Chapter 2

Automatic Affective Dynamics: An Activation–Habituation Model of Affective Assimilation and Contrast

Kari R. Irwin, David E. Huber, and Piotr Winkielman

Department of Psychology
9500 Gilman Drive
University of California, San Diego
La Jolla, CA 92093-0109
`krirwin@ucsd.edu`

Abstract. Our modeling approach seeks to better understand the computational dynamics of the affective and cognitive systems. One experimental phenomenon open to such dynamical analysis is "affective priming" -- or the influence of a prior stimulus on subsequent affective processing. In this type of procedure, ‘assimilation priming’ refers to when the response to the target is biased in favor of the prime, such as with brief or minimally attended primes. However, following long durations or highly attended primes, the response to the target is often biased against the prime, which is termed ‘contrast priming’. We present a neural dynamics model of affective priming in which this transition from assimilation to contrast occurs automatically as a result of habituation. Unlike response strategies, this transition is predicted to rise and fall in a gradual nonlinear manner as a function of prime duration. We confirmed this prediction with a speeded affect judgment task that manipulated the exposure duration of valenced images.

1 Introduction

> “In my own opinion, affective habituation is a phenomenon of the same order as sensory adaptation, and results always and only in indifference. Even if Ebbinghaus is correct, and quality passes into opposite quality, we have a sensory analogy in the case of vision: adaptation to yellow means blue-sightedness, local adaptation to green means a purple after-image.”
>
> Edward Titchener.

As proposed by Titchener in 1908, affective responses to emotionally valenced stimuli can produce habituation similar in nature to that of perceptual habituation. If this account is correct, then affect is a response that naturally and automatically rises and falls when a valenced stimulus is presented for an extended time period.

T. Nishida et al. (Eds.): Modeling Machine Emotions for Realizing Intelligence, SIST 1, pp. 17–34.
springerlink.com

By analogy, consider the response of the visual system when first stepping into bright sunlight. Initially everything is blindingly white, but habituation normalizes brightness levels and the outside world comes into view. Further habituation occurs over the course of seconds or minutes, which allows the visual system to adjust to the ambient light levels until the bright sunlight becomes comfortable. However, upon returning indoors, the opposite occurs. Initially the indoor world is too dark to see until habituation again normalizes for the ambient light levels. Indeed, it is believed that pirates on the high seas often wore a single eye patch to sidestep the slowness of habituation (Van de Water 1942). By wearing a patch while on deck, one eye was always habituated to darkness even though the other was habituated to sunlight. This way, pirates could see during brief trips to the nearly complete darkness below deck simply by flipping up their eye patch.

As noted by Titchener, perceptual habituation does not literally produce an opposite response, but instead it tempers the response to the stimulus at hand. Because habituation takes time to dissipate, this produces an apparent contrast effect when presented with a new stimulus (e.g., a return to the indoors) following extended exposure to an extreme stimulus (e.g., bright outdoor sunlight). In this chapter, we implement Titchener's suggestion by developing a dynamic neural network of habituation as applied to affective responses. Unlike the presented model, previous theories of affective priming proposed that controlled decision processes underlie contrast effects. However, automatic habituation was previously demonstrated to provide an accurate account in the domain of priming with word identification (Huber and O'Reilly 2003).

According to this account, habituation serves to normalize to the ongoing state of the world. So, in regards to affect processing, positive affect soon habituates in the presence of a stimulus that elicits a smile (e.g., an image of a puppy). By normalizing in this way, the affect system is made more sensitive to changes. Thus, exposure to a new stimulus (e.g., an image of torture), is easily detected as notably different. In this manner, normalization through habituation serves to illuminate variety in the landscape of affective events.

Although the presented affective habituation model cannot falsify strategic accounts of affective priming, we appeal to parsimony considering that these neural dynamics are known to exist, and considering that these dynamics have previously explained perceptual priming. After presenting the model equations and structure, we report model predictions as a function of prime duration. Finally, we present a new experiment with valenced images for primes and targets that confirmed these predictions.

2 The Effects of Priming: Unwarranted Influence

In psychology, the term "priming" generally refers to the overall influence an earlier stimulus has on later processes. This influence is considered to be fairly automatic in nature (occurring below the level of conscious awareness), and under some theories is thought to be driven by the overlapping perceptual or conceptual features present in earlier and later stimuli.

2.1 Perceptual and Semantic Priming

Perceptual priming is based on the physical form of the stimulus. For example, Huber, Shiffrin, Lyle, and Ruys (2001) used a two alternative forced choice procedure (2 AFC) to indirectly measure the effects of priming visual features on later perceptions. In their paradigm, a prime word was presented, followed by a briefly flashed target word (i.e., around 50 ms), then a mask, and finally a choice between two words. With this presentation sequence for repetition priming (i.e., the prime and target word are identical), it was observed that primes viewed passively for a short duration (i.e., 500 ms) resulted in an "assimilation effect" (i.e., better performance for similar targets as compared to trials with dissimilar targets). However, following primes viewed passively for a longer duration (i.e., 2,000 ms), a full reversal of these effects was observed (i.e., a 'contrast effect'). This suggests that when the prime duration is increased to certain durations, later processing can become impaired for similar targets as compared to dissimilar targets.

Unlike perceptual priming, however, in which assimilation and contrast effects are dependent on the physical form of a stimulus, conceptual priming is driven by semantic or associative features of the stimuli. Even so, the types of priming effects observed in both perceptual and conceptual priming are quite similar. For example, Meyer and Schvaneveldt (1971) showed a similar benefit based on shared prime-target meaning, such as with semantic or associative priming. In this experiment, subjects were either to press one key if both of two simultaneously presented visual letter strings were English words or to press another key otherwise. Interestingly, subjects were again faster and more accurate in responding to displays containing two semantically related words (e.g., 'bread' and 'butter') as compared to displays with two unrelated words (e.g., 'doctor' and 'butter'). The claim here is that like the automatic activation of visual features, semantic features are automatically elicited as well. In both cases, the residual activation of visual or semantic information from the first stimulus affects processing of subsequently presented items.

Priming experiments that manipulate the semantic relatedness of stimuli are used to study the representations used in language. By activating a mental construct with the mere presentation of a word, researchers are able to explore the structure by which words are stored in memory, as well as the process by which they are automatically activated and processed. The same is true for affective processing. Researchers can utilize a variety of priming procedures, presenting emotionally valenced words or images as stimuli, in order to better understand the process by which affective responses are generated and the degree to which affective states share a conceptual representation.

2.2 Affective Priming

In using traditional priming procedures with affective words or images as stimuli, the storage and retrieval of affective constructs can be targeted. And strikingly, it seems as though there are many similarities in how both neutral items (e.g., words) and affective items (e.g., emotional images) are integrated, stored, and retrieved.

A study by Fazio, Sanbonmatsu, Powell, and Kardes (1986) was one of the first to extend from previous semantic priming studies. It documented that, like semantic constructs, which are shown to be automatically activated, evaluative representations are similarly activated in an automatic way. In this study, participants were asked to indicate whether a target word was positive or negative in valence as quickly and accurately as possible. What Fazio et al. (1986) observed was that participants were often faster to make their response when the prime and target were evaluatively congruent (i.e., both positive or both negative) as opposed to when they were evaluatively incongruent. In other words, word pairs sharing only valence (i.e., an intrinsic positive or negative emotional value) produced the same type of assimilation effect as perceptually or conceptually similar words. This suggests a general effect of overlapping prime-target features, regardless of whether or not those features are valenced, and whether they are concrete or abstract in nature.

2.3 Assimilation and Contrast Effects in Priming

The field is more or less in agreement that assimilation effects in priming occur automatically. That is, the facilitated responding for similar items seen in priming occurs below the level of consciousness and as a result of the automatically activated representations (affective or not), which aid in later processing. The underlying nature of contrast priming, however, is still heavily debated.

2.3.1 Strategic Account of Contrast Effects

Traditionally, contrast in priming has been assumed to be strategic in nature. That is, while assimilation priming is likely automatic, contrast is thought to be due to a consciously implemented adjustment to the prime's influence. The implementation of such a compensatory strategic process is most likely when individuals are aware of the prime, as well as motivated and capable of correcting for the prime's unwarranted influence. These factors were confirmed in a number of priming studies. For example, the role of conscious awareness in producing contrast effects is demonstrated in a study where supraliminally presented primes were more likely to produce contrast effects than subliminal presentations of the same image (Murphy and Zajonc 1993). Additionally, Lombardi, Higgins, and Bargh (1987) observed that contrast effects were also more apparent when the primes were explicitly remembered versus primes that were not remembered. And finally, Martin, Seta, and Crelia (1990) suggested that sufficient available cognitive resources were necessary for contrast priming, given that such effects were less likely when participants were cognitively distracted.

In sum, there are numerous studies that have supported a strategic account of contrast priming, and importantly, it is not our intent to refute the role of conscious processes in mediating priming effects under some conditions. However, it is our current goal to support the relatively new stance that contrast priming could result, in part, from automatic processes as well.

2.3.2 Automatic Account of Contrast Effects

In support of the claim that an automatic mechanism may play a mediating role in affective priming, Glaser and Banaji (1999) demonstrated a full cross-over from assimilation to contrast priming simply as a function of the extremity of the prime. Participants in this study were asked to make an objective valence judgment about a series of target words (i.e., whether the word was positive or negative). In doing so, they showed a traditional assimilation effect only when both the prime and target were moderately valenced (e.g., 'butterfly' or 'fire'). Instead, if the prime word was very extreme in nature (e.g., 'love' or 'vomit'), then the effect fully reversed to contrast priming. After replicating these effects, Dijksterhuis and Smith (2002), further observed that repeated subliminal exposure to an extreme stimulus can cause it to behave as a moderately extreme stimulus. That is, the extreme stimuli, which would typically produce a contrast effect following a single exposure, instead resulted in an assimilation effect following multiple subliminal exposures.

The cross-over transitions documented by Glaser and Banaji (1999) and Dijksterhuis and Smith (2002) support the role of an automatic mechanism, because the implementation of a strategy is highly unlikely. A strategic interpretation would require that on a trial by trial basis, subjects either adopted or did not adopt a contrast strategy depending on the particular prime that was presented. Thus, in order to provide a more parsimonious and over-arching account of priming results, we propose an automatic underlying mechanism. But more specifically, we propose that the automatic mechanism at play is related to the fundamental process of habituation.

2.3.3 Automatic Contrast: An Habituation Account

In general, habituation refers to the automatic change in responsiveness of the sensory system to a constant stimulus over time. For example, perceptual phenomena like afterimages and the motion aftereffect are the result of neural adaptation of visual receptors followed by an increased sensitivity to changes in the visual field (Barlow and Hill 1963). However, in terms of affective processing, an habituation account proposes that adaptation causes the overall intensity of an affective response (e.g., the response derived from a valenced word), to gradually fade over time to a diminished level. This adaptation allows the affective system to rapidly adapt to its current state, becoming habituated to the good or the bad, and therefore more sensitive to relative changes in emotional status.

Habituation theory also postulates that it is the automatic adaptation to an affective response that produces contrast effects in priming. Suppose a case where the prime and target share a neural representation (e.g., positive valence). At short prime durations the prime is only partially processed at the time the target is presented, and the residual activation will aid in its processing (resulting in assimilation priming). However, at longer durations, the prime is fully processed, and the affective response is habituated, at the time the target is presented. Thus, the affective system is highly sensitive to changes in valence (e.g., an incongruent target), yet sluggish to reactivate the shared construct for the congruent target (resulting in contrast priming).

More generally, this theory of habituation suggests that both assimilation and contrast effects result from automatically changing levels of neural activation. Furthermore, the theory suggests that these dynamic effects of priming are linked to the degree to which a prime stimulus has been processed. Stapel, Koomen, and Ruys (2002) observed priming effects consistent with this idea. With only a manipulation of prime duration, they demonstrated a transition from assimilation to contrast as prime exposure increased from 30ms to 100ms. Our account of habituation predicts this sort of transition, since at short durations prime activation is still accumulating, thus related targets will receive some benefit. But, at longer prime durations, the prime is fully processed and habituated, making related targets more difficult to process.

These manipulations directly affect the extent to which the prime can be processed. But importantly, the results also begin to (1) provide a time course for how long affective stimuli take to process, (2) quantify the direction and magnitude of the prime's influence on later affective processes, and (3) allow postulation about underlying mechanisms that could potentially explain such non-intuitive patterns of results.

3 A Neural Habituation Account of Priming Effects

As mentioned previously, assimilation and contrast effects are evident in both perceptual and affective priming experiments, which provide reason to postulate a single automatic underlying mechanism to account for these effects. Even though these very different types of priming tasks each explore different types of processing, the results across experiments are characterized by strikingly similar effects. For example, the automatic activation of a construct from the mere presentation of a stimulus is not specific to affect, but rather functions similarly for semantic, orthographic, and other visual features. Additionally, priming effects in perceptual identification tasks exhibit the same cross-over transition from assimilation to contrast as a function of increasing prime duration, as is seen in affective priming experiments (Huber et al. 2001; Stapel et al. 2002).

The similarities across visual, semantic, and affective priming, lead us to propose that a single mechanism of activation and habituation underlies many types of processing more generally. Under this supposition, we then seek to extend an existing computational model of priming instead of developing a theory and model from scratch to account for a specific effect in affective processing.

The neural network model nROUSE (Huber and O'Reilly 2003) used residual activation and habituation to explain the priming effects observed in a word identification task. This model adequately accounts for observed perceptual priming effects with automatic mechanisms of perception and cognition. It assumes that decisions for target words are based on features (e.g., visual, orthographic, semantic), and, that both the extent to which the prime is processed and the degree to which the prime and target share features, will result in beneficial or detrimental effects on the processing efficiency of the target.

The nROUSE model of neural habituation was developed and tested within the basic processes of perception, priming, and memory. And, while it may seem a stretch to generalize such a model to more abstract affective processes, the dynamics of habituation and persistence exist throughout the brain. Thus, these dynamics should exist for all kinds of processing, including affect. Furthermore, this single account will demonstrate that the paralleled effects across domains are due to a common underlying mechanism and provide a parsimonious explanation for the observed patterns of priming effects.

The Huber and O'Reilly (2003) model in particular provides an account of habituation. It is grounded within fundamental neural processes, and, activation within the model's artificial neurons follows the same pattern of true neural activation as was observed by Tsodyks and Markram (1997). Both the observed electrophysiological recordings and the simulated firing rates show an initial accumulation of activation upon exposure to a stimulus, which peaks after some interval, and is followed by accommodation (or neural habituation). On this account, this accommodation occurs in order to reduce excess activation, conserve neural resources, and minimize any residual activation (i.e. prime persistence) after the removal of a stimulus.

There are several neural mechanisms that can explain this sort of 'activity-dependent accommodation', such as neurotransmitter depletion; however, regardless of the specific mechanism, the resultant neural behavior is well described as synaptic depression (Nelson, Varela, Sen, and Abbott 1997; Tsodyks and Markram 1997). Thus, we focus on synaptic depression (i.e., resource depletion), and use the dynamics of this process to derive a general neural network model.

Synaptic depression occurs as resources in the synapse (e.g., neurotransmitters, calcium, etc.) become depleted due to the activation of the pre-synaptic cell. Then, because of this depletion, the transferred activation between the two cells, or the overall effect of each action potential on the receiver neuron, is also greatly reduced. The dynamics of this process have been studied with electrophysiological recordings of both the pre- and post- synaptic neurons, and these recordings have helped to illustrate the magnitude and time scale of these effects. An example of such electrophysiological recordings is displayed in Figure 1 (Tsodyks and Markram 1997), which illustrates how, even with a constant activation applied directly to the pre-synaptic neuron, the post-synaptic neuron has an initial depolarization, followed by a rapid decrease to an asymptotic level after about 200ms.

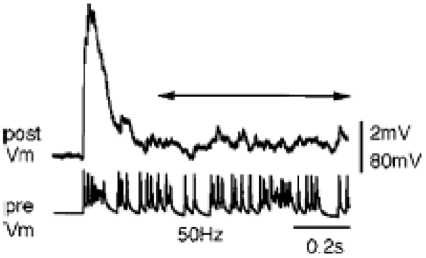

Fig. 1. Electrophysiological recordings obtained by Tsodyks and Markram (1997) with patch clamping of two connected neurons. This captures the processes of neural activation and subsequent habituation as dictated by the process of synaptic depression.

4 A Neural Network Model of Affective Priming

Using the previously described principles of neural activation and habituation, we next propose a neural network model that incorporates the dynamics of habituation and resource depletion to account for the effects observed in affective priming tasks. Previous research regarding synaptic depression has already observed several perceptual and behavioral correlates. For example, Pantic, Torres, and Kappen (2002) demonstrated that synaptic depression allows for rapid switching between memory states; Chance, Nelson, and Abbott (1998) showed that synaptic depression in the primary visual cortex allows for increased detection of moving/changing stimuli; and, Gotts and Plaut (2002) showed that damage to the synaptic depression process is a possible explanation for neuropsychological deficits in semantic processing. Because the low level process of synaptic depression correlates with several aspects of higher level functioning, we suspect that depletion of resources at the neural level may play a similar mediating role in affective processing.

4.1 Model Structure

Within the proposed model, each artificial neuron (i.e., node) individually incorporates characteristics of synaptic depression. It is the inclusion of this mechanism that produces more realistic patterns of activation for each node, in which output activations are similar to the transient dynamics observed in single cell recordings (see Figure 1; Tsodyks and Markram 1997). Each node itself is then placed within a cascading hierarchical network through which activation can travel, as seen in Figure 2 below.

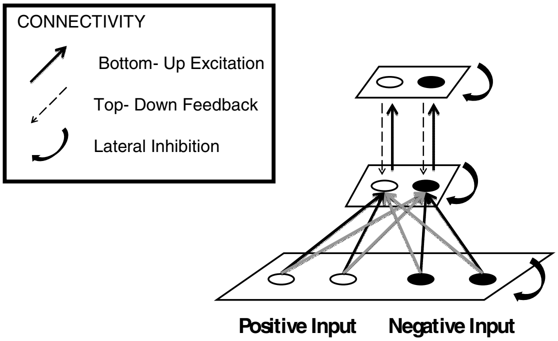

Fig. 2. The hierarchical structure for the proposed habituation model of affective processing. From bottom to top, the levels within the network represent progressively more complex features of the stimuli. There are two pathways within the network representing positive and negative valence. However, the projections from the lowest to middle layers are proportionally distributed with 85% traveling to the node of the same valence and 15% traveling to that of the opposite valence. Connections between levels in both directions are excitatory. Connections within each level are inhibitory. See text for further discussion and details.

We do not have results that constrain the exact nature of each level within the neural network in the current application of the model. So, given the lack of necessity at this point, there are no strong claims about what the hierarchical layers specifically represent or which brain regions might be involved. But rather, the structure represents the general integration process of information, in which simple visual features are combined to form a more complex representation from which an evaluative response can be made.

The model does require, however, that the integration process occurs within three distinct stages. These layers are required in order to produce both the assimilation and contrast effects like those observed in perceptual and affective priming. Assimilation effects (i.e., when the target response is biased in favor of the prime) occur as a result of source confusion, which in the model is due to converging activation between the bottom and middle layers. On the other hand, contrast effects (i.e., when the response to the target is biased in the opposite direction of the prime) emerge as a result of the habituation association after the point of convergence (i.e., the connection between the middle and highest levels). Additionally, lateral inhibition is required within each layer of the network to dampen excess activation and prevent multiple representations from being equally active at a given time.

The hierarchical network also defines two separate network streams through which activation can travel, representing both positive and negative valence. Each of these streams is made up of two nodes in the lowest visual layer, which allows for visual input from both the prime and target in the case that they are of the same valence. Then, each node in the lowest layer projects to both the positive and negative node in the middle layer, with the projected activation proportionally distributed and preferential to the node of the corresponding valence. This distributed representation is proposed in the model because valenced stimuli are often highly complex. An emotionally valenced image might easily contain both positive and negative elements when these elements are considered in isolation, even when the sum of those elements is clearly one or the other. For instance, take the image of a gun pointed at a child. The scene as a whole is obviously negative; however, the mere presence of a child's face may still activate some positive feelings on some level. This type of stimulus complexity is true for valenced visual scenes encountered in real life, as well as those used in the later presented simulation study and experiment. Thus, it seems necessary that any model of affective processing incorporate this element of proportionally distributed activation of multiple emotional states.

The balance and interaction between opponent forces of emotions (such as positive versus negative), is in line with applications of opponent process theory that are decades old. This theory was initially described by Hurvich and Jameson as a general model of neurological opponent processing for phenomena such as color vision (Hurvich & Jameson 1957). However, this model was later expanded to explain emotional behavior (Solomon & Corbit 1974). The extended model asserts that opposing emotions are paired, and that when one emotion in a pair is experienced, the other is suppressed. It also predicts an opponent reaction in the opposite direction after the primary affective response has been habituated. In effect, the theory claims that emotions modulate around a point of neutrality when stimulated, or that opponent forces can cancel each other out. This theory has been supported in studies in which subjects were asked to report feelings of different emotions (e.g., fear and pleasure) while engaging in various activities (e.g. skydiving) (Solomon and Corbit 1974). However, the currently presented affective habituation model proposes a neural mechanism by which opponent emotions may interact. With a specific network structure and dynamics of activation, this model offers falsifiable predictions and measures with which to quantify the interacting effects of opposing emotional states.

4.2 Activation Dynamics

Each individual node within the proposed network functions to capture aspects of temporal integration of information through dynamic activation updates. As initially described by Huber and O'Reilly (2003), these updates are expressed as a probability of spiking over some time interval. This expression is used as opposed to a discrete number of spikes, because it allows for each artificial neuron to be viewed as a general representation of any neuron functioning under similar inputs and outputs, simplifying the biological processes. The simulated membrane potential, *v,* in each of the nodes (neuron *i* in layer *n*) is updated as a function of weighted excitatory inputs, lateral inhibition, and leak currents as a change in potential per unit of time. The general form of this membrane potential update is:

$$\frac{\Delta v_i^n(t)}{S_n} = (1 - v_i^n)\left\{\sum_{\forall j} w_{ij} o_j^{n-1} + F \sum_{\forall k} w_{ik} o_k^{n+1}\right\} - v_i^n\left\{L + I \sum_{\forall l} o_l^n\right\} \qquad (1)$$

As can be seen in Equation 1, the weighted bottom-up and top-down excitatory inputs combine to drive the membrane potential towards the excitatory reversal potential of 1.0 (i.e., the first term in the curly brackets, multiplied by 1-v). The weights for the bottom-up connections are given a default strength of 1.0, whereas the weights for the top-down feedback are given the strength value F ($0 < F < 1$). In opposition to this excitatory input, lateral inhibition (i.e. the summation over the l unites within level n) is given a strength of I, and combines with a fixed term for leak, L. Importantly, it is the slow leak of activation that occurs within the membrane that results in residual activation (or persistence of the prime). Jointly, the facilitory and inhibitory inputs drive the membrane potential towards the reversal potential of 0 (i.e., the second term in curly brackets, multiplied by $-v$).

Importantly, not every layer within the structure is restricted to integrate information at the same speed. Instead, the overall differences between the layers are accommodated for with the parameter *Sn*. This free parameter serves to scale each of the update calculations for each of the three layers, regulating the speed at which information is integrated over time. The larger the value of Sn, the faster information is integrated, thus changing the membrane potential more quickly. Conversely, smaller values of the Sn parameter result in the layer integrating information more slowly.

Equation 2 below specifies the update calculation for synaptic output. But, unlike traditional artificial neural networks whose activation value varies only with time, this model represents the synaptic output as a product of two time varying parameters. These dynamic parameters being: (1) the amount by which the membrane potential exceeds some threshold value (Θ), and (2) a factor, a, representing available resources for that particular connection. In other words, synaptic output is the product of the on-average firing rate $(v-\Theta)$ and the effect of each action potential given the available resources (a). This dynamic calculation thus captures the process of habituation due to recent activity, and predicts how systems of neurons will be affected by this resource depletion (Abbott et al. 1997; Tsodyks & Markram 1997).

$$o = (v - \Theta)a \tag{2}$$

Equation 3 provides the update calculation for synaptic resources available in a given connection (i.e. neurotransmitters, calcium, etc), assuming mathematical constants for the parameters D (depletion) and R (resources). The process of synaptic depression is represented in the equation by two parts. First, resource depletion is driven directly by the recent firing of the pre-synaptic neuron, and secondly, the extent to which those resources have been depleted is what drives the rate of recovery.

$$\frac{\Delta a_i^n(t)}{S_n} = R(1 - a_i^n) - D o_i^n \tag{3}$$

4.3 Mapping the Model to Responding

All three of the equations above, work in conjunction with one another to continuously update neural activation and resources available in the synapse. And together, they are able to capture patterns of responding. The model assumes that visual input travels primarily in a bottom-up fashion, until it reaches layer three, at which point performance that is based on affective evaluation is achieved through the fluency responses (i.e., time-to-peak) from the output of the highest level. Within these highest nodes, target activation accumulates, but quickly reaches a peak level. It is at this point that habituation becomes strong enough to limit the accrual of additional activation. The participant's categorical response (e.g., "positive" or "negative") is assumed to be the response whose corresponding activation reaches its maximum first following the presentation of the target item (i.e., a horse race). The actual time to respond is assumed to be related to this 'time-to-peak' of the winner. Thus, this measure is used as a relative indicator of response time between conditions, as it directly corresponds to the true reaction time.

5 Testing the Model: Experimental Procedure Used for Simulation Fit and Behavioral Study

In order to test the generality of the proposed habituation theory and dynamic model, as applied to affective processing, we next present both behavioral and simulated data for an experiment that parametrically varied prime duration. Behavioral data were collected for two reasons: (1) to observe the effect of this manipulation on human behavior, and (2) to serve as a test for whether the model was able to account for these effects. Qualitatively, the theory predicts that as a function of increasing prime duration, assimilation effects will gradually flip to contrast effects (rather than an abrupt flip, such as with a strategy). However, this particular model implementation of the theory specifies a particular functional form to this transition that may, or may not, be compatible with empirical results.

The experimental procedure (see Figure 3) involved the presentation of a single prime image followed by a single target image. Each image was specifically selected according to normative ratings of positive and negative valence (Lang, Bradley, and Cuthbert 1999). Every experimental trial began with a fixation stimulus ("+"), followed by a prime picture, which was displayed for 500, 1000, 1500, 2000, or 2500 ms. Following that exposure, a target image then replaced the prime and remained on the screen until participants indicated the valence of the image with a response key. The influence of the prime on a given target response was measured by reaction time, and the time course of the observed priming effects was made possible by the incremental prime durations that were used.

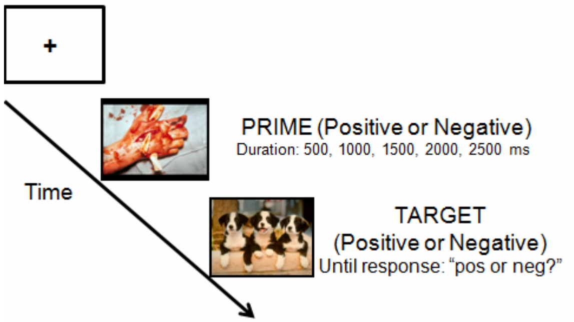

Fig. 3. Affective priming procedure used in the subsequent simulation study and behavioral experiment (Irwin, Huber, & Winkielman, in preparation)

A transition from assimilation to contrast was predicted because of the previously discussed rise and fall of neural activation upon exposure to any stimulus. When a prime is presented for only a brief duration, the level of activation is quite high at the time the target appears, thus assimilation is observed due to source confusion. However, when a prime is presented for a longer period, the meaning of the prime is more fully processed, which results in habituation to the features of the prime. This habituation makes the affective system sluggish to reactivate the shared representation (as with similar target items), yet quick to detect changes in status (as with dissimilar target items). Therefore, with increasingly long prime durations, assimilation effects are predicted to appear and disappear, and then fully reverse to contrast effects in a gradual, non-linear manner.

Additionally, as prime duration is increased to an even longer duration, it is predicted that all priming effects disappear completely. This final disappearance is a unique prediction of using a distributed coding scheme, and would not be the case if positive and negative processing occurred independently of one another. This distributed representation was used because of the complexity of the valenced stimuli, which are comprised of a number of elements ranging in affective value. Because of the this assumption, when prime exposure is sufficiently long, both the dominant response implied by an image, as well as the lesser response implied by an image, reach a level of full habituation. At that point, there is very little difference between the two responses, and so the contrast priming is reduced.

5.1 Behavioral Data: Prime Duration

Using the experimental procedure described in Figure 3, data was collected from 63 subjects, each run through 1 session comprised of 80 trials. Each individual trial included the random presentation of a single prime image for a variable duration (500, 1000, 1500, 2000, 2500 ms) and a single target image, which remained on the screen until the subject indicated its valence. In half of the trials the prime and target were congruent in valence (i.e., positive – positive or negative – negative), while in the other half, the prime and target were incongruent in valence (i.e., positive – negative or negative – positive). The indirect measure of the prime's influence was the time it took for the subject to indicate their response through a key press. The median reaction times for congruent and incongruent trials are displayed in Figure 4, as well as the relative difference between the two (incongruent trials minus congruent trials). Both plots show the observed change in target response time as a function of both prime duration and valence congruency.

The behavioral data do show the predicted non-monotonic effect. Importantly, this pattern of effects seems to confirm an automatic mechanism, as a strategic account would necessitate that subjects adopt a different strategy at each prime duration.

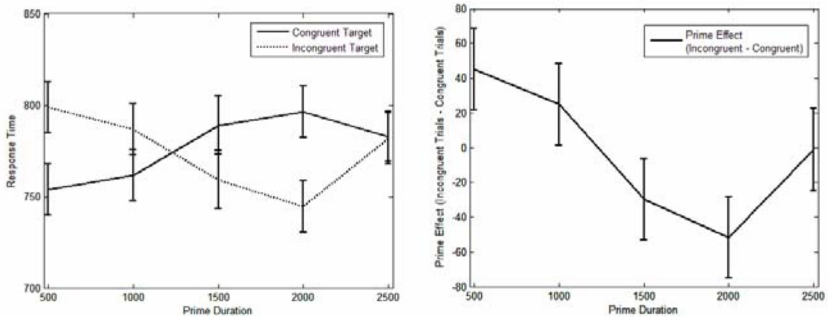

Fig. 4. Behavioral results from Irwin et al. (in preparation), which show the median correct reaction time results as a function of prime–target valence congruency and prime duration. The plot on the left shows the reaction times for congruent trials and incongruent trials separately, while the plot on the right shows the relative difference between the two priming conditions (i.e. the 'prime effect'). Error bars denote plus or minus one standard error of the mean.

5.2 Simulated Data: Prime Duration

We now present a test of the proposed habituation model, in which we assess whether or not it can account for the reported empirical findings. The simulation was run in time steps of one millisecond, where input was sent directly to the appropriate node within the lowest layer. Input was set to zero in general, but changed to one for the time steps when a visual object was presented that was the preferred input for that node. The values of the fixed parameters were left the same as those initially proposed by Huber and O'Reilly (2003). These parameters have worked well for both word and face priming and represent general cortical dynamics, so they should be applicable for affective processes as well (Huber & O'Reilly 2003; Rieth and Huber 2005). The values were: inhibition, $I = 0.3$, threshold, $\Theta = 0.15$, leak, $L = 0.15$, depression, $D = 0.324$, recovery, $R = 0.022$ and first layer's rate of integration, $S1 = 0.054$.

While these default parameters represent dynamics that have been studied and previously confirmed, the rates of integration for the middle and top layers ($S2$ and $S3$, respectively) represent novel affective elements. Thus, these parameters were adjusted to best fit the observed data. The best fitting values for these parameters were: the rate of integration for the middle level, $S2 = 0.005$, and for the highest level, $S3 = 0.03$.

The parameter specifying the proportion of activation sent to the dominant response versus the alternative response was also adjusted to best fit the observed data. Some distributed representation for the images was predicted due to their complexity (i.e. that the pictures likely evoked both a positive and negative response to some degree), however, the exact proportion of this distribution was unknown and therefore explored. For this parameter, the best fitting value was with 85% of activation being sent to the preferential node (i.e., to the 'correct' overall valence), leaving the additional 15% sent to the non-preferential node (i.e., the 'incorrect' overall valence).

Using the proposed structure and parameter values, the model captured the qualitative change in priming effects that were observed empirically. The similarities can be observed by comparing Figure 4 (i.e., the empirical data) with Figure 5 (i.e., the simulated data), which show that the observed and simulated reaction times follow a similar course as a function of prime duration and valence congruency. In moving from left to right on the graph, both figures show an initial appearance of assimilation effects at the shortest prime durations (i.e. congruent trials are responded to more quickly than incongruent trials), which diminishes and fully reverses to contrast effects as prime duration is increased (i.e. congruent trials are responded to more slowly than incongruent trials). Finally, as prime duration is increased even farther, priming effects eventually disappear altogether. The model's qualitative and quantitative fit of the data not only supports the theory of habituation as it applies to affective processing, but it also suggests that perhaps similar neural dynamics are involved in the identification and processing of both words and affective stimuli.

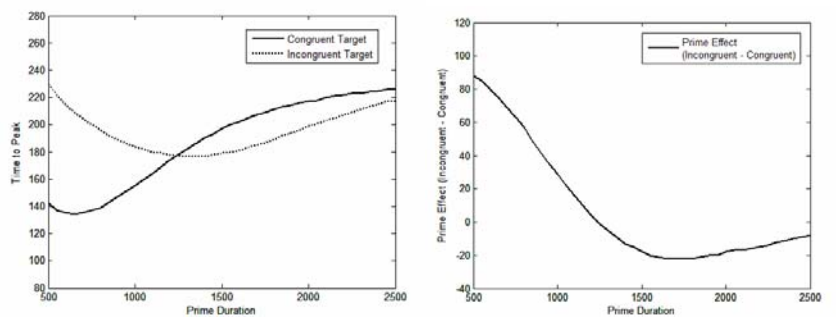

Fig. 5. Simulated data using the proposed affective habituation model. The time to peak response was used as an indicator of relative ease of processing, and directly corresponds to the observed behavioral reaction times. The graph on the left shows the time-to-peak values for congruent trials and incongruent trials separately, while the graph on the right shows the overall “prime effect”, or the relative difference between the two priming conditions.

6 General Discussion

The purpose of this chapter was to explore the dynamics underlying affective processing in order to better understand the way in which we integrate and interpret emotional information. As observed by Aristotle over 2,000 years ago, “anybody can become angry - that is easy, but to be angry with the right person and to the right degree and at the right time and for the right purpose, and in the right way - that is not within everybody's power and is not easy.” As Aristotle keenly observed, intelligent behavior requires that we respond to the right event, in the right time, and to the right degree. That is, an appropriate emotional response should only consider the relevant events at hand, and not be confused by unrelated recent events. At the same time, events unfold over time and take time to fully process. This creates a dilemma for intelligent behavior in which processes of “persistence” have to be balanced against processes of “separation.”

In addition to the problem of separation and proper attribution of events, organisms also face the task of maintaining an ability to respond to a full range of events – sensory events that range, say, in brightness, or emotional events that range in valence . As discussed in prospect theory, adaptation (or habituation) helps maintain this ability (Kahneman & Tversky 1973). Because repeated exposure to a stimulus causes the responsiveness of the sensory system to be diminished over time, habituation brings you back to a baseline level from which you are better at detecting the next change in the environment. For example, adaptation to darkness allows you detect brighter and darker things. Adaptation to happiness, allows you to distinguish events that are better or worse than your current state.

This idea of affective habituation even relates to long term adaptation to significant events over the time course of months or years. The term “hedonic treadmill” was coined by Brickman and Campbell (1971), which captures the tendency of a person to remain at a relatively stable level of happiness despite major life occurrences. For example, Brickman, Coates, and Jonoff-Bulman (1978) reported that there were only small differences in life satisfaction between lottery winners and control subjects, while Silver (1982) found that the affective experience of paraplegics was already predominantly positive after only a few weeks of their accident. These types of studies demonstrate that humans rapidly adapt to their current situation returning to some sort of set affective baseline.

Although we do not claim that the same mechanisms are directly responsible for affective habituation as it occurs over the course of milliseconds and over the course of years, it is interesting to consider some of the apparent parallels that are present. Specifically speaking to those effects of affective habituation that are seen in short term priming, however, we propose a dynamic neural network model. The model takes into account the process of habituation as it affects the perceptual and affective system's ability to parse the stream of incoming information and correctly integrate and interpret the cues at hand. Not only can this model capture the effects of priming that are observed behaviorally, but considering that the proposed automatic dynamics can handle both assimilation and contrast (and the gradual transition between the two), it also suggests that it may be premature to consider more elaborate controlled discounting theories.

It is important to note, however, that while the reported simulation provides a good account of the interaction between priming conditions and prime duration (i.e., a comparison of the right panel of Figure 4 to that of Figure 5), it does not necessarily capture the main effect of prime duration itself (i.e., a comparison of the left panel of Figure 4 to that of Figure 5). As can be observed visually, there is no main effect of prime duration in the behavioral data (i.e., the average of the two conditions is perfectly flat), however, the average of the two conditions in the simulated data exhibits a u-shaped trend. This trend emerges because of inhibition within the layers of processing, which results in a single prime duration that produces the most interference regardless of valence, as compared to other prime durations. The heightened degree of interference at this point restricts the level of activation a target response is able to accumulate, thus resulting in a response that reaches its peak value more quickly. However, at shorter and longer prime durations there are lower levels of inhibition, and therefore, the response to the target is able to reach a higher level, also taking a longer period of time. One possible explanation for this apparent lack of a prime duration effect, is that the model does not currently include an alerting mechanism to know when to respond. And, with the incorporation of such a mechanism, a better fit of the behavioral data could be possible.

Nonetheless, human behavior generally confirms the assumptions made by a habituation account, and the proposed dynamics may be useful in developing an artificial affective system. Other related models have been proposed to account for similar patterns of effects, for example, the psychophysical account by Klauer, Teige-Mocigemba, and Spruyt (2009). This model, however, is adapted from descriptive principles of perceptual processing (e.g., the Weber effect), but it does not specify the underlying mechanism. Therefore, it is limited in its predictive power. For instance, though it makes predictions regarding the manipulation of prime duration, it does not specify the exact dynamics by which the effects occur. Our account, however, proposes that it is specifically lingering activation and neural habituation, which produce various patterns of priming effects.

Whether or not it is specifically the mechanisms currently proposed, any natural or artificial system needs to encompass some sort of specific processing mechanisms, which will allow it to process the continuous stream of incoming information and appropriately adjust its affective responses. Obviously, an "emotional computer" (which will react to users' emotions by adapting its functioning) should have some "knowledge" about how long people stay upset or excited (how fast they adapt), how to structure positive or negative stimuli for maximal impact, and when a possible mis-attribution of emotion is likely to influence a decision or behavior. These traits, attainable with the automatic dynamics of the proposed habituation model, are each relevant for the artificial system to integrate affect with cognition, to learn and communicate in an interactive way, and to behave in an appropriate manner. Thus, by using a formal model that incorporates the same dynamics observed in human behavior, an artificial system's affective processing may become more realistic and appropriate over a wide range of situations.

References

Abbott, L.F., Varela, J.A., Sen, K., Nelson, S.B.: Synaptic depression and cortical gain control. Science 275, 220–224 (1997)

Barlow, H.B., Hill, R.M.: Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. Nature 200, 1345–1347 (1963)

Brickman, P., Campbell, D.T.: Hedonic relativism and planning the good society. In: Appley, M.H. (ed.) Adaptation-level theory, pp. 287–305. Academic Press, New York (1971)

Brickman, P., Coates, D., Janoff-Bulman, R.: Lottery winners and accident victims: Is happiness relative. Journal of Personality and Social Psychology 36(8), 917–927 (1978)

Chance, F.S., Nelson, S.B., Abbott, L.F.: Synaptic depression and the temporal response characteristics of V1 cells. The Journal of Neuroscience 18, 4785–4799 (1998)

Dijksterhuis, A., Smith, P.K.: Affective habituation: Subliminal exposure to extreme stimuli decreases their extremity. Emotion 2(3), 203–214 (2002)

Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., Kardes, F.R.: On the automatic activation of attitudes. Journal of Personality and Social Psychology 50, 229–238 (1986)

Glaser, J., Banaji, M.R.: When fair is foul and foul is fair: Reverse priming in automatic evaluation. Journal of Personality and Social Psychology 77(4), 669–687 (1999)

Gotts, S.J., Plaut, D.C.: The impact of synaptic depression following brain damage: A connectionist account of "access/refractory" and "degradedstore" semantic impairments. Cognitive, Affective, & Behavioral Neuroscience 2, 187–213 (2002)

Huber, D.E., Shiffrin, R.M., Lyle, K.B., Ruys, K.I.: Perception and preference in shortterm word priming. Psychological Review 108(1), 149–182 (2001)

Huber, D.E., O'Reilly, R.C.: Persistence and accommodation in short term priming and other perceptual paradigms: Temporal segregation through synaptic depression. Cognitive Science 27(3), 403–430 (2003)

Hurvich, L.M., Jameson, D.: An opponent-process theory of color vision. Psychological Review 64(6), 384–404 (1957); Irwin, K.R., Huber, D.E., Winkielman, P.: As similation and contrast in affective priming: An habituation account (in preparation)

Kahneman, D., Tversky, A.: On the psychology of prediction. Psychological Review 80(4), 237–251 (1973)

Klauer, K.C., Teige-Mocigemba, S., Spruyt, A.: Contrast effects in spontaneous evaluations: A psychophysical account. Journal of Personality and Social Psychology 96(2), 265–287 (2009)

Lombardi, W.J., Higgins, E.T., Bargh, J.A.: The role of consciousness in priming effects on categorization – assimilation versus contrast as a function of awareness of the priming task. Personality and Social Psychology Bulletin 13(3), 411–429 (1987)

Martin, L.L., Seta, J.J., Crelia, R.A.: Assimilation and contrast as a function of peoples willingness and ability to expend effort in forming an impression. Journal of Personality and Social Psychology 59(1), 27–37 (1990)

Meyer, D.E., Schvaneveldt, R.W.: Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of experimental psychology 90(2), 227–234 (1971)

Murphy, S.T., Zajonc, R.B.: Affect, cognition, and awareness – affective priming with optimal and suboptimal stimulus exposures. Journal of Personality and Social Psychology 64(5), 723–739 (1993)

Nelson, S.B., Varela, J.A., Sen, K., Abbott, L.F.: Functional significance of synaptic depression between cortical neurons. In: Bowers, J. (ed.) Computational neuroscience, pp. 429–434. Plenum Press, New York (1997)

Pantic, L., Torres, J.J., Kappen, H.J.: Associative memory with dynamic synapses. Neural Computation 14, 2903–2923 (2002)

Posner, M.I., Snyder, C.R.R.: Facilitation and inhibition in the processing of signals. In: Rabbit, P.M.A., Dornic, S. (eds.) Attention and performance, vol. V, pp. 669–682. Academic Press, New York (1975)

Rieth, C.A., Huber, D.E.: Using a neural network model with synaptic depression to assess the dynamics of feature-based versus configural processing in face identification. In: Proceedings of the 27th Annual Conference of the Cognitive Science Society, pp. 1856–1861. Erlbaum Associates, Hillsdale (2005)

Silver, R.L.: Coping with an undesirable life event: A study of early reactions to physical disability. Unpublished doctoral dissertation. Northwestern University, Evanston, IL (1982)

Stapel, D.A., Koomen, W., Ruys, K.I.: The effects of diffuse and distinct affect. Journal of Personality and Social Psychology 83(1), 60–74 (2002)

Solomon, R.L., Corbit, J.D.: An opponent-process theory of motivation: Temporal dynamics of affect. Psychological Review 81(2), 119–145 (1974)

Titchener, E.B.: Lectures on the elementary psychology of feeling and attention. McMillan, New York (1908)

Tsodyks, M.V., Markram, H.: The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proceedings of the National Academy of Science 94, 719–723 (1997)

Van de Water, M.: How to see at night. The Science News-Letter 41(23), 358–359 (1942)