Extinction of likes and dislikes:

Effects of feature-specific attention allocation

Jolien Vanaelst[*], Adriaan Spruyt[*], Tom Everaert, & Jan De Houwer

[*] These authors contributed equally to this work and therefore share joint first authorship.

Department of Psychology

Ghent University

Henri Dunantlaan 2

B-9000 Ghent, Belgium

TEL: ++32-9-264-86-13

This work was supported by the Methusalem Grant BOF16/MET_V/002 of Ghent University awarded to Jan De Houwer. Corresponding author: Adriaan Spruyt (email: Adriaan.Spruyt@UGent.be).

Abstract

The evaluative conditioning (EC) effect refers to the change in the liking of a neutral stimulus (CS) due to its pairing with another stimulus (US). We examined whether the extinction rate of the EC effect is moderated by feature-specific attention allocation (FSAA). In two experiments, CSs were abstract Gabor patches varying along two orthogonal, perceptual dimensions (i.e., spatial frequency and orientation). During the acquisition phase, one of these dimensions was predictive of the valence of the USs. During the extinction phase, CSs were presented alone and participants were asked to categorize the CSs either according to their valence, the perceptual dimension that was task-relevant during the acquisition phase, or a perceptual dimension that was task-irrelevant during the acquisition phase. As predicted, explicit valence measures revealed a linear increase in the extinction rate of the EC effect as participants were encouraged to assign attention to non-evaluative stimulus information during the extinction phase. In Experiment 1, the AMP data mimicked this pattern of results, although the effect just missed conventional levels of significance. In Experiment 2, the AMP data revealed an increase of the EC effect if attention was focused on evaluative stimulus information. Potential mechanisms to explain these findings are discussed.

Keywords: evaluative conditioning, feature-specific attention allocation, extinction, selective attention

Extinction of likes and dislikes: Effects of feature-specific attention allocation

Attitudes drive behavior (Allport, 1935). Our interpersonal interactions, the activities we pursue, the products we buy, etc., are all, to some extent, guided by our personal likes and dislikes. Because (most) attitudes are acquired during the lifetime of an organism (Walther, Nagengast, & Trasselli, 2005), it is of great theoretical and practical relevance to study the mechanisms that drive attitude acquisition and attitude change.

One way in which attitudes can be acquired or changed is through Evaluative Conditioning (i.e., EC). In a typical EC study, neutral stimuli are paired with stimuli that have a clear positive or negative meaning, which causes the valence of an initially neutral stimulus (i.e., Conditioned Stimulus or CS) to shift toward the valence of the positive or negative stimulus with which it was paired (i.e., Unconditioned Stimulus or US). This phenomenon, also referred to as the Evaluative Conditioning effect (i.e., EC effect), has now been replicated in a large number of studies (for a review, see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). From a classical conditioning perspective, one might expect that a CS will lose its valence and become neutral again when it is repeatedly presented alone during an extinction session (i.e., CS-only trials). In line with this viewpoint, Lipp, Oughton, and LeLievre (2003) demonstrated that the unpleasantness ratings of a CS that was predictive of an aversive electric stimulus (i.e., US) increased during acquisition but decreased back to neutral during extinction (see also Hofmann et al., 2010). Other researchers, however, reported evidence showing that the EC effect is highly resistant to extinction. Consider, for example, the findings of De Houwer, Baeyens, Vansteenwegen, and Eelen (2000). These authors assigned participants to either a standard EC group or an extinction group. Participants in the conditioning group were exposed to seven presentations of eight CS-US pairs each. The extinction group received the same set of CS-US pairs but these were followed by five CS-only trials. De Houwer et al. (2000) found no significant difference between the size of the EC effect

in the conditioning group and the extinction group. One way to account for these inconsistent findings is to assume that the degree to which the EC effect is sensitive to extinction is dependent upon moderating variables.

The aim of the present research was to examine the extent to which the extinction of EC effects is dependent upon Feature-Specific Attention Allocation (FSAA), that is, the amount of attention assigned to specific stimulus features such as valence, gender, size, etc. The idea that selective attention can operate at the level of specific stimulus dimensions was inspired by the Generalized Context Model (GCM) of classification (Nosofsky, 1986, 1987) according to which stimuli can be represented as points in a multidimensional space. Crucially, the structure of this multidimensional space can vary as a function of FSAA. Attending selectively to a dimension serves to stretch the space along that dimension and shrink the space along unattended dimensions (Nosofsky & Palmeri, 1997; see also Carrol & Wish, 1974; Medin & Schaffer, 1978; Medin, 1983; Nosofsky, 1984; Reed, 1972; Tversky, 1977). As a result of this attentional weighting, variations on attended stimulus dimensions become more salient relative to unattended stimulus dimensions (Medin & Schaffer, 1978). Extending this framework to semantic stimulus information, Spruyt and colleagues (Everaert, Spruyt, & De Houwer, 2013; Everaert, Spruyt, Rossi, Pourtois, & De Houwer, 2013; Spruyt, De Houwer, & Hermans, 2009; Spruyt, De Houwer, Hermans, & Eelen, 2007) suggested that (a) semantic stimulus dimensions (including stimulus valence) are processed only if and to the extent that they are selectively attended to and (b) selective attention assignment is assumed to be dependent upon current goals and task demands. In line with this reasoning, various markers of automatic evaluative stimulus processing have been found to depend on the extent to which attention is assigned to the evaluative stimulus dimension, including the dot probe effect (Everaert, et al., 2013), the emotional Stroop effect (Everaert et al., 2013), and the evaluative priming effect (Spruyt et al., 2009).

Building further on these findings, we hypothesized that the degree to which the EC effect is resistant to extinction must also depend on the degree to which the evaluative stimulus dimension is selectively attended to during extinction. A crucial element in our reasoning is the assumption that the resistance to extinction of the EC effect results from the fact that CSs, once they have acquired a clear valence, evoke a spontaneous evaluative response that is in line with the information that was acquired during the evaluative learning phase (Martin & Levey, 1978). In fact, it could be argued that this spontaneous evaluative response maintains or even strengthens the acquired valence because it co-occurs with the evoking CS and thus confirms the emotional nature of the CS (see Lewicki, Hill, & Czyzewska, 1992). According to the FSAA account, however, participants will process the valence of CSs to a lesser degree if they assign attention to another (non-evaluative) stimulus dimension. Hence, when attending to non-evaluative properties of the CSs during extinction, the presentation of a CS is no longer accompanied by a spontaneous evaluative response, which could result in novel learning that the CS is neutral.

To test this hypothesis, we manipulated the extent to which participants assigned attention to specific features of CSs, both during the acquisition phase and the extinction phase of the experiment. CSs were artificial, grayscale figures (Gabor patches, see below) that varied on two perceptually separable dimensions (i.e., spatial frequency and orientation). During the acquisition phase, selective attention for one of these dimensions (either spatial frequency or orientation) was maximized by asking participants to categorize the CSs according to this dimension (see Spruyt, Klauer, Gast, De Schryver, & De Houwer, 2014).[1] Crucially, the to-be-

[1] As pointed out by an anonymous reviewer, this procedure is similar to the procedures used in research on the so-called intradimensional-extradimensional shift effect (see George & Pearce, 1999). In line with the FSAA framework, it has been argued that the learning rate in an operant condition paradigm is higher for stimulus dimensions that were relevant (as compared to irrelevant) for training discriminations during a preceding training phase (see Sutherland & Mackintosh, 1971). Nevertheless, there is some experimental evidence that argues against the existence of this effect (e.g., Trobalon, Miguelez, McLaren, & Mackintosh, 2003). For an extensive and critical discussion, see Kattner and Green (2016).

judged dimension was predictive of the valence of the USs so that one CS category was always paired with negative USs while the other CS category was always paired with positive USs. Awareness of the contingency between the evaluative nature of the USs and variations in terms of the to-be-judged stimulus dimension of the CSs was thus maximized, thereby ensuring the emergence of a reliable EC effect (i.e., Kattner, 2012). Next, during the extinction phase, participants were asked to categorize CSs either according to the evaluative dimension (evaluative condition), the same perceptual dimension (relevant condition), or the other perceptual dimension (irrelevant condition). Participants thus assigned selective attention to valence in the evaluative condition, to a feature that was correlated with valence in the relevant condition, and to a non-evaluative semantic dimension in the irrelevant condition.

Extinction was expected to occur in the irrelevant condition but not in the evaluative condition. In fact, based on the assumption that the acquired valence of a CS can increase as the result of the repeated co-occurrence of this CS and an evaluative response (see Lewicki et al., 1992), we hypothesized that the EC effect might actually increase rather decrease over the course of the extinction phase in the evaluative condition. Finally, because participants in the relevant condition were encouraged to assign attention to a perceptual feature that was still related to valence, we expected extinction to be less pronounced in this condition relative to the irrelevant condition. Extinction was assessed by means of two different attitude measures. Evaluative ratings were administered to capture deliberate evaluations. Spontaneous evaluations were captured by the Affect Misattribution Paradigm (i.e., AMP; Payne, Cheng, Govorun, & Stewart, 2005).

Experiment 1

Method

Participants

Ninety-six students (28 men, 68 women) at Ghent University with a mean age of 21.45

years (*SD* = 4.6 years) participated for course credit or were paid €5. Three participants made a large number of errors during the acquisition phase of the experiment (i.e., 32.5 %, 37.5 %, and 52.5 %) and were therefore exposed to just 67.5 %, 62.5 %, and 47.5 %, respectively, of the CS-US pairings (see below). Because these participants were clearly outliers compared to the complete sample (*M* = 7.63 %, *SD* = 8.46 %), the data of these participants were excluded from the analyses. Two additional participants were excluded because they responded exceptionally fast (i.e., less 200 ms) on a large number of AMP trials (i.e., 15.63 % and 25.00 %). Again, these participants were clear-cut outliers in comparison with the complete sample (*M* = 1.01 %, *SD* = 4.70 %). Note, however, that the conclusions reported below were not contingent upon the inclusion or exclusion of these participants.

Materials

Based on norm data collected by Spruyt, Hermans, De Houwer, and Eelen (2002), 15 positive and 15 negative color pictures were selected to be used as USs (all 512 pixels wide and 384 pixels high). Several of these pictures originated from the International Affective Picture System (i.e., IAPS; Lang, Bradley, & Cuthbert, 1999). On a scale ranging from -5 (“very negative”) to + 5 (“very positive”), the mean valence rating of negative USs was significantly smaller than zero, $M = -3.08$, $SE = 0.20$, $t(14) = -15.34$, $p < .001$. The mean valence rating of positive USs was significantly larger than zero, $M = 2.36$, $SE = 0.18$, $t(14) = 12.78$, $p < .001$. Eight grayscale Gabor patches (384 × 384 pixels) were used as CSs. These Gabor patches varied on two orthogonal, perceptual dimensions: spatial frequency and orientation (see Figure 1). Values used for the spatial frequency dimension were 4.25, 5.5, 10.5, and 11.75 cycles. Values used for the orientation dimension were 11.25, 22.5, 67.5, and 78.75 degrees. For the AMP, 200 different Chinese pictographs were used as targets. All Chinese pictographs were presented in white and were 256 pixels wide and 256 pixels high.

An Affect 4.0 program (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010)

controlled the presentation of the stimuli as well as the registration of the responses. The experiment was run on a Dell Optiplex GX520 computer. All stimuli were presented against the black background of a 19-inch computer monitor (100 Hz).

Procedure

For the acquisition phase, participants from each experimental condition (see extinction phase) were divided in two counterbalancing groups. The first group was encouraged to assign attention to the spatial frequency of CSs (i.e., frequency group). Participants in the second group were encouraged to selectively attend to the orientation of CSs (i.e., orientation group). To manipulate FSAA, we used the procedures developed by Spruyt et al. (2014). More specifically, we asked participants to categorize the CSs in two arbitrary categories, i.e., "Category A" and "Category B". In the frequency group, participants were informed that assigning attention to "the number of lines" would help them discriminate between the two CS categories. In the orientation group, participants were informed that assigning attention to "the orientation of the lines" would be an efficient strategy to optimize their performance. The cutoff values for assigning a particular CS to either Category A or Category B were 8 cycles and 45 degrees, for the frequency and orientation group respectively.

For each participant separately, the computer program selected five positive and five negative USs from the complete list of available USs (random sampling without replacement). In the frequency group, the presentation of a positive or negative US was contingent upon the spatial frequency of the CSs. In the orientation group, the occurrence of a positive or negative US was contingent upon the orientation of the CSs. Four CSs (e.g., spatial frequency below eight cycles) were paired with all the positive USs and four CSs (e.g., spatial frequency above eight cycles) were paired with all the negative USs, leading to 40 EC trials. As such, each CS category was paired 20 times with either a positive or a negative US. The assignment of a specific CS category to a specific US category was counterbalanced across participants.

Each trial started with the presentation of a CS, which participants were asked to categorize as fast as possible. In case of an erroneous response, a 3000-ms error message (i.e., "FOUT!") appeared. In case of a correct response, the US was presented for 3000 ms. CSs were displayed until a classification response was registered and participants were asked to learn which CS belonged to which category by relying on the feedback. Participants were thus required to guess on the first trial but quickly learned to classify CSs correctly (overall error rate = 6.68 %, *SD* = 5.87 %). The inter-trial interval was drawn randomly from a uniform distribution between 1500 ms and 2500 ms. The same inter-trial interval was used throughout the rest of the study.

Following the acquisition phase, participants completed two rating phases in which each CS was presented once in a random order. First, participants were asked to provide valence ratings for all CSs using a 21-point rating scale ranging from -10 to + 10. Second, they were asked to indicate, for each picture separately, whether they thought it would be followed, at that particular moment in time, by a negative US (coded as -1), a positive US (coded as 1), or by neither of these (coded as 0). Finally, participants completed a series of 16 AMP trials, modeled after the recommendations of Payne et al. (2005). Each trial started with a 500-ms presentation of a fixation cross. Next, 500 ms after the offset of the fixation cross, a CS was presented for 75 ms, followed by a blank screen for 125 ms, and the presentation of a Chinese pictograph for 100 ms. Following the Chinese pictograph, a black-and-white masking stimulus was presented until a response was registered. Participants were instructed to press the left key if they considered the Chinese pictograph to be less pleasant than average and the right key if they considered the Chinese pictograph to be more pleasant than average.

The AMP was followed by an extinction phase in which each CS was presented alone for five times (i.e., 40 trials). Participants were divided in three conditions and were again asked to categorize the CSs. In the relevant condition, participants were asked to categorize CSs

according to the same CS dimension as in the acquisition phase. Thus, the orientation group categorized CSs according to orientation whereas the spatial frequency group categorized CSs according to spatial frequency. In the irrelevant condition, participants were encouraged to focus attention to the perceptual dimension of CSs that was irrelevant during the acquisition phase. So, participants in the orientation group were now asked to categorize CSs according to spatial frequency whereas participants in the frequency group were now asked to categorize CSs according to orientation. In the evaluation condition, participants were asked to judge the evaluative meaning of CSs (i.e., positive vs. negative). Each CS was presented until a response was registered. During the extinction phase, participants did not receive any feedback concerning their performance.

Finally, participants were again asked to provide evaluative ratings and expectancy ratings of CSs, and to complete the AMP.

Results

Acquisition effects

A significant difference was found between the expectancy ratings of positive CSs (M = 0.85, SD = 0.30) and negative CSs (M = -0.87, SD = 0.28), $F(1, 90) = 857.4$, $p < .001$, $\eta^2 = .91$. Likewise, we found a significant difference between the mean valence ratings of positive CSs (M = 2.89, SD = 3.51) and negative CSs (M = -2.77, SD = 3.72), $F(1, 90) = 65.99$, $p < .001$, $\eta^2 = .42$. The AMP data corroborated these findings. The proportion of pleasant judgments was significantly higher on positive-CS trials (M = .69, SD = .24) than on negative-CS trials (M = .32, SD = .27), $F(1, 90) = 57.62$, $p < .001$, $\eta^2 = .39$. None of these effects was qualified by an interaction with counterbalancing group (i.e., attention to either orientation or spatial frequency during the acquisition phase), all F's < 3.14, nor did any of these EC effects differ between conditions, all F's < 1.64. For ten participants, at least two of the dependent measures (i.e., expectancy ratings, valence ratings, or AMP scores) did not reveal an acquisition effect in the

expected direction (i.e., the perceived probability that CSs would be followed by a positive US should be higher for positive CSs as compared to negative CSs and positive CSs should be rated more positively than negative CSs). As it makes little sense to study extinction effects in the absence of a normal acquisition effect, the data of these eleven participants were excluded from all further analyses.

Extinction effects

For each participant, a first difference score was calculated by subtracting the (mean) post-acquisition ratings on negative CS trials from the (mean) post-acquisition ratings on positive CS trials. A second difference score was calculated by subtracting the post-extinction ratings on negative CS trials from the post-extinction ratings on positive CS trials. Finally, an extinction effect was calculated by subtracting the second difference score from the first difference score. Extinction effects were calculated for each dependent measure and were subjected to a one-way ANOVA with condition (irrelevant vs. relevant vs. evaluative) as a between-subjects factor. Because we expected a gradual decrease in the magnitude of the extinction effect from the evaluative condition over the relevant condition to the irrelevant condition, the between-subject factor "condition" was treated as a linear factor. Expectancy ratings did not reveal a significant difference in the extinction effect between the three experimental conditions, $F < 1$. The valence ratings, however, revealed a clear effect of condition, $F(1, 78) = 6.99$, $p < .01$, $\eta^2 = .08$. Next, follow-up tests were performed in which, for each condition, the post-acquisition EC effect was compared with the post-extinction EC effect. As predicted, a follow-up *t*-test showed a significant extinction effect in the irrelevant condition, $t(27) = 2.28$, $p = .03$, $d = 0.43$. As can be seen in Table 1, the EC effect was smaller after the extinction phase than after the acquisition phase. In contrast, the EC effect was more or less unaffected by the extinction procedure in both the relevant condition, $t(26) = 1.62$, $p = 0.12$, $d = 0.31$, and the evaluative condition, $t < 1$. In fact, as can be seen in Table 1, the evaluative

condition revealed a small increase in the EC effect from post-acquisition to post-extinction. The AMP data[2] mimic these results, albeit the main effect of condition just missed conventional significance levels, $F(1, 78) = 3.37$, $p = .07$, $\eta^2 = .04$. Follow-up *t*-tests revealed that the extinction effect was far from significant in both the relevant condition and the evaluative condition, t's < 1. The extinction effect was also non-significant in the irrelevant condition, but numerically there was a trend in the anticipated direction, $t(27) = 1.66$, $p = .11$, $d = 0.31$[3].

Discussion

The results of Experiment 1 are straightforward. Exposing participants to CS-only trials resulted in a reduction of the EC effect when participants attended to an irrelevant CS feature during the extinction phase of the experiment (i.e., irrelevant condition) but not when participants attended to stimulus valence (i.e., evaluative condition) or to a CS feature that was correlated with stimulus valence during the acquisition phase (i.e., relevant condition). Both the explicit and implicit valence measure revealed this data pattern. The results of Experiment 1 thus provide initial support for the hypothesis that extinction of EC effects is moderated by FSAA.

It must be noted, however, that the critical effect of condition just missed conventional significance levels for the AMP data ($p = .07$). For a number of reasons, we are inclined to attribute this null-finding to a Type-II error. First, AMP scores were numerically in perfect accordance with our expectations (see Table 1). Second, implicit measures are typically more noisy than explicit measures (e.g., Cunningham, Preacher, & Banaji, 2001). Accordingly, given that the number of AMP trials was limited to just 16 trials, it could be argued that we simply

[2] There is some debate concerning the extent to which AMP effects are driven by participants who rate the primes intentionally (see Bar-Anan & Nosek, 2012). We examined this possibility using the procedures described by Payne et al. (2013). In both experiments, AMP scores did not differ between participants who claimed that they had rated the primes intentionally and participants who did not claim that they had rated the primes intentionally, all t's < 1.

[3] When we included the ten participants who did not show an EC effect following acquisition, the valence ratings still revealed a significant effect of Condition, $F(1, 88) = 4.07$, $p < .05$, $\eta^2 = .04$. Significant effects were absent in the AMP, $F(1, 88) = 1.32$, $p = .25$, $\eta^2 = .01$, and the expectancy ratings, $F < 1$

lacked sufficient power to capture a significant modulation of the extinction effect with the AMP. Note, however, that we deliberately opted for the use of a small number of AMP trials because the AMP requires participants to assign attention to the evaluative stimulus dimension. The administration of the AMP might thus have interfered with the FSAA manipulation had we used a higher number of trials. Finally, we examined whether the extinction effects as measured by the AMP and the explicit valence measure were correlated across participants. Reassuringly, this correlation was reliable, $r = .23$, $p < .05$, adding further weight to the idea that the effects captured by the AMP were indeed meaningful.

In sum, the results of Experiment 1 are consistent with the hypothesis that FSAA is a key moderator of the extinction rate of the EC effect. The present data therefore complement and extend earlier studies showing FSAA effects in the context of EC. For example, Gast and Rothermund (2011) demonstrated that the EC effect is more likely to emerge if participants are encouraged to process evaluative stimulus information during the acquisition phase as compared to when they are encouraged to process non-evaluative stimulus information (see Olson, Kendrick, & Fazio, 2009, for similar findings in the non-evaluative domain). Not only does FSAA impact the emergence of the EC effect, Spruyt et al. (2014) showed that FSAA can also impact the generalization of the EC effect. In their studies, EC effects were found to generalize to untrained, novel stimuli that were similar to the CSs in terms of a stimulus dimension that was selectively attended to during or prior to the evaluative conditioning phase. For example, after an acquisition procedure in which (neutral) pictures of old men and young women (CSs) were paired with positive and negative USs, respectively, novel pictures of old women were also evaluated in a positive manner by participants who were encouraged to assign selective attention to the age dimension. In contrast, participants who received the exact same training regimen but were encouraged to assign attention to the gender dimension were inclined to evaluate novel pictures of old women as less favorable (for related findings, see Trobalon et

al., 2003). Taken together, it can thus be concluded that FSAA exerts a systematic impact on the functional properties of the EC effect.

Still, because a direct measurement of FSAA was not included in the design of Experiment 1, the argument that we manipulated FSAA is purely based on the fact that our experimental procedures produced the anticipated effects. One may thus argue that the findings of Experiment 1 are insufficient to conclude that our effects were causally driven by variations in FSAA. Accordingly, to further corroborate and extend the findings of Experiment 1, we conducted a replication study in which a direct measure of FSAA was administered at the very end of the experiment. More specifically, participants were presented with pairs of CSs and for each pair they were asked to judge the similarity of the two CSs. The INDSCAL algorithm, a multidimensional scaling approach that allows for the estimation of individual attention weights (Carroll & Chang, 1970; Carroll & Wish, 1974), was then used to verify whether the FSAA manipulation had been successful. Overall, participants in the relevant and irrelevant condition were expected to assign selective attention to the relevant and irrelevant stimulus dimension, respectively. Participants in the evaluative condition were expected to assign selective attention to the dimension that was relevant during the acquisition phase because the acquired valence of the CSs and the relevant dimension were perfectly confounded. Finally, we expected to find a reliable extinction effect only in participants who were able to shift their attention from the relevant to the irrelevant stimulus dimension.

Experiment 2

Method

Participants

Eighty-five female and eleven male Ghent University students with a mean age of 22.24 years (SD = 3.35 years) were paid €5 in exchange for their participation. Two participants made a large number of errors during the acquisition phase of the experiment (i.e., 37.5 % and 35.0

%) and were thus exposed to a limited number of CS-US pairings (i.e., 62.5 % and 65.0 %, respectively). Because these participants were clearly outliers compared to the complete sample (*M* = 5.76 %, *SD* = 6.09 %), their data were excluded from the analyses. We also excluded the data of one participant who was familiar with the Chinese language and knew the meaning of the Chinese ideographs used in the AMP. Finally, we excluded the data of two participants who responded exceptionally fast (i.e., less 200 ms) on a large number of AMP trials (i.e., 34.38 % and 37.50 %). Again, these participants were clear-cut outliers in comparison with the complete sample (*M* = 1.27 %, *SD* = 5.36 %). Note, however, that none of the critical extinction effects reported below were contingent upon inclusion or exclusion of these participants.

Materials and Procedure

The materials and procedures used in Experiment 2 were identical to those used in Experiment 1, except for the inclusion of a similarity judgment task at the very end of the experiment. Each trial of the similarity judgment task started with the presentation of a fixation cross in the center of the screen for 500 ms that was immediately followed by the presentation of two identical black-and-white masking stimuli (420 × 420 pixels) displayed 296 pixels apart on the left and the right side of the fixation cross. After 200 ms, the masks were replaced by two different Gabor patches that were presented for 500 ms. Each Gabor patch was then replaced with a mask that was displayed for 200 ms. Participants were asked to rate the similarity of the two Gabor patches on a four-point scale using an (AZERTY) keyboard. The scale ranged from very similar (response key “x”) over slightly similar (response key “v”) and slightly different (response key “n”) to very different (response key “;”). Participants were asked to respond within 2 seconds. If participants failed to respond within this deadline, a 300-ms message informed them that they were too slow (i.e., “TE TRAAG!”, meaning “too slow” in Dutch). The inter-trial interval varied between 500 and 1500 ms. Every possible pairing (28 pairs) of the eight CSs was presented twice, leading to a total of 56 similarity judgment trials.

Results

Acquisition effects

In accordance with Experiment 1 (see section procedure), participants quickly learned to classify CSs correctly. The overall error rate in the acquisition phase was 5.05 % (*SD* = 4.10 %). Expectancy ratings revealed a significant difference between positive CSs (*M* = 0.91, *SD* = 0.19) and negative CSs (*M* = -0.94, *SD* = 0.17), $F(1, 90) = 2830.9$, $p < .001$, $\eta^2 = .97$. In addition, a significant difference was found between the (mean) valence ratings of positive CSs (*M* = 2.84, *SD* = 2.86) and negative CSs (*M* = -3.12, *SD* = 3.65), $F(1, 90) = 86.67$, $p < .001$, $\eta^2 = .49$. The AMP data corroborated these findings. The proportion of pleasant judgments was significantly higher on positive-CS trials (*M* = .72, *SD* = .24) than on negative-CS trials (*M* = .35, *SD* = .25), $F(1, 90) = 61.68$, $p < .001$, $\eta^2 = .41$. Reassuringly, these EC effects were not moderated by the condition factor, all *F*'s < 2.12. Likewise, neither the EC effect in the valence ratings, $F(1, 89) = 3.35$, $p = .07$, $\eta^2 = .04$, nor the EC effect in the AMP data, $F(1, 89) = 1.04$, $p = .31$, $\eta^2 = .01$, was dependent upon the counterbalancing group that participants were assigned to (i.e., attention allocation to the orientation dimension or the spatial frequency dimension during the acquisition phase). The expectancy ratings, however, did show a larger difference between ratings on positive CS-trials and negative CS-trials for participants who attended to spatial frequency (*M* = 1.93, *SD* = 0.17) as compared to participants who attended to orientation (*M* = 1.76, *SD* = 0.42), $F(1, 89) = 6.13$, $p < .05$, $\eta^2 = .06$. For six participants, at least two of the dependent measures (i.e., expectancy ratings, valence ratings, or AMP scores) did not reveal an acquisition effect in the expected direction. Similar to Experiment 1, the data of these six participants were excluded from all further analyses.

Extinction effects

Individual extinction effects were calculated in the same way as for Experiment 1 (see above) and were subjected to a one-way ANOVA with condition (irrelevant vs. relevant vs. evaluative) as a between-subjects factor. As in Experiment 1, the between-subject factor

condition was treated as a linear factor.

Whereas the expectancy ratings revealed no effect of condition, $F < 1$, the valence ratings were clearly affected by the condition factor, $F(1, 82) = 11.61$, $p < .01$, $\eta^2 = .12$. As expected, follow-up *t*-tests comparing the post-acquisition EC effect with the post-extinction EC effect, revealed a reliable extinction effect in the irrelevant condition, $t(29) = 5.02$, $p < .001$, $d = 0.92$, but not in the relevant condition, $t < 1$, or the evaluative condition, $t < 1$ (for summary statistics, see Table 2).

A different degree of extinction across different conditions was not picked up, however, by the AMP, $F(1, 82) = 1.92$, $p = 0.17$, $\eta^2 = .02$. Nevertheless, given our a priori hypothesis, follow-up analyses were performed to examine the extinction effect in each experimental condition. The extinction effect was far from significant both in the irrelevant condition, $t < 1$, and the relevant condition, $t(25) = -1.16$, $p = .26$, $d = -0.13$. Interestingly, a reversed extinction effect emerged in the evaluative condition, $t(28) = - 2.76$, $p < .05$, $d = -0.51$. In line with the idea that reactivation of the CS-US relationship can strengthen the acquired valence of the CS (see Lewicki et al., 1992), the EC effect following extinction was larger, not smaller, than the EC effect after acquisition. Moreover, an exploratory analysis revealed that the effects obtained with the AMP were highly contingent on the extent to which the CSs were clear-cut instances of the experimental stimulus categories. Remember that participants were presented with eight Gabor patches, only four of which were characterized by extreme values both on the spatial frequency dimension and the orientation dimension (see Figure 2). Given that the Gabor patches were presented for just 75 ms during the AMP, one might argue that participants may have been unable to discriminate between different stimulus categories unless these categories were instantiated by salient exemplars. In line with this reasoning, the two-way interaction between stimulus salience and experimental condition was reliable, $F(1, 82) = 4.21$, $p < .05$, $\eta^2 = .05$. Follow-up analyses confirmed that the anticipated moderation of the extinction effect was

reliable when restricting the analyses to the AMP trials with salient stimuli, $F(1, 82) = 6.11$, $p < .05$, $\eta^2 = .08$. Follow-up *t*-tests revealed that this effect was mainly driven by an increase of the EC effect from post-acquisition ($M = .47$, $SD = .49$) to post-extinction ($M = .72$, $SD = .40$) in the evaluative condition, $t(28) = -3.36$, $p < .01$, $d = -.62$. In contrast, there was no change in the EC effect from post-acquisition to post-extinction the irrelevant condition, $t < 1$, or the relevant condition, $t(25) = -1.74$, $p = .09$, $d = -.34$. When the analyses were restricted to the AMP trials with non-salient stimuli, there was no indication whatsoever for a differential extinction rate across different experimental conditions, $F < 1$. A similar moderation by stimulus salience was not picked up by the explicit measures. Moreover, a similar analysis of the data obtained in Experiment 1 revealed no effects of stimulus salience either, all *t*'s < 1. Caution is thus in order when interpreting these results.

Manipulation Check: FSAA after extinction

The similarity data were entered into the SPSS V22.0 statistical package (SPSS Inc., 1997) and were analyzed using the INDSCAL algorithm (Carroll & Chang, 1970; Carroll & Wish, 1974). Given the use of two-dimensional stimuli (i.e., Gabor patches that varied in terms of orientation and spatial frequency), the analysis was constrained to two dimensions. The algorithm reached convergence after 11 iterations (i.e., S-stress decrease smaller than 0.0001). The eventual representation had an S-stress of .19, and explained, on average, 79.1% of the variance of each individual participant. In sum, both measures of model fit showed that a two-dimensional model reached an acceptable fit. As can be seen in Figure 2, the dimensions of this two-dimensional space correspond to the orientation dimension and the spatial frequency dimension.

Individual attention weights were coded such that negative values signaled selective attention for the orientation dimension and positive values signaled selective attention for the spatial frequency dimension. We thus expected the mean attention weight to be negative in

participants who were required to focus attention on the orientation dimension during the extinction phase. Likewise, the mean attention weight was expected to be positive in participants who were required to focus attention on the spatial frequency dimension. Participants who were asked to evaluate the CSs during the extinction phase (i.e., evaluative condition) were expected to focus their attention upon the stimulus dimension that was predictive of the valence of the USs during the acquisition phase. Accordingly, depending on whether the orientation dimension or the spatial frequency dimension was predictive of the valence of the USs, we expected the attention weights in this group to be negative or positive, respectively.

Individual attention weights were subjected to a 2 (FSAA group: attention to orientation versus attention to spatial frequency) × 3 (condition: irrelevant condition vs. relevant condition vs. evaluative condition) ANOVA. While the main effect of FSAA reached significance, $F(1, 79) = 9.26$, $p < .05$, $\eta^2 = .10$, a significant interaction between FSAA group and condition revealed that the impact of the FSAA manipulation was dependent upon condition, $F(2, 79) = 10.46$, $p < .01$, $\eta^2 = .21$. As can be seen in Table 3, the mean attention weights were in line with the FSAA manipulation both in the relevant condition, $F(1, 24) = 14.22$, $p < .001$, $\eta^2 = .37$, and the evaluative condition, $F(1, 27) = 11.46$, $p < .01$, $\eta^2 = .30$. In the irrelevant condition, however, the mean attention weights were numerically in the opposite direction, albeit not significantly so, $F(1, 28) = 3.94$, $p = 0.06$, $\eta^2 = .12$. We will discuss this observation at length shortly.

Extinction effects as a function of FSAA

Based on individual attention weights obtained through the MDS approach, participants were divided in two groups. Participants were assigned to the successful group if individual attention weight concurred with the FSAA manipulation during the extinction phase. Participants were assigned to the unsuccessful group if their attention weights revealed attention assignment opposite to the FSAA manipulation during the extinction phase. For each of the

three dependent measures (i.e., expectancy ratings, explicit valence ratings, and AMP scores), extinction scores were subjected to a two-way ANOVA with condition (irrelevant condition vs. relevant condition vs. evaluative condition) and manipulation check (successful vs. unsuccessful FSAA manipulation) as between-subjects factors. Results showed that the interaction between manipulation check and condition was neither reliable for the expectancy ratings, $F < 1$, nor for the evaluative ratings, $F(1, 79) = 2.40$, $p = .13$, $\eta^2 = .03$, or the AMP scores, $F < 1$.

Discussion

Taken together, the present results replicate the results obtained in Experiment 1. As in Experiment 1, the evaluative ratings revealed that the extinction rate of the EC effect was dependent upon variations in FSAA during the extinction phase. In line with our hypotheses, exposure to a series of CS-only trials resulted in a reliable drop of the EC effect only if participants were encouraged to assign selective attention to a stimulus dimension that was orthogonal to evaluative stimulus information (i.e., the irrelevant condition).

The results obtained with the AMP extend this observation in two ways. First, while the overall analyses of the AMP data failed to reveal a clear-cut effect of the FSAA manipulation on the extinction rate of the EC effect, the AMP scores did reveal a significant increase in the EC effect from post-acquisition to post-extinction in the evaluative condition. This observation is in perfect accordance with the hypothesis that reactivation of the CS-US relationship can strengthen the acquired valence of the CS (see Lewicki et al., 1992). Second, exploratory analyses showed that the effect of FSAA was reliable if the analyses were restricted to CSs that were clear-cut instances of the experimental stimulus categories. This data pattern is readily accounted for if one assumes that the very brief presentation time of the CSs (i.e., 75 ms) prevented participants from discriminating between different stimulus categories unless these categories were instantiated by salient exemplars.

Importantly, the present experiment included a direct assessment of FSAA at the very end of the experiment. As anticipated, both in the relevant condition and evaluative condition, participants were inclined to assign attention to the stimulus dimension that was task-relevant during the preceding extinction phase. In the irrelevant condition, however, the mean attention weights were numerically in the opposite direction. That is, participants who were asked to judge the orientation dimension during the extinction phase were, on average, inclined to assign attention to the spatial frequency dimension. Likewise, if the extinction phase required participants to judge the CSs in terms of spatial frequency, attention weights captured at the very end of the experiment were indicative of selective attention for the orientation dimension. To account for this (unexpected) data pattern, the order in which participants completed the different measures is key. Remember that, immediately prior to the measurement of FSAA, participants were asked to provide expectancy ratings for each of the CSs. For participants in the irrelevant condition, by definition, the stimulus dimension that was predictive of the USs during the acquisition phase was orthogonal to the stimulus dimension that was task-relevant during the extinction phase. It can thus be argued that the requirement to retrieve knowledge about the CS-US relationship may have triggered selective attention for the stimulus dimension that was task-relevant during the acquisition phase, as was evidenced by the FSAA scores obtained in this group.

Clearly, this reasoning implies that the FSAA scores obtained at the very end of the experimental sessions were not an accurate reflection of selective attention assignment during the extinction phase. It is therefore not surprising that, despite our initial hypotheses, no relationship was found between the FSAA scores and the extent to which our experimental manipulation impacted the extinction rate of the EC effect in different conditions. Still, the pattern of results obtained with the FSAA measure is important, for two reasons. First, it shows that FSAA is highly volatile and that relatively subtle procedural details can be sufficient to

induce changes in FSAA. Given that FSAA has been shown to modulate automatic evaluative stimulus processing (Everaert, Spruyt, & De Houwer, 2011; Spruyt et al., 2009; 2012), spatial attention allocation (Everaert, et al., 2013), as well as the generalization of EC effect (Spruyt et al., 2014), it could thus be worthwhile to scrutinize in future research the (situational and/or person specific) factors that determine (the flexibility of) FSAA. Second, both in Experiment 1 and Experiment 2, the AMP was administered after the explicit ratings. It may thus be hypothesized that stronger AMP effects may have occurred had we administered the explicit ratings after the AMP. Corroborating this idea, Vanaelst, Spruyt, and De Houwer (2016) did obtain reliable FSAA effects using the AMP in a study in which the AMP was used to capture spontaneous preferences immediately after an FSAA manipulation treatment. Further research would be needed, however, to substantiate this interpretation.

General Discussion

Whereas some researchers have argued that the EC effect is highly resistant to extinction, others have argued that the repeated presentation of a CS in the absence of a US does lead to a reduction of the EC effect. Based on the FSAA framework developed by Spruyt and colleagues (Everaert et al., 2013; Spruyt et al., 2007, 2009, 2012; Spruyt & Tibboel, 2015; Spruyt, 2014), we hypothesized that the extinction rate of the EC effect is dependent upon the degree to which selective attention is assigned to the evaluative stimulus dimension during extinction. To test this hypothesis, we conducted two experiments in which participants were asked to focus their attention on different aspects of CSs during the extinction phase. Participants were either asked to assign selective attention to the evaluative tone of the CSs (i.e., evaluative condition), to a (perceptual) stimulus dimension that was related to stimulus valence (i.e., relevant condition), or to a (perceptual) stimulus dimension that was unrelated to stimulus valence (i.e., irrelevant condition). Both experiments were identical, except for the fact that, in Experiment 2, an attempt was made to include a direct measure of FSAA at the end of the extinction phase.

In line with our expectations, the explicit valence ratings obtained in Experiment 1 revealed a reduction of the EC effect in the irrelevant condition, but not in evaluative condition or the relevant condition. The same pattern was observed for the AMP data, albeit the effect just missed conventional significance levels (p = .06). Experiment 2 corroborated and extended these findings in two ways. First, as in Experiment 1, the explicit valence ratings revealed a significant reduction of the EC effect in the irrelevant condition only. Second, the results obtained with the AMP suggested that the FSAA effect was moderated by the extent to which the CS were clear-cut instances of the experimental stimulus categories. The moderation of the extinction effect by FSAA was reliable for salient but not for non-salient stimuli.

Interestingly, our findings hint at the possibility that the impact of FSAA on the extinction rate of the EC effect might be two-fold. First, assigning selective attention to a stimulus dimension that is orthogonal to valence seems to promote a rapid decay of the EC effect. Second, both in Experiment 1 (i.e., evaluative ratings) and Experiment 2 (i.e., AMP), a reliable increase (not a decrease) of the EC effect was observed in the evaluative condition. That is, repeated evaluation of a CS seems to result in a strengthening of its evaluative tone (see Lewicki et al., 1992). Further research would be required, though, to substantiate the generality of this interesting finding.

Likewise, further research would be needed to shed light on the mechanism(s) underlying our effects. In fact, at least five different accounts can be given for the observation that the magnitude of the EC effect was affected by the nature of the classification task that was performed during the extinction phase of the experiment. As a first possibility, given that a clear-cut data pattern was obtained with the evaluative ratings but not with the AMP, one might simply argue that our results resulted from demand effects. For two reasons, however, this possibility seems unlikely. First, the anticipated pattern of results was present in the evaluative ratings but not in the US expectancy ratings. An explanation in terms of demand effects would

predict the same pattern of results for both explicit measures as there is no obvious reason why participants would use their assumptions about the critical hypotheses to strategically bias their evaluative ratings but not their US expectancy ratings. Second, additional linear mixed effect analyses showed that the EC effect in Experiment 1 (but not in Experiment 2) researched significance even if the US expectancy ratings were not in line with the actual CS-US pairings during the acquisition phase (see Pleyers, Corneille, Luminet, & Yzerbyt, 2007)[4]. It must be noted, however, that the vast majority of the US expectancy ratings did correspond with the actual CS-US parings (i.e., 90 %), so these results should be interpreted with caution.

A second explanation of our findings implies that the association between a CS and its corresponding summary evaluation is subject to decay only if and to the extent that evaluative stimulus processing is hampered. Because the degree to which evaluative stimulus processing occurs is known to depend upon FSAA (see above) and assuming that decay of associations is promoted by events in which CSs are experienced without a concurrent emotional response, this account can readily explain why significant extinction emerged in the irrelevant condition only.

Third, it might be argued that occasion setting was responsible for the extinction effect observed in the irrelevant condition (see Rydell & Gawronski, 2009). According to this viewpoint, the requirement to assign attention to different stimulus properties during the acquisition and extinction phase of the experiment resulted in the formation of contextualized

[4] To examine whether the EC effect was dependent upon (explicit) US expectancy, valence ratings were subjected to a linear mixed effects model. Fixed effects were CS type (positive vs. negative), US expectancy (in line with actual pairings vs. not in line with actual pairings), and the interaction between these factors. Participants and stimuli were defined as crossed random effects. The mixed-model *F* tests were computed using Kenward-Roger's adjusted degrees of freedom (Kenward & Roger, 1997). The analysis revealed a strong in interaction between CS and US expectancy both in Experiment 1, $F(1, 1486.07) = 252.15$, $p < 0.001$, and in Experiment 2, $F(1,1526.02) = 102.43$, $p < 0.001$. In Experiment 1, however, a significant EC effect emerged irrespective of whether the US expectancy ratings were in line with the actual CS-US pairings, $F(1, 1197.26) = 785.32$, $p < 0.001$, or were not in line with the actual CS-US pairings, $F(1, 241.93) = 24.18$, $p < .001$. In Experiment 2, a reliable EC effect was found only if the US expectancy ratings were in line with the actual CS-US pairings, $F(1, 1297.57) = 1142.80$, $p < 0.001$.

representations of CSs. Accordingly, it might be hypothesized that the EC effect might have surfaced again had we tested participants under conditions that promoted selective attention for the stimulus dimension that was task-relevant during acquisition (i.e., ABA renewal). The results obtained with the FSAA measure suggest that this possibility is certainly a viable route for future studies.

A fourth account is based on exemplar-based models of categorization and memory (Hintzman, 1984; Smith & Zarate, 1992). According to these models, when memory is probed with a target stimulus (e.g., a CS), memory traces of specific objects, persons, or experiences contribute to the overall memory response as a function of their similarity to the target stimulus. The stronger the overlap between a target stimulus and a particular exemplar representation, the stronger the influence of that exemplar representation on the memory response. Crucially, FSAA is assumed to determine the weight of each stimulus dimension in the computation of the similarity between the target stimulus and the exemplar representations (Smith & Zarate, 1992). One can thus expect a significant reduction of the EC effect for two reasons. First, because positive and negative CS categories were equivalent in terms of the stimulus dimension that was task-irrelevant during the acquisition phase, both positive and negative exemplars contributed to the overall memory response as soon as participants focused their attention on this stimulus dimension. Second, given that evaluative stimulus processing is reduced under conditions that promote selective attention for a neutral stimulus dimension, exemplar information stored during the extinction phase must have been relatively neutral in the irrelevant condition as compared to the relevant condition and the evaluative condition. In sum, the net memory response at the time of testing would thus be based on a mixture of neutral and non-neutral memory traces, thereby reducing the EC effect.

Finally, our findings may also be accounted for in terms of a propositional model of EC (see De Houwer, 2009). According to this framework, EC effects are mediated by propositional

knowledge about the CS-US relation. One might thus argue that the emergence of an extinction effect must involve the formation of new propositions about the (absence of a) relationship between CSs and USs and/or the valence of the CSs. Crucially, such a corrective process would be more likely to occur in the irrelevant condition as compared to the relevant condition and the evaluative condition because participants experience the CS without a concurrent evaluative response only in the former condition.

In sum, while our findings suggest that the EC effect can be reduced by an extinction regimen that requires participants to assign attention to a stimulus dimension that is unrelated to stimulus valence, it remains an open question how this effect can be accounted for at the mental-process level. Nevertheless, our findings are important as EC procedures are increasingly used to modify likes and dislikes in applied settings (e.g., Houben, Schoenmakers, & Wiers, 2010). It seems particularly interesting, for example, to verify whether the outcome of exposure treatment programs is dependent upon the extent to which patients are encouraged to assign selective attention to nonevaluative stimulus information. It thus seems a viable approach to further scrutinize the underlying mechanisms and operating conditions of the extinction effect reported in the present paper.

References

Allport, G. W. (1935). Attitudes. In C. Murchinson (Ed.), *A handbook of social psychology* (pp. 798–844). Worcester, MA, US: Clark University Press.

Bar-Anan, Y., & Nosek, B. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality & Social Psychology Bulletin*, *38*, 1194–208. doi:10.1177/0146167212446835

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*, 283–319. doi:10.1007/BF02310791

Carroll, J. D., & Wish, M. A. (1974). Multidimensional perceptual models and measurement methods. In E. C. Carterette & M. . Friedman (Eds.), *Handbook of perception (Vol. 2). Psychophysical judgement and measurement* (Volume 2., pp. 391–447). New York: Academic Press.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit Attitude Measures: Consistency, Stability, and Convergent Validity. *Psychological Science*, *12*, 163–170. doi:10.1111/1467-9280.00328

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*, 1–20. doi:10.3758/LB.37.1.1

De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture-picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*, 237–242. doi:10.1037//0097-7403.26.2.237

Everaert, T., Spruyt, A., & De Houwer, J. (2011). On the (un)conditionality of automatic attitude activation: the valence proportion effect. *Canadian Journal of Experimental Psychology*, *65*, 125–32. doi:10.1037/a0022316

Everaert, T., Spruyt, A., & De Houwer, J. (2013). On the malleability of automatic attentional biases: effects of feature-specific attention allocation. *Cognition & Emotion*, *27*, 385–400. doi:10.1080/02699931.2012.712949

Everaert, T., Spruyt, A., Rossi, V., Pourtois, G., & De Houwer, J. (2013). Feature-specific attention allocation overrules the orienting response to emotional stimuli. *Social Cognitive and Affective Neuroscience*, *9*, 1351–1359. doi:10.1093/scan/nst121

Gast, A., & Rothermund, K. (2011). What you see is what will change: evaluative conditioning effects depend on a focus on valence. *Cognition & Emotion*, *25*, 89–110. doi:10.1080/02699931003696380

George, D. N., & Pearce, J. M. (1999). Acquired distinctiveness is controlled by stimulus relevance not correlation with reward. *Journal of Experimental Psychology. Animal Behavior Processes*, *25*, 363–373.

Hintzman, D. L. (1984). MINERVA 2 : A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*, 96–101. doi:10.3758/BF03202365

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: a meta-analysis. *Psychological Bulletin*, *136*, 390–421. doi:10.1037/a0018916

Houben, K., Schoenmakers, T. M., & Wiers, R. W. (2010). I didn't feel like drinking but I don't know why: the effects of evaluative conditioning on alcohol-related attitudes, craving and behavior. *Addictive Behaviors*, *35*, 1161–1163. doi:10.1016/j.addbeh.2010.08.012

Kattner, F. (2012). Revisiting the relation between contingency awareness and attention: Evaluative conditioning relies on a contingency focus. *Cognition & Emotion*, *26*(932363961), 166–175. doi:10.1080/02699931.2011.565036

Kattner, F., & Green, C. S. (2016). Transfer of Dimensional Associability in Human Contingency Learning, *42*(1), 15–31. doi:10.1037/xan0000082

Kenward, M. G., & Roger, J. H. (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics*, *53*, 983–997. doi:10.2307/2533558

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *International Affective Picture System (IAPS): Instruction Manual and Affective Ratings*. *The Center for Research in Psychophysiology. University of Florida, Gainesville, Florida*. The center for research in psychophysiology, University of Florida.

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious Acquisition of Information. *American Psychologist*, *47*, 796–801. doi:00000487-199206000-00013

Lipp, O. V, Oughton, N., & LeLievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation*, *34*, 219–239. doi:10.1016/S0023-9690(03)00011-0

Martin, I., & Levey, A. B. (1978). Evaluative Conditioning. *Advances in Behaviour Research*

and Therapy, *1*, 57–102. doi:10.1016/0146-6402(78)90013-9

Medin, D. L. (1983). Structural principles of categorization. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, Cognition, and Development: Interactional Analyses* (pp. 203–230). Hillsdale, NJ: Erlbaum.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238. doi:10.1037//0033-295X.85.3.207

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 104–114. doi:10.1037/0278-7393.10.1.104

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology*, *115*, 39–61. doi:00004785-198603000-00004

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *13*, 87–108. doi:10.1037/0278-7393.13.1.87

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300. doi:0033-295X/97/S3.00

Olson, M. A., Kendrick, R. V, & Fazio, R. H. (2009). Implicit learning of evaluative vs. non-evaluative covariations: The role of dimension accessibility. *Journal of Experimental Social Psychology*, *45*, 398–403. doi:10.1038/2061099b0

Payne, B. K., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, C. D., & Lundberg, K. B. (2013). Intention invention and the affect misattribution procedure: reply to Bar-Anan and Nosek (2012). *Personality & Social Psychology Bulletin*, *39*, 375–86. doi:10.1177/0146167212475225

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–93. doi:10.1037/0022-3514.89.3.277

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*, 130–144. doi:10.1037/0278-7393.33.1.130

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*, 382–407. doi:10.1016/0010-0285(72)90014-X

Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition & Emotion*, *23*, 1118–1152. doi:10.1080/02699930802355255

Smith, E. ., & Zarate, M. A. (1992). Exemplar-Based Model of Social Judgment. *Psychological Review*, *99*, 3–21. doi:10.1037/0033-295X.99.1.3

Spruyt, A. (2014). Attention please: Evaluative priming effects in a valent/non-valent categorisation task (reply to Werner & Rothermund, 2013). *Cognition & Emotion*, *28*, 560–569. doi:10.1080/02699931.2013.832653

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: a free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology*, *57*, 36–45. doi:10.1027/1618-3169/a000005

Spruyt, A., De Houwer, J. De, & Hermans, D. (2009). Modulation of automatic semantic priming by feature-specific attention allocation. *Journal of Memory and Language*, *61*, 37–54. doi:10.1016/j.jml.2009.03.004

Spruyt, A., De Houwer, J., Everaert, T., & Hermans, D. (2012). Unconscious semantic activation depends on feature-specific attention allocation. *Cognition*, *122*, 91–95. doi:10.1016/j.cognition.2011.08.017

Spruyt, A., De Houwer, J., Hermans, D., & Eelen, P. (2007). Affective Priming of Nonaffective Semantic Categorization Responses. *Experimental Psychology*, *54*, 44–53. doi:10.1027/1618-3169.54.1.44

Spruyt, A., Hermans, D., De Houwer, J., & Eelen, P. (2002). On The Nature of the Affective Priming Effect: Affective Priming of Naming Responses. *Social Cognition*, *20*, 227–256. doi:10.1521/soco.20.3.227.21106

Spruyt, A., Klauer, K. C., Gast, A., De Schryver, M., & De Houwer, J. (2014). Feature-Specific Attention Allocation Modulates the Generalization of Recently Acquired Likes and Dislikes. *Experimental Psychology*, *61*, 85–98. doi:10.1027/1618-3169/a000228

Spruyt, A., & Tibboel, H. (2015). On the automaticity of the evaluative priming effect in the valent/non-valent categorization task. *PloS One*, *10*, e0121564. doi:10.1371/journal.pone.0121564

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. London/New York: Academic Press.

Trobalon, J. B., Miguelez, D., McLaren, I. P. L., & Mackintosh, N. J. (2003). Intradimensional and Extradimensional Shifts in Spatial Learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *29*(2), 143–152. doi:10.1037/0097-7403.29.2.143

Tversky, A. (1977). Features of Similarity. *Psychological Review*, *84*, 327–352. doi:10.1037/0033-295X.84.4.327

Vanaelst, J., Spruyt, A., & De Houwer, J. (2016). How to Modify (Implicit) Evaluations of Fear-Related Stimuli: Effects of Feature-Specific Attention Allocation. *Frontiers in Psychology*, *7*, 1–9. doi:10.3389/fpsyg.2016.00717

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition & Emotion*, *19*, 175–96. doi:10.1080/02699930441000274

Table 1

Mean valence ratings, AMP scores, expectancy ratings, EC effects, and extinction effects as a function of Condition, moment, and CS type in Experiment 1 (SD's in parentheses).

	Acquisition			Extinction			
	CS type			CS type			
Condition	Positive	Negative	Difference	Positive	Negative	Difference	Extinction-effect
			Expectancy ratings				
Irrelevant	.93 (.20)	-.96 (.11)	1.89 (0.28)	0.87 (.22)	-0.91 (.21)	1.78 (0.40)	.12 (.32)
Relevant	.85 (.30)	-.84(.30)	1.69 (0.56)	0.87 (.30)	-0.92 (.22)	1.79 (0.49)	-.09 (.43)
Evaluative	.93 (.18)	-.93 (.17)	1.87 (0.33)	0.88 (.22)	-0.93 (.15)	1.81 (0.31)	.06 (.19)
			Valence ratings				
Irrelevant	2.73 (3.59)	-2.54 (4.18)	5.28 (7.22)	1.30 (2.92)	-1.22 (2.20)	2.53 (4.53)	2.75 (6.39)
Relevant	3.39 (3.17)	-3.61 (3.28)	7.00 (5.49)	3.27 (3.12)	-2.64 (3.61)	5.91 (5.82)	1.09 (3.50)
Evaluative	4.36 (2.57)	-4.04 (2.49	8.39 (4.35)	4.26 (3.22)	-4.91 (2.88)	9.17 (5.84)	-0.78 (4.25)

	AMP scores						
Irrelevant	.72 (.21)	.27 (.28)	.45 (.46)	.65 (.21)	.36 (.27)	.29 (.44)	.16 (.51)
Relevant	.68 (.24)	.33 (.27)	.35 (.45)	.70 (.23)	.34 (.30)	.36 (.51)	-.01 (.26)
Evaluative	.75 (.25)	.31 (.27)	.44 (.48)	.76 (.28)	.27 (.30)	.49 (.54)	-.05 (.46)

Note. Difference scores were calculated by subtracting values for negative CSs from values scores for positive CSs.

Table 2

Mean valence ratings, AMP scores, expectancy ratings, EC effects, and extinction effects as a function of Condition, moment, and CS type in Experiment 2 (SD's in parentheses).

	Acquisition			Extinction			
	CS type			CS type			
Condition	Positive	Negative	Difference	Positive	Negative	Difference	Extinction-effect
			Expectancy ratings				
Irrelevant	.89 (.17)	-.96 (.13)	1.85 (0.26)	.87 (.22)	-.85 (.24)	1.72 (0.40)	.13 (.40)
Relevant	.93 (.13)	-.97 (.11)	1.90 (0.20)	.89 (.21)	-.94 (.16)	1.84 (0.32)	.07 (.34)
Evaluative	.96 (.12)	-.94 (.11)	1.90 (0.19)	.90 (.22)	-.93 (.15)	1.83 (0.29)	.07 (.24)
			Valence ratings				
Irrelevant	2.94 (2.21)	-3.25 (2.95)	6.19 (4.62)	1.87 (1.82)	-1.99 (2.76)	3.86 (4.29)	2.33 (2.55)
Relevant	2.84 (2.59)	-2.58 (3.61)	5.41 (5.76)	2.38 (1.99)	-2.43 (3.11)	4.82 (4.93)	0.60 (3.46)
Evaluative	3.53 (3.32)	-4.57 (3.60)	8.10 (6.58)	3.99 (2.95)	-4.31 (3.39)	8.30 (6.13)	-0.20 (2.53)

	AMP scores						
Irrelevant	.71 (.23)	.34 (.26)	.38 (.44)	.72 (.27)	.32 (.24)	.40 (.47)	-.03 (.44)
Relevant	.75 (.24)	.38 (.23)	.37 (.45)	.76 (.21)	.27 (.29)	.50 (.48)	-.13 (.55)
Evaluative	.76 (.22)	.26 (.23)	.50 (.40)	.85 (.21)	.16 (.21)	.69 (.40)	-.19 (.37)

Note. Difference scores were calculated by subtracting values for negative CSs from values scores for positive CSs.

Table 3

Mean attention weights as a function of FSAA during the extinction phase and condition in Experiment 2 (SD's in parentheses).

Condition	Attention to frequency	Attention to orientation
Irrelevant	-0.21 (0.98)	0.39 (0.66)
Relevant	0.66 (0.68)	-0.59 (0.99)
Evaluative	0.43 (0.73)	-0.76 (1.12)

Note: Positive weights signal selective attention for the spatial frequency dimension whereas negative weights signal selective attention for the orientation dimension

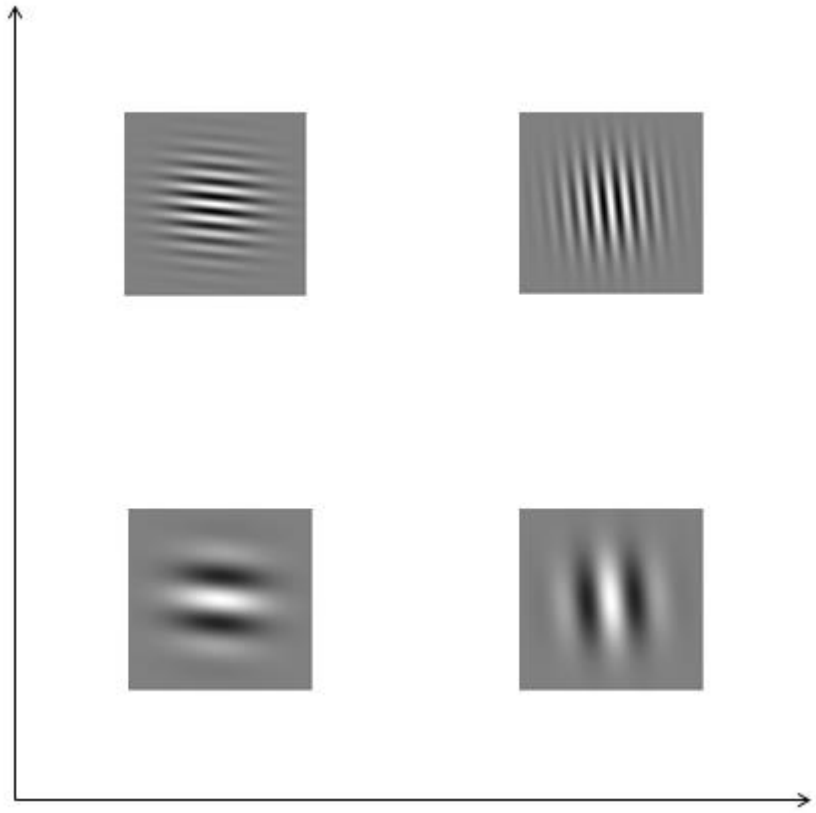

Figure 1. Examples of Gabor Patches.

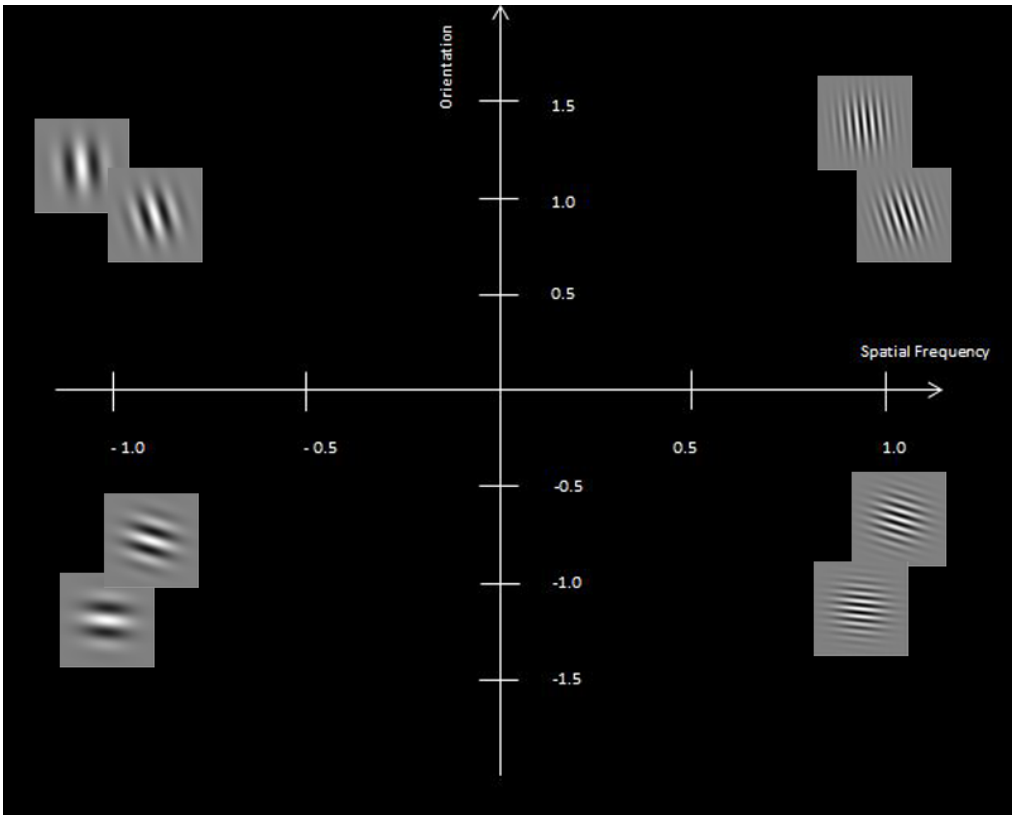

Figure 2. Multidimensional representation derived from participants' similarity judgments.