

## **Mechanisms Underlying Effects of Approach-Avoidance Training on Stimulus Evaluation**

Pieter Van Dessel<sup>1</sup>

Andreas B. Eder<sup>2</sup>

Sean Hughes<sup>1</sup>

<sup>1</sup>Ghent University, Belgium

<sup>2</sup>University of Würzburg, Germany

*In press, JEP:LMC*

Author note: Correspondence concerning this article should be addressed to Pieter Van Dessel, Ghent University, Department of Experimental-Clinical and Health Psychology, Henri Dunantlaan 2, B-9000 Ghent (Belgium). E-mail: Pieter.vanDessel@UGent.be.

### **Abstract**

Over the past decade an increasing number of studies across a range of domains have shown that the repeated performance of approach and avoidance (AA) actions in response to a stimulus leads to changes in the evaluation of that stimulus. The dominant (motivational-systems) account in this area claims that these effects are caused by a rewiring of mental associations between stimulus representations and AA systems that evolved to regulate distances to positive and negative stimuli. In contrast, two recently forwarded alternative accounts postulate that AA effects are caused by inferences about the valence of actions and stimuli (inferential account) or a transfer of valenced action codes to stimulus representations (common-coding account). Across four experiments we set out to test these three competing accounts against one another. Experiments 1-3 illustrate that changes in stimulus evaluations can occur when people perform valenced actions that bear no relation to a distance regulation, such as moving a manikin upwards or downwards. The observed evaluative effects were dependent on the evaluative implication of the instructed movement goal rather than whether the action implied a movement towards or away from the stimuli. These results could not be explained with a rewiring of associations to motivational systems. Experiment 4 showed that changes in stimulus evaluations occurred after participants passively observed approach-avoidance movements, supporting an explanation in terms of cognitive inferences.

*Keywords:* approach-avoidance; attitudes; action framing; inferential account; common valence coding

### **Mechanisms Underlying Effects of Approach-Avoidance Training on Stimulus Evaluation**

It has been recognized for decades that behavior is shaped by likes and dislikes (Allport, 1935). Hence, understanding how preferences are acquired and how they can be changed is an important endeavor for psychological research. One paradigm that has shown great promise for changing preferences involves the performance of approach and avoidance (AA) actions. When participants repeatedly approach a certain stimulus and avoid another stimulus they typically develop a preference for the approached stimulus over the avoided stimulus (Laham, Kashima, Dix, Wheeler, & Levis, 2014). These AA training effects have been observed across a variety of stimuli, from pictures of unfamiliar faces (Woud, Maas, Becker, & Rinck, 2013), to social groups (Kawakami, Phillips, Steele, & Dovidio, 2007), and unhealthy foods (Zogmaister, Perugini, & Richetin, 2016). AA training may have important applied potential insofar as it can lead to changes in alcohol consumption (Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011), smoking behavior (Wittekind, Feist, Schneider, Moritz, & Fritzsche, 2015), and fear responses (Jones, Vilensky, Vasey, & Fazio, 2013).

Researchers have traditionally explained why changes in stimulus evaluations occur following AA actions by appealing to two rudimentary motivational systems that predispose individuals to approach appetitive and avoid aversive stimuli (Cacioppo, Priester, & Berntson, 1993; Elliot, Eder, & Harmon-Jones, 2013). According to this model, positive (appetitive) stimuli automatically trigger a behavioral tendency to approach, whereas negative (aversive) stimuli automatically trigger a tendency to avoid (e.g., Chen & Bargh, 1999; Solarz, 1960). In-line with this idea, Krieglmeier, Deutsch, De Houwer, and De Raedt (2010) showed that participants moved a virtual manikin (representing the self) faster towards positive words and away from negative words on a computer screen. These effects were obtained even when the movement was irrelevant

for the task at hand. Such findings suggest that AA movements are mediated by specialized motivational systems that make them distinct from other classes of actions (Krieglmeyer, Deutsch, & De Houwer, 2013; for a discussion see also Eder, Rothermund, & Hommel, 2016). More important within the context of the present paper, some researchers have postulated that the link between AA behavior and motivational systems is bidirectional. These researchers proposed that repeatedly approached or avoided stimuli acquire a positive or negative valence through their associations with AA motivational systems (Neumann, Förster, & Strack, 2003) or with valenced concepts that are inherently tied to these systems (e.g., the self; see Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011; Markman & Brendl, 2005). As a result, people evaluate approached stimuli more favorably and avoided stimuli more unfavorably after AA training (Laham et al., 2014; Kawakami et al., 2007).

Recently, two alternative models have been offered to explain the impact of AA training on stimulus evaluations. The first is the *inferential account* of AA training effects which suggests that people sometimes infer stimulus evaluations from their actions (Van Dessel, De Houwer, & Gast, 2016; similar to self-perception theory, Bem 1972; Koriat, Ma'ayan, & Nussinson, 2006). Specifically, participants who repeatedly approach or avoid a certain stimulus might generate propositions about these stimulus-action relations. After repeatedly approaching one stimulus, participants may infer that it is positive, and after repeatedly avoiding a second stimulus they may infer that it is negative. People might make these inferences because they have learned that positive objects have typically been approached and that negative objects have typically been avoided in the past. Once participants infer information about the valence of a stimulus, the activation of this mental proposition may impact not only explicit (i.e., non-automatic) but also implicit (i.e., automatic) evaluative responses towards the stimulus (De Houwer, 2014). Unlike motivational-

systems theories, this inferential account assumes that AA training effects result from general (inferential) mechanisms of evaluative learning that are not specifically tied to approach and avoidance. Any information that allows one to infer the valence of a stimulus can produce changes in stimulus evaluations.

A second alternative to the motivational systems account is the *common-coding account* (Eder & Rothermund, 2008; Eder & Klauer, 2009). Derived from the Theory of Event Coding (TEC; Hommel, Müsseler, Aschersleben, & Prinz, 2001), this account proposes that perceptions and actions are represented by networks of distributed feature codes (so called “event files”) that represent, among other perceivable features, the positivity and negativity of an event. Given that approach behavior is typically motivated by the expectation of a desired (positive) action consequence, and avoidance by the expectation to prevent an undesired (negative) consequence, feature networks representing these actions should incorporate affective codes representing the positivity or negativity of the action target (Eder & Hommel, 2013). When an AA action is repeatedly performed in response to a neutral stimulus, an event file is created that binds together (1) the stimulus representation and (2) the affective code that is part of the action representation (Eder, Müsseler, & Hommel, 2012). The link to the evaluative code in the event file then relates the stimulus to a particular valence, biasing the person’s evaluation of the stimulus towards the valence of the action (for evidence see Eder & Klauer, 2007, 2009). Importantly, an evaluative action code can be formed on the basis of different features of a planned action, such as the intrinsic valence of produced action effects (e.g., positive and negative perceptions; Eder, Rothermund, De Houwer, & Hommel, 2015), the evaluative implication of the intended action effect (e.g., turning an aversive sound on or off; Eder, Rothermund, & De Houwer, 2013), or verbal descriptors of a movement goal (e.g., moving a lever towards and away from the body; Eder & Rothermund, 2008).

Framing an action in positive terms (e.g., ‘approach’, ‘smile’, ‘hug’, etc.) and negative terms (e.g., ‘avoid’, ‘frown’, ‘kill’, etc.) might assign positive and negative valences to the action representation that can transfer to stimulus representations bound to the same event file after sufficient training. According to this theory, a training of any valenced action can produce a systematic change in stimulus evaluation, even when the trained action is unrelated to approach and avoidance defined by distance changes.

In the current research, our aim was to shed light on the mental mechanisms that underlie AA training effects on stimulus evaluations. This question is important given that AA training is currently one of the most popular methods for changing (implicit) stimulus evaluations, in part due to the fact that AA training has been found to influence even difficult to change behaviors (e.g., implicit prejudice: Kawakami et al., 2007; addictive behaviors: Wiers et al., 2011) and that effects seem to occur under some of the conditions of automaticity (e.g., unintentionally: Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016; unconsciously: Kawakami et al., but see: Van Dessel, De Houwer, Roets, & Gast, 2016). AA training effects have recently been observed across many different domains in psychology, including social psychology (e.g., racial evaluations: Phillips et al., 2011), clinical psychology (e.g., alcohol: Wiers et al., 2011; cigarettes: Wittekind et al., 2015; social anxiety: Taylor & Amir, 2012) and educational psychology (e.g., mathematics: Kawakami, Steele, Cifa, Phillips, & Dovidio, 2008). By testing the predictions of the three competing accounts of AA training effects we are bound to learn more about the underlying mechanisms of these effects, which is especially useful for understanding the conditions under which AA training procedures can establish optimal effects. For instance, if AA training effects are the result of inferential mechanisms rather than motivational systems mechanisms, then better effects might be obtained when targeting specific inferences via AA training rather than establishing and

overtraining contingencies between stimuli and distance-change movements (see Van Dessel, Hughes, & De Houwer, 2017). The current studies also speak directly to the moderators of AA training effects. For instance, we will learn whether AA training effects require actual distance change movements or can also be obtained on the basis of other actions or mere observation of such actions. This information has direct relevance to existing therapies that are already using AA procedures with the aim of changing attitudes or pathological behaviors.

### Study Overview

We report four experiments that tested predictions of motivational-systems, common-coding, and inferential accounts. Because these competing accounts all have a high degree of flexibility and because many of the assumptions underlying these accounts have never been specified, we believe it is difficult, if not impossible, to distinguish between these accounts on the basis of a few individual experiments. However, testing predictions derived from these accounts will allow us to further *constrain* these accounts (i.e., force them into adopting specific assumptions without which they cannot account for our effects) and to have greater confidence in the assumptions that survive this process.

In Experiment 1, we used an adapted version of the manikin task designed by De Houwer, Crombez, Baeyens, and Hermans (2001) in which participants are instructed to move a stick figure (i.e., a manikin) up or down on the screen in response to the presentation of a stimulus. Participants repeatedly moved the manikin up in response to two nonwords and down in response to two other nonwords. We varied the initial position of the manikin (above or below the nonword) such that for one nonword the upward movement implied an action *towards* the nonword, while for a second nonword an upward movement implied an action *away* from the nonword. Similarly, for a third nonword the downward movement implied an action *towards* the nonword while the downward

movement implied an action *away* from a fourth nonword (Figure 1; see also Krieglmeyer et al. 2010, who used the same set-up to distinguish between common-coding and motivational-systems explanations of the effects of stimulus valence on AA action performance). We then measured participants' implicit and explicit evaluations of each of the four nonwords. By comparing evaluations of the nonwords we could test two separate effects: (1) the effect of the instructed movement goal (i.e., instructions to move the manikin upwards and downwards) and (2) the effect of actual distance change (i.e., whether the action constituted a movement towards or away from the stimulus). Whereas the common-coding account only predicts effects of the instructed movement goal, motivational-systems accounts only predict effects of actual distance change.

Experiment 2 replicated Experiment 1 using a new set of social stimuli (i.e., fictitious social groups) and extended that work by manipulating how actions were evaluatively framed. Participants either received instructions to (1) *move the manikin up* in response to the names of members from one fictitious social group and *down* in response to the names of members from another fictitious social group or (2) *approach* the names of members from one fictitious social group and *avoid* the names of members from another fictitious social group. In this way we sought to test whether the valence of the evaluative action frame (positive: up/approach vs. negative: down/avoid) and the type of evaluative action frame (up and down vs. approach and avoid) moderated evaluative responses towards the two social groups. Whereas the common-coding account predicts a main effect of the valence of the evaluative action frame, the other accounts predict specific interactions with the type of evaluative action frame. In Experiment 3, we extended our work to two separate tasks that did not involve moving a manikin on the screen. We examined whether action training effects on stimulus evaluations are stronger when the actions are directly related to AA motivations (i.e., involve towards/away movements) than when the actions do not

have any relation to approach or avoidance. Specifically, we compared a task that required participants to move towards or away from social group members to a task that involved moving those names up and down. Motivational-systems accounts predict an effect of the former task only, whereas the common-coding and inferential account predict effects of both tasks. Finally, Experiment 4 examined whether the mere observation of AA actions can lead to changes in stimulus evaluation. This prediction was derived from the inferential account and is not compatible with common-coding or motivational-systems accounts. Predictions of the three competing accounts are delineated in more detail in the respective introductions of the experiments.

For all experiments, the study design, data-analysis plan, and the hypotheses derived from the common-coding and inferential account of AA training were pre-registered on the Open Science Framework website prior to data-collection. The pre-registered plans as well as all data, experimental and analytic scripts are available at <https://osf.io/s9kb5/>. Any deviation from pre-registration is noted in the main text. We report all manipulations and measures for the different studies. The sample size of the experiments was determined prior to the data collections and pre-registered together with the study design. For Experiment 1, we stopped the data-collections when at least 250 participants had completed all measures of the experiment to ensure that we would have sufficient statistical power to detect even small effects after excluding data of participants on the basis of the pre-registered criteria (necessary sample size = 156 to have power = 0.80 to find a small effect of  $d_z = 0.20$  at alpha = .05). For Experiments 2 and 3, we stopped data-collection when 400 participants had completed all measures of the experiment to ensure that we would have sufficient statistical power to detect even small effects in these between-subject designs (necessary sample size = 387 to have power = 0.80 to find a small effect of  $\eta^2 = 0.02$  at alpha = .05). For Experiment 4, the sample size was determined using sequential Bayes hypothesis testing. Analyses

were performed when 100 participants had completed the study and sample size was increased by steps of 50 participants until a decisive Bayes factor larger than 10 was obtained for the critical *t*-test analyses (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). In accordance with previous studies and with our pre-registered data analysis plan, we excluded the data of participants who were unable to remember the correct stimulus-action contingencies in all four experiments. Importantly, including the data of these participants reduced the overall magnitude of some of the reported effects, but did not change the statistical significance of any of the reported effects in any of the experiments. Our research has been approved by the local ethics committee of Ghent University (approval number 2016/63).

### **Experiment 1**

Participants were instructed that they should move a manikin up in the presence of two nonwords and down in the presence of two other nonwords. In the manikin task, a manikin was presented either above or below the nonword. Depending on the relative location of the manikin (see Figure 1), an up response resulted either in a movement towards one of the nonwords (up-towards word) or away from another nonword (up-away word), whereas a down response resulted either in a movement towards one nonword (down-towards word) or away from another nonword (down-away word). After the manikin task, explicit evaluations of the four nonwords were measured with an evaluative rating task and implicit evaluations were measured with two versions of the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998).

Based on a motivational-systems account of AA training effects, nonwords should be evaluated more favorably after a training task that involves moving the self towards the nonword (approach) compared to a training task that involves moving the self away from the nonword (avoidance). Such actions represent approach and avoidance actions because they regulate the

distance between the symbolic self and the stimulus (Markman & Brendl, 2005; Förster, 2001). The instructed movement goal (up vs. down) should not influence stimulus evaluations, because running upwards and downwards is not a biologically relevant motion for motivational systems tuned to appetitive/aversive stimuli.

In contrast, the common-coding account predicts that stimulus evaluations should systematically change with the training of any valenced action that is linked to a stimulus representation in the event file after training (Eder & Klauer, 2009). Because performing an upward movement is a positive action frame and performing a downward movement is a negative action frame (for corresponding ratings see Eder & Rothermund, 2008), nonwords should be evaluated more favorably after being linked to upwards movements relative to downwards movements. Furthermore, changes in evaluations should not be affected by manikin movement towards and away from the word because these visualizations of the movement are irrelevant for the intended action and response valence should be most determined by the goal of the response (Memelink & Hommel, 2013; see also Phaf, Mohr, Rotteveel, & Wicherts, 2014).

Finally, according to an inferential account, changes in evaluations of the nonwords should depend on the types of inferences that participants make when they register the stimulus-action relations (Van Dessel, De Houwer, & Gast, 2016). An effect of distance change should occur if movements towards or away from a nonword lead to the formation of information about the contingencies between nonwords and AA actions and facilitate inferences about the evaluative properties of the nonwords. The account, as it currently stands, does not make clear (directional) predictions about the presence or absence of an effect of the instructed movement goal (i.e., up vs. down). Though it incorporates the possibility that the performance of (valenced) actions can produce changes in stimulus evaluations due to inferential mechanisms even when these

actions do not constitute AA actions (Van Dessel, De Houwer, & Smith, 2017) it has not been specified which other actions are likely to produce effects on stimulus evaluations.

## **Method**

**Participants.** A total of 269 English-speaking volunteers participated online via the Prolific Academic website (<https://prolific.ac>). The experiment was programmed in Inquisit 4.0 and hosted via Inquisit Web (Millisecond Software, Seattle, WA). In-line with the standard treatment of data in web-based studies on AA training effects (e.g., Van Dessel, De Houwer, Gast, Smith, et al., 2016), we excluded data from participants who (a) did not fully complete all questions and tasks (41 participants; i.e., 15.24%), (b) had IAT error rates for any of the IATs above 30% across the entire task, or above 40% for any of the critical IAT blocks (13 participants; i.e., 4.83%), (c) completed more than 10% of IAT trials faster than 400ms (20 participants; i.e., 7.43%), or (d) made at least one error on the questions that probed memory for the manikin task instructions (18 participants; i.e., 6.69 %). In correspondence with Eder and Rothermund (2008), we also decided to remove the data of participants who had error rates above 30% for ‘label trials’ (see below for a description of those trials) because this would be a strong indication of internal action recoding. No participants needed to be excluded based on this criterion. Analyses were performed on the data of 177 participants (85 women, mean age = 30,  $SD = 12$ ).

**Manikin Task Instructions.** Participants were told that the experiment would involve four words: UDIBNON, BAYRAM, LOKANTA, and ENANWAL. They were then informed that during the first task they would see a stick figure (i.e., a manikin) on the screen and that they would have to move this manikin using the up and down keys of the computer keyboard. The instructions also asked them to memorize the following three guidelines: First, they should imagine that the manikin represented them and that when they moved the manikin up or down it was them who were moving. Second, they should move the manikin up when they would see the word BAYRAM

or ENANWAL and move the manikin down when they would see the word UDIBNON or LOKANTA (the assignment of the words to the up and down actions was counterbalanced across participants). Third, on some trials they would see the words UP or DOWN appear on the screen and on these trials they should respond accordingly (e.g., by pressing the up key when seeing the word UP). These ‘label trials’ were implemented to discourage participants from internally recoding the up and down actions in other terms (e.g., as movements towards and away from the nonwords). As discussed in Eder & Rothermund (2008), participants are prone to select the set of stimulus and response codes that most effectively maximize performance on the task. Because participants are required to quickly respond to the label trials and can most efficiently do so when they use the up and down label response codes they should be inclined to use up and down action framing throughout the entire task.

After reading the instructions, participants indicated which action they would perform when they would see the nonword UDIBNON, BAYRAM, ENANWAL, and LOKANTA by selecting “move the manikin up”, “move the manikin down”, or “I don’t remember”. They proceeded to the manikin task when they selected the correct response for all nonwords.

**Manikin Task.** The manikin task consisted of two blocks of 100 trials. During each block, each of the four nonwords was presented on 20 trials while the words UP and DOWN were each presented on 10 trials. Each trial started with the presentation of the manikin, followed (after 500ms) by the presentation of a nonword in training trials or the word UP or DOWN in label trials. The words and nonwords appeared in the center of the screen, while the manikin appeared in the upper or lower half of the screen. For two of the nonwords, the manikin always appeared above the nonword (i.e., the up-away and the down-towards words). For the other two nonwords, the manikin always appeared below the nonword (i.e., the up-towards and the down-away words). Whenever the participant made a correct response by pressing the up or down key on the keyboard, the

manikin moved up or down on the screen. Duration of the manikin movement was 1200ms. An incorrect response produced error feedback. Participants had to make a correct response on each trial to complete the task. The manikin did not move in a label trial and the trial ended after a correct response. The next trial started 100ms later.

**IAT.** After completion of the manikin task, half of the participants first performed two IATs measuring implicit evaluations of the four nonwords. The other participants first completed an evaluative rating task. In one IAT, the up-towards word and the down-away word were evaluated (IAT1). The other IAT involved evaluation of the up-away word and the down-towards word (IAT2). The order of the two IATs was counterbalanced across participants. The IATs followed the procedure described in more detail by Van Dessel, Gawronski, Smith, and De Houwer (2017). In each of the IATs, participants categorized eight attribute words (e.g., wonderful, evil) as ‘positive’ or ‘negative’ and two nonwords (e.g., UDIBNON and BAYRAM) as their respective names (e.g., as ‘Udibnon’ or ‘Bayram’). To avoid that the target stimuli were classified only on the basis of simple perceptual features, the nonwords were presented in different font types (Arial Black and Fixedsys), capitalizations (uppercase and lowercase), and sizes (16pt and 18pt), resulting in 8 different stimuli for each nonword. This procedure is often used for the measurement of implicit evaluations of nonword stimuli with an IAT (see Zanon, De Houwer, Gast, & Smith, 2014). The attribute words were always presented in Arial Black, font size 16, uppercase. The IATs consisted of three practice blocks and two experimental blocks. Participants began with 20 practice trials sorting the target words and 20 practice trials sorting positive and negative stimuli. Next, participants completed 56 trials in which one nonword (e.g., UDIBNON) and positive stimuli shared one response key and the other nonword (e.g., BAYRAM) and negative stimuli shared another response key (or vice versa). Participants then practiced sorting target words on 40 trials

with a reversed response key assignment. Finally, participants completed a second set of 56 trials in which one nonword (e.g., BAYRAM) shared a response key with negative and the other nonword (e.g., UDIBNON) shared a response key with positive. On each trial, a word or nonword was presented in the center of the screen until the participant pressed one of the two valid keys (i.e., F or J). If the response was correct, the stimulus disappeared and the next stimulus was presented 400 ms later. If the response was incorrect, the word was replaced by a red 'X' for 200 ms. The next word appeared 400 ms after the red 'X' was removed from the screen. For each IAT, an IAT-score was calculated on the basis of participants' performance using the D4-algorithm (Greenwald, Nosek, & Banaji, 2003) so that positive scores indicate a preference for the nonword in response to which they had performed an up movement in the manikin task over the nonword in response to which they had performed a down movement in the manikin task. The Spearman-Brown corrected split-half reliability of the IAT scores, calculated on the basis of an odd-even split, was  $r(175) = .80$  for IAT1 and  $r(175) = .74$  for IAT2.

**Evaluative rating task.** In the evaluative rating task, participants were asked to rate how pleasant or unpleasant they found each of the four nonwords. Participants gave their ratings on a Likert scale ranging from 1 (*extremely unpleasant*) to 9 (*extremely pleasant*).

**Stimulus-action contingency questions.** After the evaluation tasks, participants were asked to complete a manipulation check for each nonword. They first answered what action they had performed in the presence of the word UDIBNON, BAYRAM, ENANWAL, or LOKANTA by selecting an option from a dropdown menu with "move the manikin up", "move the manikin down", and "I don't remember" as possible answers. Participants then answered a follow-up question that asked whether performing this action resulted in a movement of the manikin towards or away from the nonword. Possible answers were (1) "the manikin moved towards the word", (2) "the manikin moved away from the word", (3) "I don't remember/I did not notice".

**Additional questions.** Participants then answered three additional questions: (1) They rated the words “up” and “down” on a scale ranging from 1 (*extremely negative*) to 9 (*extremely positive*). (2) They indicated to what extent they had imagined that *they* were the manikin in the manikin task on a scale ranging from 1 (*not at all*) to 9 (*I always imagined this*). (3) They indicated on what basis they had given the evaluative ratings for the nonwords. Possible answers were: “I responded based on what I thought the experimenter wanted me to do”, “I responded based on what I learned about the stimuli earlier in the experiment”, “I responded based on my feelings about the stimuli”, and “I do not know why I gave ratings the way I did”. We included this final question to control for demand characteristics that might bias the observed effects (see Sharpe & Whelton, 2016).<sup>1</sup>

## Results

**Explicit ratings.** Explicit ratings of the four nonwords were subjected to a 2 x 2 repeated measures analysis of variance (ANOVA) with *Vertical Movement* (manikin moved up/down in response to the nonword) and *Distance Change* (manikin moved towards/away from the nonword) as within-subjects factors. We observed a main effect of Vertical Movement,  $F(1,529) = 44.00$ ,  $p < .001$ . Participants preferred nonwords in response to which they made an up action ( $M = 5.53$ ,  $SD = 1.41$ ) over nonwords in response to which they made a down action ( $M = 4.77$ ,  $SD = 1.41$ ),  $d_z = 0.40$ , 95% confidence interval of the difference in ratings (CI diff) = [0.48, 1.05]. The Bayes factor, calculated in accordance with Rouder, Speckman, Sun, Morey, and Iverson (2009), indicates that the data provide strong evidence for this effect,  $BF_1 > 10000$ . The main effect of Distance Change was not significant,  $F(1,529) = 0.09$ ,  $p = .77$ . Participants did not prefer the nonword they moved towards ( $M = 5.17$ ,  $SD = 1.19$ ) over the nonword they moved away from ( $M = 5.13$ ,  $SD = 1.13$ ),  $d_z = 0.03$ , 95% CI diff = [-0.15, 0.22]. The Bayes factor indicated strong evidence for the

---

<sup>1</sup> In total, 9, 37, 29, and 8 participants could be classified as demand compliant in Experiments 1, 2, 3, and 4 respectively. Excluding the data of these participants did not change the statistical significance of any of the reported effects for any of the experiments.

null hypothesis (i.e., that Distance Change does *not* influence explicit ratings),  $BF_0 = 11.16$ . There was no interaction between Vertical Movement and Distance Change,  $F(1,529) = 0.86$ ,  $p = .36$ .

**IAT scores.** We performed a one-tailed paired  $t$ -test comparing IAT1 and IAT2 scores. Because IAT1 scores indicate a preference for the up-towards word over the down-away word and IAT2 scores indicate a preference for the up-away word over the down-towards word, this comparison provides a test of the effect of Distance Change. We did not observe more positive IAT1 scores ( $M = 0.03$ ,  $SD = 0.44$ ) compared to IAT2 scores ( $M = 0.05$ ,  $SD = 0.39$ ),  $t(176) = -0.29$ ,  $p = .61$ ,  $d_z = -0.02$ , 95% CI diff =  $[-0.09, 0.07]$ . The Bayes factor indicated strong evidence for the null hypothesis,  $BF_0 = 14.80$ . We also performed a one-tailed one-sample  $t$ -test on the means of the two IAT scores to test for the effect of Vertical Movement on implicit evaluations. This mean score was significantly larger than zero ( $M = 0.04$ ,  $SD = 0.31$ ),  $t(176) = 1.74$ ,  $p = .042$ ,  $d_z = 0.13$ . However, the Bayes factor indicated anecdotal evidence for the null hypothesis that Vertical Movement does not influence IAT scores,  $BF_0 = 1.43$ .<sup>2</sup>

**Correlational and mediation analyses.** We deviated from our pre-registered data-analysis plan by also performing correlational and mediation analyses. These analyses tested how effects on implicit and explicit evaluations relate to participants' answer to the self-report questions. First, we created separate indices of the effect of Distance Change and Vertical Movement for explicit ratings and IAT scores. For explicit ratings, the Distance Change index was calculated by averaging the ratings of the avoid-up word and the avoid-down word and subtracting this score from the average ratings of the approach-up word and the approach-down word. The Vertical Movement

---

<sup>2</sup> For the sake of clarity, we deviated from our pre-registered data-analysis plan by performing additional analyses with item-based linear mixed effects models as implemented in R package lme4 (Bates, Maechler, Bolker, & Walker, 2014) on IAT scores and explicit rating scores. This approach allowed us to control for (and test) possible effects of counterbalancing factors. Overall, these analyses produced similar results to the ANOVA and  $t$ -test analyses (see Appendix).

index was calculated by averaging the ratings of the approach-down word and the avoid-down word and subtracting this score from the average ratings of the approach-up word and the avoid-up word. For IAT scores, the Distance Change index was calculated by subtracting the score of IAT2 from the score of IAT1. The Vertical Movement index was calculated by averaging the IAT1 and IAT2 scores. Correlations were significant for the two Distance Change indices,  $r(175) = .25$ ,  $p < .001$ , and for the two Vertical Movement indices,  $r(175) = .29$ ,  $p < .001$ .

Participants indicated that they liked the word ‘up’ ( $M = 6.74$ ,  $SD = 1.32$ ) more than they liked the word ‘down’ ( $M = 4.17$ ,  $SD = 1.42$ ),  $t(176) = 15.37$ ,  $p < .001$ ,  $d_z = 1.16$ , 95% CI diff = [2.24, 2.90]. Participants’ mean difference in rating of the words correlated positively with the Vertical Movement index for explicit ratings,  $r(175) = .23$ ,  $p = .002$ , but not with any of the other indices,  $r_s < .09$ ,  $p_s > .25$ . To further investigate the relationship between up-down label ratings and evaluative word ratings, we performed mediation analyses with the lavaan package (version 0.5-16; Rosseel, 2012). We used the bootstrap method to estimate standard errors for the effects. The indirect effect of Vertical Movement on explicit word ratings with label word ratings as a mediator was statistically significant,  $\beta = .40$ ,  $Z = 3.32$ ,  $p = .001$ , 95% CI of  $\beta = [.17, .64]$ ,  $R^2_{ind} = 0.33$ . The direct effect of Vertical Movement was marginally significant after controlling for label word ratings,  $\beta = .22$ ,  $Z = 1.68$ ,  $p = .093$ , 95% CI of  $\beta = [.01, .50]$ ,  $R^2_{dir} = 0.06$ .

Participants indicated that, in the manikin task, they imagined themselves to be the manikin to a moderate extent ( $M = 5.16$ ,  $SD = 2.20$ ). This index did not correlate significantly with any of the indices of Distance Change and Vertical Movement,  $r_s < .11$ ,  $p_s > .18$ .

Participants correctly reported whether the manikin moved towards or away from the four nonwords for only slightly more than one word on average ( $M = 1.41$ ,  $SD = 1.32$ ). Only 17 participants (i.e., 6.32%) indicated the correct contingency for all nonwords. The number of correct

classifications did not correlate significantly with any of the Vertical Movement or Distance Change indices,  $r_s < .11$ ,  $p_s > .16$ .

## **Discussion**

Experiment 1 provides evidence that (a) participants acquire an explicit preference for nonwords after training of an upward movement relative to a downward movement, and that (b) this effect does not depend on a distance-regulating function of the actions as approach or avoidance. These conclusions are consistent with the predictions derived from the common-coding account of AA training effects. Specifically, actions that are coded evaluatively produce changes in stimulus evaluations by their cross-modal links in an event file, while distance change does not influence stimulus evaluations when such changes are unimportant for the trained movement (Memelink & Hommel, 2013). Our findings are difficult to reconcile with a motivational-systems account of AA training effects that views the training of AA movements as being central for a change of associations with motivational systems. We obtained strong evidence that repeatedly moving a symbolic self towards and away from a word did not affect the evaluation of that word.<sup>3</sup> Furthermore, it is difficult to see how this account can explain the strong effect of the upward and downward movements on stimulus evaluation.

The inferential account of AA training effects also did not predict the current results. First, an effect of distance change was expected. Importantly, however, this null finding can be explained on the basis of the observation that only very few participants (17 participants, i.e., 6.32% of the total sample) registered the contingencies between AA actions and stimuli. This is assumed to be a crucial factor in AA training effects from an inferential perspective (see Van Dessel, De Houwer, & Gast, 2016). Furthermore, it is possible that an AA movement provides no inferential cue when

---

<sup>3</sup> Strong evidence for this null effect was also observed when analyses only included the data of 99 participants (i.e., 55.93% of the total sample) who indicated that they strongly imagined themselves to be the manikin.

it is irrelevant for the training task. Second, the inferential account did not make clear predictions about the presence (or absence) of an effect of Vertical Movement. One possible (albeit post-hoc) explanation of this effect is that participants infer that stimuli are positive when the presentation of the stimulus allows them to move up (e.g., they might think of ‘moving up in the world or social hierarchy’ which is arguably a positive event). This accords with the assumption that the performance of (valenced) actions can produce changes in stimulus evaluations even when these actions do not constitute AA actions (see also Van Dessel, De Houwer, & Smith, 2017; Bem, 1972). This explanation also fits with the results of the mediation analyses indicating that liking of the words ‘up’ and ‘down’ mediates the effect of Vertical Movement on explicit evaluations. It is important to note, however, that the mediation results do not provide clear support for inferential accounts because the observed mediation is also consistent with other accounts (e.g., evaluations of action frames should also be predictive of training effects according to common-coding accounts). We therefore did not include predictions about mediation analyses in our pre-registered hypotheses for any of the performed experiments.

Although the results of Experiment 1 provide clear support for the contribution of general (evaluative) learning mechanisms in AA training effects, two limitations should be noted. First, though Bayes factors provided strong evidence for an effect of up and down movements on explicit evaluations, they did not support the hypothesis that these movements affect implicit evaluations. This limitation could indicate that training valenced movements only influences the formation of explicit evaluations, while training distance changes could influence both explicit and implicit evaluations (e.g., Kawakami et al., 2007; but see also Huijding, Muris, Lester, Field, & Joesse, 2011). An alternative explanation is that procedural details may have impeded the detection of an effect on the IAT. Due to the 2x2 within-subjects design with four different stimuli, the effect of distance change could only be estimated by computing the mean of two separate IAT scores and

comparing this mean score to zero. This comparison might be problematic because the zero point on an IAT is often arbitrary (i.e., it does not reflect the zero point on an underlying psychological dimension; for a discussion see Blanton & Jaccard, 2006).

A second possible limitation is the low relevance of distance changes for the participant. Movements towards or away from meaningless nonwords might not have been interpreted as motivationally relevant actions. If this precluded activation of the motivational systems of approach and avoidance, it could provide an explanation for why we did not observe an effect of distance change on stimulus evaluations.

To address these potential limitations, a second experiment was conducted examining training effects on evaluations of two fictitious social groups (for an argumentation why AA motivation is important in the context of social groups see Kawakami et al., 2007). Instructions for the response training were now manipulated between groups, with one group receiving approach-avoidance instructions and a second group receiving up-down instructions. This change in design allowed us to directly compare effects of task-relevant and task-irrelevant distance changes during training.

## **Experiment 2**

Two factors were manipulated between participants: (1) how the action was framed in the instructions (i.e., approach-avoid versus up-down), and (2) the initial position of the manikin (above or below the names). Half of the participants were instructed to move a manikin upwards in response to members of one fictitious social group and downwards in response to members of another fictitious social group (up-down action framing). The other participants were instructed to approach members of one fictitious social group by moving towards them with the manikin and to avoid members of another fictitious social group by moving away from them (approach-avoid

action framing). This set-up allowed us to examine effects of the valence of the instructed actions (positive: up/approach vs. negative: down/avoid) on implicit and explicit evaluations as a function of action framing (up and down vs. approach and avoid). Furthermore, the initial position of the manikin (above or below the names) was now manipulated between-participants. For half the participants in the up-down action framing condition, an up action implied approach whereas a down action implied avoidance. For the other participants in this condition, the up action implied avoidance and the down action implied approach. In the approach-avoid action framing condition, the approach and avoidance action implied an upwards movement for half of the participants and a downward movement for the other half of the sample. This set-up allowed us to examine whether changes in evaluations depend on the uninstructed meaning of the action as either (1) approach or avoid in the up-down action framing condition or (2) an up or down movement in the approach-avoid action framing condition.

According to the common-coding account, participants should prefer the social group that they performed a positively framed action for (i.e., approach/ move up) over the group that they performed a negatively framed action for (i.e., avoid/ move down). This account would also argue that the effect of the valence of the instructed action should not be moderated by how an action is framed (approach-avoid vs. up/down) or by the initial location of the manikin (below vs. above the stimulus). The inferential account would also predict that the valence of the instructed action should influence stimulus evaluations. Yet, unlike the common coding account, it would predict that this effect should be qualified by how an action is framed. Participants should be more likely to infer the valence of a stimulus from the trained action with approach-avoid relative to up-down instructions because (1) people typically approach stimuli that they consider positive and avoid stimuli that they consider negative (rather than move up or down in response to positive or negative

stimuli) and (2) this propositional information is crucial for inferring stimulus valence. Finally, a motivational-systems account would predict changes in evaluations after participants repeatedly perform actions effecting a distance decrease (approach) or increase (avoidance) between the symbolic self and the stimulus reacted to. Specifically, stimuli in both action framing conditions should be liked more after they are paired with approach actions and less after pairings with avoidance actions.

### **Method**

**Participants.** A total of 402 English-speaking volunteers participated online via the Prolific Academic website. We excluded data from participants who (a) did not fully complete all questions and tasks (11 participants; i.e., 2.74%), (b) had IAT error rates above 30% across the entire task, or above 40% for any of the critical IAT blocks (22 participants; i.e., 5.47%), (c) completed more than 10% of IAT trials faster than 400ms (3 participants; i.e., 1.49%), (d) made at least one error on the questions that probed memory for the manikin task instructions (21 participants; i.e., 5.22%), or (e) had more than 30% errors on label trials in the manikin task (2 participants; i.e., 0.99%). Analyses were performed on the data of 343 participants (154 women, mean age = 32,  $SD = 11$ , 171 in the up-down framing condition and 172 in the approach-avoid framing condition).

**Procedure.** The procedure of Experiment 2 was identical to Experiment 1 with the following exceptions. First, participants were informed that the experiment would involve the names of members of two social groups called the Luupites and the Niffites. They were told that all the names of Luupites have two consecutive vowels in them and end with “lup”. They were then shown two examples of Luupites’ names (i.e., Loomalup, Ageelup). Subsequently, participants were told that all the names of Niffites would contain two consecutive consonants and end with “nif.” Again, this statement was followed by two Niffites names (i.e., Borrinif, Kennunif).

Second, half of the participants were instructed to move the manikin *up* when they would see the name of a Niffite and to move the manikin *down* when they would see the name of a Luupite (up-down framing condition). The other half of the participants were instructed to *approach* the names of Luupites by moving the stick figure towards a Luupite and to *avoid* the names of Niffites by moving the stick figure away from a Niffite (approach-avoid framing condition). The assignment of the stimuli to the actions was counterbalanced across participants and across conditions.

Third, in each trial of the manikin task, participants saw the manikin and one of four possible names of a member of Niffites or Luupites. Half the participants in each of the framing conditions always saw the manikin appear above the names whereas the other participants always saw the manikin appear below the names. On label trials, the words UP/DOWN or APPROACH/AVOID were presented depending on the action framing condition.

Fourth, following the manikin task, participants performed an IAT that measured implicit evaluations of Niffites and Luupites and an explicit rating task that measured explicit evaluations of Niffites and Luupites. The order of IAT and rating task was counterbalanced. The IAT procedure was identical to Experiment 1 with the only exception that target words consisted of the four Niffites and Luupites names that were also used in the training task (see also Van Dessel, De Houwer, Gast, Smith, et al., 2016). The evaluative rating task was identical to Experiment 1 with the exception that participants now rated how pleasant or unpleasant they found the two social groups (i.e., Niffites and Luupites).

Participants answered the same additional questions as in Experiment 1 along with an additional rating of the words ‘approach’ and ‘avoid’.

## **Results**

**Explicit rating scores.** Explicit rating scores were calculated by subtracting the pleasantness rating of Luupites from the pleasantness rating of Niffites, such that higher scores indicate a stronger preference for Niffites over Luupites. These scores were subjected to a 2 (*Action Framing Condition*: up-down, approach-avoid) x 2 (*Affective S-R Mapping*: positive response to Niffites [up-movement/approach] and negative response to Luupites [down-movement/avoid] or vice versa) x 2 (*Manikin Location*: above the names, below the names) between-subjects ANOVA. The factor *Task Order* (IAT first, rating task first) was not included in the ANOVA because it did not significantly improve model fit. We observed a main effect of Affective S-R Mapping,  $F(1,335) = 75.22, p < .001$ . Importantly, this effect was qualified by a significant interaction between Affective S-R Mapping and Action Framing Condition,  $F(1,335) = 75.22, p < .001$ . Replicating previous research (e.g., Van Dessel, De Houwer, Gast, Smith, et al., 2016), participants in the approach-avoid framing condition exhibited a stronger preference for Niffites over Luupites when they had approached Niffites and avoided Luupites ( $M = 1.68, SD = 2.67$ ) than when they had approached Luupites and avoided Niffites ( $M = -1.53, SD = 2.07$ ),  $t(170) = 8.79, p < .001, d = 1.34$ , 95% CI diff = [2.50, 3.95],  $BF_1 > 10000$ . Participants in the up-down framing condition exhibited a stronger preference for Niffites over Luupites when they had moved the manikin upwards in response to Niffites and downwards in response to Luupites ( $M = 0.33, SD = 2.02$ ) than vice versa ( $M = -0.59, SD = 1.94$ ),  $t(169) = 3.03, p = .003, d = 0.46$ , 95% CI diff = [0.32, 1.52],  $BF_1 = 10.95$ . Importantly, the effect of Affective S-R Mapping was larger in the approach-avoid framing condition than in the up-down framing condition,  $d_{diff} = 0.88$ . There were no other main or interaction effects,  $F_s < 0.89, p_s > .34$ . The interaction of Affective S-R Mapping and Action Framing Condition was significant when the manikin appeared above the names and when the manikin appeared below the names,  $F_s > 10.22, p_s < .003$ .

**IAT scores.** The IAT score was calculated using the D4-algorithm such that higher scores

indicate a stronger preference for Niffites over Luupites. The Spearman-Brown corrected split-half reliability of the IAT score was  $r(341) = .91$ . The  $2 \times 2 \times 2$  ANOVA on IAT scores revealed a main effect of Affective S-R Mapping,  $F(1,335) = 149.71, p < .001$ , and a main effect of Action Framing Condition,  $F(1,335) = 6.46, p = .011$ . Importantly, these main effects were qualified by a significant interaction,  $F(1,335) = 14.45, p < .001$ . Participants in the approach-avoid framing condition exhibited a stronger preference for Niffites over Luupites when they had approached Niffites and avoided Luupites ( $M = 0.32, SD = 0.48$ ) than when they had approached Luupites and avoided Niffites ( $M = -0.49, SD = 0.44$ ),  $t(170) = 11.65, p < .001, d = 1.77, 95\% \text{ CI diff} = [0.68, 0.96], \text{BF}_1 > 10000$ . Participants in the up-down framing condition exhibited a stronger preference for Niffites over Luupites when they had moved the manikin upwards in response to Niffites and downwards in response to Luupites ( $M = 0.00, SD = 0.51$ ) than vice versa ( $M = -0.43, SD = 0.44$ ),  $t(169) = 5.93, p < .001, d = 0.91, 95\% \text{ CI diff} = [0.29, 0.58], \text{BF}_1 > 10000$ . The effect of Affective S-R Mapping was larger in the approach-avoid framing condition than in the up-down framing condition,  $d_{\text{diff}} = 0.86$ . There were no other main or interaction effects,  $F_s < 1.07, p_s > .30$ . The interaction between Affective S-R Mapping and Action Framing Condition was significant when the manikin appeared above the names and when it appeared below the names,  $F_s > 4.19, p_s < .043$ .

**Correlational analyses.** For these analyses, we calculated an IAT score and explicit rating score that indicated a preference for the social group in response to which participants had performed a positive action (up/approach) over the social group in response to which participants had performed a negative action (down/avoid). These two scores correlated significantly for participants in the up-down framing condition,  $r(169) = .24, p = .002$ , but not for participants in the approach-avoid framing condition,  $r(170) = .11, p = .15$ .

Participants liked the word ‘up’ ( $M = 6.50$ ,  $SD = 1.47$ ) more than they liked the word ‘down’ ( $M = 3.97$ ,  $SD = 1.35$ ),  $t(342) = 18.92$ ,  $p < .001$ ,  $d_z = 1.02$ , 95% CI diff = [2.26, 2.79]. The mean difference in the word ratings was positively correlated with the effect in the explicit ratings in the up-down action framing condition,  $r(170) = .15$ ,  $p = .044$ , but not with the effect in the IAT scores,  $r(170) = -.09$ ,  $p = .24$ . Participants also liked the word ‘approach’ ( $M = 6.46$ ,  $SD = 1.29$ ) more than the word ‘avoid’ ( $M = 3.21$ ,  $SD = 1.40$ ),  $t(342) = 27.61$ ,  $p < .001$ ,  $d_z = 1.49$ , 95% CI diff = [3.02, 3.49]. The mean difference in the word ratings correlated positively with the effect in the explicit ratings in the approach-avoid action framing condition,  $r(170) = .24$ ,  $p = .001$ , but not with the effect in the IAT scores,  $r(170) = .07$ ,  $p = .39$ . Notably, the mean difference in the evaluative ratings of the words ‘approach’ and ‘avoid’ was significantly bigger than the mean difference in the ratings of the words ‘up’ and ‘down’,  $t(342) = 4.86$ ,  $p < .001$ , 95% CI diff = [0.43, 1.02]. This difference indicates a stronger valence of the words ‘approach’ and ‘avoid’ in comparison to the words ‘up’ and ‘down’.

## Discussion

The results of Experiment 2 corroborate and extend those of Experiment 1 by showing that (1) a training of up and down actions in response to members of fictitious social groups can produce changes in implicit and explicit evaluations of these social groups and (2) these effects are not moderated by an AA function of these actions in terms of distance-regulations. These results are in-line with predictions of the common-coding account and the inferential account yet strongly contrast with predictions of motivational-systems accounts.

Experiment 2 also showed that the training procedure produced larger changes in liking when the trained actions were framed as approach and avoidance relative to up and down movements. This result was predicted by the inferential account: the performance of AA actions in response to a stimulus provides a better basis for making inferences about the evaluative properties

of the stimulus. In contrast, the common-coding account did not predict an effect of the type of action framing. That said, the common-coding account could explain this effect on the basis of the result that participants rated action labels related to approach/avoidance as more strongly valenced than up-down action labels. Evaluative transfer from the action to the stimulus representation should be larger the stronger the valence of the action, explaining a bigger training effect with AA action framing. In-line with this idea, the self-reported valence of the action frame words correlated with the size of evaluative training effects on explicit (but not implicit) evaluations.

### **Experiment 3**

In Experiment 2 we observed that actions produced the biggest changes in stimulus evaluations when framed as approach and avoidance. Though this results fits with predictions of the inferential account, it could also indicate that motivational-systems mechanisms do play a partial role in AA training effects in specific conditions (e.g., when a distance change is clearly disambiguated). To test this idea, Experiment 3 compared effects of actions that were framed as a movement towards and away from social group members with effects of actions that were framed as moving group members up and down. The former actions are framed in AA terms, whereas the latter actions are framed in evaluative terms without relation to approach and avoidance. Deviating from a manikin display used in Experiments 1 and 2, the stimuli (names of social group members) are now moved upwards and downwards on the computer screen in the up-down action framing condition, whereas key presses in the towards-away action framing condition zoomed the stimuli towards and away from the physical self/body.

In the up-down framing condition, half of the participants now received instructions to move members of one fictitious social group upwards and members of another fictitious social group downwards. Participants pressed a key with an up or down arrow when a name was presented

at the centre of the screen, producing a movement of the name in upwards and downwards directions on the screen. In the towards-away framing condition, instructions indicated that participants should move social groups either towards or away from them. Participants pressed the same keys as in the up-down framing condition, but this time a key press gradually increased or decreased the size of the name, creating the illusion that the name approaches or recedes from the participant. Such a zooming effect is often used in AA research to disambiguate AA actions (e.g., Van Dessel, De Houwer, Gast, Smith, et al., 2016; Wiers et al., 2011; Wittekind et al., 2015).

The common-coding account predicts an effect of the affective S-R mapping: participants should prefer the social group linked to a positive action relative to the group linked to a negative action. Because towards and away action words are typically not rated as more strongly valenced than up and down action words (at least in a German-speaking sample: for corresponding results see Eder & Rothermund, 2008), no interaction with type of action framing was expected. The inferential account also predicts an effect of affective S-R mapping. However, this effect might be qualified by an interaction with how the action is framed. Participants might more easily infer stimulus valence on the basis of their own movements towards or away from a stimulus than on the basis of up or down movements performed by a stimulus. Finally, according to a motivational-systems account, a training effect should only be observed in the towards-away framing condition.

## **Method**

**Participants.** A total of 400 English-speaking volunteers participated online via the Prolific Academic website. We excluded data from participants who (a) did not fully complete all questions and tasks (4 participants; i.e., 1.00%), (b) had IAT error rates above 30% across the entire task, or above 40% for any of the critical blocks (31 participants; i.e., 7.75%), (c) completed more than 10% of IAT trials faster than 400ms (3 participants; i.e., 0.75%), (d) made at least one error on the questions that probed memory for the action task instructions (42 participants; i.e., 10.50%), or (e)

had more than 30% errors on label trials in the action task (7 participants; i.e., 1.75 %). Analyses were performed on the data of 313 participants (157 women, mean age = 33,  $SD = 12$ ).

**Procedure.** The procedure of Experiment 3 was largely identical to Experiment 2 with a few exceptions. First, the instructions for the action task specified that (a) participants should move the names of members of one social group *up* by pushing the up-arrow key and move members of the other social group *down* by pushing the down-arrow key (up-down action framing), or that (b) participants should move the names of members of one social group *towards* themselves by pushing the down-arrow key and the members of the other social group *away* from them by pushing the up-arrow key (towards-away action framing).

Second, on each trial of the action task, the name of a member of the Niffites or Luupites was presented on the screen. After participants emitted the correct response, the name moved up or down (in the up-down framing condition) or zoomed towards or away from them (in the towards-away framing condition). The visual action effects consisted of three action frames that each lasted 100 ms and gradually move the name to the final destination on the screen. On label trials, the words UP and DOWN or TOWARDS and AWAY were presented (depending on the action framing condition), respectively; key presses in the label trials produced no visual effects.

At the end of the experiment participants answered an additional question that asked how positive (or negative) they considered each of the two actions.

## Results

**Explicit rating scores.** Explicit rating scores were calculated in accordance with Experiment 1 and were subjected to a 2 (*Action Framing Condition*: up-down, towards-away) x 2 (*Affective S-R Mapping*: positive response to Niffites [up/towards-movement] and negative response to Luupites [down/away-movement] or vice versa) between-subjects ANOVA. We observed a main effect of Affective S-R Mapping,  $F(1,309) = 6.76, p = .010$ , but no interaction of

Affective S-R Mapping and Action Framing Condition,  $F(1,309) = 0.12, p = .73$ . Participants in the towards-away framing condition exhibited a stronger preference for Niffites over Luupites if they had zoomed Niffites names towards themselves and Luupites names away from them ( $M = 0.39, SD = 2.24$ ) than vice versa ( $M = -0.28, SD = 1.94$ ),  $t(155) = 2.01, p = .046, d = 0.32, 95\% CI diff = [2.50, 3.95], BF_1 = 1.10$ . Participants in the up-down framing condition exhibited a stronger preference for Niffites over Luupites when Niffites names were moved upwards and Luupites names downwards ( $M = 0.26, SD = 1.58$ ) than vice versa ( $M = -0.57, SD = 2.13$ ),  $t(154) = 2.76, p = .007, d = 0.44, 95\% CI diff = [0.23, 1.42], BF_1 = 5.41$ . The effect size of the training effect was not larger with a towards-away framing than with an up-down framing,  $d_{diff} = -0.12$ .

**IAT scores.** Reliability of the IAT score was  $r(311) = .89$ . We performed a 2 (*Action Framing Condition*) x 2 (*Affective S-R Mapping*) x 2 (*IAT Block Order: Niffites and positive first, Luupites and positive first*) ANOVA on IAT scores. The factor IAT Block Order was included because inclusion of this factor significantly improved model fit. The ANOVA revealed a significant main effect of Affective S-R Mapping,  $F(1,305) = 11.26, p < .001$ . The interaction with Action Framing Condition was not significant,  $F(1,305) = 0.19, p = .67$ . Participants in the towards-away framing condition exhibited a stronger preference for Niffites over Luupites after zooming Niffites names towards and Luupites names away from them ( $M = 0.18, SD = 0.50$ ) than after a reverse training ( $M = -0.34, SD = 0.42$ ),  $t(155) = 7.02, p < .001, d = 1.12, 95\% CI diff = [0.37, 0.66], BF_1 > 10000$ . Participants in the up-down framing condition exhibited a stronger preference for Niffites over Luupites after a training to move Niffites names upwards and Luupites names downwards ( $M = 0.00, SD = 0.56$ ) compared to a reverse training ( $M = -0.43, SD = 0.46$ ),  $t(154) = 5.14, p < .001, d = 0.82, 95\% CI diff = [0.26, 0.58], BF_1 > 10000$ . The effect size of the training

effect was not significantly larger with a towards-away framing than with an up-down framing,  $d_{diff} = 0.30$ .<sup>4</sup>

## Discussion

The results of Experiment 3 showed that (1) repeatedly moving members of fictitious social groups up and down can produce changes in implicit and explicit evaluations of these social groups and (2) effect size was not significantly different in the condition with zooming effects and up-down effects. These results are consistent with predictions made by the common-coding account and provide further support for the idea that the evaluative action frame is causally involved in AA training effects. Results contrast with predictions made by motivational-systems accounts which assume that an evaluative transfer from actions to stimuli would take place after a training of AA responses but not with other types of valenced actions.

Recall that the inferential account predicted that a stronger effect would emerge when actions were framed as movement towards-away rather than up-down. One may speculate (post-hoc) about why this moderation was not found. Specifically, three factors might have reduced the effect of towards/away movements. First, participants might have recoded towards/away actions to up/down button presses for a performance benefit (by facilitating more rapid and accurate presses of the arrow-keys). In support of this, participants were significantly faster and more accurate on label trials in the up-down framing condition than in the towards-away framing condition (see Appendix). Second, the zooming feature might have reduced the effectiveness of the training procedure. A previous study showed that participants often consider a stimulus movement towards them to be unpleasant (Hsee, Tu, Lu, & Ruan, 2014). This might have reduced

---

<sup>4</sup> Correlational analyses revealed significant correlations between participants' difference in liking ratings of the positive and negative action and the effect of Affective S-R Mapping on implicit and explicit evaluation scores for participants in both action framing conditions (see Appendix).

liking of the stimulus that was zoomed towards. Finally, it is possible that moving something up and down provides a stronger cue for inferences about stimulus valence than moving it towards or away from the self. Because the results of Experiment 3 (and the previous experiments) provided evidence against a motivational-systems account but did not allow us to distinguish between common-coding and inferential accounts of AA training effects, Experiment 4 set out just to do that.

### **Experiment 4**

In Experiment 4, we address the novel question of whether changes in stimulus evaluations occur when participants do not perform AA actions but merely observe AA-related movements. In-line with Experiment 2, on each trial of an AA task, the name of a member of a fictitious social group was presented on-screen and a manikin either moved towards or away from the name. Importantly, however, participants did not control the manikin themselves (i.e., they did not press keys) but simply observed the manikin moving on the screen. Participants were explicitly informed that the manikin was moved by the computer, and they were instructed to merely watch these manikin movements.

According to the common-coding account, an AA action must be executed or cognitively planned for AA training to establish a link between the representation of the action and the representation of the action target (the stimulus). In-line with this assumption, Eder and Rothermund (2008, Experiment 5) observed no facilitative effect of a congruency relation between affective stimuli and valenced action labels when participants were not ready to enact the action referred to by the verbal label. Hence, the passive observation of manikin movements without simulation or engagement in own action planning should not produce a change in stimulus evaluation.

In contrast, the inferential account predicts that performance of an AA action is not required for a change in stimulus evaluation. When a person learns that a certain stimulus is approached, this should trigger an inference about the valence of the stimulus (e.g., “*Stimulus A is approached and approached things are typically positive, and hence, stimulus A is positive*”, Van Dessel, De Houwer, & Gast, 2016), even when the person is not responsible for the action. Because the passive observation of AA movements can lead to the acquisition of the propositional knowledge that a stimulus is approached or avoided, this should be sufficient to produce a change in stimulus evaluation.

### **Method**

**Participants.** A total of 152 English-speaking volunteers participated online via the Prolific Academic website. We excluded the data from participants who (a) did not fully complete all questions and tasks (1 participant; i.e., 0.66%), (b) had IAT error rates above 30% across the entire task, or above 40% for any of the critical blocks (4 participants; i.e., 2.63%), (c) completed more than 10% of IAT trials faster than 400ms (4 participants; i.e., 2.63%), (d) made at least one error on the questions that probed contingency memory (12 participants; i.e., 7.89%), or (e) indicated that they did not believe that a chance algorithm determined selection of the groups that the manikin was running towards or away from (14 participants; i.e., 10.69%). Analyses were performed on the data of 117 participants (61 women, mean age = 38,  $SD = 12$ ).

**Procedure.** The procedure of Experiment 4 was identical to the condition in Experiment 2 with approach-avoid action framing and manikin location below the word, except for the following differences. First, instructions for the AA task specified that participants would see a manikin approach or avoid the names of a Niffite or Luupite in each trial and that they should pay careful attention to what group the manikin was moving towards or away from. Instructions also made clear that the computer is responsible for the manikin movements, not the participant or any another

person, and that a random computer algorithm has selected the group that would be approached or avoided by the manikin. Latter information was included to prevent participants from construing manikin movements as intentional actions of another social agent (i.e., the manikin) and from engaging in simulating the observed movement (Kunz, Creem-Regehr, & Thompson, 2009).

Second, on each trial of the AA task, the name of a member of the Niffites or Luupites was presented on the screen for 690ms (the mean latency in the AA training task of Experiment 2) followed by the manikin movement. The manikin moved upwards (i.e., approach movement) when a Niffites name was presented and downwards (i.e., avoidance movement) when a Luupites name was presented (or vice versa). There were no label trials in this version of the AA task.

Third, three additional questions about the AA task were presented after the evaluation measures. (1) Two contingency memory questions asked participants to indicate what action they had observed when they saw names of Niffites or Luupites. Possible options were: “The manikin approached Niffites/Luupites”, “The manikin avoided Niffites/Luupites”, or “I don’t remember”. (2) Participants also indicated whether they had understood that a chance algorithm determined what groups the manikin was running towards and what group the manikin was running away from (possible options: “Yes, I understood that” or “No, I believed the selection has a deeper meaning”) and (3) what they believed to be the purpose of the study (open response format).

## Results

**Explicit rating scores.** The explicit rating scores were subjected to a 2 (*Stimulus-Action Mapping*: Niffites approached/Luupites avoided vs. Niffites avoided/Luupites approached) x 2 (*Task Order*) between-subjects ANOVA. We observed a main effect of Stimulus-Action Mapping,  $F(1,113) = 18.54, p < .001$ . Participants exhibited a stronger preference for Niffites over Luupites when they had observed the manikin approach Niffites and avoid Luupites ( $M = 0.81, SD = 2.32$ )

than when they had observed the manikin approach Luupites and avoid Niffites ( $M = -1.12$ ,  $SD = 2.68$ ),  $t(115) = 4.16$ ,  $p < .001$ ,  $d = 0.77$ , 95% CI diff = [1.01, 2.85],  $BF_1 = 335.92$ . There was a significant interaction between Stimulus-Action Mapping and Task Order,  $F(1,113) = 9.15$ ,  $p = .003$ . The Stimulus-Action Mapping effect was significant when participants first provided the explicit ratings and then performed the IAT,  $F(1,56) = 25.47$ ,  $p < .001$ , but not when they performed the IAT first,  $F(1,57) = 0.86$ ,  $p = .36$ .

**IAT scores.** The Spearman-Brown corrected split-half reliability of the IAT score was  $r(115) = .88$ . A 2 (*Stimulus-Action Mapping*) x 2 (*Task Order*) x 2 (*IAT Order*) ANOVA on IAT scores revealed a main effect of IAT Order,  $F(1,123) = 13.47$ ,  $p < .001$ . More importantly, we also observed a main effect of Stimulus-Action Mapping,  $F(1,123) = 10.15$ ,  $p = .002$ . Participants exhibited a stronger preference for Niffites over Luupites when they had approached Niffites and avoided Luupites ( $M = -0.07$ ,  $SD = 0.50$ ) than when they had approached Luupites and avoided Niffites ( $M = -0.33$ ,  $SD = 0.45$ ),  $t(115) = 2.87$ ,  $p = .005$ ,  $d = 0.53$ , 95% CI diff = [0.08, 0.43],  $BF_1 = 7.43$ . There were no other main or interaction effects,  $F_s < 1.84$ ,  $p_s > .17$ .

**Additional analyses.** Inspection of participants' responses to the question asking to what extent participants had imagined themselves to be the manikin revealed that a total of 47 participants (i.e., 40.17% of the entire sample) reported that they had imagined themselves to be the manikin to some extent. Because this could indicate that some participants still engaged in own action planning, we deviated from our pre-registered data-analysis plan by performing ANOVA's on implicit and explicit evaluations that included participants' response to this question (Manikin Identification) as a covariate. These analyses revealed a main effect of Stimulus-Action Mapping on both implicit and explicit evaluations,  $F_s > 7.73$ ,  $p_s < .007$ . The interaction between Stimulus-Action Mapping and Manikin Identification was non-significant on explicit rating scores,  $F(1,113) = 1.14$ ,  $p = .29$ , and marginally significant on IAT scores,  $F(1,113) = 3.30$ ,  $p = .072$ . The effect of

Stimulus-Action Mapping on IAT scores was bigger when participants imagined themselves to be the manikin to a *lesser* extent. Note that this trend could have occurred because participants who did not identify with the manikin were more attentive in general such that they better knew that they should not identify with the manikin and also showed a bigger observation-based AA effect. Finally, we also performed analyses on the data of only those 70 participants who indicated no identification with the manikin. Crucially, these analyses revealed the Stimulus-Action Mapping effect both on explicit rating scores,  $t(68) = 3.30$ ,  $p = .002$ ,  $d = 0.79$ , 95% CI diff = [0.63, 2.56],  $BF_1 = 21.13$ , and IAT scores,  $t(68) = 3.98$ ,  $p < .001$ ,  $d = 0.95$ , 95% CI diff = [0.20, 0.61],  $BF_1 = 138.51$  (see Appendix for a complete description of these results).

## **Discussion**

The results of Experiment 4 reveal that a passive observation of AA movements in response to the members of fictitious social groups can produce robust changes in implicit and explicit evaluations of those social groups. These effects were only predicted by the inferential account and provide further support for the idea that inferences are causally involved in AA effects. The observation-based AA effect cannot be explained by the common-coding account which assumes that action effects on stimulus evaluation require the performance or cognitive planning of valenced actions. Results also do not fit well with motivational-systems accounts for which motivations to perform AA movements play a central role.

## **General Discussion**

Four experiments examined the mental mechanisms that underlie effects of AA training on stimulus evaluation. In Experiment 1, training participants to move a manikin (representing the self) up and down in response to a nonword led to a change in explicit evaluations of the nonwords regardless of whether the movements implied a distance decrease (approach) or increase

(avoidance). Experiment 2 showed that evaluative action framing is causally involved in AA training effects. Both the training of upwards-downward and approach-avoidance (AA) movements in response to fictitious social groups was effective in changing implicit and explicit evaluations of the social groups when the movements were framed in these (evaluative) terms. However, evaluative effects were stronger with AA compared to up-down action framing. Experiment 3 provided additional evidence that the impact of stimulus-action relations on stimulus evaluations is not restricted to approach and avoidance effects such that moving stimuli upwards and downwards can change evaluations of those stimuli even when there is no distance change involved. Finally, Experiment 4 showed that changes in stimulus evaluations can occur even when AA movements are passively observed without engaging in own action. Overall, these results provide strong evidence that AA training effects are mediated by general mechanisms of evaluative learning (rather than processes tied to specific motivational AA systems) and are most parsimoniously explained with cognitive inferences.

### **Theoretical implications**

Our findings are not in-line with predictions of motivational-systems accounts of AA training effects that explain these effects with a (re-)training of associations to motivational systems promoting approach and avoidance (defined as distance regulations; Neumann et al., 2003). In Experiments 1-3, training people to perform upwards and downwards movements in response to neutral stimuli affected evaluations of these stimuli, although these movements were unrelated and/or orthogonal to AA movements in terms of distance regulations. Experiment 4 even shows that AA effects do not require self-generated AA actions to be performed at all, which contradicts the idea that distance change to the self is a key determinant of AA effects (see Phills et al., 2011). To accommodate our results, these traditional accounts of AA training would need to make a number of additional assumptions. Most importantly, it seems necessary to postulate that another

process, one that does not depend on the activation of motivational systems of approach-avoidance, also contributes to AA training effects and that this other process can produce (strong) effects of up-down trainings and observations on stimulus evaluations. Note that it is also possible that motivational systems play a small role in attitudinal effects that are obtained on the basis of distance change movements with manikins but play a bigger role in effects of other movements (e.g., full-body movements).

The current findings are better explained by alternative accounts that propose more general (evaluative) learning mechanisms as a causal explanation for AA training effects. First, the common-coding account (Eder & Klauer, 2009) can explain the results of Experiments 1 to 3. In these experiments, changes in stimulus evaluations were found after participants performed up-down actions that were unrelated to approach and avoidance in terms of a distance regulation (cf. Eder & Rothermund, 2008). The size of evaluative transfer from the trained action to the stimuli correlated with the self-reported valence of the action frame words, which fits with the idea that affective features of actions are transferred to stimuli in a common-coding domain after training. This correlation was observed for effects on explicit measures in Experiments 1, 2 and 3 and on the implicit IAT measure in Experiment 3 (but not Experiments 1 and 2). The results of Experiment 4 could however not be explained by the common-coding account. According to this account, action planning is necessary for a training-induced AA effect. If participants did not engage in action planning in Experiment 4, then there should have been no liking change after the observation phase. It is possible that the manikin movement in Experiment 4 triggered a simulation of the observed movement to some extent, despite the clear information that the computer was controlling the manikin (see Kunz et al., 2009). However, our results indicated that observation-based AA effects are not stronger when participants imagine themselves to be the manikin to a greater extent.

Yet, it is still possible that participants did not accurately indicate the extent to which they imagined to be the manikin or that participants imagined moving toward or away from the social groups along with the manikin even if they didn't imagine themselves being the manikin. To address this, future research might suppress action simulation of observed movements more actively with the use of an interfering motor task (e.g., Springer et al., 2011) or might describe the manikin as a member of a disliked out-group to minimize the extent to which participants identify with the manikin.

The current results are most parsimoniously explained via the inferential account (Van Dessel, De Houwer, & Gast, 2016). According to this account, the observed changes in liking were mediated by inferences about the evaluative properties of stimuli that were derived from propositions concerning the trained actions. This account argues that the way in which an action is framed should influence the types of propositions that emerge about it (and thus the subsequent evaluative inferences). For instance, evaluative inferences will be made on the basis of AA actions whenever (1) action framing instructions disambiguate the meaning of these actions and (2) specific stimuli are then related to the actions (see Van Dessel, De Houwer, Gast, Smith, et al., 2016). This is consistent with the observed effects in the AA framing conditions of Experiments 2 and 3. However, in situations where there is no AA framing of (distance-regulating) actions participants may not construe those actions as approach and avoidance. Moreover, participants may not register the contingency relation between stimuli and AA actions (as in Experiment 1) which will prevent them from exhibiting AA effects (see Van Dessel, De Houwer, & Gast, 2016). Note that the inferential account assumes that the performance of actions other than AA can also produce changes in stimulus evaluations if those actions lead to inferences about evaluative properties of target stimuli. Such evaluative inferences may be facilitated when actions are framed in positive or

negative terms. For instance, participants may think of the word ‘up’ as positive because it is often used as a synonym for happy (and of the word ‘down’ as negative because it is often used as a synonym for unhappy). As a result, the action of moving up or down in response to a stimulus (Experiment 2) or moving a stimulus up or down (Experiment 3) may facilitate evaluative inferences (e.g., ‘the stimulus helps me or the manikin to be up/happy so it must be positive’ or ‘the stimulus itself is up/happy so it is positive’) and produce changes in stimulus evaluations. Experiment 4 further showed that the mere observation of AA actions can also produce changes in implicit and explicit evaluations, a prediction made by the inferential account but not by either motivational-systems accounts or the common-coding account.

It is also possible that a joint contribution of inferential processes and motivational or common-coding processes to AA training effects could explain the current results. For instance, both common-coding and inferential processes might be involved in standard AA training effects (as reported in the approach-avoid framing condition Experiment 2) whereas only inferential processes might be involved in observation-based AA effects (as reported in Experiment 4). In-line with this idea, additional analyses that compared results of Experiments 2 and 4 revealed that the experience-based AA effects observed in Experiment 2 were stronger than the observation-based AA effects observed in Experiment 4 (see Appendix). Similarly, it is possible that motivational-systems are involved in standard AA training effects but not in training effects that do not contain AA action framing (e.g., in the up-down framing condition of Experiment 2). Note, however, that the inferential account also predicted smaller effects in the non-standard AA conditions (of Experiments 2 and 4) because inferences about stimulus valence should be more likely to occur with AA framing than with up-down framing (Experiment 2) and with performed actions than with observed actions (Experiment 4). It is also noteworthy that effect sizes observed

in the non-standard AA conditions were larger in absolute magnitude than the estimated benefit in effect size in standard AA conditions compared to non-standard AA conditions. Hence, it seems plausible that a (very) large part of the variance in (standard) AA training effects might be due to inferential influences.

Note that a joint contribution of processes could also explain AA training effects that emerge under other conditions than the conditions that were examined in the current experiments. For instance, AA motivational systems might contribute to AA effects only under specific conditions such as when stimuli are highly appetitive to participants (and that participants might therefore have a stronger tendency to approach or avoid; see Wiers, Rinck, Dictus, & Van den Wildenberg, 2009). Future studies are required to obtain a good estimate of the separate contributions of the proposed mechanisms to AA effects (that occur under different conditions).

### **Practical implications**

The present results have important practical implications. First, they indicate that AA actions are not the only actions that can be used to change stimulus evaluations. Moving up or down in response to a stimulus (Experiment 1 and 2) and moving a stimulus up or down (Experiment 3) also influence stimulus evaluation. Our results indicate that training many different valenced actions should produce changes in liking. This recommendation is consistent with modern embodiment accounts that associate a variety of behaviors with positive and negative valence (Laird & Strout, 2007; Niedenthal, 2007). It also seems that actually performing (valenced) actions is not necessary for obtaining changes in stimulus evaluations. Merely observing those actions can already lead to (strong) effects. Observation-based learning of stimulus-action contingencies might provide a new, promising route for changing stimulus evaluation.

Second, our research provides information about possible boundary conditions of AA training effects by showing that these effects depend on the evaluative framing of actions. The same action can produce opposite effects depending on how the goals of the trained actions are represented in the cognitive system (Experiments 1 and 2). This is consistent with recent work indicating that instructions might be a crucial factor in AA training effects. For instance, Van Dessel, De Houwer, Gast, and Smith (2015) obtained evidence that AA effects can also be obtained following mere instructions about AA actions, in the absence of actually performed actions. The fact that many AA training studies do not use task instructions that disambiguate the valence of AA actions might explain why some studies obtained no or even reversed effects (e.g., Becker Jostmann, Wiers, & Holland, 2015; Vandebosch & De Houwer, 2011; for a discussion see Mertens, Van Dessel, & De Houwer, 2017).

Third, our results suggest that AA training can have benefits compared to other paradigms that involve the repeated pairing of valenced actions and stimuli. Experiment 2 found that a training of button presses had stronger effects on implicit and explicit stimulus evaluations when these actions were framed as approach-avoidance than when framed as up-down movements. At the same time, not every training paradigm involving AA action framing is created equal. An AA training procedure with a zoom effect (Experiment 3) produced smaller effects on stimulus evaluations compared to a manikin movement task (Experiment 2). This difference might indicate that an explicit reference of the training procedure to approach/avoidance goals is more effective. Alternatively, it is possible that a zooming effect is a problematic feature that reduces the AA action effects (see also Vandebosch & De Houwer, 2011). Future work is needed because AA training studies often use different action instructions and visualizations to disambiguate AA movements. It may be interesting to directly compare effects of other AA training procedures to end up with

the best possible procedure for changing stimulus evaluations. One study found that effects of a manikin AA training task on racial prejudice were similar in magnitude to effects of AA training with joystick movements (Van Dessel, Gast, Roets, & De Houwer, 2017), but it can be important to perform such comparisons also in other contexts.

One important issue that we did not address in our studies is whether the current results relate to AA training effects on clinically relevant outcomes such as alcohol consumption (Wiers et al., 2011) or smoking behavior (Wittekind et al., 2015). According to the present research, a training of consumptive behaviors might be similarly influenced by the movement goals that were implemented for the AA training and could critically involve inferential processing. It should be noted, however, that the present research exclusively examined effects on evaluations of nonword stimuli and fictitious social groups. Additional research is needed to examine whether our findings generalize to behavioral outcomes and real-life objects with pre-existing attitudes.

### **Concluding remarks**

Taken together, our findings challenge the idea that AA training effects on evaluations of novel stimuli are caused by a rewiring or reconfiguration of associations to motivational systems specialized for approach and avoidance. Stimulus evaluations can be changed via procedures that relate valenced actions to stimuli, and the outcome of the training procedure critically depends on the movement goals for the training phase. These results offer a new perspective on AA training effects, providing evidence that general learning mechanisms can cause evaluative transfer effects from actions to stimuli. Results further indicate that inferential processes might be strongly involved in these effects. We hope that this knowledge will help improve action training procedures in clinical and other domains.

### References

- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology*. Worcester, Mass: Clark University Press.
- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer Jr. (Ed.), *Advances in social cognition*, Vol. 10. (pp. 1–61). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. <http://CRAN.R-project.org/package=lme4>
- Becker, D., Jostmann, N. B., Wiers, R. W., & Holland, R. W. (2015). Approach avoidance training in the eating domain: Testing the effectiveness across three single session studies. *Appetite*, 85(April 2016), 58–65. doi:10.1016/j.appet.2014.11.017
- Beckers, T., De Houwer, J., & Eelen, P. (2002). Automatic integration of non-perceptual action effect features: The case of the associative affective Simon effect. *Psychological Research*, 66, 166-173.
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. 6. (pp. 1–62). New York: Academic Press.
- Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, 61, 27–41. doi:10.1037/0003-066X.61.1.27
- Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes II: Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*, 65, 5–17. doi:10.1037/0022-3514.65.1.5
- Darwin, C. (1997). *The expression of the emotions in man and animals*. New York: Oxford University Press. (Original work published in 1872).

- De Houwer, J. (2014). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass*, 8, 342-353. doi:10.1111/spc3.12111
- De Houwer, J., Crombez, G., Baeyens, F., & Hermans, D. (2001). On the generality of the affective Simon effect. *Cognition & Emotion*, 15, 189–206. doi:10.1080/0269993004200051
- De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, 10, 230-241.
- Eder, A. B., & Hommel, B. (2013). Anticipatory control of approach and avoidance: An ideomotor approach. *Emotion Review*, 5(3), 275–279. doi:10.1177/1754073913477505
- Eder, A. B., & Klauer, K. C. (2007). Common valence coding in action and evaluation: Affective blindness towards response-compatible stimuli. *Cognition & Emotion*, 21, 1297–1322.
- Eder, A. B., & Klauer, K. C. (2009). A common-coding account of the bidirectional evaluation–behavior link. *Journal of Experimental Psychology: General*, 138, 218–235. doi:10.1037/a0015220
- Eder, A. B., Müsseler, J., & Hommel, B. (2012). The structure of affective action representations: temporal binding of affective response codes. *Psychological Research*, 76, 111–118. <https://doi.org/10.1007/s00426-011-0327-6>
- Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis)match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General*, 137, 262–281. doi:10.1037/0096-3445.137.2.262
- Eder, A. B., Rothermund, K., & De Houwer, J. (2013). Affective compatibility between stimuli and response goals: A primer for a new implicit measure of attitudes. *PLoS ONE*, 8(11), e79210. doi:10.1371/journal.pone.0079210

- Eder, A. B., Rothermund, K., De Houwer, J., & Hommel, B. (2015). Directive and incentive functions of affective action consequences: an ideomotor approach. *Psychological Research, 79*, 630–649. doi:10.1007/s00426-014-0590-4
- Eder, A. B., Rothermund K., & Hommel, B. (2016). Commentary: Contrasting motivational orientation and evaluative coding accounts: on the need to differentiate the effectors of approach/avoidance responses. *Frontiers in Psychology, 7*, Article 163.
- Elliot, A. J., Eder, A. B., & Harmon-Jones, E. (2013). Approach–avoidance motivation and emotion: Convergence and divergence. *Emotion Review, 5*, 308–311. <https://doi.org/10.1177/1754073913477517>
- Förster, J. (2001). Success/failure feedback, expectancies, and approach/avoidance motivation: How regulatory focus moderates classic relations. *Journal of Experimental Social Psychology, 37*, 253–260. doi:10.1006/jesp.2000.1455
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480. doi:10.1037/0022-3514.74.6.1464
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216. doi: 10.1037/0022-3514.85.2.197
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): a framework for perception and action planning. *Behavioral Brain Sciences, 24*, 849–937. doi:10.1017/S0140525X01000103

- Huijding, J., Muris, P., Lester, K. J., Field, A. P., & Joesse, G. (2011). Training children to approach or avoid novel animals: Effects on self-reported attitudes and fear beliefs and information-seeking behaviors. *Behaviour Research and Therapy*, *49*, 606–613. <https://doi.org/10.1016/j.brat.2011.06.005>
- Hsee, C., Tu, Y., Lu, Z., & Ruan, B. (2014). Approaching aversion: Negative reactions toward approaching stimuli. *Journal of Personality and Social Psychology*, *106*, 699–712. doi:10.1037/a0036332
- Jones, C. R., Vilensky, M. R., Vasey, M. W., & Fazio, R. H. (2013). Approach behavior can mitigate predominately univalent negative attitudes: Evidence regarding insects and spiders. *Emotion*, *13*, 989-996. doi:10.1037/a0033164
- Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial evaluations and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology*, *92*, 957-971. doi:10.1037/0022-3514.92.6.957
- Kawakami, K., Steele, J. R., Cifa, C., Phills, C. E., & Dovidio, J. F. (2008). Approaching math increases math = me and math = pleasant. *Journal of Experimental Social Psychology*, *44*, 818–825. doi:10.1016/j.jesp.2007.07.009
- Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*, 36–69. doi:10.1037/0096-3445.135.1.36
- Krieglmeyer, R., Deutsch, R., De Houwer, J. & De Raedt, R. D. (2010). Being moved: Valence

- activates approach-avoidance behavior independently of evaluation and approach-avoidance intentions. *Psychological Science*, *21*, 607–613. doi:10.1177/0956797610365131
- Krieglmeyer, R., Houwer, J. D., & Deutsch, R. (2013). On the nature of automatically triggered approach–avoidance behavior. *Emotion Review*, *5*, 280-284.
- Kunz, B. R., Creem-Regehr, S. H., & Thompson, W. B. (2009). Evidence for motor simulation in imagined locomotion. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1458–1471. <https://doi.org/10.1037/a0015786>
- Laham, S. M., Kashima, Y., Dix, J., Wheeler, M., & Levis, B. (2014). Elaborated contextual framing is necessary for action-based attitude acquisition. *Cognition & Emotion*, *28*, 1119-1126. doi:10.1080/02699931.2013.867833
- Laird, J. D., & Strout, S. (2007). Emotional behaviors as emotional stimuli. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment*. (pp. 54–64). New York: Oxford University Press.
- Markman, A. B., & Brendl, C. M. (2005). Constraining theories of embodied cognition. *Psychological Science*, *16*, 6-10. doi: 10.1111/j.0956-7976.2005.00772.x
- Memelink, J., & Hommel, B. (2013). Intentional weighting: a basic principle in cognitive control. *Psychological Research*, *77*(3), 249–259. <https://doi.org/10.1007/s00426-012-0435-y>
- Mertens, G., Van Dessel, P., & De Houwer, J. (2017). Approach-avoidance and fear: Intersecting regularities modulate effects of approach-avoidance training on stimulus evaluations. Manuscript under review.
- Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*, 1002–1005.

- Neumann, R., & Strack, F. (2000). Approach and avoidance: The influence of proprioceptive and exteroceptive cues on encoding of affective information. *Journal of Personality and Social Psychology, 79*, 39–48. doi: 10.1037/0022-3514.79.1.39
- Neumann, R., Förster, J., & Strack, F. (2003). Motor compatibility: The bidirectional link between behavior and evaluation. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion*. (pp. 371–391). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Phaf, R. H., Mohr, S. E., Rotteveel, M., & Wicherts, J. M. (2014). Approach, avoidance, and affect: a meta-analysis of approach-avoidance tendencies in manual reaction time tasks. *Frontiers in Psychology, 5*. <https://doi.org/10.3389/fpsyg.2014.00378>
- Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and Blacks with approach behaviors. *Journal of Personality and Social Psychology, 100*, 197–210. doi: 10.1037/a0022159
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. URL: <http://www.jstatsoft.org/v48/i02/>
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237. doi: 10.3758/PBR.16.2.225
- Schönbrodt, F., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*. Advance online publication. doi:10.1037/met0000061

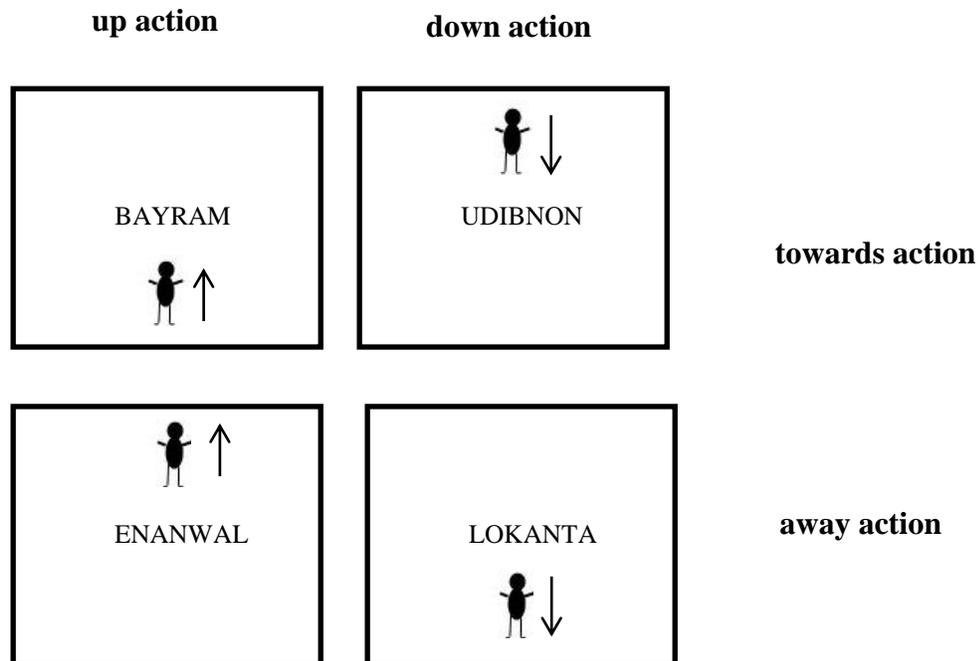
- Sharpe, D., & Whelton, W. J. (2016). Frightened by an Old Scarecrow: The remarkable resilience of demand characteristics. *Review of General Psychology*.
- Solarz, A. K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of Experimental Psychology*, *59*, 239–45. doi:10.1037/h0047274
- Springer, A., Brandstädter, S., Liepelt, R., Birngruber, T., Giese, M., Mechsner, F., & Prinz, W. (2011). Motor execution affects action prediction. *Brain and Cognition*, *76*, 26–36. <https://doi.org/10.1016/j.bandc.2011.03.007>
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220-247. doi:10.1207/s15327957pspr0803\_1
- Vandenbosch, K., & De Houwer, J. (2011). Failures to induce implicit evaluations by means of approach-avoid training. *Cognition and Emotion*, *25*, 1311-1330. doi:10.1080/02699931.2011.596819
- Van Dessel, P., De Houwer, J., & Gast, A. (2016). Approach-avoidance training effects are moderated by awareness of stimulus-action contingencies. *Personality and Social Psychology Bulletin*, *42*, 81-93. doi:10.1177/0146167215615335
- Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. T. (2015). Instruction-based approach–avoidance effects: Changing stimulus evaluation via the Mere Instruction to Approach or Avoid Stimuli. *Experimental Psychology*, *62*, 161-169. doi:10.1027/1618-3169/a000282
- Van Dessel, P., De Houwer, J., Gast, A., Smith, C. T., & De Schryver, M. (2016). Instructing implicit processes: When instructions to approach or avoid influence implicit but not

- explicit evaluation. *Journal of Experimental Social Psychology*, 63, 1-9.  
doi:10.1016/j.jesp.2015.11.002
- Van Dessel, P., De Houwer, J., Roets, A., & Gast, A. (2016). Failures to Change Stimulus Evaluations by means of Subliminal Approach and Avoidance Training. *Journal of Personality and Social Psychology*, 110, e1-e15. doi:10.1037/pspa0000039
- Van Dessel, P., De Houwer, J., & Smith, C. T. (2017). Relational information moderates approach-avoidance instruction effects on implicit evaluation. *Acta Psychologica*. Advance online publication. doi:10.1016/j.actpsy.2017.03.016
- Van Dessel, P. Gast, A., Roets, A., & De Houwer, J. (2017). Effects of approach-avoidance instructions and approach-avoidance training on racial prejudice. Manuscript in preparation.
- Van Dessel, P., Gawronski, B., Smith, C. T., & De Houwer, J. (2017). Mechanisms underlying approach-avoidance instruction effects on implicit evaluation: Results of a preregistered adversarial collaboration. *Journal of Experimental Social Psychology*, 69, 23-32.  
doi:10.1016/j.jesp.2016.10.004
- Van Dessel, P., Hughes, S. & De Houwer, J. (2017). Consequence-based approach-avoidance: a new and improved method for changing behavior. Manuscript under review.
- Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science*, 22, 490–497. doi:10.1177/0956797611400615

- Wiers, R. W. Rinck, M., Dictus, M. & Van den Wildenberg, E. (2009). Relatively strong automatic appetitive action-tendencies in male carriers of the OPRM1 G-Allele. *Genes Brain & Behavior*, 8, 101-106.
- Wittekind, C. E., Feist, A., Schneider, B. C., Moritz, S., & Fritzsche, A. (2015). The approach-avoidance task as an online intervention in cigarette smoking: A pilot study. *Journal of Behavior Therapy and Experimental Psychiatry*, 46, 115–120. doi:10.1016/j.jbtep.2014.08.006
- Woud, M. L., Maas, J., Becker, E.S., & Rinck, M. (2013). Make the manikin move: Symbolic approach-avoidance responses affect implicit and explicit face evaluations. *Journal of Cognitive Psychology*, 25, 738-744. doi:10.1080/20445911.2013.817413
- Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning? *Quarterly Journal of Experimental Psychology*, 67, 2105-2122. doi: 10.1080/17470218.2014.907324
- Zogmaister, C., Perugini, M., & Richetin, J. (2016). Motivation modulates the effect of approach on implicit preferences. *Cognition and Emotion*, 5, 890-911. doi:10.1080/02699931.2015.1032892

### **Acknowledgments**

Pieter Van Dessel is supported by a Post-Doctoral fellowship of the Scientific Research Foundation, Flanders (FWO-Vlaanderen). Sean Hughes is supported by Methusalem Grant BOF16/MET\_V/002 of Ghent University. Andreas Eder is supported by a grant from the German Research Foundation (DFG ED201/3-1). The authors would like to thank Jan De Houwer for his comments on an earlier draft of this manuscript.



**Figure 1.** Illustration of the four experimental conditions in Experiment 1. Participants were instructed to move the manikin upwards in response to two nonwords and downwards in response to two other nonwords. Initial placement of the manikin was fixed for each nonword such that for one nonword the up action constituted a movement towards the nonword (e.g., BAYRAM), for one nonword the up action constituted a movement away from the nonword (e.g., ENANWAL), for one nonword the down action constituted a movement towards the nonword (e.g., UDIBNON), and for one nonword the down action constituted a movement away from the nonword (e.g., LOKANTA).

## Appendix

### Additional analyses for Experiment 1

**Linear mixed effects analyses.** For explicit ratings, we fitted a model that included *Participant* as random factor and *Vertical Movement* and *Distance Change* as fixed factors. We also included the random factor of *Word* (BAYRAM, UDIBNON, LOKANTA, ENANWAL) to control for variance due to the specific word that was evaluated. Including this random factor significantly improved model fit,  $\chi^2(1) = 46.82, p < .001$ . We did not include the fixed factor of *Task Order* (IAT first, explicit rating task first) because this factor did not improve model fit,  $\chi^2(4) = 3.43, p = .49$ . Similar to the ANOVA results, we observed a main effect of Vertical Movement,  $\chi^2(1) = 46.91, p < .001$ , but no main effect or interaction effect involving the factor Distance Change,  $\chi^2s < 0.50, ps > .48$ .

For IAT scores, we fitted a model that included *Participant* as random factor and the fixed factor of *IAT type* (IAT 1, IAT2). We included the random factor of *IAT Score* (12 possible scores: IAT score representing a preference of UDIBNON over BAYRAM / of LOKANTA over ENANWAL / ...) because this factor significantly improved model fit,  $\chi^2(1) = 8.23, p = .004$ . We did not include the fixed factors of *IAT order* (IAT1 first, IAT2 first) and *Task Order* because they did not improve model fit,  $\chi^2s < 0.85, ps > .65$ . In accordance with the *t*-test analyses, we observed no main effect of IAT type,  $\chi^2(1) = 0.25, p = .86$ . The overall intercept of the model was also non-significant,  $\chi^2(1) = 1.16, p = .28$ .

### Additional analyses for Experiment 2

**Manikin task trials.** Overall, participants' mean reaction time (RT) was 865 ms ( $SD = 372$ ). Participants made an error on 5.98% of the trials ( $SD = 9.10\%$ ). Participants in the up-down framing condition made 2.91% errors in the label trials ( $SD = 5.54\%$ ) and their mean RT was 769

ms ( $SD = 299$ ). Participants in the approach-avoid framing condition made more errors on label trials ( $M = 6.03\%$ ,  $SD = 8.00\%$ ),  $t(364) = 4.33$ ,  $p < .001$ , and were slower ( $M = 889$  ms,  $SD = 389$ ),  $t(364) = 3.37$ ,  $p < .001$ .

**Correlational analyses.** For these analyses, we calculated an IAT score and explicit rating score that indicated a preference for the social group in response to which participants had performed a positive action (up/approach) over the social group in response to which participants had performed a negative action (down/avoid). These two scores correlated significantly for participants in the up-down framing condition,  $r(166) = .30$ ,  $p < .001$ , but not for participants in the approach-avoid framing condition,  $r(170) = .11$ ,  $p = .15$ .

Participants indicated that, in the manikin task, they imagined themselves to be the manikin to a moderate extent ( $M = 5.81$ ,  $SD = 2.14$ ). This index did not correlate significantly with the IAT or explicit rating scores for participants in either of the two framing conditions,  $r_s < .04$ ,  $p_s > .63$ .

Participants indicated that they liked the word ‘up’ ( $M = 6.50$ ,  $SD = 1.47$ ) more than they liked the word ‘down’ ( $M = 3.97$ ,  $SD = 1.35$ ),  $t(342) = 18.92$ ,  $p < .001$ ,  $d_z = 1.02$ , 95% CI diff = [2.26, 2.79]. The mean difference in ratings of the words correlated positively with the explicit rating score for participants in the up-down action framing condition,  $r(170) = .15$ ,  $p = .044$ , but not with any of the other scores,  $r_s < .06$ ,  $p_s > .47$ . Participants also indicated that they liked the word ‘approach’ ( $M = 6.46$ ,  $SD = 1.29$ ) more than they liked the word ‘avoid’ ( $M = 3.21$ ,  $SD = 1.40$ ),  $t(342) = 27.61$ ,  $p < .001$ ,  $d_z = 1.49$ , 95% CI diff = [3.02, 3.49]. Participants’ mean difference in rating of the words was significantly larger than participants’ mean difference in ratings of the words ‘up’ and ‘down’,  $t(342) = 4.86$ ,  $p < .001$ , 95% CI diff = [0.43, 1.02]. The mean difference in rating of the words correlated positively with the explicit rating score for participants in the

approach-avoid action framing condition,  $r(170) = .24$ ,  $p = .001$ , but not with any of the other scores,  $r_s < .07$ ,  $p_s > .31$ .

Participants most often reported correctly what action the manikin had made in response to each of the fictitious groups that they did not receive instructions about ( $M = 1.69$ ,  $SD = 0.65$ ) (approach/avoid for participants in the up-down framing condition and up/down for participants in the approach-avoid framing condition). In total, 274 participants (i.e., 79.88%) indicated the correct contingency for both groups. This score did not correlate significantly with any of the implicit or explicit evaluation scores in any of the framing conditions,  $r_s < .11$ ,  $p_s > .16$ .

### **Additional analyses for Experiment 3**

**Manikin task trials.** Overall, participants mean RT was 932 ms ( $SD = 379$ ). Participants made an error on 5.93% of the trials ( $SD = 7.36\%$ ). Participants in the up-down movement task made 2.39% errors in the label trials ( $SD = 4.78\%$ ) and their mean RT was 744 ms ( $SD = 248$ ). Participants in the zooming task made more errors on label trials ( $M = 7.82\%$ ,  $SD = 11.76\%$ ),  $t(360) = 5.70$ ,  $p < .001$ , and were slower ( $M = 1196$  ms,  $SD = 1039$ ),  $t(360) = 5.61$ ,  $p < .001$ .

**Correlational analyses.** We calculated an IAT score and explicit rating score that indicated a preference for the social group in response to which participants had performed a positive action (up/towards) over the social group in response to which participants had performed a negative action (down/away). These two scores correlated significantly for participants in the up-down movement task,  $r(154) = .33$ ,  $p < .001$ , but not for participants in the zooming task,  $r(155) = .09$ ,  $p = .25$ .

Participants in the up-down movement task indicated that they liked the ‘up’ action ( $M = 6.17$ ,  $SD = 1.29$ ) more than they liked the ‘down’ action ( $M = 4.65$ ,  $SD = 1.49$ ),  $t(155) = 8.95$ ,  $p < .001$ ,  $d_z = 0.72$ , 95% CI diff = [1.18, 1.85]. Participants’ mean difference in ratings of the actions

correlated positively with the explicit rating score,  $r(154) = .25, p = .002$  and with the IAT score,  $r(154) = .23, p = .004$ . Participants in the zooming task indicated that they liked the ‘towards’ action ( $M = 6.28, SD = 1.29$ ) more than they liked the ‘away’ action ( $M = 4.15, SD = 1.44$ ),  $t(156) = 11.09, p < .001, d_z = 0.89, 95\% \text{ CI diff} = [1.75, 2.51]$ . Participants’ mean difference in rating of the positive and negative actions was significantly bigger than participants’ mean difference in rating of the positive and negative actions for participants in the up-down movement task,  $t(311) = 2.40, p = .017, 95\% \text{ CI diff} = [0.11, 1.12]$ . Participants’ mean difference in ratings of the towards and away actions correlated positively with the explicit rating score,  $r(155) = .17, p = .034$  and with the IAT score,  $r(155) = .15, p = .063$ .

Participants in the up-down movement task indicated that they liked *the word* ‘up’ ( $M = 6.46, SD = 1.37$ ) more than they liked the word ‘down’ ( $M = 4.30, SD = 1.57$ ),  $t(155) = 11.38, p < .001, d_z = 0.91, 95\% \text{ CI diff} = [1.78, 2.53]$ . Participants’ mean difference in ratings of the action words correlated positively with the explicit rating score,  $r(154) = .26, p < .001$  and with the IAT score,  $r(154) = .28, p < .001$ . Participants in the zooming task indicated that they liked the word ‘approach’ ( $M = 6.57, SD = 1.23$ ) more than they liked the word ‘avoid’ ( $M = 3.11, SD = 1.46$ ),  $t(156) = 19.61, p < .001, d_z = 1.57, 95\% \text{ CI diff} = [3.10, 3.80]$ . These participants’ mean difference in rating of the words ‘approach’ and ‘avoid’ correlated positively with the explicit rating score,  $r(155) = .14, p = .074$  but not with the IAT score,  $r(155) = .11, p = .19$ , and was significantly bigger than participants’ mean difference in rating of the words ‘up’ and ‘down’ for participants in the up-down movement task,  $t(311) = 5.02, p < .001, 95\% \text{ CI diff} = [0.79, 1.81]$ .

#### **Additional analyses for Experiment 4**

We performed separate analyses on explicit rating scores and IAT scores of the subset of 70 participants who indicated that they did not imagine themselves to be the manikin.

**Explicit rating scores.** The ANOVA revealed a main effect of Stimulus-Action Mapping,  $F(1,66) = 10.10, p = .002$ . Participants exhibited a stronger preference for Niffites over Luupites when they had observed the manikin approach Niffites and avoid Luupites ( $M = 0.82, SD = 2.05$ ) than when they had observed the manikin approach Luupites and avoid Niffites ( $M = -0.78, SD = 1.98$ ),  $t(68) = 3.30, p = .002, d = 0.79, 95\% \text{ CI diff} = [0.63, 2.56], BF_1 = 21.13$ . There was no interaction between Stimulus-Action Mapping and Task Order,  $F(1,66) = 0.78, p = .38$ .

**IAT scores.** The Spearman-Brown corrected split-half reliability of the IAT score was  $r(68) = .88$ . The ANOVA revealed a main effect of IAT Order,  $F(1,62) = 5.38, p = .024$ . More importantly, we also observed a main effect of Stimulus-Action Mapping,  $F(1,62) = 13.49, p < .001$ . Participants exhibited a stronger preference for Niffites over Luupites when they had approached Niffites and avoided Luupites ( $M = 0.03, SD = 0.46$ ) than when they had approached Luupites and avoided Niffites ( $M = -0.38, SD = 0.37$ ),  $t(68) = 3.98, p < .001, d = 0.95, 95\% \text{ CI diff} = [0.20, 0.61], BF_1 = 138.51$ . There were no other main or interaction effects,  $F_s < 0.89, p_s > .35$ .

### Comparative Analyses for Experiments 2 and 4

We performed separate analyses comparing explicit rating scores and IAT scores of (1) the 85 participants in Experiment 2 in the approach-avoid action framing with manikin location below the word and (2) the 70 participants in Experiment 4 who indicated that they did not imagine themselves to be the manikin.

**Explicit rating scores.** The 2 (*Experiment*: Experiment 2 vs. Experiment 4) x 2 (*Stimulus-Action Mapping*: Niffites approached and Luupites avoided or vice versa) x 2 (*Task Order*) ANOVA revealed a main effect of Stimulus-Action Mapping,  $F(1,151) = 39.35, p < .001$ . Participants exhibited a stronger preference for Niffites over Luupites when they had approached Niffites and avoided Luupites or observed the manikin approach Niffites and avoid Luupites ( $M =$

1.33,  $SD = 2.37$ ) than when they had approached Luupites and avoided Niffites or observed the manikin approach Luupites and avoid Niffites ( $M = -0.99$ ,  $SD = 2.03$ ),  $t(153) = 6.43$ ,  $p < .001$ ,  $d = 1.04$ , 95% CI diff = [1.59, 3.01],  $BF_1 > 10000$ . Importantly, there was only a marginally significant interaction between Stimulus-Action Mapping and Experiment,  $F(1,151) = 3.24$ ,  $p = .074$ , indicating a stronger effect of Stimulus-Action Mapping in Experiment 2 than in Experiment 4,  $d_{diff} = 0.44$ .

**IAT scores.** The 2 (*Experiment*) x 2 (*Stimulus-Action Mapping*) x 2 (*Task Order*) x 2 (*IAT Order*) ANOVA revealed a main effect of Stimulus-Action Mapping,  $F(1,139) = 66.81$ ,  $p < .001$ . Participants exhibited a stronger preference for Niffites over Luupites when they had approached Niffites and avoided Luupites or observed the manikin approach Niffites and avoid Luupites ( $M = 0.19$ ,  $SD = 2.37$ ) than when they had approached Luupites and avoided Niffites or observed the manikin approach Luupites and avoid Niffites ( $M = -0.41$ ,  $SD = 2.03$ ),  $t(153) = 8.19$ ,  $p < .001$ ,  $d = 1.32$ , 95% CI diff = [0.47, 0.75],  $BF_1 > 10000$ . Importantly, there was a significant interaction between Stimulus-Action Mapping and Experiment,  $F(1,139) = 8.44$ ,  $p = .004$ , indicating a stronger effect of Stimulus-Action Mapping in Experiment 2 than in Experiment 4,  $d_{diff} = 0.70$ .