the other one (see also Samuelson & Zekhauser, 1988). This effect is directly explained by Tversky and Kahneman (1991) by appeal to the loss-aversion function. Because losses are weighted more than gains, participants who evaluate their choices with the already-owned object serving as the reference point decline the exchange. For the control participants, either the values may be computed relative to the neutral reference (Tversky & Kahneman, 1991), or each option can be used as a reference for the other options (Tversky & Simonson, 1993). In either case, there is no reference bias, consistent with the nearly equal choice fractions in this case.

Preference for improvements over trade-offs. Participants are offered the possibility to exchange an option they just received (the reference) for one of two more valuable options of roughly equal value, *A* and *B*. For half of the participants, the reference is similar to but of less value than *A* (e.g., option *F* in Figure 1), and for the other half, the reference is similar to but of less value than *B*. The majority of participants (more than 65%) preferred to exchange the reference object for the similar option that dominates it (Tversky & Kahneman, 1991). This follows the principle that losses are weighted more heavily than gains because the similar choice involves a small improvement, whereas the dissimilar choice involves a trade-off in which a large gain is outweighed by a large loss.

Advantages and disadvantages: Small ones are preferred over large. Participants imagine making a choice between two items (jobs), *A* and *B*, to replace a reference item (a present job that is being terminated). Again, for half of the participants, the reference is similar to *A*, and for the other half, it is similar to *B*. Unlike the case just considered, the reference is not a dominated option. Instead, it is a relatively extreme option on the indifference line with the similar and dissimilar alternatives. For example, the reference similar to *A* would correspond to S_e in Figure 1. Also, note that the reference itself cannot be chosen; it is described as no longer available. Even in this case, most participants (about 70%) chose the option similar to the reference. Tversky and Kahneman (1991) explained this finding, too, by appeal to loss aversion. The similar option involves small gains and losses, whereas the distant one involves large gains and losses. Because losses are weighted more than the corresponding gains, the combination of the larger gain and loss is less preferred than the combination of the smaller gain and loss.

In summary, the three effects described above, along with the compromise and attraction effects, are all directly explained by loss aversion. We have already discussed how the compromise and attraction effects can be addressed in the DFT_{DDI} model; we now consider the three additional effects. The model can address the improvement/trade-off effect, without assuming an explicit asymmetric value function as proposed by Tversky and Kahneman (1991), because it can be viewed simply as an instance of the attraction effect. The dominated reference option boosts the activation of the similar option, via distance-dependent activation by negated inhibition. However, this mechanism cannot account for the endowment/status-quo effect. Distance-dependent activation by negated inhibition cannot be responsible for the tendency for participants to prefer the owned object: Because distance is symmetric, the two objects must inhibit each other to an equal extent. Instead, some other principle must be applied to address loss aversion in this situation. Busemeyer and Townsend (1993) accounted for this effect by proposing that the initial preference state for the owned alternative is greater than that for the other, not owned, alternative. Although such an account deserves consideration, it is worth noting that justification for assuming that there is an initial preference for the owned alternative is not clear, unless one appeals to something very much like loss aversion. Even as an implementation of loss aversion, the proposed mechanism may not be robust enough to account for all instances of the effect. The impact of the initial preference will diminish and eventually vanish as a decision maker deliberates (Busemeyer & Townsend, 1993).[2] Thus, the appeal to an initial preference for the owned object may not account for the effect in situations in which the participants are given time to deliberate before deciding, as in Knetsch and Sinden (1984). In their experiment, participants were first given a gift (a lottery ticket) and were then told that they would have the option to keep the gift (and play in the lottery) or exchange it for a cash amount. A control group was offered the option of buying the lottery ticket for the same cash amount. Participants were then invited to leave the room in which the choice was offered and to stop at a "cash desk" outside to discuss the options before finalizing their decision. This procedure was adopted so that partici-

[2] For example, the effect of the initial preference vanishes over time in the simulation illustrated in Figure 7 of Busemeyer and Townsend (1993), in which the response criterion is assumed to increase with the time deadline. Even if the deadline was not increased, any effects of initial preference will necessarily be located at the fast part of the reaction time distribution.

pants would not influence each other in their choices. However, it also seems to ensure that the participants are given considerable time to deliberate, which should allow the initial preference state to dissipate. Similar considerations arise in an intriguing example of the status-quo effect lasting over several years in choices among alternative car insurance policies in New Jersey and Pennsylvania (Kahneman, Knetsch, & Thaler, 1991). These problems do not arise in the approach we advocate, in which loss aversion operates to provide a stable preference for the previously owned alternative, without erosion of the effect over time.

Finally, let us consider the advantage–disadvantage situation. Here, the reference is not dominated by the other options, as it is in the improvement/trade-off case; so, its valence is not generally negative and does not boost the activation of the similar option by negated inhibition. Moreover, the previously held job is not available as a choice alternative; so, it is not clear that it should actually be treated as an option in the decision process. If it were treated as an option, then for consistency with the DFT_{DDI} model's account of the status-quo effect, this option should receive an initial positive preference, but this would if anything lead to inhibition of the similar option. In the absence of such an initial positive preference, the correlation mechanism in the DFT_{DDI} could operate to influence choices, but in this case, it would if anything reduce, rather than enhance, choices of the similar alternative, as in the similarity effect.[3] Thus, if the DFT_{DDI} theory is to account for the effect, it would appear that yet another principle will have to be added to address the data.[4]

In summary, work within the DFT does not rely on asymmetric value functions and has instead offered a range of different mechanisms to account for several different scenarios that have been used to motivate a direct appeal to the principle of loss aversion. Thus, distance-dependent activation by negated inhibition addresses only some of the effects. We now consider further difficulties with this mechanism.

Treatment of Distance Dependency in DFT_{DDI}

One feature of the distance-dependent inhibition assumption used in Roe et al. (2001) is that no specific function has been introduced that specifies the exact way in which inhibition varies with distance between alternatives. This approach has the advantage of avoiding unnecessary overspecification. At the same time, it introduces considerable model freedom, and there are reasons for uncertainty about the existence of a satisfactory distance-dependent function that obeys the stated principle: "The basic idea is that the strength of the lateral interconnection between a pair of options is a decreasing function of the distance between the two options" (Roe et al., 2001, p. 374). In the main simulations of the similarity, attraction, and compromise effects presented in Figures 4, 7, and 12 in Roe et al., the value of the inhibition between options A and similar options (S_e and D) was set to the same value as that between option A and the compromise option C (.025); lesser inhibition was used only for more distant alternatives (the strength of inhibition between B and any of A, S_e, D, R, and F was set to .001). Thus, the inhibition between alternatives A and C is no less than that between alternative A and more proximal options (S and D), even though alternative C is shown as lying quite a bit farther from A than any of the other mentioned alternatives (see Figure 1). The absolute differences used in Figure 1 are arbitrary, and the compromise and similarity effects are generally obtained in quite distinct experiments. Thus, we cannot be sure that alternative C in compromise studies is in fact less similar to the A and B alternatives used there than the S alternatives in similarity studies are to the A and B alternatives that they use. Even so, it is potentially problematic for the DFT_{DDI} account if it could produce the right magnitudes for the various effects only when the compromise alternative C is effectively as close to both A and B as A is to both of the S alternatives and the various dominated alternatives D, R, and F.

In this context, it is worth noting that there is a tension in the model between the accounts it offers for the similarity and the compromise effects. The similarity effect decreases with the strength of the lateral inhibition between the similar alternatives, whereas the compromise effect increases with strength of lateral inhibition between the compromise and noncompromise alternatives. Thus, it is not clear that it will be possible to account for the actual magnitudes of the compromise and similarity effects if the alternatives in the compromise situation are far enough apart to mandate lower inhibition than that operating between the similar alternatives.

The points we have made in this section do not count as conclusive arguments against the DFT_{DDI} approach. It may be possible to specify a particular, consistently applied function relating distance to strength of inhibition that allows the similarity, compromise, and attraction effects to be captured at the same time. However, the DFT_{DDI} approach combines the use of a distance-dependent inhibition function with the propagation of negative activation to account for the attraction effect. We now turn to additional issues that arise in the reliance on propagation of negative activation.

Propagation of Negative Activations

The DFT_{DDI} approach uses a completely linear dynamical system, in which units can take on positive and negative activations, both of which can propagate via interconnections. This allows the development of closed-form mathematical solutions, which is highly desirable. However, it contrasts with the practice in many connectionist models and other biological neural network models. A basic operating principle of most of these models is that they make use of some form of nonlinearity in the function that deter-

[3] Exactly what effect the inclusion of an alternative that cannot be chosen might have on preferences is not clear. Two possibilities present themselves. (a) When it comes time to choose, only the available alternatives are considered. In this situation, the correlations between the reference object and the similar alternative would not influence choice probabilities. (b) When it comes time to choose, all participating options are considered, but if an unavailable option is chosen, the choice is rejected, so that a second choice must be made. This would tend to produce a disadvantage for the similar option.

[4] Busemeyer and Johnson (in press) have suggested that the DFT_{DDI} can account for this situation by incorporating a third attribute called *availability* to distinguish the reference from the available options. By virtue of its unavailability, the reference would have a lower overall valance than the other alternatives, causing it to become dominated and thereby allowing it to activate the similar alternative by negated inhibition. This assumption is not necessary in our approach, and we think it raises additional difficulties. Space constraints prevent a fuller consideration.

mines the output that the units in the network generate on the basis of their inputs. In part, this is done because the computations that can be performed by a neural network are severely limited unless there is at least one intermediate layer of nonlinear units between inputs and outputs (Rumelhart, Hinton, & McClelland, 1986). For this reason, some type of nonlinearity is generally supposed as a generic aspect of the framework. Many models are further influenced by the fact that neurons communicate by sending action potentials at some rate that is intrinsically bounded below by 0. To achieve this, they may transform the net input they receive from other units (a linear sum) according to a nonlinear function bounded below by 0, or they may propagate the activation value on the basis of the net input only if it is greater than or equal to 0. The latter approach is used widely in the models of Grossberg (1988), in a class of models called *interactive activation* models (McClelland & Elman, 1986; McClelland & Rumelhart, 1981) and in our leaky, competing integrator model of perceptual identification (Usher & McClelland, 2001).

In addition to the neural motivation, the propagation of negative activations can have undesirable consequences in networks with mutual inhibitory connections among units that stand for competing alternatives. The difficulty arises, for example, in the interactive activation model of visual letter and word recognition (McClelland & Rumelhart, 1981), in which there are units for letter features in each of four display locations, units for letters in each of the same display locations, and units for words that span the four letter positions. Mutually inconsistent units within the same level have mutually inhibitory connections (thus, e.g., units for alternative letters in the first position are mutually inhibitory). In this model, allowing the propagation of negative activations was initially tried, but with deleterious consequences. One problem is that the inhibition of a particular unit results in excitation by negated inhibition of all of the unit's competitors. For example, activation of a word results in inhibition of all other words. If initially all word units are at a resting activation of 0, there is no problem, but as soon as one or a few words receive excitation from the letter level, they then inhibit the vast majority of words, and as their activations go below 0, they all suddenly begin to excite each other because of activation by negated inhibition. They then all become activated together, at which point they then inhibit each other, creating an oscillation in which informational differences in the patterns of activation are quickly eliminated. Note that the problem does not arise if activations below 0 do not propagate. Then, units that are excited can inhibit other units and excite units for the letters they contain, but units that are inhibited below 0 do not all send each other activation by negated inhibition.

The propagation of negative activations is not prevented in all neural network models, and we do not wish to suggest that the idea is somehow intrinsically incorrect. However, the same problem with the propagation of negative activations that arises in the interactive activation model would prove problematic for the DFT_{DDI} if it were extended to situations in which there are several similar alternatives all simultaneously competing. In that case, the problem of oscillations seen in the interactive activation model would also arise. These considerations, when taken together with the issues noted above, contribute to our suggestion that it may be worth considering whether the phenomena captured by the DFT_{DDI} model could be captured in a model that does not rely on the distance-dependent propagation of negative activations.

LCAs with Loss-Aversion Value Function

The model we explore here is based on the LCA model previously introduced (Usher & McClelland, 2001) to account for perceptual identification in situations involving two or more choice alternatives. This model shares many assumptions with the DFT_{DDI} model, including the use of leaky, competing units that integrate intrinsically noisy or stochastic information. Our model was intended as a simplification of a more complex neurophysiological process that captures the dynamics of ensembles of neurons thought to collectively represent psychological variables such as the states of activation of the various alternatives in a choice situation (see, e.g., Usher & Niebur, 1996). To address multialternative, multiattribute choice, we adopt a further assumption incorporated in multiattribute versions of the DFT (Busemeyer & Diederich, 2002; Diederich, 1997, 2003; Roe et al., 2001) as well as in a previous neural network model for multiattribute decision making (Usher & Zakay, 1993). This assumption, whose precursor was also used in Tversky's (1972) EBA model, involves a stochastic process of switching attention between the attributes of the choice alternatives.[5] As already shown by Tversky (1972), this assumption can explain the similarity effect. However, our model differs from the DFT_{DDI} in that it follows Tversky and Kahneman (1991) and Tversky and Simonson (1993) in assuming the existence of an asymmetric value function displaying loss aversion. The use of such a value function provides explanations for the compromise and attraction effects, avoiding the need to invoke distance-dependent activation by negated inhibition. Although our model does use lateral inhibition, activations do not propagate below 0; so, there is no boosting by negated inhibition, and the strength of lateral inhibitory interactions is uniform rather than distance dependent.

In Figure 2, the model is illustrated for situations involving a choice between three alternatives (A1, A2, and A3), characterized by their values on two dimensions (labeled Q for performance quality and E for economy). The model operates as follows. At each time iteration, one dimension is chosen randomly to be the focus of attention. The input to each of the LCA units is determined by an input preprocessing stage, on the basis of nonlinear transformations of the differences between all pairs of alternatives on the chosen dimension. Thus, unlike in Tversky and Simonson (1993), the loss-aversion value function is applied separately within each dimension.

Preprocessing Stage

The characteristics of this stage follow Tversky and Simonson (1993) in assuming that in multialternative choice situations, when participants are faced with options that do not provide an explicit reference, they evaluate the options in relation to each other (i.e., each option is being used as a reference point in the evaluation of each other option; the explicit reference situation is addressed below). For example, in the three-alternative case, inputs I_i to the leaking accumulators are governed by these equations:

[5] In EBA, the attributes are sampled without replacement. We adopt the *sampling with replacement* procedure used by Roe et al. (2001; see also, Diederich, 1997, 2003, and Usher & Zakay, 1993).

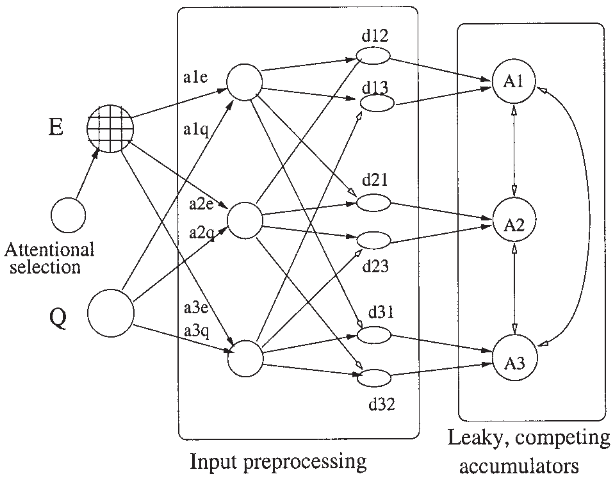

Figure 2. Model scheme for a choice between three options (A1, A2, and A3) characterized by two dimensions, Q and E, respectively. The solid-headed arrows correspond to excitation, and the open-headed arrows correspond to inhibition. At every time step, an attentional system stochastically selects the activated dimension (E in this illustration). The input-item units in the second layer represent each option according to its weights on both of the dimensions and project into difference-input units (d) in the third layer (ellipses; together, the second and third layer have the function of input preprocessing). This layer converts the differences via an asymmetric nonlinear value function before transmitting them to the leaky, competing accumulators, A_i, in the choice layer.

$$I_1 = V(d_{12}) + V(d_{13}) + I_0, \quad (1)$$

$$I_2 = V(d_{21}) + V(d_{23}) + I_0, \quad (2)$$

and

$$I_3 = V(d_{31}) + V(d_{32}) + I_0, \quad (3)$$

where d_{ij} is the differential (advantage or disadvantage) of option i relative to option j, computed on the chosen dimension; V is the nonlinear advantage function; and I_0 is a positive constant that can be seen as promoting the available alternatives into the choice set. Without this constant, the input to each accumulator unit would always be negative because of the loss-averse value function. The nonlinear advantage function (see Figure 3) is chosen, consistent with Tversky and Kahneman (1991) and Tversky and Simonson (1993), to provide diminishing returns for high gains or losses and aversion for losses relative to the corresponding gains:

$$V(x) = z(x),\ x > 0 \quad (4)$$

and

$$V(x) = -\{z(|x|) + [z(|x|)]^2\},\ x < 0. \quad (5)$$

Here $z(x) = \log(1 + x)$ is a function whose slope at the origin is unity and decreases monotonically with gains. Notice that, as proposed by Tversky and Simonson (1993), the value function for losses is a convex function of the corresponding gains.[6] Moreover, as in the graphical displays of the context-advantage function (Tversky & Kahneman, 1991), the value function has a higher slope in the domain of losses than in the domain of gains, providing an advantage for similar options and penalizing dissimilar option pairs (see Figure 3). This is the essential component that enables us to account for the attraction, compromise, and loss-aversion effects.

We assume that when the options for choice are framed relative to a reference (this may include an option the participant is required to give up to choose a new option in its place), the value function is evaluated relative to this reference (Kahneman & Tversky, 1984; Tversky & Kahneman, 1991). For example, in the case in which a previously held (terminating) job must be exchanged for one of two (this generalizes straightforwardly to n-choice) other (available) jobs, we have

$$I_1 = V(d_{1R}) + I_0 \quad (6)$$

and

$$I_2 = V(d_{2R}) + I_0, \quad (7)$$

where d_{1R} and d_{2R} are the advantages and/or disadvantages of the two options relative to the reference, R.

These features make our model similar to the context-dependent advantage model (Tversky & Simonson, 1993). However, instead of obtaining a single preference value for every pair of choices, the model assumes that gains and losses are estimated on each dimension separately, combining the assumptions of the context-dependent advantage model with that of the EBA model.

Leaky-Integration Process

The activation values of the leaky, competing choice units, A_i, integrate the input subject to decay, according to the following iterative procedure (see Usher & McClelland, 2001, for the differential equation version):

$$A_i(t + 1) = \lambda A_i(t) + (1 - \lambda)[I_i(t) - \beta \sum_{j \neq i} A_j(t) + \xi_i \cdot (t)]. \quad (8)$$

Here, λ is the neural decay constant, β is the global inhibition parameter, ξ corresponds to a normally distributed noise term with zero mean and $SD = \sigma$, and I_i corresponds to the inputs as previously indicated. These equations are further supplemented by resetting negative activations to zero. In Usher and McClelland (2001), we showed that this is quantitatively indistinguishable from an alternative formulation in which units are allowed to take on negative activations but these activations are not propagated (no negative firing rates).

Details of the Implementation

Option representations. As in DFT_{DDI}, we consider sets of options characterized by two attributes (or dimensions). This includes sets of three options to examine the similarity, compromise, and attraction effects (Roe et al., 2001), as well as choices between two options relative to a reference to account for loss-aversion situations (Tversky & Kahneman, 1991). The dimensions were scaled within the interval (0, 1). Furthermore, all the options (except for the dominated alternatives, D, R, or F) were chosen on

[6] Consider a gain of size x, a loss of size $-x$, and the values $V(x)$ and $V(-x)$ of the associated gain and loss. The convex function we use has the property that $V(-x) = f[V(x)]$, where $f(x) = -(x + x^2)$.

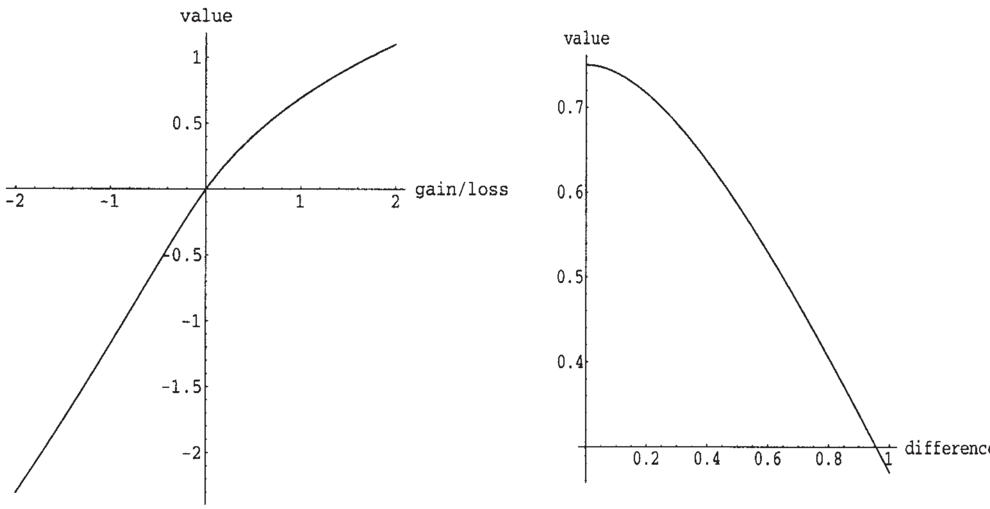

Figure 3. Left: Nonlinear value function, similar to the one used in the reference-dependent model of riskless choice (Tversky & Simonson, 1993). Right: Average input to the model obtained by averaging gains and losses in the value function of A and adding a constant, I_0 = .75.

a diagonal line, corresponding to equal binary preferences, and the weight of (or the time spent on) each of the two attributes was equal. The options A and B were set to (.15, .85) and (.85, .15). The similar options S_e and S_c were set to (.1, .9) and (.2, .8). The compromise option C was set to (.5, .5). The dominated options were set as follows: D was set to (.1, .8), R to (.05, .85), and F to (.15, .75). For the inferior (but not strictly dominated) option I, several cases were considered with attribute values (x, .9), with different values of $x < .1$ (see Figure 1).

Model parameters. The parameter values used in the simulations were $\sigma = .2$ and $I_0 = .75$. The value of the leak parameter was set to $\lambda = .94$, as in Roe et al. (2001). Also, as in that model, we varied the value of the inhibition parameter to explore its effect on the choice patterns.

Simulation procedure. Simulation sets of 1,000 trials were run for the following six scenarios: (a) loss aversion (status quo; A, B with A as reference), (b) loss aversion (high vs. low advantage–disadvantage; A, C with S_e as reference), (c) similarity compromise (A, B, S_c), (d) similarity extreme (A, B, S_e), (e) compromise (A, B, C), and (f) attraction (A, B, D). We also tested the difference between range and frequency decoys by replacing (in f) D with R or F, respectively, and we tested the attraction scenario with a choice between A and B, relative to an explicit reference.

The choice fraction, $P_i(t)$, is computed by running the simulation for 500 iterations and measuring the alternative whose activation is the highest at t. This traces the choice probabilities that would arise from the use of an externally controlled stopping rule at each time point t. Because the actual value of t is not known and may be variable, the model's final choice frequencies for a given alternative are obtained by averaging $P_i(t)$ over the interval $100 < t < 500$, corresponding to the assumption that the choice is precipitated at a random instant between $t = 100$ and $t = 500$; initial transient effects within the first 100 time steps are not included. The basic effects we present are independent of these details, and we provide the full $P_i(t)$ curves when they are informative. Because of the importance of the magnitude of inhibition in the accounts for some of the effects offered by Roe et al. (2001), we explicitly consider to what extent the accounts offered in our model depend on the strength of inhibition in the presentation of the results below.

Results and Discussion

Reference effects: Loss aversion. The choice probability for the A alternative for inhibition in the range (0, .75) is displayed (see Figure 4) for the status-quo situation (option A serving as a reference in a choice between A and B) and for the job situation (with S_e as reference in a choice between A and C). Consistent with the choice data (Knetsch, 1989; Tversky & Kahneman, 1991), we observed a strong bias for participants to choose the option favored by the status quo or to choose the option similar to the reference. (In both situations the choice probability is 50% when the reference is located at mid-distance between the choice options.) In the model, the effect is the outcome of the loss-aversion value function, which penalizes the option with both a large advantage and a

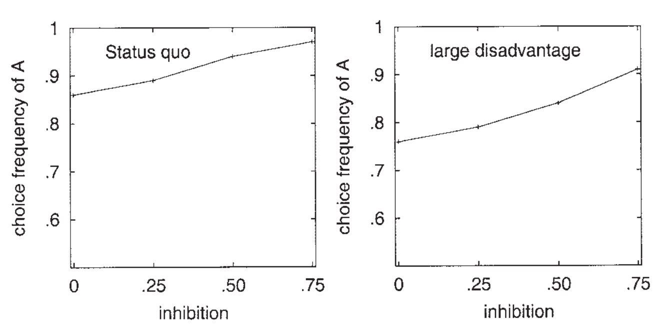

Figure 4. Choice probability for choosing option A (1,000 trials per data point) as a function of the inhibition parameter. Left: Status-quo effect (A is reference in choice between A and B). Right: The job scenario (S_e is the reference in choice between A and C; see the text for further information).

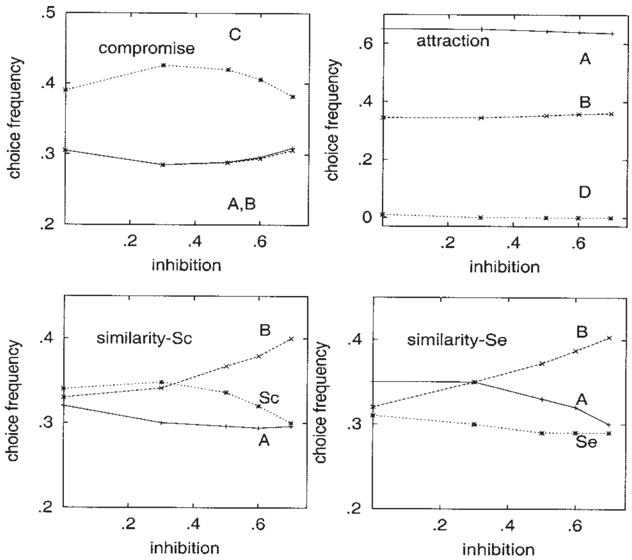

Figure 5. Simulation results (1,000 trials per data point). Top left: Compromise (*A*, *B*, *C*), the compromise option *C* wins. Top right: Attraction (*A*, *B*, *D*), the *A* alternative (similar to *D*) wins. Bottom row: Similarity [left: (A, B, S_c); right: (A, B, S_e)], the dissimilar alternative *B* wins. The *y*-range in the four panels is different to allow all the patterns of qualitative effects to be seen.

large disadvantage relative to the reference. The increase in the level of inhibition has the effect of increasing the competition between the two choice units, amplifying the magnitude of the effect.

Similarity, compromise, and attraction effects. We turn now to the scenarios c–f, involving violations of independence, which were accounted for in Roe et al. (2001). Here, we rely on Equations 1–3, corresponding to a choice among three alternatives without an explicit reference. In Figure 5, the global choice frequencies are shown as a function of the lateral inhibition β for these four scenarios (the simulations were done at $\beta = \{0, .25, .5, .6, .7\}$ and are linearly interpolated at points in between). For a range of inhibition values ($.4 < \beta < .7$), all three effects are obtained.

The explanation of the compromise and attraction effects is a direct outcome of the loss-aversion advantage function, as in the context-advantage model (Tversky & Simonson, 1993). In both conditions, the distant options (*B* in the similarity condition and both *A* and *B* in the compromise condition) are penalized by the asymmetry in the loss-aversion value function. As a result, the option *C* (having fewer distant alternatives) is preferred in the compromise condition, and *A* is preferred over *B* (which has two distant alternatives as opposed to one for *A*) in the attraction condition. For the attraction effect, although the dominated option *D* takes almost no shares, it does attract the choice pattern toward the dominant option *A*. A similar result (not shown) is obtained if the dominated option is used as a reference according to Equations 4–7, capturing the choice preference for improvements relative to trade-offs. The degree of inhibition has a nonmonotonic effect on the magnitude of the compromise effect. At low values, the competition between the options has the effect of increasing the choice sensitivity as the inhibition and the leak balance each other (Usher & McClelland, 2001) and therefore enhance the choice in favor of the compromise, whereas higher levels of inhibition diminish the effect.

The similarity effect is illustrated in the bottom panels of Figure 5. Note that for the similarity effect, there is also a slight sign of a compromise effect, such that *A* gains a little relative to *B* and *S* when it is "inside" *S* (similarity: S_e) and loses a little when it is "outside" *S* (similarity: S_c). The explanation of the similarity effect and the effect of inhibition on it is simple. Here, whereas the loss-aversion function penalizes the dissimilar option, *B*, the correlation between the activations of the similar options *A* and *S* (see Figure 7, top panel, in *The Importance of Leaky Integration* section, below) helps it. As in the EBA model (Tversky, 1972) and in the DFT, the similar options share high activation during the same choice intervals, and thus, they share their choices. If, for example, the loss-aversion effect is to give the *A* option a share of less than 66% for the choice set (*A*, *B*, *D*), when *D* is substituted by *S* and assuming that *A* and *S* are now splitting their shares, a small advantage for the dissimilar option *B* results. This advantage is further amplified by the inhibition. As this increases, the two similar options, *A* and *S*, compete to a higher degree, to the advantage of the dissimilar option *B*. Note that the higher degree of competition between *A* and *S* arises even though our model does not use distance-dependent inhibition. It occurs in this case because the activation of the similar alternatives covaries as attention switches between dimensions.

Additional dynamic effects can be observed for the choice pattern in the compromise and the attraction conditions. These are illustrated in Figure 6, which shows the choice probabilities for the different alternatives, $P_i(t)$, for choices made at times varying from 0 to 190 iterations of the computer simulation. The left panel shows the evolution of the choice preference in the compromise condition. We can observe that it takes about 30 iteration steps for the compromise option to dominate the choice. This is because early on one or the other of the extremes is likely to dominate, but as the activations are integrated, the fluctuations in the extreme options are averaged out, leading to an advantage for the compromise. In the attraction condition (right panel), we see that it takes about 10 iterations for the similar option, *A*, to dominate the dissimilar option, *B*. Early on, *D* shares some of the choices with *A* (as with the similarity effect). As the noise is averaged out with

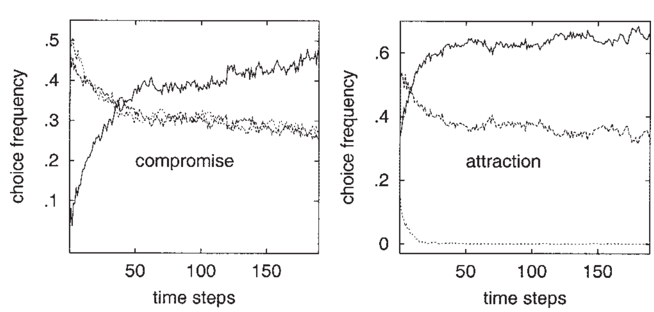

Figure 6. Time-dependent choice preference for the compromise effect (left: *C*, solid line, upper curve; *A* and *B*, dashed–dotted lines) and for the attraction effect (right: *A*, solid line, upper curve; *B*, dashed line, middle curve; *D*, dotted line, lower curve). The inhibition parameter is $\beta = .25$. All other parameters are as given in the text.

Table 1
Effects of Inferior Decoy, I, on the Choice Between the A and B Options

x	*P*(*A*)	*P*(*I*)	*P*(*B*)
.10	.33	.29	.37
.08	.49	.13	.37
.06	.58	.05	.36
.05	.61	.03	.36

integration time, the amount of choices for the dominated option decreases, and the similar options come to dominate the choice. Our model (as well as the DFT) thus predicts the emergence of the compromise effect (after an early stage in which the opposite effect would be obtained) and an enhancement of the attraction effect with time. (In particular, the model predicts that some participants will show a reversal of their choices as the deliberation progresses in the compromise condition.) Experimental results seem to confirm these predictions, indicating that, as decision makers are encouraged to deliberate longer, the magnitude of the attraction and the compromise effects increases (Dhar, Nowlis, & Sherman, 2000; Simonson, 1989).

We considered two additional effects reported in the literature and accounted for by the DFT model. First, we tested the impact of the difference between range and frequency decoys on the magnitude of the attraction effect. Consistent with Huber et al. (1982; and as in the Roe et al., 2001, model), we found a larger attraction effect for the (*A*, *B*, *R*) set ($P(A) = .66$, $P(B) = .34$) than for the (*A*, *B*, *F*) set ($P(A) = .63$, $P(B) = .37$). Second, we examined the impact on the shares of the *A* and *B* options of changing the third option from an equal competitor, S_e, with values (.1, .9) to an inferior but not strictly dominated option, *I*, with values (*x*, .9), for $x < .1$. The results are presented in Table 1.

Consistent with the choice data (Huber et al., 1982), we found that transforming a competitor into an inferior option dramatically alters the shares between the two similar options (*A* and *I*) but not the share of the dissimilar option. The DFT model predicts a modest decline in the shares of *B* (of 5%) for the same change in the shares of *A* (Roe et al., 2001, Table 4).

The Importance of Leaky Integration

In addition to stochastic alternation of attention, the DFT and our current model also share another important assumption. In both

Table 2
Choice Probabilities in the Model With Perfect Integration

	Option in the choice set					
Effect	*A*	*B*	*D*	*C*	S_c	S_e
Attraction	.87	.13	.00			
Compromise	.07	.07		.84		
Similarity[a]	.22	.12			.65	
Similarity[b]	.70	.12				.18

Note. For each effect, there were three options in the choice set.
[a] This row shows probabilities for choice options for the similarity effect in the compromise condition. [b] This row shows probabilities for choice options for the similarity effect in the extreme condition.

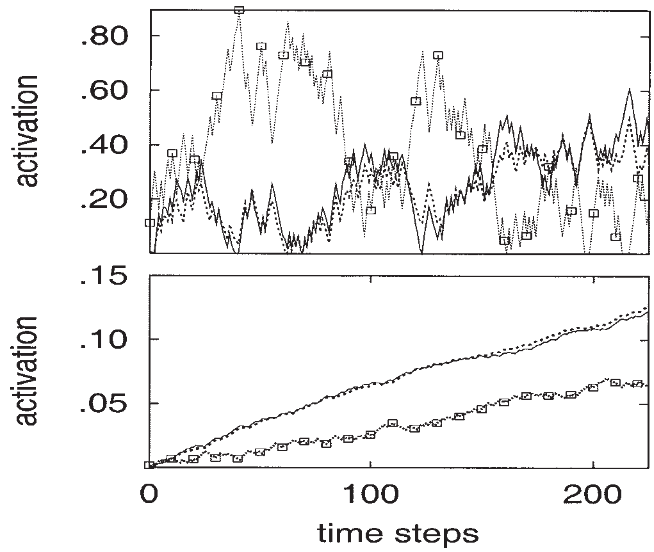

Figure 7. Activation trajectories for one trial in the similarity condition for the leaky integrator with competition (top) and for the perfect integrator (bottom). *A*, solid line; *B*, dotted line with open squares; S_c, dashed line without symbols.

models, the choice is modeled as a sequential sampling process with leakage of information over time. This Ornstein–Uhlenbeck (OU) diffusion process can be distinguished from classical diffusion processes in which the samples of information are integrated without loss (e.g., Ratcliff, 1978). The importance of the OU process has been argued both within the DFT framework (Busemeyer & Townsend, 1993; Diederich, 1997; Townsend & Busemeyer, 1995) as applied to decision making and in our LCA model (Usher & McClelland, 2001) for the domain of perceptual choice. We now examine how the feature of leaky versus perfect integration affects the multiattribute effects described here. To do this, we ran the same simulations, but with the parameter λ set to the value of .999 (corresponding effectively to perfect or lossless integration). The results are given in Table 2.

We observed that although the attraction and the compromise effects occur (at magnitudes that are beyond the range of experimental data), there is a total reversal of the similarity effect (the dissimilar option gets only 12% of the shares).[7] To help understand the reversal of the similarity effect, we show in Figure 7 the activation trajectories for one trial of the simulation each for the case of leaky integration ($\lambda = .94$; top panel) and for the case of perfect integration (bottom panel). We observed that unlike with the leaky integration, in which the dissimilar option (dotted line with open squares) has extended time intervals in which it dominates the option set, with the perfect integration the dissimilar option is dominated throughout the simulation. This is a result of

[7] In a separate set of simulations, we tested whether the nonleaky integrator model can account for the effect when the asymmetry in the value function is diminished. To do this, we parameterized the magnitude of this asymmetry by α in the convex function $f(x) = x + \alpha x^2$. For $\alpha = .29$, we obtain a relatively small compromise effect (of 4%). Even in this situation, however, the perfect integrator model is unable to account for the similarity effect.

the asymmetric value function. The leakage of information is thus essential in accounting for the similarity effect, once a loss-aversion advantage function is assumed. This is because it makes the activation of the choice units dependent on a recency-based temporal window. As the attributes are stochastically sampled, there are time windows in which the dissimilar choice unit receives a stronger input (when the supporting attribute is sampled), and this unit makes a recovery and dominates the option set. With a perfect integration, the advantages and disadvantages that correspond to the sequential sampling of the attributes are integrated, and they average out, leading to the advantage of the S_c option because of the compromise effect.

Evaluations and New Predictions

We have proposed a model that shares a number of important assumptions with the DFT framework. These include leaky integration and lateral inhibition that triggers choice competition. The model also shares with the DFT model the stochastic sampling of attributes (see also Busemeyer & Diederich, 2002; Diederich, 1997; Diederich & Busemeyer, 1999; Usher & Zakay, 1993). The models make different assumptions, however, about the principles of processing relating to inhibition. Roe et al. (2001) used distant-dependent inhibition and relied on activation by negated inhibition, whereas in our model, inhibition is distance independent and there is no propagation through inhibitory connections when activations go below 0. Instead, we rely on the asymmetric value functions, in which losses are a convex function of gains (Tversky & Simonson, 1993), which have played a central role in previous work in decision making. The inclusion of loss aversion in the model enables us to account for a large amount of data, such as the status-quo and other reference effects, which motivated Tversky and colleagues (Tversky & Kahneman, 1991; Tversky & Simonson, 1993) to rely on this mechanism in their decision-making theory.

Like the $\mathrm{DFT_{DDI}}$, our model can account for violations of independence of irrelevant alternatives (i.e., similarity, attraction, and compromise) that have challenged other theories of multiattribute decision making. In addition, this model as well as the $\mathrm{DFT_{DDI}}$ accounts for some more subtle effects, such as an increased attraction effect with range—rather than frequency—decoys and the preserved shares of the competitor *B* when the option similar to *A*, S_e, is transformed into an inferior option (*I*). The models also make the same prediction regarding the increase in the magnitude of compromise effects with reaction time and choice reversals under time pressure.

Despite these similarities, the way in which the DFT (as implemented in Roe et al., 2001) and the present model account for many of the effects is not the same, and this is reflected in a number of differences and diverging predictions. The main difference concerns the way in which the attraction and the compromise effects are explained.

Consider first the attraction effect. Whereas in Roe et al. (2001) the dominating option (*A*) wins because of an additional boost from the similar dominated option, in our model the effect is due to the cost suffered by the dissimilar option (*B*), which is penalized by the asymmetric value function. Our model predicts therefore that the shares of the *B* option are almost preserved when a competitive option such as S_c is transformed into a dominated option such as *D* or *F*; because *B* is almost equidistant relative to S_c, *F*, and *D*, its shares are preserved and the attraction is due only to a redistribution of the shares between the two similar options: $P(B|A, B, S_c) = .38$; $P(B|A, B, F) = .37$; $P(B|A, B, D) = .36$. In $\mathrm{DFT_{DDI}}$, however, transforming the competitor option into a dominated option generates a specific boosting for the activation of the similar option, *A*. This has the effect that the shares of the dissimilar option are diminished to a larger extent (approximately 13% in Roe et al., 2001). Experimental studies by Huber et al. (1982) have reported that the shares of the *B* competitor are not affected by transforming S_e into an inferior option *I*. As we saw, both our and the DFT model were able to account for those results, although the preservation is within 1% in the results we presented above compared with 5% in Roe et al. (2001, Table 2). As the experimental precision may not suffice to test this small difference in the models' predictions, future studies should focus on choices for options sets (*A*, *B*, *S*) and (*A*, *B*, *D*), in which the difference between the predictions is higher as a result of the fact that *D* is strictly dominated by *A* (inferior on both dimensions).

A more fundamental difference between the models is their accounts of the compromise effect. In our model, the extreme options are penalized by the value function to the benefit of the compromise. In the DFT model, conversely, the effect is not driven by differential levels of activation (as the attraction effect) but by correlations due to the local inhibition: The activations of the extremes are correlated in time because they both compete with the compromise. Although both mechanisms account for the compromise effect, the correlation mechanism is weaker relative to the boost of activation. This makes the magnitude of the compromise effect only 4% ($P(C) = .37$, in Roe et al., 2001, relative to a large attraction effect of 19%). In the model we have presented here, the magnitude of the compromise effect is larger (about 10%), which is more consistent with the choice data (Simonson, 1989).

Moreover, because of their different nature, the two types of mechanisms lead to several qualitative differences in their predictions for new choice situations. Consider first the impact of the distance between the options on the magnitude of the compromise effect. Because the anticorrelations in the $\mathrm{DFT_{DDI}}$ are driven by lateral inhibition that decays with distance, the effect should decrease with distance. In our model, conversely, the loss aversion increases with distance (see Figure 3). As a result, we found that the compromise effect shows an increasing relationship. To show this, we symmetrically changed the distance, *d*, of the extremes *A* and *B* on each attribute from the compromise *C* from .35 (in the previous simulations) to .25 and .15. The results are as follows: For $d = .15$, $P(C) = .33$; for $d = .25$, $P(C) = .40$; and for $d = .35$, $P(C) = .42$. (This simulation corresponds to an inhibition of $\beta = .5$. Other β values yield similar results; e.g., $\beta = .6$ gives $P(C) = .35$, for $d = .15$; $P(C) = .41$, for $d = .25$; and $P(C) = .41$, for $d = .35$.)

Second, future experiments could directly test the correlation hypothesis. Consider, for example, a situation in which a participant chooses one of the extremes, say *A*, following a response signal at *t*. According to the correlation hypothesis, the option with the next highest activation at *t* is the other extreme, *B*. To test this, one could carry out an experiment in which on a subset of trials, the choice alternative *A* is declared unavailable at the instant that it is chosen, and the participant must then make a second speeded choice. Assuming that the participant selects the alternative that is next most active at the same instant as the first selection, the DFT model should predict that for speeded choices, $P(B|A) > P(C|A)$, whereas the converse is predicted in our model. Finally, it may also be possible to test the correlation hypothesis without a second response by comparing the reaction time for choices elicited by a response signal for compromise versus extreme options. Under the assumption that the time to resolve the choice is larger for options whose activation values (at the moment when the response signal is received) are similar, one should expect longer reaction times when either of the correlated options, *A* or *B*, is chosen than when alternative *C* is chosen.

Conclusion

We have extended our LCA model of perceptual choice to address preferential choice situations, incorporating attentional switching between attributes (following Roe et al., 2001) and loss aversion. With these extensions, we offer an account for several effects that have been captured within DFT by relying on distant-dependent activation by negated inhibition. We do not wish to suggest in any way that the overall DFT should be rejected. DFT models have been used to account for many important effects in decision making, such as violations of stochastic dominance and effects of time pressure (Diederich, 2003; Diederich & Busemeyer, 1999), and our model, which has been successfully applied to many phenomena in perceptual choice, is in many ways very similar to the DFT. For these reasons, we do not view our model as a competitor to the DFT. Instead, we view it as suggesting some constructive amendments to the DFT framework. We have proposed that the use of distance-dependent activation by negated inhibition should be replaced with an explicit reliance on the principle of loss aversion. Although both of these alternative mechanisms can account for basic aspects of the violation of independence effects, the magnitude of the effects may favor our approach. Moreover, we have demonstrated that all these effects, as well as the status-quo effect and two other effects arising when an explicit reference is provided, can be accounted for with the use of a loss-averse value function. It remains to be seen whether a similarly adequate account can be achieved within the DFT with a specified monotonically decaying distance function. We have also suggested that it may be useful to incorporate a nonlinearity of the kind used in many neural network models to avoid undesirable consequences that can arise from activation by negated inhibition when there are many suppressed alternatives. In summary, our LCA model with loss aversion can be viewed as a version of the DFT, incorporating loss aversion and truncation of activations at 0 instead of distant-dependent inhibition and propagation of negative activations.

Although our proposed model can be seen as falling within the same overall framework as DFT, the processes that enable it to account for choice patterns in decision making are different in important ways from the processes that take place in the model proposed by Roe et al. (2001). In addition to the differences noted above, the two models make several distinct empirical predictions. Future tests of these predictions, whichever way they turn out, will enhance our understanding of the dynamics of decision making and will contribute to the ongoing process of uncovering the principles of human decision making.

References

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences, 43,* 345–370.

Busemeyer, J. R., & Johnson, J. G. (in press). Computational models of decision making. In D. Koehler & N. Harvey (Eds.), *Handbook of judgment and decision making.* Cambridge, MA: Blackwell.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory. *Psychological Review, 100,* 432–459.

Debreu, G. (1960). Review of R. D. Luce: Individual choice behavior. *American Economic Review, 50,* 186–188.

Dhar, R., Nowlis, S. M., & Sherman, S. J. (2000). Trying hard or hardly trying: An analysis of context in choice. *Journal of Consumer Research, 9,* 189–200.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology, 41,* 260–274.

Diederich, A. (2003). Multi-attribute decision field theory account of decision making under time pressure. *Psychological Bulletin & Review, 10,* 157–166.

Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic dominance principle of decision making. *Psychological Science, 10,* 353–359.

Grossberg, S. (1988). *Neural networks and natural intelligence.* Cambridge, MA: MIT Press.

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.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss-aversion and status-quo bias. *Journal of Economic Perspectives, 5,* 193–206.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision making under risk. *Econometrica, XLVII,* 263–291.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39,* 341–350.

Kahneman, D., & Tversky, A. (2000). *Choices, values and frames.* Cambridge, England: Cambridge University Press.

Knetsch, J. L. (1989). The endowment effect and evidence of nonreversible indifference curves. *American Economic Review, 79,* 1277–1284.

Knetsch, J. L., & Sinden, J. A (1984). Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of values. *Quarterly Journal of Economics, 99,* 507–521.

Marley, A. A. J. (1989). Random utility family that includes many "classical" models and has closed form choice probabilities and choice reaction times. *British Journal of Psychology, 42,* 13–36.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of contexts effect in letter perception. *Cognitive Psychology, 23,* 1–44.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

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., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processes. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 45–77). Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.

Samuelson, W., & Zekhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1,* 7–59.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research, 16,* 158–174.

Sjoberg, L. (1977). Choice, frames and similarity. *Scandinavian Journal of Psychology, 18,* 103–115.

Townsend, J. T., & Busemeyer, J. R. (1995). Dynamic representation of decision-making. In R. F. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 101–120). Cambridge, MA: MIT Press.

Tversky, A. (1972). Elimination by aspects. *Psychological Review, 79,* 281–299.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference dependent model. *Quarterly Journal of Economics, 107,* 1039–1061.

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

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592.

Usher, M., & Niebur, E. (1996). Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience, 8,* 311–327.

Usher, M., & Zakay, D. (1993). A neural network model for attribute based decisions processes. *Cognitive Science, 17,* 349–396.

Received August 7, 2002
Revision received November 18, 2003
Accepted November 18, 2003 ■