The Effect of Opposition Relations on Evaluation

Tal Moran and Yoav Bar-Anan

Ben-Gurion University of the Negev, Beer-Sheva, Israel

Brian A. Nosek

University of Virginia and Center for Open Science, USA

We examined the moderating role of the type of relation between target objects and affective stimuli on attitude formation. We focused on the difference between opposition relations – relations that link the target object to stimuli of one valence but outcomes of the opposite valence (e.g., Bob prevented war) and similarity relations (link the target to stimuli and outcomes of the same valence; e.g., Bob brought peace). Experiments 1-4 suggested that opposition relations create less extreme evaluations than similarity relations. Experiment 5 failed to find support for the hypothesis that opposition relations induce less extreme evaluations because they are more difficult to process. Experiment 6 provided evidence that opposition relations induce less extreme evaluations because they form associations between the target object and the opposite valence. Because opposition relations are common in the social environment (e.g., doctors cure cancer, policemen stop crime), these findings have broad practical and theoretical implications.

The Effect of Opposition Relations on Attitude Formation

Forming attitudes often requires learning the relations between the targets and other stimuli. People like policemen for *stopping* criminals and grandmothers for *bringing* presents. People dislike bullies for *taking* lunch money and politicians for *increasing* social conflict. But does it matter what relation led to the inferred evaluation? Is the evaluation of someone who *ends* negative events as positive as the evaluation of someone who *starts* positive events? Is the evaluation of *taking away* pleasant stuff from people as negative as *giving* unpleasant stuff? The present research tested the moderating role of the relation type on the extremity of the formed evaluation.

There are number of ways in which target objects can be related to affective stimuli. We focus on the distinction between relations that indicate that the two related objects hold *similar* valence, and relations that indicate that the two related objects hold *opposite* valence. For simplicity, we call those *similarity* and *opposition* relations. Similarity relations (e.g., *start*, *cause* and *give*) link a target to stimuli of a certain valence and to an outcome of similar valence. For instance, *increase* is a similarity relation because it relates the target (e.g., a politician) to a negative stimulus (e.g., unemployment) and leads to negative outcome (poverty). On the other hand, opposition relations (e.g., *end*, *prevent*, and *take*) link a target to stimuli of a certain valence but to an outcome of the opposite valence. For instance, *decrease* is an opposition relation because it links the target (e.g., a politician) to a negative stimulus (e.g., unemployment) but conveys a positive effect of the target (prosperity). The present research tested the hypothesis that evaluations of targets that have opposition relations with affective stimuli are less extreme than evaluations of targets that have similarity relations with affective stimuli.

Theoretical Background

A fundamental aspect of human cognition is the ability to learn and manipulate concepts defined by systematic relationships among multiple objects (Hummel & Holyoak, 2005). Relational thinking involves the ability to recognize analogies between different situations and to form schemas, or relationally defined concepts (Gick & Holyoak, 1983). Relational thinking underlies cognitive abilities such as the ability to learn and use roles (e.g., Lovett & Anderson, 2005), the ability to comprehend visual scenes (e.g., Green & Hummel, 2003), and the ability to

understand and learn language (e.g., Gentner & Loewenstein, 2002). Research on relational thinking has focused on investigating the human ability to understand relations, how this ability changes with development (e.g., Richland, Morrison, & Holyoak, 2006), and how people represent and reason about different types of relations (Corral & Jones, 2014; Hummel & Holyoak, 2005; Doumas, Hummel, & Holyoak, 2008).

The distinction between similarity and opposition relations resembles the distinction between coordination and opposition relations, in research on operant learning (Dymond et al., 2008; Hayes, Barnes-Holmes, & Roche, 2001; Whelan & Barnes-Holmes, 2004). An assumption of the Relational Frame Theory (RFT; Hayes, et al., 2001) is that relating is an action that involves responding to one event in terms of another, and that humans and non-humans alike can learn to relate stimuli in a number of different ways. The most fundamental type of relational responding according to the RFT is *coordination*. Coordination relations resemble identity, sameness or similarity of stimulus events. Opposition relations, on the other hand, resemble contradiction or contrast. There is evidence that people respond in accordance with similarity and opposition relations (Whelan & Barnes-Holmes, 2004). In the context of evaluative learning, previous research has focused on people's ability to understand relational information, and change their evaluation according to that information (Fiedler & Unkelbach, 2011; Forderer & Unkelbach, 2012; Gawronski & Walther, 2008; Gawronski, Walther, & Blank, 2005).

The Effect of Opposition Relations on Evaluation

At the center of the present research is the hypothesis that opposition relations form less extreme evaluations than similarity relations. We base this prediction on two factors: the link to opposite valence and the relative difficulty in understanding opposition relations.

The link to opposite valence. Opposition relations link the target object to stimuli with valence opposite of the target's real valence, whereas similarity relations link the target object to stimuli of similar valence. There are a few lines of research that suggest that linking objects to affective stimuli leads to an assimilative evaluative effect: the evaluation of the object becomes more similar to the evaluation of the affective stimuli. Research on Evaluative Conditioning (EC) suggests that when a neutral stimulus repeatedly occurs in temporal proximity to an affective stimulus, people's subsequent evaluation of the neutral stimulus becomes more similar to their

evaluation of the affective stimulus (De Houwer, Thomas, & Baeyens, 2001; Walther, Weil, & Dusing, 2011). EC might attenuate the evaluation suggested by opposition relations: people should like physicians because they cure illness, but the co-occurrence between physicians and illness might reduce that liking.

There are also other lines of research with evidence that linking an object to affective stimuli leads to evaluative assimilation. According to the spontaneous trait transference (STT) effect, communicators are often perceived as having traits that they merely describe in others (Carlston & Skowronski, 2005; Skowronski, Carlston, Mae, & Crawford, 1998). For instance, reading a message about the laziness of another person would lead to a negative evaluation of the informer. This effect was explained as the result of an association formed between the trait concepts and the communicator who described the behavior. In a different line of research, people transferred their (dis)liking of one person to another person, only because both were members of the same social group (Ranganath & Nosek, 2008; see also Ratliff, Swinkels, Klerx, & Nosek, 2012). Those lines of evidence suggest that linking an object with affective stimuli leads to an assimilative effect on the evaluation of the object, even if other information suggests that there is no good reason to infer similarity between the object and the affective stimuli. That possibility is important for the case of opposition relations because opposition relations indicate that the target has valence that is opposite of the valence of the affective stimuli linked with the target.

Processing difficulty. The contradiction between a link to stimuli of one valence and to an outcome of the opposite valence might cause difficulty in processing information conveyed with opposition relations. In order to infer the target's valence from an opposition relation, people must reverse the valence of the affective stimuli linked to the target object. A politician whose actions helped to decrease poverty is positive *because* poverty is negative. A politician whose actions helped to increase prosperity is positive *because* prosperity is positive. Inference of the politician's valence might be easier in the latter example because it does not require any reversal. The complexity involved in inference from opposition relations might limit people from reaching strong (or extreme) conclusions regarding the evaluation of the target.

Some support for this reasoning comes from studies about negation. When people learn that previous information is false, they need to reverse that information in order to infer the correct conclusions. Past research demonstrated that, under certain circumstances, processing negations is not as efficient as processing affirmations (Deutsch, Gawronski, & Strack, 2006; Gilbert, 1991; Gilbert, Tafarodi, & Malone, 1993; Mayo, Schul, & Burnstein, 2004). For instance, Deutsch, Gawronski and Strack (2006) found that it takes less time to correctly identify the valence of affirmed (e.g., *a party*) versus negated (e.g., *no party*) versions of positive and negative words. In addition, Gilbert, Tafarodi, and Malone (1993) found that participants who were under cognitive load when they learned that a statement about a target was false formed attitudes that disregarded the statement's falsehood (e.g., disliking Richard despite learning that the statement "Richard is a crook" is false).

Although both inference from false statements (negation) and inference from opposition relation involve valence reversal, they are not identical processes. First, understanding negation often requires people to first believe the statement is true (Gilbert et al., 1993). In contrast, learning opposition relations does not require entertaining the opposite relation (e.g., learning that Richard *stops* crooks does not require first believing that Richard *helps* crooks). Second, when people learn false statements like *Richard is not a crook*, they learn that Richard and crooks are *not* related in reality, whereas learning opposition relations like *Richard stops crooks* suggests that Richard and crooks are, in fact, related. Last, the processing disadvantage of negation disappears when the negation is easily translated to a clear opposite affirmative alternative (Mayo, Schul, & Burnstein, 2004); or when accessible background beliefs suggest that the information presented is not true (Richter, Schroeder, & Wöhrmann, 2009). However, these conditions are not relevant in the case of processing opposition relations.

To summarize, research about negation provides evidence that, under most circumstances, reversal of information consumes cognitive resources and might attenuate evaluative learning, but it does not provide direct evidence that inference from opposition relations is more difficult than inference from similarity relations, or that this difference in processing difficulty has any effect on the extremity of attitudes induced by opposition versus similarity relations.

More support for the hypothesis that opposition relations might form less extreme evaluations because it is more difficult to process opposition relations, comes from Relational Frame Theory (RFT; Hayes, et al., 2001). According to RFT, coordination relations represent the most fundamental type of relational responding, the first relational responding learned during development. The learning of opposition relations on the other hand, develops only after the learning of coordination relations. As such it is possible that the less practiced opposition relations are more difficult to process and might have a weaker impact on the formation of evaluations.

Previous Evidence

The main focus of previous evaluation research on relational learning was the question whether people take relational information into account when they form impressions. Förderer and Unkelbach (2012) found that people prefer targets that like positive stimuli over target that like negative stimuli, and prefer targets that dislike negative stimuli over targets that dislike positive stimuli. Gawronski et al. (2005) found that people prefer targets liked by positive people over targets liked by negative people, and prefer targets disliked by negative people over targets disliked by positive people. More relevant to the present research, both studies found a main effect of co-occurrence, over and above the effect of the specific relation: people reported more liking of targets that were linked with positive stimuli over targets that were linked to negative stimuli[1]. A possible weakness of this past evidence is that these studies used relational qualifiers that have an affective meaning of their own (i.e., *love* is positive and *hate* is negative). Therefore, a rival account for those findings is that negative relations have a weaker effect on attitude formation than positive relations.

More relevant evidence comes from research that used similarity versus opposition relations to examine the sensitivity of automatic versus deliberate evaluation to validity information (Moran & Bar-Anan, 2013). Participants observed target characters that started or ended pleasant or unpleasant sounds. Because *end* is an opposition relation, the link between the characters that ended affective events and the events that they ended was not valid evidence that

---

[1] The information regarding the co-occurrence effect in Förderer and Unkelbach's research was obtained via personal communication (S. Förderer, June, 6, 2013).

they have similar valence. Compatible with the prevalent theoretical notion that validity information has a stronger influence on automatic evaluation than on deliberate evaluation, automatic evaluation was more sensitive than deliberate evaluation to the link between the characters and the events. Specifically, although people reported preference for characters that ended unpleasant events over pleasant events, they showed automatic preference for characters that ended pleasant music over characters that ended unpleasant screams.

In the present context, the important question regarding Moran and Bar-Anan's (2013) research is whether *deliberate* evaluation toward targets that ended affective events was less extreme than toward targets that started affective events. To test that for the present investigation, we examined whether the effect of the target's role (whether the target helped or harmed the participant) on the target's deliberate evaluation was moderated by the relation type. With both relations, we expected more positive evaluation of helpers (characters that started positive events and characters that ended negative events) than harmers (characters that started negative events and characters that ended positive events). However, we expected a stronger effect of character's role when the characters started affective events than when the characters ended events.

Table 1
Experiments 1-2 in Moran and Bar-Anan (2013): Evaluation as a function of target's role and relation type

Target's role		Positive	Negative	Positive	Negative		
Relation type		Similarity	Similarity	Opposition	Opposition	Preference for the helper over the Harmer	
	N					Similarity	Opposition
Exp. 1	32	7.63 (1.96)	2.25 (1.98)	6.16 (2.37)	4.56 (2.18)	5.38 (3.34)	1.59 (3.69)
Exp. 2	59	7.54 (1.86)	2.76 (2.07)	6.64 (2.07)	4.88 (1.93)	4.78 (3.38)	1.76 (2.83)

The average deliberate evaluations of all targets in Moran and Bar-Anan's (2013) two experiments are presented in Table 1. The predicted moderation was strong in both experiments, η_p^2s = 0.50, 0.47, ps <.001, .001, in Experiments 1 and 2, respectively. In Experiments 1 and 2, respectively, the effect of target's role (helper vs. harmer) was stronger when the targets started sounds, η_p^2s = 0.72, 0.66, ps < .001, .001, than when the targets ended sounds, η_p^2s = 0.16, 0.28,

*p*s = .02, .001. In summary, the results from Moran and Bar-Anan's research provide initial evidence in support of our main hypothesis.

The Present Research

The present investigation had two goals. The first goal was to extend previous initial evidence in support of the assumption that opposition relations induce less extreme evaluation than similarity relations. We tested this assumption in Experiments 1-4 using different types of opposition and similarity relations, and found that opposition relations consistently induced less extreme evaluations than similarity relations. The second goal was to evaluate our hypotheses about the factors that contribute to that effect. We first tested the hypothesis that opposition relations induce less extreme attitudes because they are more difficult to process than similarity relations. Experiment 5 tested whether limiting participants' opportunity to learn the evaluative information would increase the effect of relation type on the extremity of the induced preference (i.e., would increase the difference in induction potency between opposition and similarity relations). In Experiment 6, we tested the hypothesis that opposition relations are less potent than similarity relations because opposition relations form a link between the target objects and stimuli of opposite valence. We tested whether increasing participants' focus on the stimuli linked to the targets would increase the effect of relation type on the extremity of the induced evaluation[2].

Experiments 1-3

To test the effect of opposition relations on attitude formation, we exposed participants to four targets that helped or harmed the participants by having similarity or opposition relations with positive or negative stimuli. In each experiment, we presented information about four families of creatures. One family helped the participants by having a similarity relation with a positive stimulus (e.g., started a pleasant stimulus); one family helped participants by having an opposition relation with a negative stimulus (e.g., ended an unpleasant stimulus); one family harmed the participants by having a similarity relation with a negative stimulus (e.g., started an

---

[2] We report all data exclusions, manipulations, and measures, and how we determined our sample sizes. In all experiments, decisions to stop collecting data did not depend on the obtained results. To see the materials and data of the whole project (Experiments 1-6) visit https://osf.io/cegda/.

unpleasant stimulus); and one family harmed participants by having an opposition relation with a positive stimulus (e.g., ended a pleasant stimulus). After learning this information, participants evaluated each family.

We tested whether opposition relations induce less extreme attitudes than similarity relations. Each study compared one action that implied a similarity relation (*start*, *give*, and *allow*, in Experiments 1–3, respectively) to one action that implied an opposition relation (*end*, *take*, and *prevent)*. For instance, we compared the preference for a family that ended the presentation of unpleasant stimuli over a family that ended the presentation of pleasant stimuli (i.e., a preference induced by an opposition relation) to the preference for a family that started the presentation of pleasant stimuli over a family that started the presentation of unpleasant stimuli (i.e., a preference induced by a similarity relation). If opposition relations induce less extreme attitudes than similarity relations, then the preference that was induced by an opposition relation should be smaller than the preference that was induced by a similarity relation.

Method

Participants. The participants in Experiments 1-3, and in Experiment 5, volunteered to participate on the internet at the Project Implicit research website (Nosek, 2005). Experiments 1-3 were online at the same time, but participating in one made participants ineligible to be randomly assigned to the others. The experiments were available for participants for a week— enough time to achieve more than 95% statistical power for detecting effects considerably smaller than the moderation effect found in previous research (Moran & Bar-Anan, 2013). In all the experiments, we included only participants who completed all the dependent measures. In Experiment 1, these were 68 (65% women, M_{age} = 26.2, SD_{age} = 10.3) participants, in Experiment 2, these were 39 (59% women, M_{age} = 28.9, SD_{age} = 11.7) participants, and in Experiment 3, these were 50 (75% women, M_{age} = 23.4, SD_{age} = 9.7) participants[3].

---

[3] The studies were removed from the website after a week because their overall completion rate (53%, 157/296) was below 60%, the minimum rate allowed in Project Implicit. Thus, the decision to stop data collection did not depend on the obtained results. Because the participants are volunteers, completion rate in Project Implicit tends to be low, typically 60%-65%. We suspect that the reason for the lower completion rate in the present studies is that, unlike most studies in Project Implicit, it did not involve prejudice toward known social groups.

Materials. The target stimuli were 16 drawings of alien creatures divided equally into four families (taken from Moran & Bar-Anan, 2013). Families differed in their color (green, purple, red, and yellow) and their head shape. In Experiment 1, the positive stimulus was a drawing of gold bars presented alongside images of puppies and smiling babies, and the negative stimulus was a drawing of garbage piles presented alongside images of threatening dogs and crying babies. In Experiments 2 and 3, the affective stimuli were images of puppies and cockroaches.

Learning procedure. The experiments always started with the learning procedure. Participants played a game in which they met alien creatures from four different families and had the goal of learning what action is typical of each family. On each trial of the learning task, a creature appeared in spatiotemporal proximity to a positive or a negative stimulus and performed an action that either helped or harmed the participant. The creature helped or harmed the participants by having similarity or opposition relation with the affective stimulus. Each creature belonged to one of the families. Participants knew in advance that each family always performed exactly the same action (e.g., gave puppies to the participant), and their task was to learn which action each family performed. The family-action match was counterbalanced between participants.

In Experiment 1, the similarity relation was *starts* and the opposition relation was *ends*. The four actions were *starting the appearance of gold bars*, *starting the appearance of garbage piles*, *ending the appearance of gold bars*, and *ending the appearance of garbage piles*. The procedure consisted of gold trials and garbage trials. Gold trials started with the appearance for 1500ms of a creature from the "start-gold" family. Then the gold-bars image started appearing on the screen for 3 to 5 (the number varied between trials) consecutive one-second presentations, separated by a blank screen for 200ms. The presentation of the gold bars continued until a member of the "end-gold" family appeared (for 1500ms). Garbage-piles trials were identical but with the garbage-piles image instead of the gold-bars image. After the appearance of an "end" creature, the screen was blank for two seconds, and then a "start" creature appeared, starting the next trial. The learning phase contained two blocks of 12 randomly ordered trials (6 gold trials and 6 garbage trials).

In Experiment 2, the similarity relation was *give*s and the opposition relation was *take*s. The actions were *giving puppies*, *giving cockroaches*, *taking puppies away*, and *taking cockroaches away*. Each trial started with a one-second fixation image, followed by a three-second presentation of a creature on the left side of the screen, a puppy or a cockroach on the right side of the screen, and the word "gives" or "takes" between them (e.g., *creature gives puppy*). Each family appeared 12 times (randomized trial order).

Experiment 3 replicated Experiment 2 with different relations: *allows* and *prevents*. The actions were: *allow the participant to get puppies*, *allow the participant to get cockroaches*, *prevent the participants from getting puppies*, and *prevent the participants from getting cockroaches*. Instead of "gives" and "takes," the text between the creature and the object was "allows you to get" or "prevents you from getting".

Evaluation. After the learning task, participants reported on a 9-point scale how much they liked each family of creatures (1=dislike strongly, 9=like strongly).

Memory. After the evaluation, participants indicated the action each family performed during the game with four questions, one family at a time. The response options listed the four possible actions, of which the participants chose one.

Design. The design in each experiment was 2 (relation type: similarity, opposition; within participants) X 2 (target's role: positive [helped participants], negative [harmed participants]; within participants).

Results and Discussion

Evaluation. The upper part of Table 2 presents the evaluation scores for each target type, for each Experiment. We submitted the evaluation scores to a 2 (relation type: similarity, opposition) x 2 (target's role: positive [helped participants], negative [harmed participants]) ANOVA. The ANOVA showed a main effect of target's role in all three experiments: Experiment 1: $F(1, 67) = 84.55$, $p < .001$, $\eta_p^2 = .55$; Experiment 2: $F(1, 38) = 53.83$, $p < .001$, $\eta_p^2 = .58$; Experiment 3: $F(1, 49) = 143.38$, $p < .001$, $\eta_p^2 = .74$. That main effect reflected more liking of the targets that helped participants ($Ms = 4.68, 4.77, 5.08$, $SDs = 0.98, 0.97, 1.07$, in Experiments 1, 2 and 3, respectively), than of targets that harmed participants ($Ms = 2.79, 2.36, 2.04$, $SDs = 1.23, 1.28, 0.99$). In Experiment 2, there was also main effect of relation type, $F(1,$

38) = 7.14, p = .011, η_p^2 = .15, reflecting more liking of the targets that gave stimuli (M = 3.85, SD = 0.74) than of targets that took away stimuli (M = 3.28, SD = 0.90).

Table 2
Experiment 1-3: Evaluation as a function of target's role and relation type

Target's role		Positive	Negative	Positive	Negative		
Relation type		Similarity	Similarity	Opposition	Opposition	Preference for the helper over the Harmer	
Experiment 1	N					Similarity	Opposition
All participants	68	4.87 (1.18)	2.65 (1.54)	4.50 (1.18)	2.94 (1.38)	2.22 (2.07)	1.56 (1.87)
Perfect memory	56	4.95 (1.23)	2.52 (1.53)	4.59 (1.17)	2.73 (1.27)	2.43 (2.06)	1.86 (1.89)
Perfect comprehension	33	5.39 (0.79)	2.00 (1.06)	5.06 (0.75)	2.24 (1.06)	3.39 (1.64)	2.82 (1.55)
Experiment 2							
All participants	39	5.21 (0.95)	2.49 (1.50)	4.33 (1.53)	2.23 (1.37)	2.72 (2.04)	2.10 (2.27)
Perfect memory	26	5.19 (0.98)	2.31 (1.38)	4.46 (1.53)	2.00 (1.20)	2.88 (2.08)	2.46 (2.23)
Perfect comprehension	21	5.33 (0.86)	1.81 (0.98)	4.90 (1.18)	1.76 (0.89)	3.52 (1.75)	3.14 (1.90)
Experiment 3							
All participants	50	5.28 (1.14)	2.00 (1.26)	4.88 (1.22)	2.08 (1.07)	3.28 (1.94)	2.80 (1.96)
Perfect memory	35	5.40 (0.95)	1.77 (1.21)	5.00 (1.03)	2.06 (0.97)	3.63 (1.77)	2.94 (1.70)
Perfect comprehension	31	5.52 (0.85)	1.42 (0.67)	5.06 (1.03)	1.90 (0.79)	4.10 (1.22)	3.16 (1.49)

Notes. Perfect comprehension: showed perfect memory and preference for each of the positive targets over each of the negative targets.

Most importantly, in all three experiments the effect of target's role was further moderated by relation type, Experiment 1: $F(1, 67)$ = 7.25, p = .009, η_p^2 = .09, 90% CI [.01, .21]; Experiment 2: $F(1, 38)$ = 8.35, p = .006, η_p^2 = .18, 90% CI [.03, .34]; Experiment 3: $F(1, 49)$ = 5.02, p = .029, η_p^2 = .09, 90% CI [.004, .23]. These interactions reflected a stronger effect of target's role in the similarity relation condition, Experiment 1: $F(1, 67)$ = 78.09, p < .001, η_p^2 = .53; Experiment 2: $F(1, 38)$ = 69.34, p < .001, η_p^2 = .64; Experiment 3: $F(1, 49)$ = 143.19, p < .001, η_p^2 = .74, than in the opposition relation condition, Experiment 1: $F(1, 67)$ = 47.16, p < .001, η_p^2 = .41; Experiment 2: $F(1, 38)$ = 33.50, p < .001, η_p^2 = .46; Experiment 3: $F(1, 49)$ = 102.17, p < .001, η_p^2 = .67.

Memory. The results so far support our assumption that opposition relations induce less extreme evaluation than similarity relations. According to the *learning-difficulty* hypothesis, one factor that contributes to this effect is that it is more difficult to understand information in the form of opposition relations than in the form of similarity relations. In search for support of that assumption we tested whether people showed less accurate memory of the creatures' roles when the roles involved an opposition relation with the affective stimuli than when the roles involved a similarity relation.

84% of the participants in Experiment 1, 67% of the participants in Experiment 2, and 71% of the participants in Experiment 3 had accurate memory regarding the action each family performed during the game. We conducted a logistic regression analysis predicting memory accuracy from the relation type and target's role. Contrary to our prediction, in all three experiments, all the effects were small (all $zs < 1.79$) and none were reliably larger than zero (all $ps > .07$).

Further, when we repeated the evaluation analyses only with participants who showed perfect memory of the roles, relation type still moderated evaluation extremity with similar effect sizes to those found with the whole sample (see Table 2, and full results in the online supplement). The moderation effects among participants who showed excellent comprehension of the evaluative information go against the hypothesis that opposition relations were less effective than similarity relations in inducing attitudes because people were less likely to understand information conveyed with opposition relations than information conveyed with similarity relations. We revisit that hypothesis in Experiment 5.

Experiments 1-3 provide strong support for the hypothesis that opposition relations form less extreme evaluations than similarity relations. For the sake of generality, in Experiment 4, we tested that hypothesis with a different procedure – one that does not involve fictional creatures as targets, presents no repetitions, and provides the behavioral information at the same time of judgment.

Experiment 4

Participants read two behaviors performed by each of 16 target persons, and evaluated the person based on the behaviors. Four persons performed positive behaviors described with

similarity relations (e.g., *Increased the company's profits*), four persons performed negative behaviors described with similarity relations (e.g., *Increased the company's debts*), four persons performed positive behaviors described with opposition relations (e.g., *Decreased the company's debts*), and four persons performed negative behaviors described with opposition relations (e.g., *Decreased the company's profits*).

Method

  Participants. 49 student participants in an Israeli university completed the study in cubicles (69% women, $M_{age} = 23.98$, $SD_{age} = 2.18$). We do not remember our sample size plans. The decision to stop the experiment did not depend on the results.

  Materials. We used 32 behaviors. Half of the 32 behaviors had a positive outcome, and half had a negative outcome. Half of the negative and half of the positive behaviors described a similarity relation between the man and an affective stimulus, and the other half described an opposition relation between the man and an affective stimulus. Sixteen American male names were paired with the 32 behaviors: two behaviors of the same outcome and the same relation-type described each man. We used American names to minimize prior experience with the names among this Israeli sample. Table 3 presents translation of the behaviors from Hebrew.

  Procedure. In each trial of the task, participants saw a man's name, and two behaviors printed below the name. At the bottom of the screen the question "*How much do you like this man?*" appeared with a scale of 1 (Dislike strongly) to 9 (Like strongly). In other words, rating occurred when all the information was available on the same screen. After the participant responded, the next trial started.

  The task had four blocks. Block 1 was a practice block with four trials (two men were present with positive behavior and two with negative behaviors). Block 2 had 16 trials. Each trial showed one man with two behaviors of the same type (e.g., described opposition relations with positive stimuli). Four trials presented positive behaviors that described similarity relations with positive stimuli, four trials presented negative behaviors that described similarity relations with negative stimuli, four trials presented positive behaviors that described opposition relations with negative stimuli, and four trials presented two negative behaviors that described opposition relations with positive stimuli. The eight behaviors of each of the four types were randomly

assigned to the four relevant trials (i.e., behaviors were paired randomly). The order of the 16 trials was random. Blocks 3 and 4 showed the same trials as Blocks 1 and 2, to increase measurement reliability. We counterbalanced between participants the assignment of names to the behavior types.

Table 3

Experiment 4: The behaviors presented in the experiment as a function of behavior type.

Similarity relation with positive stimuli	Similarity relation with negative stimuli	Opposition relation with positive stimuli	Opposition relation with negative stimuli
Started a fun game for the neighborhood children next to his home	Started a big fight between the neighborhood children next to his home	Ended the fun game for the neighborhood children next to his home	Ended the big fight between the neighborhood children next to his home
Renewed the relationship with an acquaintance who was previously nice to his family	Renewed the relationship with an acquaintance who was previously mean to his family	Ended the relationship with an acquaintance who was previously nice to his family	Ended the relationship with an acquaintance who was previously mean to his family
Turned on the joyful music because it was pleasant to his wife.	Turned on the annoying music because it was unpleasant to his wife.	Turned off the joyful music although it was pleasant to his wife.	Turned off the annoying music although it was unpleasant to his wife.
Drastically increased the profits of his company	Drastically increased the debts of his company	Drastically reduced the profits of his company	Drastically reduced the debts of his company
Gave a nice toy to the baby	Gave a dangerous toy to the baby	Took away a nice toy from the baby	Took away a dangerous toy from the baby
Led to his son's success in school	Led to his son's failure in school	Prevented his son's success in school	Prevented his son's failure in school
Enrolled his children to the best school in the neighborhood	Enrolled his children into the worst school in the neighborhood	Pulled his children from the best school in the neighborhood	Pulled his children from the worst school in the neighborhood
Added fun activities to the classwork	Added boring activities to the classwork	Removed the fun activities from the classwork	Removed the boring activities from the classwork

Notes. The behaviors are translated from Hebrew.

Design. The design was 2 (relation type: similarity, opposition; within participants) X 2 (target's role: positive, negative; within participants).

Results and Discussion

Eight trials measured evaluation for each of the four conditions. We averaged the ratings for each condition ($\alpha > .86$). Table 4 presents those scores. A 2 (relation type: similarity, opposition) x 2 (target's role: positive, negative) ANOVA on the evaluation scores found a strong main effect of target's role, $F(1, 48) = 314.86$, $p < .001$, $\eta_p^2 = .86$, reflecting more liking of the positive targets ($M = 7.15$, $SD = 1.08$) than of the negative targets ($M = 2.72$, $SD = 0.90$). This strong effect is not surprising given that the information was available to the participants

during the evaluation. There was also an unexpected main effect of relation type, $F(1, 48) = 8.16$, $p = .005$, $\eta_p^2 = .15$, reflecting more liking of the targets that were described with opposition relations ($M = 5.03$, $SD = 0.53$) than of targets that were described with similarity relations ($M = 4.84$, $SD = 0.53$).

Most importantly, the effect of target's role was moderated by relation type, $F(1, 48) = 18.83$, $p < .001$, $\eta_p^2 = .28$, 90% CI [.11, .42]. As expected, the interaction reflected a stronger effect of target's role when the targets were described with similarity relations, $F(1, 48) = 328.46$, $p < .001$, $\eta_p^2 = .87$, than when the targets were described with opposition relations, $F(1, 48) = 265.73$, $p < .001$, $\eta_p^2 = .84$. Despite having all of the information and plenty of time to form and report their evaluations, participants nonetheless showed reliably weaker preferences when the information involved opposition relations than when it involved similarity relations.

Table 4
Experiment 4: Evaluation as a function of target's role and relation type

Target's role	Positive	Negative	Positive	Negative	Preference for the helper over the Harmer	
Relation type	Similarity	Similarity	Opposition	Opposition	Similarity	Opposition
	7.18 (1.14)	2.49 (0.95)	7.11 (1.09)	2.96 (0.98)	4.69 (1.81)	4.15 (1.78)

In Experiments 1-4, across different pairs of relations, the similarity relations were always more effective than opposition relations in inducing preferences for positive over negative targets. These results replicate and extend results found with a different paradigm (Moran & Bar-Anan, 2013). With these results we are now more confident that relations that indicate that the target object is related to stimuli of opposite valence induce less extreme attitudes than relations that indicate that the target object is related to stimuli of similar valence. In the next two experiments, we tested possible moderators of this effect, as a method to uncover factors that might underlie the effect.

The fact that Experiments 1-3 found the effect of opposition relations even among participants who remembered the relational information perfectly provides evidence against the hypothesis that learning difficulty is behind the effects found in those experiments. These people showed no indication they misunderstood the information in any of the conditions, and still

showed less extreme evaluations after learning opposition relations. Moreover, Experiment 4 showed the effect of opposition relations even when the relational information was available to the participants during the evaluation. These results indicate that learning difficulty is probably not the *only* factor that contributes to our finding. Yet, learning difficulty could still influence the effect. Therefore, as a different strategy for finding evidence in support of the learning difficulty account, Experiment 5 tested whether limiting participants' opportunity to learn the evaluative information would increase the effect found in the previous experiments.

Experiment 5

Experiment 5 replicated Experiment 2, adding a manipulation of the number of trials that presented the information about each target. If learning difficulty contributes to the effects observed in Experiments 1-4, then the effect of opposition relations on deliberate evaluation should increase as the number of learning trials decreases.

Method

Participants. We planned to collect at least 600 participants for high statistical power (we did not conduct power estimates in advance). We inadvertently obtained many more participants before stopping collection and analyzing the data. 891 participants completed all the measures (55% women, $M_{age} = 29.18$, $SD_{age} = 12.09$).

Learning procedure. The learning procedure was identical to Experiment 2. The only change was that we manipulated between participants the number of learning trials. Each of the four families appeared in 3 trials, 6 trials, or 12 trials.

Evaluation. After the learning task, participants reported on an 11-point scale how much they liked each family of creatures (1=dislike strongly, 11=like strongly)[4].

Memory. The memory measure was identical to the measure used in Experiments 2.

---

[4] For exploratory reasons, all participants completed a Brief IAT (Sriram & Greenwald, 2009) after the deliberate evaluation measure. The results were very similar to those found with deliberate evaluation (see full details in https://osf.io/zji9b/).

Design. The design was 3 (number of learning trials: 3, 6, 12; between participants) X 2 (relation type: similarity, opposition; within participants) X 2 (target's role: positive, negative; within participants).

Results and Discussion

Memory. 53% of the participants in the 3 trials condition, 62% of the participants in the 6 trials condition, and 77% of the participants in the 12 trials condition displayed a perfectly accurate memory regarding the action each family performed during the game. We conducted a logistic regression analysis predicting memory accuracy from the relation type, target's role and number of learning trials (represented by two dummy variables). Number of learning trials provided the expected effect: memory accuracy was better in the 12-trials condition ($M = 0.90$, $SD = 0.31$) than in the 6-trial condition ($M = 0.81$, $SD = 0.28$), $b = 0.70$, $p < 0.001$, which was better than the memory accuracy in the 3-trials condition, ($M = 0.74$, $SD = 0.31$), $b = -0.38$, $p = 0.008$. These differences indicate that we manipulated learning difficulty successfully.

The logistic regression also revealed a significant effect of relation, $b = 0.75$, $p < .001$, reflecting better memory for roles involving similarity relations ($M = 0.84$, $SD = 0.28$) than for roles involving opposition relations ($M = 0.79$, $SD = 0.33$). The effect of relation was moderated by target's role, $b = -0.71$, $p < .001$, reflecting an effect of relation only for positive creatures, $b = 0.73$, $p < .001$, but not for negative creatures, $b = 0.03$, $p = .64$. These results provide some support for the assumption that opposition relations are more difficult to learn than similarity relations. We also found that, overall, memory was more accurate for negative than for positive roles, $b = 0.95$, $p < .001$.

Evaluation. Table 5 presents the evaluation scores for each target type, by number of trials. We submitted the evaluation scores to a 3 (number of learning trials) x 2 (relation type) x 2 (target's role) ANOVA. The ANOVA found a main effect of target's role, $F(1, 888) = 1213.21$, $p < .001$, $\eta_p^2 = .57$, reflecting more liking of the targets that helped participants ($M = 8.58$, $SD = 2.10$), than of targets that harmed participants ($M = 3.96$, $SD = 2.37$). There was also a main effect of relation type, $F(1, 888) = 50.51$, $p < .001$, $\eta_p^2 = .05$, reflecting more liking of targets that gave stimuli ($M = 6.52$, $SD = 1.34$), than of targets that took away stimuli ($M = 6.02$, $SD = 1.58$). We also found an interaction between the number of learning trials and role, $F(2, 888) = 5.81$, p

$= .003$, $\eta_p{}^2 = .01$. The effect of role was the strongest in the 12 learning trials condition, $F(1, 297)$ $= 524.19$, $p < .001$, $\eta_p{}^2 = .63$, weaker in the 6 learning trials condition, $F(1, 302) = 429.90$, $p <$ $.001$, $\eta_p{}^2 = .58$, and weakest in the 3 learning trials condition, $F(1, 289) = 285.77$, $p < .001$, $\eta_p{}^2 =$ $.49$. The fact that the effect of role increased as the number of learning trials increased is more evidence that we successfully manipulated learning difficulty.

Replicating the results of Experiments 1-4, we found significant interaction between target's role and relation type, $F(1, 888) = 106.25$, $p < .001$, $\eta_p{}^2 = .10$, 90% CI [.07, .13]. The interaction reflected stronger effect of role in the similarity relation condition, $F(1, 890) =$ 1388.60, $p < .001$, $\eta_p{}^2 = .60$, than in the opposition relation condition, $F(1, 890) = 682.58$, $p <$ $.001$, $\eta_p{}^2 = .43$. Most importantly, in contrast to the learning difficultly hypothesis, the number of learning trials did not moderate the interaction between role and relation type, $F < 1$, $\eta_p{}^2 < .001$. Further, if there is any effect at all, it was opposite to the *learning-difficulty* hypothesis: the interaction between relation type and target's role *decreased* when the number of trials decreased: $\eta_p{}^2 = .13$ with 12 learning trails, $\eta_p{}^2 = .12$ with 6 learning trials, and $\eta_p{}^2 = .08$ with 3 learning trials[5].

Table 5
Experiment 5: Evaluation as a function of target's role, relation type, and number of learning trials

Target's role		Positive	Negative	Positive	Negative		
Relation type		Similarity	Similarity	Opposition	Opposition	Preference for the helper over the Harmer	
	N					Similarity	Opposition
3 trials	290	8.84 (2.44)	4.17(2.83)	7.69 (2.89)	4.37 (2.87)	4.67 (4.36)	3.32 (4.87)
6 trials	303	9.19(2.08)	3.80(2.75)	8.07 (2.74)	3.87 (2.68)	5.40 (4.19)	4.19 (4.46)
12 trials	298	9.40 (2.09)	3.69(2.67)	8.28 (2.68)	3.85 (2.70)	5.71 (4.04)	4.42 (4.26)

The results of Experiment 5 suggest that learning difficulty does not contribute to the effect of relation type on evaluation extremity. Although we found some evidence that the learning of opposition relations was more difficult than the learning of similarity relations (reflected by worse memory accuracy for opposition relations), we could not find evidence that

---

[5] As in Experiments 1-3, we repeated the evaluation analyses with participants who showed perfect memory of families' roles. The results were similar: number of learning trials did not moderate the effect of opposition relation on evaluation (see full results in the online supplement).

this difficulty is the reason that opposition relations attenuated the induced evaluation. When participants had fewer trials for learning the relational information, the difference between the evaluations induced by similarity versus opposition relation did not increase.

In Experiment 6 we turned our attention to the other factor that, according to our hypothesis, attenuates the effect of opposition relations on evaluation extremity. That factor is the association between the target object and the affective stimulus to which it is related. According to that hypothesis, opposition relations link the target object to stimuli of opposite valence. Upon evaluation, the valence of those stimuli is activated and pushes the evaluation away from the target's valence that was inferred from the relation. Experiment 6 tested this hypothesis.

Experiment 6

In Experiment 6 we manipulated participants' focus during the learning of similarity and opposition relations. We previously used this method to examine the effect of associative information on automatic evaluation (Moran, Bar-Anan, & Nosek, 2015). Participants either focused on the stimuli that co-occurred with the targets or on the outcome of the relation between the targets and the stimuli. If one reason for the attenuated effect of opposition relations on evaluation is the valence of the co-occurring stimuli, focus on the co-occurring stimuli should further attenuate the effect of opposition relations on evaluation, whereas focus on the outcome inferred from the relation should weaken the attenuation (i.e., cause more extreme evaluation).

Method

Participants. 73 student participants in an Israeli university completed the study in cubicles (82% women, $M_{age} = 22.92$, $SD_{age} = 1.28$). We planned to run 24 participants per condition, based on the effect sizes found with a similar study in our lab that measured automatic evaluation (Moran, Bar-Anan, & Nosek, 2015). We inadvertently collected data from one participant more than planned.

Materials. Target stimuli were identical to the targets used in Experiments 1-3 and 5. The affective stimuli during learning were a relaxing musical melody and an unpleasant human scream (Moran and Bar-Anan, 2013).

Design. The design was 3 (focus: inference, co-occurrence, no-focus; between participants) X 2 (relation type: similarity, opposition; within participants) X 2 (target's role: positive, negative; within participants).

Learning procedure. Instructions informed participants that they would listen a few times to two different sound segments and that four families would each have one constant role throughout the task: starting one of the segments or ending one of the segments. We counterbalanced between participants which family performed each of the four roles. In the *outcome focus* condition we instructed participants to learn whether each family is good or bad for them. In the *co-occurrence focus* condition we instructed participants to learn to which sound each family was related. There were no specific instructions in the control condition.

The learning included 12 music (positive sound) trials and 12 scream (negative sound) trials, separated by a soft ticking sound that played for 10–15 seconds. Each trial in the task began with a presentation of the "starting" creature, appearing in silence for 500ms. Next, the auditory stimulus began, playing for a randomly determined duration of 10–30 seconds. The "starting" creature remained on the screen for the first two seconds of the playback. Then the screen turned 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.

The 24 trials were divided to four blocks, each with three trials for each sound. After each block, participants completed a questions block, in which they answered one question about each of the four families. In the *focus on inference* condition, the question was "did this family help you or harm you?", and participants could choose "this family helped me" or "this family harmed me". In the *co-occurrence focus* condition the question was "was this family related to positive or negative sound?" and participants could choose "this family was related to positive sound" or "this family was related to negative sound". Feedback at the end of the trial indicated whether the answer was correct. There were no questions blocks in the control condition.

Evaluation. After the learning task, participants reported on a 9-point scale how much they liked each creature in a total of 16 trials – four creatures from each family (1=dislike strongly, 9=like strongly).

Memory. After the evaluation, 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 and Discussion

Memory. 100% of the participants in *inference focus* condition, 100% of the participants in *co-occurrence focus* condition, and 91% of the participants in the control condition displayed a perfectly accurate memory regarding the action each family performed during the task. The high accuracy indicates that the focus conditions did not impair people's ability to learn what action each family performed throughout the task.

Evaluation. Table 6 presents the evaluation scores for each target type, for each focus condition. We submitted the evaluation scores to a 3 (focus) x 2 (relation type) x 2 (target's role) ANOVA. A main effect of target's role, $F(1, 70) = 186.10$, $p < .001$, $\eta_p^2 = .72$, reflected more liking of targets that helped participants ($M = 6.71$, $SD = 1.57$), than of targets that harmed participants ($M = 3.41$, $SD = 1.38$). The role effect was moderated by focus condition, $F(2, 70) = 16.59$, $p < .001$, $\eta_p^2 = .32$, reflecting a strong role effect in the *inference focus* condition, $F(1, 23) = 137.61$, $p < .001$, $\eta_p^2 = .85$, weaker in the control condition, $F(1, 23) = 56.41$, $p < .001$, $\eta_p^2 = .71$, and the weakest in the *co-occurrence focus* condition, $F(1, 24) = 16.68$, $p = .0004$, $\eta_p^2 = .41$.

Table 6
Experiment 6: Evaluation as a function of target's role, relation type and focus condition

Target's role		Positive	Negative	Positive	Negative		
Relation type		**Similarity**	**Similarity**	**Opposition**	**Opposition**	**Preference for the helper over the Harmer**	
Focus	**N**					**Similarity**	**Opposition**
Outcome	24	8.32 (1.15)	1.76 (1.52)	7.07 (1.97)	3.57 (1.53)	6.56 (2.30)	3.50 (2.62)
No Focus	24	7.72 (1.56)	2.03 (1.79)	5.57 (2.21)	4.63 (1.66)	5.69 (2.83)	0.95 (2.76)
Co-occurrence	25	7.07 (1.34)	2.83 (1.61)	4.57 (1.96)	5.59 (1.43)	4.24 (2.50)	-1.02 (2.61)

Replicating the previous finding, the interaction between target's role and relation type was significant, $F(1, 70) = 139.03$, $p < .001$, $\eta_p^2 = .66$, 90% CI [.55, .73]. The interaction

reflected stronger effect of role in the *similarity relation* condition, $F(1, 72)= 301.32$, $p < .001$, $\eta_p^2 = .80$, than in the *opposition relation* condition, $F(1, 72)= 8.70$, $p = .0043$, $\eta_p^2 = .10$. Most importantly, supporting our hypothesis, focus further moderated the interaction between target's role and relation type, $F(2, 72)= 3.22$, $p = .04$, $\eta_p^2 = .08$, 90% CI [.0008, .17][6]. As predicted, the moderation by relation type was stronger in the *co-occurrence focus* condition, $F(1, 24)= 64.99$, $p < .001$, $\eta_p^2 = .73$, 90% CI [.53, .80], than in the *outcome focus*, $F(1, 23)= 34.03$, $p < .001$, $\eta_p^2 = .59$, 90% CI [.34, .71], and than in the *control* condition, $F(1, 23)= 42.99$, $p < .001$, $\eta_p^2 = .50$, 90% CI [.41, .75]. To illustrate that important effect, compatible with the pattern we predicted, the difference between the preferences induced with similarity versus opposition relation was $M = 5.26$, $SD = 3.26$, in the *co-occurrence focus* condition, $M = 4.74$, $SD = 3.54$, in the control condition, and $M = 3.06$, $SD = 2.57$, in the *outcome focus* condition. A focus on the *outcome* linked to the targets decreased the relation type effect, whereas a focus on the *stimuli* linked to the targets increased the effect. Interestingly, the effect of the linked valence was so strong in the *co-occurrence focus* condition that the family that ended pleasant music was rated more positively than the family that ended unpleasant screams (Table 6).

These results support the hypothesis that the valence of the linked stimuli contributes to the attenuation of the evaluation induced by opposition relations in comparison to the evaluation induced by similarity relations. Moreover, in this experiment, focus on co-occurrence during learning of opposition relations resulted in preference for the family that harmed the participants over the family that helped the participants. This means that if people focus on the co-occurrence when they learn relational information in the form of opposition relations, they might form deliberate evaluations opposite of the target's effect on them. Importantly, this effect was found despite participants' perfect memory in the co-occurrence focus condition regarding the relational information (the role of each of the four families during the learning task).

General Discussion

Human understanding is enriched by the ability to specify a wide range of relations between objects. Knowing that a person is linked to sorrow is of little value without relational

---

[6] Although barely significant, we consider these results reliable because we found a similar effect ($\eta_p^2 = .06$, $p = .002$) in a previous experiment, in which the self-report ratings were an exploratory measure (see supplement materials of Moran et al., 2015).

information: Does the person cause sorrow? End sorrow? Is filled with sorrow? Is unable to feel sorrow? There are many possible relations between entities, but not much theory and evidence on how different types of relations influence evaluative learning. In six experiments, we identified two types of relations that have different effects on evaluation. Relations that link a target to stimuli of a certain valence but to an outcome of the opposite valence (opposition relations) induce less extreme evaluations than relations that link a target to stimuli of a certain valence and to an outcome of the same valence (similarity relations).

We hypothesized two (non-competing) reasons for this effect, and found evidence for one of them. One hypothesis was that the effect is due to learning difficulty: it is more difficult to process information in the form of opposition relations than in the form of similarity relations. When information is hard to process, it might have a less extreme effect on learning, especially if some people fail to properly understand the implications of the information. However, we found very little evidence that opposition relations were more difficult to comprehend than similarity relations: Only in Experiment 5 (but not in the other experiments) participants' memory of information provided in the form of similarity relation was more accurate than their memory of information provided in the form of opposition relations. Further, the effect of relation type on evaluation extremity did not decrease among the subsamples of participants who showed strong evidence that they understood the relational information. Finally, the effect was found even when all the relevant information was available to participants at the time of evaluation (Experiment 4).

In Experiment 5 we manipulated learning difficulty directly by varying the number of learning trials. We reasoned that if opposition relations are less potent in attitude induction because they are more difficult to comprehend, their relative potency would further decrease when participants have only three or six learning trials, rather than 12. We found strong evidence that overall learning was indeed more difficult when participants had fewer trials to learn: memory accuracy decreased, and so did the overall induced preferences (Table 5). However, the effect of relation type—the difference in evaluation extremity induced by opposition versus similarity relations—did not increase. Post-hoc power analysis reveals that the results of Experiment 5 are not for lack of statistical power. That experiment had a probability of 98% to detect a moderation effect of $\eta_p^2 = .02$. Therefore, we conclude that learning difficulty is

probably not a contributing factor to our finding that opposition relations induce less extreme evaluations than similarity relations.

We found better evidence in support of a different reason for our findings. Our second hypothesized reason was that, unlike similarity relations, opposition relations link the target object to affective stimuli of opposite valence. The inference from the relation *the target ends unpleasant sounds* suggests that the target is positive. However, the mere link between the target and unpleasant sounds pushes the evaluation to the negative pole, thus attenuating the extremity of the positive evaluation. To test that assumption, in Experiment 6 we manipulated participants' focus, during learning, between the two pieces of information that, we hypothesized, push the evaluation to two different directions: we focused participants either on the outcome of the targets' action (ending unpleasant sounds leads to a *positive* outcome) or on the linked stimulus (the unpleasant sounds). Compatible with our hypothesis, we found that a focus on the link with the affective stimulus pushed the evaluation to the valence of the affective stimulus, whereas a focus on the outcome of the relation between the target and the affective stimulus pushed the evaluation to the valence of the outcome. By controlling people's focus on the link between the target object and the affective stimuli we were able to increase or decrease the moderating effect of relation type on evaluation extremity. Importantly, participants in those conditions showed perfect memory of the targets' actions. It was only their evaluation—not their understanding of the information—that was affected by the focus manipulation. In conclusion, the results of the present research suggest that the reason for the attenuation of evaluations induced by opposition relations is that opposition relations link the target object to valence opposite to the valence suggested by inference from the relational information.

Implications for understanding deliberate evaluation

The finding that opposition relations induce relatively attenuated evaluation might have interesting implications on evaluation theory. Many contemporary evaluation models assume that the automatic activation of the valence associated with the target object serves as the initial judgment at the beginning of deliberate evaluative processes. However, that initial judgment can be replaced by a different judgment if the automatically activated evaluation does not seem valid (Cunningham et al., 2007; Fazio, 1990, 2007; Gawronski & Bodenhausen, 2006; Petty, Briñol, & DeMarree, 2007). Consistent with this, we observed that evaluation of targets with opposition

relations to affective stimuli almost always reflected the targets' true valence and not the valence of the affective stimuli that was linked to them. For instance, in Experiments 2 and 5, participants disliked creatures who took cute puppies from them, despite the association between the targets and puppies. On the other hand, our main finding might suggest that the initial judgment, reflecting the valence of the associated stimuli, was *adjusted* by the deliberate evaluative processes rather than completely *replaced*. Deliberate evaluation did not completely ignore the link between the target and the valence of the related affective stimuli. The valence of the affective stimuli attenuated the target's evaluation. In other words, false implications of evaluative associations influenced deliberate evaluation even when people took into account the true evaluative meaning of that target-stimuli relation: participants knew that the creatures were associated with cute puppies only because the creatures harmed the participants by taking the cute puppies away, and yet the positive valence of the cute puppies attenuated the negative evaluation of the creatures. Further investigations could examine how and when false implications of associated valence influence deliberate evaluation and could extend present theory of deliberate evaluative processes.

Coda

Our results suggest that causing enjoyment or affliction is more effective in attitude induction than ending enjoyment, bringing relief or preventing affective outcomes. This provides a basis for specific predictions about likeability of people or objects as a function of their roles in everyday life. Friends that are the source of fun might be better liked than those that relieve pain; politicians stopping popular reform may be less hated than those leading unpopular reform; and, drugs that stop illnesses may be liked less than those that enhance well-being. In persuasion, opposition relation messages may be less effective in changing attitudes than those using similarity relation messages. The generalizability of the present results to these applications is unknown, but could be pronounced.

References

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of personality and social psychology*, *89*(6), 884-898.

Corral, D., & Jones, M. (2014). The effects of relational structure on analogical learning. *Cognition*, *132*(3), 280-300.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, *25*(5), 736-760.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *Spanish Journal of Psychology, 10*, 230-241.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of Likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853–869.

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: negation as reflective operation. *Journal of personality and social psychology*, *91*(3), 385-405.

Doumas, L. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological review*, *115*(1), 1-43.

Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2008). Derived avoidance learning: Transformation of avoidance response functions in accordance with same and opposite relational frames. *The Psychological Record, 58,* 269-286.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior; The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75-109). San Diego, CA; Academic Press.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition, 25,* 603-637.

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion, 25*, 639-656.

Förderer, S. & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS – US relations. *Cognition and Emotion, 26,* 534-540

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and prepositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44,* 59-127.

Gawronski, B., & Walther, E. (2008). The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin*, *34*(9), 1276-1289.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology*, *41*(6), 618-626.

Gentner, D., & Loewenstein, J. (2002). Relational language and relational thought. In J. Byrnes & E. Amsel (Eds.), Language, literacy, and cognitive development (pp. 87–120). Mahwah, NJ: Erlbaum

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive psychology*, *15*(1), 1-38.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*, 107–119.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. Journal of Personality and Social Psychology, 65, 221-233.

Green, C. & Hummel, J.E. (2004). Relational perception and cognition: Implications for cognitive architecture and the perceptual-cognitive interface. In B.H. Ross (Ed.): The Psychology of Learning and Motivation, Vol 44, (pp. 201-226). San Diego, CA: Academic Press.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.

Hummel, J. E., & Holyoak, K. J. (2005). Relational Reasoning in a Neurally Plausible Cognitive Architecture An Overview of the LISA Project. *Current Directions in Psychological Science*, *14*(3), 153-157.

Lovett, M. C., & Anderson, J. R. (2005). Thinking as a production system. In K. J. Holyoak & R. Morrison (Eds.), The Cambridge handbook of thinking and reasoning (pp. 401–429). New York: Cambridge University Press

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs. "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology, 40,* 433–449.

Moran, T., & Bar-Anan, Y. (2013). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion, 27,* 743-752.

Moran, T., Bar-Anan, Y. & Nosek, B. A. (2015). Processing Goals Moderate the Effect of Co-occurrence on Automatic Evaluation. *Journal of Experimental Social Psychology, 60*, 157-162.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134,* 565-584.

Petty, R. E., & Briñol, P. (2006). A meta-cognitive approach to "implicit" and "explicit" evaluations: Comment on Gawronski and Bodenhausen (2006). *Psychological Bulletin, 132*, 740-744.

Petty, R. E. Briñol, P., & DeMarree, K. G. (2007). The Meta-Cognitive Model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition, 25,* 657-686.

Ranganath, K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science*, *19*(3), 249-254.

Ratliff, K. A., Swinkels, B. A., Klerx, K., & Nosek, B. A. (2012). Does one bad apple (juice) spoil the bunch? Implicit attitudes toward one product transfer to other products by the same brand. *Psychology & Marketing*, *29*(8), 531-540.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology, 94*, 249–273

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: background knowledge permits fast and efficient validation of information. *Journal of personality and social psychology*, *96*(3), 538-558.

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of personality and social psychology*, *74*(4), 837-848.

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology*, *56*, 283–294.

Walther, E., Weil, R., & Düsing, J. (2011). The role of evaluative conditioning in attitude formation. *Current Directions in Psychological Science*, *20*(3), 192-196.

Whelan, R., & Barnes-Holmes, D. (2004). The transformation of consequential functions in ccordance with the relational frames of same and opposite. *Journal of the Experimental Analysis of Behavior, 82,* 177-195.