The Effect of Object-Valence Relations on Automatic Evaluation

Tal Moran and Yoav Bar Anan
Ben-Gurion University, in the Negev, Be’er Sheva

Abstract

Two experiments tested the effect of co-occurrence of a target object with affective stimuli on automatic evaluation of the target when the relation between the target and the affective stimuli suggests that they have opposite valence. Participants learned about targets that ended an unpleasant noise or a pleasant music. The valence of such targets is opposite to the valence of the affective stimuli that co-occur with them. Participants reported preference for targets that ended noise over targets that ended music, but automatic evaluation measures revealed the opposite preference. This suggests that automatic evaluation is sensitive to co-occurrence between stimuli more than to the relation between the stimuli, and that relational information has a stronger influence on deliberate evaluation than on automatic evaluation. These conclusions support the Associative-Propositional Evaluation model (Gawronski & Bodenhausen, 2006), and add evidence regarding the sensitivity of the Evaluative-Conditioning effect to relational information.

Author’s note: This project was supported by grant from the European Union (PIRG06-GA-2009-256467) to Yoav Bar-Anan. Correspondence concerning this article should be sent to Tal Moran, tmo@post.bgu.ac.il.
The Effect of Object-Valence Relations on Automatic Evaluation

A central theme in contemporary attitude research is investigating the differences between automatic and deliberate evaluation (Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000). One of the main differences that have been suggested pertains to the effect of mental evaluative associations on automatic versus deliberate evaluation. Evaluative associations are memory links between the attitude object and evaluative stimuli and responses. The Associative-Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011) suggests that automatic evaluation always reflects the activation of evaluative associations, whereas deliberate evaluation reflects evaluative associations only if these are considered valid evaluative evidence.

The APE model posits that automatic evaluation is determined by the overall valence of the concepts that are activated in response to the attitude object (i.e., the activation of evaluative associations; Gawronski & Bodenhausen, 2011, p. 62). For instance, if an object is associated with positive concepts, then the automatic evaluation of the object would be positive. In contrast, deliberate evaluation reflects the activated valence only if it is considered valid evidence that the attitude object possesses that valence. For example, false slurs against a politician may form an association between the politician and negative valence. Upon deliberate evaluation of the politician the activation of negative valence would be rejected because it is not valid evidence that the politician is negative. In contrast, according to the APE model, the slurs should induce negative automatic evaluation.

In support of the APE model’s assumptions, a few studies found that when the activation of associations implied an evaluation that did not fit other valid evaluative evidence, the associations had stronger influence on automatic than on deliberate evaluation (Petty, Tormala, Brinol, & Jarvis, 2006, Study 1; Prestwich, Perugini, Hurling, & Richetin, 2010; Ranganath & Nosek, 2008; Rydell & McConnell, 2006; Rydell, McConnell, Mackie, & Strain, 2006). However, in these studies participants were never explicitly told that the information that formed the associations (e.g., a repeated presentation of the target object with positive stimuli; Petty et al., 2006) was not valid evaluative evidence. Instead, a separate learning episode added explicitly valid evaluative information that was either consistent or inconsistent with the evaluation implied by the activation of the associations.

Peters and Gawronski (2011) manipulated directly the validity of the information that was supposed to form the evaluative associations. In their studies, participants read positive and negative behavioral descriptions about four target persons. The behavioral description regarding two of the targets was followed by explicit information that the description was false and the behavior was actually uncharacteristic of the target person. According to the APE model, appearing with positive behavioral descriptions should increase automatic liking even if the descriptions are presented as false. Refuting that prediction, most of Peters and Gawronski’s studies did not find any difference between automatic and deliberate evaluation. When the descriptions were immediately flagged as false information, both automatic and deliberate evaluations showed the same preference for the person who appeared with negative behaviors over the person who appeared with positive
behaviors. Peters and Gawronski found divergent automatic and deliberate evaluations only when participants learned what statements were false after they had already read all the behavior descriptions. In that condition, automatic evaluation showed no preference between the person who appeared with false negative behaviors and the person who appeared with false positive behaviors, whereas deliberate evaluation showed the expected preference (see Gregg, Banaji, & Seibt, 2006, for similar procedure and results). That is, the findings supported the assumptions of the APE model only when the information about the validity of the co-occurrence information was delayed.

Peters and Gawronski (2011, p. 567) concluded that when people can assess quickly the validity of the evaluation implied by object-valence co-occurrence, the validity information is incorporated into the mental representation of the object’s evaluation, thus influencing automatic and deliberate evaluation similarly. In other words, when people knew that the behavioral description that was attributed to the target was false information, they did not form an association between the target and the valence of the behavior. Instead, they formed an association between the target and the valence implied by the information that the behavior is uncharacteristic of that person. However, there might be a different reason that prevented the formation of an association between the targets and the valence of the false behavioral descriptions that co-occurred with them. Peters and Gawronski did not manipulate only the validity of the evaluation that behaviors implied. They also manipulated the validity of the information that the target and the behaviors co-occurred. Therefore, perhaps it was not the false validity of the evaluation implied by the behaviors that prevented the formation of an association between the target and the valence of the behaviors. Instead, perhaps evaluative associations did not form simply because participants knew that the target and the behaviors did not co-occur. If that is the case, then perhaps true co-occurrence information would form evaluative associations even if the evaluation that the activation of these associations would imply is false.

Like Peters and Gawronski, we conducted the present research as a direct test of the effect of the validity of the evaluation implied by the activation of evaluative associations on automatic versus deliberate evaluation. However, unlike Peters and Gawronski, we did not manipulate the validity of the co-occurrence information. The co-occurrence information was always true. We manipulated only the validity of the evaluation implied by the activation of the associations that the co-occurrence was expected to form. In the present studies, all the participants learned about four families of creatures. One family started an unpleasant noise, one started a pleasant music, one ended unpleasant noise and one ended pleasant music. That is, we provided co-occurrence information (with what affective stimuli the target co-occurs) and relation information (the relation between the target and the affective stimuli). Using relations instead of false statements allowed us to manipulate only the validity of the evaluation implied by the activation of the associations that the co-occurrence was expected to form, and not the validity of the co-occurrence information itself. The creatures that ended the affective events indeed co-occurred with the events but the relation between the events and the creatures implied that the creatures and the events had the opposite valence. After the learning episode, we measured participants’ deliberate and automatic evaluation (we repeated the same experiment twice, each time with a different automatic evaluation measure).
We expected the co-occurrence between the creatures and the affective events to form associations regardless of the relation. According to the APE model, the activation of that association should later influence automatic evaluation. For instance, upon automatic evaluation of a creature that ended an unpleasant noise, the negative valence of the noise should be activated, causing a more negative automatic evaluation of the creature. In contrast, upon deliberate evaluation people should recognize that the negative valence is activated because the creature ended the unpleasant noise, and not because the creature is negative. That relational information should cause a more positive deliberate evaluation of the creature. In other words, the present research tested whether, as the APE model would predict, relational information has a stronger influence on deliberate evaluation than on automatic evaluation.

Although we designed our research to test the APE model, it is also relevant to Evaluative Conditioning (EC) research. According to De Houwer (2007), EC is a change in the evaluation of a neutral stimulus (the conditioned stimulus, CS) after co-occurring with affective stimuli (the unconditioned stimulus, US). Whereas in most EC studies there is no information about the relation between the co-occurring stimuli, our procedure explains how the stimuli relate to each other. Therefore, the present research can be viewed as a test whether relational information moderates EC. In that respect, our attitude induction is a conceptual replication of a study by Förderer and Unkelbach (2012) who presented CS-US pairs simultaneously with information about whether the CS loves or loathes the US, and measured self-reported evaluation of the CS’s. They found that relational information moderates EC: participants reported preference for the CS that loathed negative US’s over the CS that loathed positive US’s.

Förderer and Unkelbach interpreted their finding as support for the propositional account for EC (De Houwer, 2009; Mitchell, De Houwer, & Lovibond, 2009). That account contends that EC is the result of the formation of propositions (a belief that the CS and the US co-occur) rather than automatic formation of a CS-US associations. According to Förderer and Unkelbach, the fact that relations determined the EC effect in their study supports the assumption that EC is mediated by propositional processes because propositional processes can encode relations whereas associative processes cannot. Our research extends Förderer and Unkelbach’s study by testing whether relational information moderates EC when evaluation is measured indirectly, with automatic evaluation measures. If, as the APE model would predict, automatic evaluation would be sensitive to the CS-US co-occurrence despite the evaluative meaning of the relational information, it might be difficult for the propositional account of EC to explain the results.

**Method**

**Participants**
 Israeli student participants completed the study in separate cubicles. There were 32 participants in Experiment 1 (61% Women, $M_{age} = 23.1$, $SD_{age} = 1.08$), and 59 participants in Experiment 2 (57% Women, $M_{age} = 24.3$, $SD_{age} = 1.80$).

**Materials**
 The target stimuli were 16 drawings of male alien creatures divided equally into four families. Families differed in their color (green, purple, red, and yellow) and their head shape. The affective stimuli during learning were a relaxing musical
melody and a horrifying human scream. The automatic evaluation measures also required evaluative stimuli: 6 positive words (In translation from Hebrew, Experiment 1 used Fun, Happy, Peace, Joy, Love and Wonderful; Experiment 2 used Pleasure, Happy, Peace, Delight, Love and Wonderful) and 6 negative words (Experiment 1: Sorrow, Pain, Failure, Evil, Threat and Bad; Experiment 2: Grief, Pain, Failure, Evil, Threat and Bad).

**Learning**

The learning included 10 melody trials and 10 scream trials, ordered randomly and separated by a soft ticking sound that played for 10–15 seconds. Each trial began with a presentation of the “starting” creature, appearing in silence for 500ms. Next, the auditory stimulus began, playing for a randomly selected duration of 10–30 seconds. The “starting” creature remained on the screen for the first two seconds of the playback. Then the screen remained blank until the “ending” creature appeared for the last two seconds of the auditory stimulus’s playback and remained on the screen for another 500ms of silence. When the screen turned blank, the ticking sound commenced until the next trial started. We told participants in advance that the creatures started and ended the stimulus. We instructed participants to learn which family performed each of the four actions for a later memory test. We counterbalanced between participants which family performed each of the four actions.

**Automatic evaluation**

After the learning task, participants completed the measure of automatic evaluation. Experiment 1 used the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). In the IAT, participants sort stimuli to categories using two computer keys. In the critical blocks, participants respond with the left key to stimuli of two categories (e.g., "Green creatures" and "Good words"), and with the right key to stimuli of two other categories (e.g., "Purple creatures" and "Bad words"). Because "Green creatures" and "Good words" share the same key, people who like the green creatures and dislike the purple creatures respond more quickly. The IAT also included blocks in which green creatures and bad words share the same key. The difference between the average response latencies in these two types of blocks is interpreted as automatic preference.

We measured the automatic preference between each two families that shared the same relation. Participants completed two IATs—one for the "starting" families (comparing the family that started music with the family that started the scream) and one for the "ending" families (comparing the family that ended the music with the family that ended scream). The order of the two IATs was counterbalanced and did not moderate the results. The seven-block IAT followed the procedure described in Nosek, Greenwald, and Banaji (2005). For each of the two IATs, we counterbalanced the order of the blocks such that sometimes the helping creatures shared a key with good words first (i.e., on blocks 3–4), and sometimes the harming creatures shared a key with good words first.

The only difference between the experiments was that Experiment 2 measured automatic evaluation with the Sorting Paired Features task (SPF; Bar-Anan, Nosek, & Vianello, 2009). In the SPF, two pairs of category labels appear in each of the four corners of the screen. In each trial, participants sort pair stimuli (e.g., a green creature and the word wonderful) to one of four category pairs (e.g., green creatures+good
words, green creatures+bad words, purple creatures+good words, purple creatures+bad words). Faster sorting indicated a stronger association between the pair categories. Therefore, a person who prefers purple creatures over green creatures will be faster on trials that present purple creatures with good words or green creatures with bad words than on trials that present purple creatures with bad words or green creatures with good words. We used eight blocks of 40 trials. Four nonconsecutive blocks (e.g., 1, 3, 5, and 7) presented the families that started stimuli and the other four blocks presented the families that ended stimuli. The order of the blocks was counterbalanced between participants and did not moderate the results.

**Deliberate evaluation**

After the automatic evaluation measure, an image of all the creatures in each family was presented (one image at a time), and participants reported on a 9-point scale how much they liked the family (1=dislike strongly, 9=like strongly).

**Memory**

After the rating, participants indicated the action each family performed during the game (in four questions: one family at a time). The response options listed the four possible actions, of which the participants chose one.

**Results**

**Deliberate Evaluation**

The mean ratings of all four families in the two experiments are presented in Table 1. We tested the results with a 2 (Associated valence: positive, negative) x 2 (Validity: true [start], false [end]) repeated measures ANOVA. Both experiments found a main effect for the associated valence, $F(1, 31) = 31.84$, $p < .0001$, $\eta_p^2 = 0.50$, $d = 2.03$, in Experiment 1, and $F(1, 58) = 52.00$, $p < .0001$, $\eta_p^2 = 0.47$, $d = 1.89$, in Experiment 2. That main effect reflected an overall preference for the families that were associated with pleasant music ($Ms = 6.09, 6.21, SDs = 1.21, 1.25$, in Experiments 1 and 2, respectively) over the families associated with unpleasant noise ($Ms = 4.20, 4.70, SDs = 1.28, 1.34$). Peters and Gawronski (2011) found similar main effect for associated valence on deliberate evaluation in all their studies. This suggests that deliberate evaluation is not completely immune to the co-occurrence information that is not valid evaluative evidence. We investigate that sensitivity further in a separate research (Moran, Bar-Anan, & Nosek, 2012).

As predicted by the APE model, the effect of associated-valence was moderated by the validity condition, $F(1, 31) = 44.24$, $p < .0001$, $\eta_p^2 = 0.58$, $d = 2.39$, in Experiment 1, and $F(1, 58) = 88.31$, $p < .0001$, $\eta_p^2 = 0.60$, $d = 2.47$, in Experiment 2. The interaction reflected opposite preferences in the two validity conditions. Deliberate evaluation showed preference for the family associated with positive stimulus over the family associated with negative stimulus in the true validity condition (when the families started the stimulus), $F(1, 31) = 82.95$, $p < .0001$, $\eta_p^2 = 0.72$, $d = 3.27$, in Experiment 1, and $F(1, 58) = 117.71$, $p < .0001$, $\eta_p^2 = 0.66$, $d = 2.85$, in Experiment 2; but the opposite preference for the family associated with negative stimulus over the family associated with positive stimulus in the false
validity condition (when the families ended the stimulus), $F(1, 31) = 5.97$, $p = .02$, \(\eta^2_p = 0.16\), $d = 0.88$, in Experiment 1, and $F(1, 58) = 22.88$, $p < .0001$, \(\eta^2_p = 0.28\), $d = 1.26$, in Experiment 2. In other words, deliberate evaluation always showed preference for the families that helped the participants over the families the harmed them. In Experiment 2 there was also a main effect of validity, $F(1, 58) = 11.47$, $p = .001$, \(\eta^2_p = 0.16\). That main effect reflected an overall preference for the families that stopped stimuli ($M=5.76$, $SD=1.42$) over the families that started stimuli ($M=5.15$, $SD= 1.01$).

### Table 1. Evaluations of the target families (SD in parentheses)

<table>
<thead>
<tr>
<th></th>
<th><strong>Self-reported</strong></th>
<th><strong>Automatic</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive Stimulus</td>
<td>Negative Stimulus</td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Started</td>
<td>7.63 (1.96)</td>
<td>2.25 (1.98)</td>
</tr>
<tr>
<td>Stopped</td>
<td>4.56 (2.18)</td>
<td>6.16 (2.37)</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Started</td>
<td>7.54 (1.86)</td>
<td>2.76 (2.07)</td>
</tr>
<tr>
<td>Stopped</td>
<td>4.88 (1.93)</td>
<td>6.64 (2.07)</td>
</tr>
</tbody>
</table>

*Note. For self-reported attitudes, each cell is the combination of the relation on the left and the stimulus at the top. For instance, the 4.88 at the bottom line is the mean attitude toward the family that ended the pleasant music. For automatic evaluation, positive D scores represent preference for the family associated with positive valence over the family associated with negative valence. The internal consistency is Cronbach’s Alpha, computed from the scores of separate blocks in each task.*

### Automatic Evaluation

Unlike with the self-report measures, there is no reliable method to compute the automatic evaluation of each of the four targets. The only reliable measures were the preferences between each two families that were compared in each IAT or SPF. We computed IAT $D$ scores (Greenwald, Nosek, & Banaji, 2003) in Experiment 1, and SPF $D$ preference scores (Bar-Anan et al., 2009) in Experiment 2, such that scores above zero reflected automatic preference for the family associated with positive valence over the family associated with negative valence. T-tests found no difference between the two validity conditions, $t < 1$, $d = 0.04$, in Experiment 1, and $t < 1$, $d = 0.02$, in Experiment 2. As predicted by the APE model, the automatic evaluation always showed preference for the families that were associated with positive valence. In the *true* validity condition, the IAT preference in Experiment 1 was $M = 0.23$, $SD = 0.42$, $t(31) = 3.02$, $p = .005$, $d = 0.54$, and the SPF preference in Experiment 2 was $M = 0.12$, $SD = 0.41$, $t(58) = 2.25$, $p = .02$, $d = 0.29$. In the *false* validity condition the IAT preference in Experiment 1 was $M = 0.21$, $SD = 0.5$, $t(31) = 2.40$, $p = .02$, $d = 0.42$, and the SPF preference in Experiment 2, was $M = 0.11$, $SD = 0.34$, $t(58) = 2.39$, $p = .02$, $d = 0.32$.

### Direct comparison of the effect of validity on deliberate versus automatic evaluation

We conducted a $2 \times 2$ repeated measures ANOVA on the standardized preference scores (for the family associated with positive valence over the family associated with negative valence).
Following Peters and Gawronski (Peters, personal communication, April 3, 2012), we standardized the rating of each stimulus family separately (e.g., a standardization based on the ratings of the purple family, across participants). Then, we computed preference scores from the standardized explicit scores. For the IAT or SPF, we standardized separately the score for each comparison of the same two family stimuli (e.g., green versus purple). Figure 1 presents the mean standardized scores.

We found a main effect of validity, $F(1, 31) = 27.90, \ p < .0001, \ \eta^2_p = 0.47$, in Experiment 1, and $F(1, 58) = 37.36, \ p < .0001, \ \eta^2_p = 0.39$, in Experiment 2. This effect reflected a stronger preference for the families associated with pleasant music in the true validity condition ($Ms = 1.17, 1.12, SDs = 0.76, 1.12$, in Experiments 1 and 2, respectively) than in the false validity condition ($Ms = -0.04, -0.12, SDs = 0.80, 0.82$). As predicted by the APE model, the effect of validity was moderated by the evaluation condition, $F(1, 31) = 43.72, \ p < .0001, \ \eta^2_p = 0.58$, in Experiment 1, and $F(1, 58) = 46.37, \ p < .0001, \ \eta^2_p = 0.44$, in Experiment 2. This confirmed statistically the previously reported divergence between deliberate and automatic evaluation:

![Figure 1](image-url)
validity affected deliberate evaluation, $F(1, 31) = 44.24, p < .0001, \eta_p^2 = 0.59$, in Experiment 1, and $F(1, 58) = 88.31, p < .0001, \eta_p^2 = 0.60$, in Experiment 2. Automatic evaluation on the other hand, was not affected by the validity condition ($Fs < 1$).

**Memory**

Memory accuracy was high (Experiment 1: 94%, Experiment 2: 95%), and validity had no effect on memory accuracy, $ts < 1$. The results remained significant when including only participants who remembered correctly the roles of all four families.

**General Discussion**

Two experiments found that targets that ended affective events induced opposite deliberate and automatic evaluations. People reported liking of targets that ended an unpleasant noise over targets that ended a pleasant music but the automatic evaluation measures found the opposite results: preference for the targets that ended the pleasant music over the targets that ended the unpleasant noise. These results suggest that co-occurrence has a stronger influence than relational information on automatic evaluation, and that relational information influences deliberate evaluation more than it influences automatic evaluation.

The relational information in our research served as a manipulation of the validity of the evaluation implied by the evaluative associations that the co-occurrence information formed. Therefore, the present results support the APE model’s assumptions (Gawronski & Bodenhausen, 2006, 2011) that the validity of the evaluation implied by the activation of associations influences deliberate evaluation more than it influences automatic evaluation. Previous studies that tested similar questions manipulated the validity of the co-occurrence information (Gregg et al., 2006; Peters & Gawronski, 2011). Participants observed co-occurrence between the target object and affective stimuli, followed by a verbal indication whether the co-occurrence was true information or false. When the validity information was delayed, validity influenced deliberate evaluation more than it influenced automatic evaluation. When validity information was provided immediately, validity had the same strong effect on automatic and deliberate evaluation. Peters and Gawronski (2011) inferred that temporal separation between the co-occurrence information and the information about its validity allows for divergent automatic and deliberate evaluations. Yet, in our research, a single learning episode that allowed quick assessment whether the co-occurrence is valid evidence that the co-occurring stimuli are of the same valence induced opposite automatic and deliberate evaluations.

To reconcile the previous and present results, we suggest that automatic evaluation is formed by co-occurrence information that is considered true at the time of learning, regardless of the evaluative implication of the co-occurrence. If the co-occurrence information is considered false, it is less effective in forming associations in people’s memory. The present results suggest that true co-occurrence between an object and affective stimuli forms associations between the object and the affective stimuli even if people know that the relation between the object and the stimuli implies that they are of opposite valence. Later, the activation of these associations
influences automatic evaluation. Whereas deliberate evaluation corrects for that effect and reflects mostly the actual relation between the target and the associated affective stimulus, automatic evaluation is less sensitive to the relational information.

We caution that our failure to detect any influence of relational information on automatic evaluation is a null effect, and that it is limited by the particular automatic evaluation measures used in this research. The present results were obtained by comparing objects that shared a relation but differed in the valence with which they co-occurred. Perhaps an effect of relational information on automatic evaluation could be detected with other measures or with other comparisons. Finally, it is also possible that concrete differences in procedures and materials caused results different from Peters and Gawronski’s (2011) findings.

**Implications for Evaluative Conditioning**

The present research is relevant to research on EC because we used co-occurrence of the target stimuli with affective stimuli to form evaluative associations. Our results replicated Förderer and Unkelbach’s (2012) findings that relational information reverses the EC effect that is typically observed when no relational information is provided. That is, participants reported preference for a CS that co-occurred with a negative US (the family that stopped unpleasant noise) over the CS that co-occurred with a positive US (the family that stopped pleasant music).

However, our automatic evaluation measures found no evidence for the moderation of EC by relational information. This result is problematic for the propositional account of EC (De Houwer, 2009; Mitchell, De Houwer, & Lovibond) because participants probably did not believe in the proposition that the CS that ended the unpleasant noise is unpleasant, and yet they showed a stronger association between that CS and unpleasant than between the CS and pleasant. Participants probably believed in two propositions regarding the CS’s that ended the US’s: the CS co-occurred with the US, and the CS had the opposite valence of the US. To explain our findings with the propositional account, further research will need to find why the first proposition had a stronger influence on automatic evaluation than the latter.

**Summary**

The present research provided a direct test of the APE model’s main assumption: the validity of the evaluation implied by the activation of evaluative associations has a stronger influence on deliberate evaluation than on automatic evaluation (Gawronski & Bodenhausen, 2006, 2011). We tested that assumption by investigating the influence of relational information on automatic versus deliberate evaluation. The relational information varied the validity of the evaluation implied by the activation of evaluative associations while keeping the validity of the co-occurrence information that formed the associations true. In support of the APE model, we found that relational information had a stronger effect on deliberate than on automatic evaluation.
References


