Mere exposure effects - Merely total activation?

Bruce F. Katz
School of Cognitive and Computing Sciences
University of Sussex
Brighton BN1 9QH UK
brucek@cogs.susx.ac.uk

Abstract

The mere exposure effect, in which subjects prefer items they have previously been exposed to over unexposed items, is explained as the effect of competitive learning in a connectionist network. This type of unsupervised learning will cause the network to respond more strongly to patterns on which it has been trained. If it is assumed that positive affect is proportional to total activation, then the mere exposure effect is a direct consequence of this process. The addition of a habituation rule, with a dishabituating recovery element, can also explain factors which reduce or enhance the effect. These include the effect of exposure count, display presentation sequence, the complexity of the patterns, the effect of a delay after presentation, and finally, the effects of varying exposure duration. In the case of this last factor, in addition to showing that very short exposure durations can enhance the effect, the model reveals why it may be possible to respond positively to a stimulus that one cannot recall perceiving.

Introduction

The repeated presentation of an unreinforced and unclassified stimulus will cause subjects to prefer this stimulus over unexposed stimuli; this is known as the mere exposure effect (Zajonc, 1968). This effect is perhaps the most robust in the literature on aesthetic preference. It has been found with a number of stimulus types including nonsense words, meaningful words, Chinese characters, photographs, music, and people (Harrison, 1977). Bornstein (1989) carried out a meta-analysis on 208 published studies on the mere exposure effect between the years 1968 and 1987. He found a combined significance of $p < .0000001$, and a fail-safe N of 33,047. That is, there would have to be this many unpublished studies with zero effect size to render the combined probability insignificant.

In addition to demonstrating the consistency of the mere exposure effect, Bornstein's analysis revealed a number of factors which serve to enhance or reduce the size of the effect. This paper will treat five of these:

1) Number of exposures

Bornstein found that the exposure effect was reduced in studies with large number of exposures. A number of studies have shown that preference is an inverted U-shaped curve as a function of exposure count. For example, Kail and Freeman (1973) found an increase followed by a decrease in rated attractiveness of ideographs as a function of number of exposures. Brentar, Neuendorf, and Armstrong (1994) have found a similar result in response to songs. This is consistent with the common pattern whereby one initially plays a newly acquired piece of music at high frequency, but one finds one's attraction to it decline after many repeated listenings. These results indicate that in addition to whatever is causing the exposure effect, some sort of habituation eventually sets in causing the preference for the over-exposed stimulus to decline.

2) Homogeneous vs. heterogeneous display

Berlyne (1970) stressed the importance of the presentation sequence in determining the size of the exposure effect. Homogeneous display consists of exposing a subject to a given stimulus a number of times, followed by the the presentation of the next stimulus a number of times, etc. Heterogeneous display is achieved by alternating the stimuli during each exposure. Berlyne found that heterogeneous display created a larger exposure effects with high-frequency stimuli. Bornstein's (1989) meta-analysis revealed that the combined homogeneous experiments yielded no exposure effect, but that the combined heterogeneous experiments showed a highly significant effect ($p < .0000001$).

3) Complexity

Berlyne and his colleagues (1974) have stressed the importance of complexity on aesthetic preference. In particular, they showed that while ratings of interestingness increase with increasing complexity, affective ratings such as liking form an inverted U as a function of this variable. They also demonstrated that complex stimuli exhibit a less steep rise in affect as a function of exposure, and Berlyne (1970) has also demonstrated less steep declines with complex stimuli as a function of exposure. Bornstein (1989) claims that six of nine studies have shown greater exposure effects with complex stimuli than with simple ones, two found no difference, and one study favoured simple stimuli. In summary, there is some support for the claim that complex stimuli produce stronger exposure effects, although may take more presentations to exhibit such effects.

4) Delay after exposure

Studies directly studying the effect of delay after exposure have produced conflicting results (Harrison, 1977; Bornstein, 1989). However, Bornstein's (1989) meta-analysis revealed a significant "sleeper" effect. The exposure effect was greater if the ratings were completed after all the stimuli were presented, rather than immediately after each stimulus presentation. A forced delay after the the presentation of all stimuli also resulted in a more consistent effect than immediate ratings.

5) Exposure time

Bornstein's (1989) meta-analysis also showed that the exposure effect is more consistent when stimuli are briefly presented than when they are presented for long periods of

time. Bornstein and D'Agostino (1992) have tested this hypothesis directly by using stimuli of 5-ms and 500-ms. As expected, the former produced greater exposure effects. In addition, recognition ratings for the briefly exposed stimuli did not differ from chance. Apparently, the mere exposure effect can be achieved without recognition, and may even be enhanced by subliminal presentation (Bornstein, 1989). This is discussed further in the final discussion.

An effect as important and robust as mere exposure has naturally attracted a number of theoretical treatments. Three of those are now briefly discussed:

1) Opponent process models

These models propose two affective systems, positive and negative, acting in opposition (Solomon & Corbit, 1974). The initial response to a novel stimulus is assumed to be negative. With repeated exposure and greater familiarity, however, the negative affective response is weakened, permitting the antagonistic positive affective system to have greater input in determining the overall affective state. Despite some evidence for the sort of rebound effects such a model would predict, two problems remain. First, it requires that one's initial response to a novel stimulus invariably to be negative, which appears *prima facie* to be false. Second, it does little, in itself, to explain the five variables modulating the effect described above.

2) Arousal models

Most closely associated with D.E. Berlyne, arousal models postulate that positive affect is an inverted U-shaped curve as a function of the arousal potential of the stimulus. Berlyne (1971) suggested that a complex stimulus, initially somewhere to right of the inflexion point on this curve, becomes subjectively less complex with repeated exposures. Hence, it becomes more liked as it comes closer to the apex of the inverted U. This would also explain why simple stimuli become less well-liked with repeated exposures. However, the model does less well in predicting inverse relation between exposure duration and the size of the exposure effect, and the role of delay on the effect, and it is not clear how one could operationalize the model to incorporate these auxiliary effects.

3) Two-process models

Two-process models suggest a familiarity effect is counterbalanced by a habituation effect (Bornstein, 1989). Initially, exposure to a stimulus causes it to become less threatening, and therefore preferred to a larger extent. However, eventually boredom will set in, causing the subject to lose interest in the now overly familiar item. Thus, one can explain the eventual downturn in affective response with repeated presentations. The tendency for homogeneous presentation, and long exposures to quash the effect can be explained along similar grounds. Delay should decrease boredom, and therefore will increase the exposure effect. Finally, presumably one becomes less bored with complex stimuli than with simple ones, explaining the effect of this variable.

However, two problems remain. First, familiarity, which forms the basis for the first process in the two-process theory, does not seem to be a requisite of the mere exposure effect in that subliminal stimuli cause an exposure effect, and may be superior to supra-liminal stimuli in doing so (Bornstein, 1989). Second, one would like to know how to operationalize the notions of familiarity and boredom in order to make predictions concerning the interactions between the various modulating factors.. The purpose of this paper is propose a two-process connectionist model that meets these objections.

A Connectionist Model

The model rests on two unsupervised learning rules. Thus, it is consistent with the fact that mere exposure effect occurs in the absence of a teacher. In addition, the associative character of the proposed rules is consistent with the fact that the mere exposure effect is seen as far down on the phylogenetic scale as insects (Bornstein, 1989). Before, presenting the model in detail, however, a means of measuring the affective response of the model must be proposed.

The fundamental assumption of this work is that positive affect is a monotonic function of cortical activity Thus, this measure contrasts with optimal arousal theories (Berlyne, 1970), which propose a downturn in affect with over-arousal. One justification for this measure is that it is consistent with the traditional aesthetic principle of unity in diversity (Martindale, 1984). The more competing representations the network is able to maintain, the higher the overall activity of the network. Conversely, low activity implies either low diversity in the input stimulus or the inability of the network to represent the diverse aspects of a complex stimulus at once. I have shown how this measure is useful in understanding the unification of incongruities in humourous stimuli (Katz, 1993), and how it may be used to measure the worth of simple melodies (Katz, 1994).

The mere exposure effect follows immediately from this premise acting in conjunction with an unsupervised learning regime such as competitive learning (Rumelhart, & Zipser, 1986). Exposure to a stimulus causes the weights to realign such that future presentation of the stimulus will provide more activity to the classification layer. Eventually, this realignment results in super-threshold activity in this layer, and the organism prefers those stimuli that trigger such activity over those that do not. A habituation effect must be postulated in order to provide for the eventual downturn in activity with over-exposure.

Figure 1 shows how these two processes interact in the proposed model. The model consists of two layers, an input layer consisting of a grid of units, and a classification layer, consisting of a set of winner-take-all clusters. Each such cluster consists of excitatory connections from units to themselves, and inhibitory connections to all other units in the cluster. Excitatory connections also form between active units in the input layer and active units in the competitive clusters as the classification process occurs. Inhibitory connections form between mutually active units in the input grid to form a novelty filter (Kohonen, 1987), i.e., a sub-system which provides more activity to novel stimuli and less to frequently presented stimuli. This filter provides the habituation effect necessary to reduce activity provided to the classification layer.

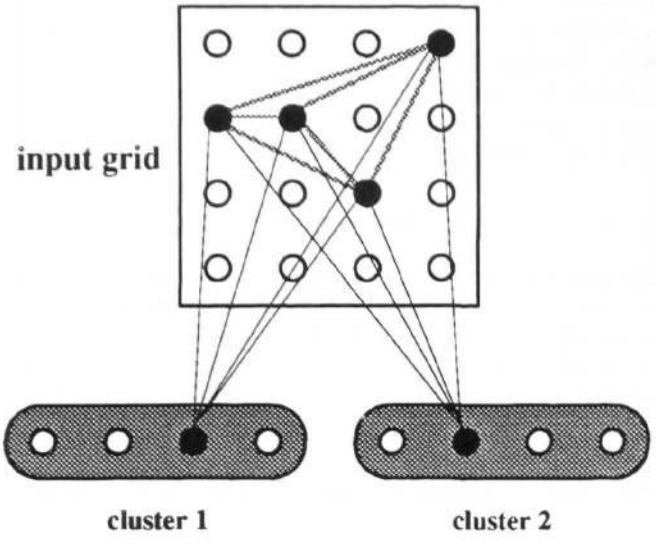

Figure 1. The model. Units in the input layer connect in an excitatory fashion to units in the classification layer (solid lines). They form inhibitory connections between themselves to form a novelty filter (shaded lines).

Learning between the input and classification layers is governed by the competitive rule

$$\Delta w_{ij} = \lambda_1 (a_j / \Sigma a_k) [a_i / \Sigma a_l - w_{ij}], \quad (1)$$

where w_{ij} is the weight between unit i in the input layer and unit j in the classification layer, a_i and a_j are the activities of units i and j respectively, Σa_k is the sum of the activity of all the units in the cluster of which j is a member, Σa_l is the sum of the activities of all units in the input layer, and λ_1 is the learning rate. Equation 1 reduces to the discrete competitive rule (Rumelhart & Zipser, 1986)

$$\Delta w_{ij} = \lambda_1 [1 / \Sigma a_l - w_{ij}], \quad (2)$$

when there is only a single winner j at full activation (1.0) in the output layer (i.e., $(a_j / \Sigma a_k) = 1$), and when all input units are also at unit activation. It is not possible to use Equation 2, however, for two reasons. First, a clear winner must emerge gradually in each cluster in order to simulate the gradual increase in activity and therefore affect as entailed by the exposure effect; selecting a single winner artificially would mean that the activity in the classification layer was constant. The term $(a_j / \Sigma a_k)$ in Equation 1 accommodates the "soft" winner-take-all, or contrast enhancement competitive network which permits multiple activity at relaxation. Second, because of the novelty filter, full activity cannot be guaranteed in the input layer. The term $(a_i / \Sigma a_l)$ in Equation 1 maintains constant learning despite lowered activity in the input layer (this will prove important for low exposure durations discussed in section 3.5).

The novelty filter is governed by

$$\Delta w_{ij} = - \lambda_2 a_i a_j, \quad \text{when } a_i \text{ and } a_j > \varepsilon, \text{ and} \quad (3)$$
$$\Delta w_{ij} = + \lambda_3, \quad \text{otherwise.}$$

The first part of Rule 3 ensures that mutually active units form an anti-Hebbian, inhibitory connection, causing the activity of such units to be reduced. The second part of the rule enables the system to dishabituate when the activity of these units are decoupled. This part of the rule is for recovery from habituation only; weights between input units are not allowed to creep above 0.0.

Relaxation in the network follows the typical network rule

$$a_i = S (\Sigma w_{ij} a_j), \quad (4)$$

where

$$S(x) = 1 / (1 + e^{- (x - \theta)/T}), \quad (5)$$

is the sigmoid output function. Update is accomplished asynchronously to prevent oscillation in the input layer and the competitive clusters.

In summary, unsupervised learning in the network follows two simple rules. The classification rule in Equation 1 is essentially a normalized Hebbian rule, and the filter in Rule 3 is essentially an anti-Hebbian rule with a restorative element. In the following simulations, it will be shown that these rules, in conjunction with the assumption that positive affect is proportional to the amount of activity registered by the classification layer, provide results in accord with the experimental data associated with the mere exposure effect.

Simulation Results

Five sets of simulations are now presented, corresponding to the five main exposure effects described in the introduction. Except where noted, the following parameters are in force. Learning rates in Rules 1 and 3, λ_1, λ_2, and λ_3 are all set to 0.2; ε in Rule 3 is set to 0.001. The threshold for all units θ is 0.9, and the temperature T is 0.1. These two parameters help ensure that only a single winner emerges in each cluster once learning has occurred. Excitatory recurrent weights in the competitive network are 0.5, and lateral inhibitory weights are -0.5. Five clusters consisting of four units each are used. Initial weights between layers are set at random such that the sum of all weights to a given classification unit sum to 1.0. Input stimuli consist of randomly generated stimuli on a 5 by 5 grid with each grid element having a 50% probability of being on. Changing the parameters within reasonable limits does not alter the qualitative form of the results to be presented.

Simulation 1: Number of Exposures

In this simulation, a single input was repeatedly exposed to the network. The graph in Figure 2 show the mean activity in the cluster layer as a function of the number of exposures with full habituation (i.e., $\lambda_2 = \lambda_3 = 0.2$, as usual) and for reference, no habituation in the input layer (i.e., $\lambda_2 = \lambda_3 = 0.0$). Both curves are the averages of the curves obtained

over 10 trials. With no habituation, the exposure effect is monotonic, and the winning units in the classification layer asymptotically approach 1.0 as the weights become aligned to the input vectors. A similar effect can be seen when habituation is in place, but at after 6 exposures the activity in the input layer causes a decline in activity achieved in the classification layer, asymptotically approaching 0.0 as the number of exposures increase. Thus, in common with all two-factor theories, the network model accounts for the inverted U relating affect to exposure number. The following four sections show how providing heterogeneous display, increasing stimulus complexity, inserting a delay after exposure, and decreasing exposure time can overcome some of the habituation to provide a stronger or longer lasting exposure effect.

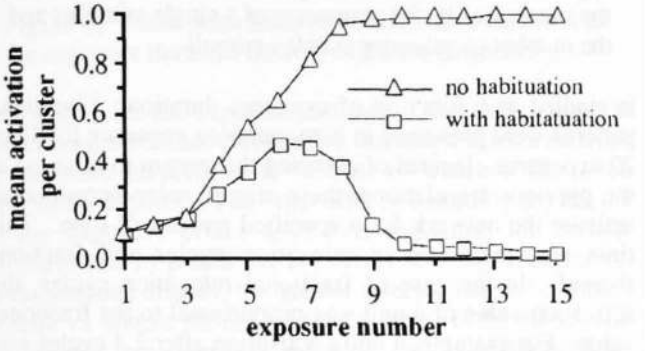

Figure 2. Mean activation per cluster for a single stimulus as a function of exposure number, with and without habituation in place.

Simulation 2: Homogeneous/Heterogeneous Display

In this simulation, 5 input stimuli were presented to the network in a homogeneous display sequence or a heterogeneous display sequence. In the former case, a given stimulus was presented for the specified number of exposures, followed by the next stimulus presented in this manner. Activity in the classification layer was measured for all 5 stimuli at the end of this sequence. Heterogeneous display meant that stimuli 1-5 were presented sequentially, and this process was repeated for the specified number of exposures, after which network activity was measured. All stimuli were presented in the same order across presentations in both cases.

Heterogeneous and homogeneous display involve equal numbers of presentations of a given stimulus for a given exposure number. Despite this, the graph in Figure 3 shows radically different results for the two presentation types (each data point represents the average of ten trials). In accord with the human experimental results showing that homogeneous presentation yields weak exposure effects, these data show a small exposure effect for low exposure frequency and then a decline as exposure number increases.

Two factors contribute to the lack of exposure effect for high frequency homogeneous presentations. First, successive presentation of a single stimulus results in habituation; this lowers the activity in the input layer for the

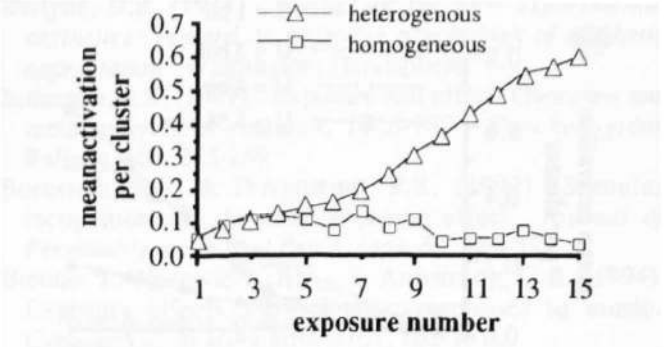

Figure 3. Mean activation per cluster for multiple stimuli as a function of presentation sequence and exposure number.

last few stimuli presented. Earlier stimuli are able to recover from this habituation because other stimuli are interposed between their original exposure and the final test of their activity. However, because of this interval, they fall prey to the second factor. The interposed stimuli will share some of the same winners as the earlier stimuli, but will cause the weights to be realigned in accord with these later patterns. This lessens the response to the earlier stimuli. Heterogeneous presentation counteracts these factors by permitting the dishabituation to occur because successive stimuli will have non-overlapping features. Furthermore, heterogeneous presentation ensures that all stimuli have been presented a relatively short time before testing, ameliorating the effect of the second factor.

Simulation 3: Complexity

Stimulus complexity in this simulation was operationalized in a manner similar to that of Berlyne (1974). Simple stimuli are assumed to differ from each other in relatively few ways, while complex stimuli differ along a number of differing features. Four levels of complexity over five features were tested here. The first, simplest level was created by allowing one feature to take two possible equiprobable values; all the other features took one value only. Thus, there were a total of 2 possible stimuli, with an uncertainty of $U = \log(2) = 1$ bit. In the next level, 3 features took 2 possible values, the other 2 features took only one value, resulting in an uncertainty of $U = \log(8) = 3$ bits. For the next level of complexity, 4 features took 2 values resulting in an uncertainty of $U = \log(16) = 4$ bits. The last, highest level of complexity consisted of 4 features taking 2 values, and 1 feature taking three, with an uncertainty of $U = \log(48) = 5.58$ bits.

Five stimuli were chosen according to a given complexity level and presented to the network in a heterogeneous fashion. The graph in Figure 4 shows activity in the classification layer as a function of the number of times that these sets were presented. In accord with the experimental data, larger exposure effects were found with higher complexity (high uncertainty) stimulus sets after 15 presentations of the set. Less interstimulus similarity within a set, and therefore lower habituation explains this

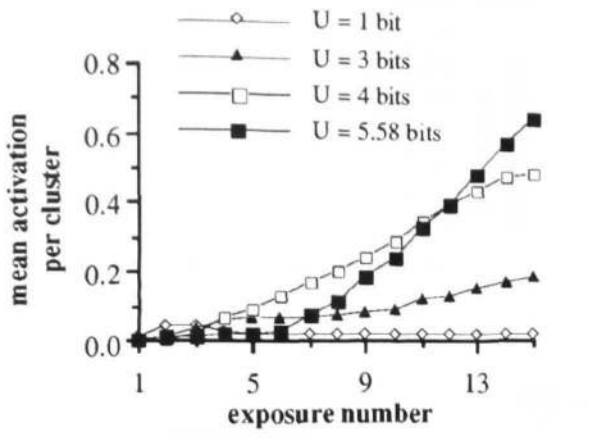

Figure 4. Mean activation per cluster as a function of the uncertainty U of the stimulus set and exposure number.

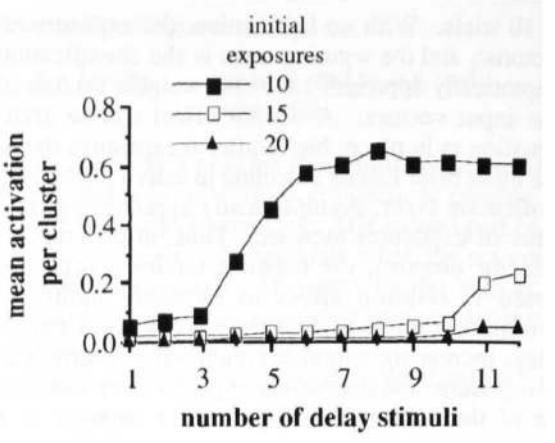

Figure 5. Mean activation per cluster as a function of the number of initial exposures of a single stimulus and the number of subsequent delay stimuli.

result. Also in accord with experimental data, high complexity initially exhibited a less steep rise in activity as a function of exposure number. This is the effect of low complexity stimuli sharing more winners because they are more similar, causing a faster rise in this curve before habituation sets in. In order to extend these results to experiments with natural stimuli such as simple and complex pieces of music, it must be assumed that the raw input is transformed into feature space before it is operated on (see Katz, 1994 for an example of how exposure effects can be demonstrated with melodies of varying complexity).

Simulation 4: Delay after exposure

In this simulation, a single stimulus was presented repeatedly to the network. A variable number of randomly generated stimuli were then presented, after which the network's response to the original stimulus was measured, to simulate a delay between exposure and measurement of affect. Figure 5 shows these results as a function of the number of delay stimuli and the number of presentations of the original stimulus; each data point is the average of ten trials. For all three exposure frequencies, there is a rise in network response as a function of the delay. The reason for this is that repeated presentation of the original stimulus results in habituation, but the delay stimuli result in dishabituation, restoring the input layer to its original response. However, as the results also show, this restoration works best if the original stimulus was not highly over-exposed. The reason for this is that high exposure results in a near complete dampening of input layer activity. This causes no clear winner to emerge, and therefore affects classification learning. These results do not explain the fact that the exposure effect is augmented by mere time delay (Bornstein 1989), although this could be possibly explained by a passive dishabituation effect.

Simulation 5: Exposure duration

In the final simulation, activity in the classification layer is studied as a function of exposure duration. Five input patterns were presented in homogeneous sequence for 10 or 20 exposures. Instead of allowing the system to relax, as in the previous simulations, these stimuli were permitted to activate the network for a specified amount of time. This time was measured in relaxation cycles and fractions thereof. In the case of fractional relaxation cycles, the activation value of a unit was proportional to the fractional value. For example, a unit's activation after 1.4 cycles was the activation value after a single relaxation cycle plus 0.4 of the difference between this value and what it would have been if 2 relaxation cycles had taken place. The graph in Figure 6 shows the results of measuring classification activity after the network was exposed to the patterns (the network was allowed to relax in this testing phase); each data point is the average of ten trials. In accord with the experimental data, both initial exposure frequencies show increased activity for shorter exposure times. Lack of habituation makes the lower exposure frequency presentation somewhat more effective. These results occur because short exposure durations result in less activity in the input units, and by Rule 3 less habituation. Learning in the normalized competitive Rule 1 is not affected by low activity in the input layer. Thus, low exposure duration is favoured. However, learning with very low exposure duration, though resulting in higher activity in the clusters, does not reliably produce a clear winner, when fully exposed to the original pattern. This occurs because the contrast enhancement mechanism provided by the winner-take-all clusters does not have time to suppress the activity of the losing units, and therefore they are subject to the learning process in addition to the winner. This often results in all units becoming active in a cluster to a small extent when the network is exposed to the pattern in the test phase; this is discussed further in the next section.

Discussion

In summary, two simple learning rules, in conjunction with an activation measure of affect, yield simulation results

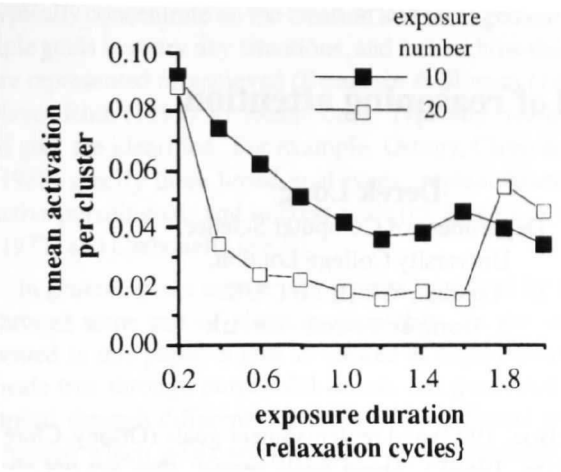

Figure 6. Mean activation per cluster as a function of the exposure duration for two exposure frequencies.

in accord with the experimental data. The basic exposure effect occurs because competitive learning causes a network to be more responsive to stimuli to which it has been previously exposed. The other effects, including the fall off in the exposure effect with over-exposure, the greater effectiveness in producing the effect of heterogeneous vs. homogeneous display, the greater effectiveness of complex stimuli vs. simple stimuli, and the increase in the effect with delay and low exposure duration can be explained by competitive learning acting in conjunction with a novelty filter. The network model is also capable of making predictions about the interactions between these variables, which is not necessarily possible in models which are less completely specified.

In particular, the model may reveal why exposure to subliminal stimuli can produce an exposure effect in the absence of recognition. On the face of it, this is a strange result - subjects are responding positively to stimuli they have been exposed to, and at the same time claiming they have never seen the stimuli. One explanation is that two systems are subserving perception, one affective, and one cognitive (Zajonc, 1980), and that the affective system is amenable to subliminal effects, while the cognitive one only works in conjunction with awareness. The model proposed here provides an alternative explanation. Recall that that low exposure duration resulted in an exposure effect, but that it did so by activating all units in a cluster to a small extent, rather than producing a clear winner. It is possible that subjects say they do not recognize the stimulus because it does not produce these winners, as a supra-liminal stimulus would. However, the net activity in this layer is still greater than for an unencountered stimulus. In effect, the warm glow of affective response occurs without Titchener's warm glow of recognition.

References

Berlyne, D.E. (1970). Novelty, complexity and hedonic value. *Perception and Psychophysics*, 8, 279-286.

Berlyne, D.E. (1971). *Aesthetics and psychobiology*. New York: Appleton.

Berlyne, D.E. (1974). *Studies in the new experimental aesthetics: Toward an objective psychology of aesthetic appreciation*. Washington: Hemisphere.

Bornstein, R.F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-987. *Psychological Bulletin*, 106, 265-289.

Bornstein, R.F. & D'Agostino, P.R. (1992) Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology*, 63, 545-552.

Brentar, J., Neuendord, K.A., & Armstrong, G.B. (1994). Exposure effects and affective responses to music. *Communication Monographs*, 61, 161-181.

Harrison, A.A. (1977). Mere exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press.

Kail, R.V., & Freeman, H.R. (1973). Sequence redundancy, rating dimensions and the exposure effect. *Memory and Cognition*, 1, 454-458.

Katz, B. F. (1993). A neural resolution of the incongruity-resolution and incongruity theories of humour. *Connection Science*, 5, 59-75.

Katz, B. F. (1994). An ear for melody. *Connection Science*, 6, 299-324.

Kohonen, T. (1987). *Self-organization and associative memory*. Springer-Verlag.

Martindale, C. (1984). The pleasure of thought: A theory of cognitive hedonics. *Journal of Mind and Behavior*, 5, 49-80.

Rumelhart, D.E., and Zipser, D. (1986). Feature discovery by competitive learning. In D.E.Rumelhart and James L McClelland (Eds.), *Parallel Distributed Processing, vol. 1*. Cambridge: MIT Press.

Solomon, R.L. & Corbit, J.D. (1974). An opponent process theory of motivation: I. The temporal dynamics of affect. *Psychological Review*, 89, 119-145.

Zajonc, R.B. (1968). The attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9, 1-27.

Zajonc, R.B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151-175.