Failures to Change Stimulus Evaluations by means of Subliminal Approach and Avoidance Training

Pieter Van Dessel¹
Jan De Houwer¹
Arne Roets¹
Anne Gast²

¹Ghent University, Belgium
²University of Cologne, Germany

Authors 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. Pieter Van Dessel is supported by a Ph.D. fellowship of the Scientific Research Foundation, Flanders (FWO-Vlaanderen). Jan De Houwer is supported by Methusalem Grant BOF09/01M00209 of Ghent University and by the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office (IUAPVII/33). Anne Gast is supported by an Emmy-Noether-Grant from the German Science Foundation (GA 1520/2-1). The research in this paper has been supported by Grant FWO12/ASP/275 of FWO - Vlaanderen.
Abstract

Previous research suggests that the repeated performance of approach and avoidance (AA) actions in response to a stimulus causes changes in stimulus evaluations. Kawakami, Phills, Steele, and Dovidio (2007) and Jones, Vilensky, Vasey, and Fazio (2013) provided evidence that these AA training effects occur even when stimuli are presented only subliminally. We also examined whether reliable AA training effects can be observed with subliminal stimulus presentations but added more sensitive checks of perceptual stimulus discriminability. Three experiments, including a direct replication of the study by Kawakami et al. (2007), failed to provide any evidence for effects of subliminal AA training on implicit or explicit evaluations. Bayesian analyses indicated that our data provide robust evidence that subliminal AA training does not cause changes in evaluations. In contrast, we observed changes in evaluations when participants were provided with (either correct or incorrect) information about the stimulus-action contingencies in the subliminal AA training task and when participants performed a supraliminal AA training task that allowed participants to detect these contingencies. These findings support the idea that contingency awareness is necessary for the occurrence of AA training effects.

Keywords: subliminal, approach, avoidance, contingency awareness, evaluation
Failures to Change Stimulus Evaluations by means of Subliminal Approach and Avoidance Training

It has been recognized for many decades that a person’s behavior is largely determined by his/her likes or dislikes (Allport, 1935). Accordingly, understanding how preferences are formed and how they can be influenced is a fundamental research area in psychological science. Prior research showed that the repeated execution of approach and avoidance (AA) actions in response to a stimulus can cause changes in stimulus evaluations. When participants repeatedly approach one stimulus and avoid another stimulus, one typically observes a preference for the approached stimulus over the avoided stimulus (e.g., Kawakami, Steele, Cifa, Phillips, & Dovidio, 2008; Laham, Kashima, Dix, Wheeler, & Levis, 2014, Woud, Maas, Becker, & Rinck, 2013; but see Vandenbosch & De Houwer, 2011). The first demonstration of this effect was provided by Kawakami, Phillips, Steele, and Dovidio (2007). They found that participants who repeatedly approached photographs of Black people and avoided photographs of White people exhibited more positive evaluations of Black relative to White people on the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998).

In two of the four experiments reported by Kawakami et al. (2007), participants performed the AA movements in response to the words ‘approach’ or ‘avoid’. Immediately preceding the presentation of these words, photographs of Black and White people were presented under conditions that limited the conscious detection of these stimuli (i.e., presentation in between two masking stimuli and for a duration of 23 ms). When photographs of Black people were consistently paired with approach movements and photographs of White people with avoidance movements, participants reported more positive implicit evaluations of Black relative to White people (Experiment 2) and showed more immediacy and openness for communication
when interacting with a Black confederate (Experiment 4). The authors concluded that AA training effects may occur outside of participants’ awareness of the contingencies in the training task.

Recently, a study by Jones, Vilensky, Vasey, and Fazio (2013) provided corroborative evidence for effects of AA training with subliminally presented stimuli. They observed changes in participants’ explicit evaluations of insects when approach behaviors, performed in response to the presentation of the word ‘TOWARD’, were repeatedly paired with masked images of insects, presented for 13 ms (Experiment 1). Moreover, when participants’ repeatedly performed approach behaviors in response to subliminally presented images of spiders, they exhibited more positive implicit evaluations of spiders (Experiment 2) and reported reduced anxiety ratings when encountering live spiders (Experiment 3).

The observation of subliminal AA training effects is in line with the idea that AA training effects are due to automatic association formation mechanisms. This idea entails that the repeated pairing of an AA action and a stimulus automatically results in the gradual formation of associative links between the representations of the stimulus and positively or negatively valenced representations (Phills, Kawakami, Tabi, Nadolny, & Inzlicht, 2011; Woud et al., 2013).

Recent evidence has, however, challenged the idea that AA training effects are (exclusively) the result of automatic associative learning processes. In two studies that were conducted at our laboratory, we provided evidence that AA training effects are moderated by awareness of the stimulus-action contingencies (Van Dessel, De Houwer, & Gast, 2015). When participants repeatedly performed AA actions in response to novel face stimuli, they exhibited a preference for the approached stimuli on implicit and explicit measures of evaluation – but only when they were able to correctly identify what action they had performed most often in response
to the stimuli. We obtained no evidence that AA training caused changes in stimulus evaluations in the absence of contingency awareness.

The observation that contingency awareness moderates AA training effects does not fit well with the idea that automatic association formation underlies these effects. To accommodate these findings it seems necessary that traditional association-formation accounts make a number of additional assumptions (e.g., that the formation of associations depends on specific boundary conditions such as whether sufficient attention is attributed to the identity of the stimulus; see Van Dessel, De Houwer, & Gast, 2015). In contrast, alternative, propositional accounts of AA training may more easily explain these results. These accounts might, for instance, entail that participants who acquire stimulus-action contingency information may elaborate on this information and infer that the approached stimulus is positive (because they typically approach good things). Once this proposition is formed, this may influence both explicit and implicit stimulus evaluation (see De Houwer, 2014). In line with this account, a recent set of studies showed that typical AA training effects occurred even when participants did not perform AA actions but were merely instructed that they would later on have to perform these actions (Van Dessel, De Houwer, Gast, & Smith, 2015). Participants who received instructions to approach one fictitious social group (e.g., Niffites) and avoid another fictitious social group (e.g., Luupites) exhibited a preference for the former group both on implicit and explicit measures of evaluation.

Although AA learning via instructions suggests that conscious knowledge of stimulus-action contingencies is sufficient for acquiring AA effects, it does not exclude the possibility that, under certain circumstances, AA training effects may occur in the absence of contingency awareness (e.g., as the result of the automatic formation of associations). Therefore, it is important to actively establish evidence or seek to confirm potential evidence for AA training.
effects that occur in the absence of contingency awareness. Demonstrating that AA training effects can occur after subliminal stimulus presentations provides the strongest case that contingency awareness is not a necessary condition for these effects to occur. Hence, subliminal AA training effects are theoretically important because they provide strong evidence for automatic association formation models of AA training. Such evidence would strongly constrain propositional accounts of AA training. For instance, to explain subliminal AA training effects, these accounts would have to assume that propositions (e.g., ‘I’d like to approach Black people’) can be formed even if one is unaware of the perceptual stimulation that initiated the formation of this proposition (e.g., photographs of Black people).

As we noted above, two sets of studies provided support for the existence of subliminal AA training effects (Kawakami et al., 2007; Jones et al., 2013). If the conclusion drawn from these studies (i.e., that AA training effects can occur after subliminal presentations) is valid, these studies present a very strong case for the possibility of AA training effects in the absence of contingency awareness and hence for automatic association formation as an underlying mechanism. In our opinion, however, there are two reasons why this conclusion can currently not be drawn.

First, both studies suffer from an important methodological limitation: because they did not include objective measures of stimulus visibility they could not assure that every presentation of the stimulus was indeed presented below the threshold of conscious awareness (‘subliminally’). To establish whether participants were aware of the rapidly presented stimuli, Kawakami et al. (Experiments 2 and 4), and Jones et al. (Experiment 1) used a funnel debriefing procedure to question participants for stimulus awareness after the experiment. When participants expressed any awareness of the identity of the stimuli, their data were excluded from
the analyses. In the subliminal perception literature, however, there is a lot of debate about whether such subjective self-report measures provide an accurate and reliable method to determine awareness (for a summary of this discussion, see Snodgrass, Bernat, & Shevrin, 2004). For instance, a simple lack of confidence may contribute to someone's reluctance to report having been aware of a particular stimulus, despite having some subjective experience of it (Cleeremans, 2001). Jones et al. did not use any awareness tests in Experiments 2 and 3, but they did conduct a pilot test in which 12 participants first completed the approach spiders training and then indicated for 40 stimuli, 10 of which were spider pictures, whether they had seen these stimuli during the training task. None of the participants reported seeing any of the spider pictures, suggesting that the pictures had not been perceived consciously. This procedure, however, may suffer from the same limitations as the funnel debriefing procedure (e.g., issues related to memory recall or a lack of confidence). Moreover, it does not take into account that there may be individual differences in people’s ability to detect rapidly presented pictures. The authors address this limitation in their discussion section where they state that they “cannot rule out the possibility that some presentations of images were correctly identified or that any awareness that occurred contributed to the effect on attitude change.” (p.995).

Second, even if one assumes that the method is suitable to exclude conscious awareness of the stimuli, the empirical support provided by the two studies might not be sufficient to yield substantial evidence for the hypothesis that AA training is indeed possible with subliminal stimulus presentation. In order to investigate this issue, we performed Bayesian analyses. Bayesian analyses provide a Bayes Factor that denotes the weight of evidence provided by the data for competing hypotheses. As such, Bayes Factors can provide an indication of how strongly the data support either the null hypothesis (BF$_0$; reflecting the absence of a significant
effect) or the alternative hypothesis ($BF_1$; reflecting the presence of a significant effect). $BF$ scores can be computed for both the null and the alternative hypothesis. $BF$ scores smaller than 1, between 1 and 3, and between 3 and 10, respectively designate ‘no evidence’, ‘anecdotal evidence’, and ‘substantial evidence’ for either the null or the alternative hypothesis (Jeffreys, 1961). For instance, when $BF_0 = 10$ (and $BF_1 = 1/10 = .10$) the observed data are 10 times as likely to have occurred under the null hypothesis than under the alternative hypothesis, providing substantial evidence for the null hypothesis. When $BF_1 = 2.5$ (and $BF_0 = 1/2.5 = .40$) the observed data are 2.5 times as likely to have occurred under the alternative hypothesis than under the null hypothesis, providing anecdotal evidence for the alternative hypothesis. We reanalyzed the critical t-tests reported in Jones et al. (2011), and Kawakami et al. (2007) for all experiments that showed effects of subliminal AA training on stimulus evaluations by computing a Bayes factor ($BF$) according to the procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009). Though the obtained $BF$ scores always provided evidence in favor of the alternative hypothesis (that subliminal AA training causes significant changes in implicit evaluations), the evidence was only anecdotal (Kawakami et al. Experiment 2: $BF_1 = 2.31$; Jones et al. Experiment 1: $BF_1 = 1.64$; Jones et al. Experiment 2: $BF_1 = 1.26$). This attests that replication is warranted to establish the robustness of the subliminal AA training effect. We also assessed the available evidence across these experiments by performing a Bayes factor meta-analysis (Rouder & Morey, 2011). This analysis did provide strong evidence for subliminal AA training effects, $BF_1 = 46.46$ (Table 1). However, this value may be artificially inflated because no unsuccessful attempts have ever been reported, possibly due to the file-drawer problem.

We performed the current research in order to gain more information on whether AA training effects can be observed with subliminal stimulus presentations. Most importantly, we
included more sensitive checks of perceptual stimulus discriminability that are based on participants' forced-choice decisions regarding the identity of the stimulus by using d' measures. These measures are widely used in studies with subliminal stimulus presentations because they provide an objective assessment of stimulus discriminability (Merikle, Smilek, & Eastwood, 2001). If we would find robust AA training effects also when these measures indicate that participants are unable to discriminate the stimuli, this would constitute important evidence that AA training can change evaluations in the absence of awareness of the stimuli and thus stimulus-action contingencies. If, on the other hand, AA training effects strongly depend on participants’ ability to discriminate the stimuli, this would be consistent with the idea that AA training effects necessarily involve stimulus awareness.

In this paper, we report three experiments. Experiments 2 and 3 examined subliminal AA training effects for evaluations of novel stimuli, whereas Experiment 1 was an exact replication of the experiment by Kawakami et al. (2007, Experiment 2) who found subliminal AA training effects on evaluations of Black and White social groups. This experiment was selected for replication because it provided the ‘strongest’ evidence in favor of subliminal AA training effects ($BF_1 = 2.31$). We used Kawakami et al.’s exact procedures and materials and extended upon Kawakami et al. by (1) including a d’ measure of stimulus perceptibility to examine whether stimulus presentations were indeed subliminal, (2) including also explicit measures of evaluations to assess subliminal AA training effects, and (3) supplementing the data analytic strategies that Kawakami et al. used for the investigation of the AA training effects with Bayesian analyses.

**Experiment 1**

**Method**
Participants. Sixty-two native Dutch-speaking undergraduates (49 women) participated in exchange for a monetary reward of 7 euros. This sample size was determined by performing a power analysis according to the procedures recommended by Cohen (1988) with the aid of G-Power software (version 3.1.). We ensured that the power to obtain an effect size of \( d = 0.67 \) (the effect size observed by Kawakami et al., 2007) was greater than .80 (achieved power = .84). All participants had normal or corrected-to-normal vision and were naive with respect to the purpose of the experiment. Participants were randomly assigned to an approach Blacks training condition where they received training to approach subliminally presented photographs of Black people and avoid subliminally presented photographs of White people or to a control condition where they received training to make a leftward or rightward movement in response to the subliminal presentation of photographs of Black or White people\(^1\).

Apparatus and Materials. The experiment was programmed and presented using the Direct RT Empirisoft Software package (DirectRTv2012) on a Tori PC with a 19-inch monitor (85 Hz refresh rate). The stimuli for the AA training task and the IAT (i.e., black-and-white photographs of 30 Black faces, 30 White faces, and 48 moonscapes) and the script for the subliminal AA training task were provided to us by one of the authors of Kawakami et al. (2007).

Procedure. Participants were seated at a desk in an individual cubicle in front of a computer with a keyboard and joystick (Logitech Wingman) attached to it. After participants had given informed consent, they were informed that they would perform a series of unrelated tasks. Participants in the approach Blacks training condition were instructed to pull the joystick toward themselves when the word ‘approach’ was presented and to push the joystick away from themselves when the word ‘avoid’ was presented. Participants in the control condition were
instructed to push the joystick to the right when presented with the word ‘right’ and to push the joystick to the left when presented with the word ‘left’.

The AA training task consisted of ten blocks of 48 trials. On each trial, a forward mask consisting of a photograph of a moonscape was presented for 300 ms and followed by a photograph of a Black or White person’s face presented for 23.52 ms (two refresh cycles). A backward moonscape mask was then presented for 35.29 ms (three refresh cycles) and followed by the word “approach” or the word “avoid” (in the approach Blacks training condition) or the word “left” or the word “right” (in the control condition). In the approach Blacks training condition, a Black face was always followed by the word approach and a White face was followed by the word avoid. In the control condition, for half the participants, the word “left” followed a Black face and the word “right” followed a White face. For the other half of the participants of the control condition, the contingencies of left/right word and Black/White face were reversed. The word remained on screen until participants responded by moving the joystick. After participants made a correct response, a blank screen was presented for 1000 ms before the start of the next trial. After incorrect responses, participants were presented with a blank screen for 100 ms, followed by the presentation of a red X in the middle of the screen for 800 ms. Another blank screen was presented for 100 ms before the start of the next trial.

Participants then performed an IAT where they categorized photographs of six Black and six White faces which had not been used in the AA training task along with positive and negative words. In accordance with standard IAT procedures (Greenwald et al., 1998), participants were presented with five blocks of trials. Participants started with two practice blocks, one in which they categorized the photographs into the categories Black or White and one in which they categorized words as positive or negative. Categorization was done by pushing a left (‘Q’) or
right key (‘M’) on an AZERTY keyboard. The practice blocks were followed by a critical block in which participants categorized both words and photographs. The practice block with Black and White faces was then repeated, but the response assignments were reversed. Finally, the critical block was repeated with the reversed response assignment for the Black and White faces. Each critical block consisted of 60 trials in which a word or photograph was presented in the center of the screen until the participant pressed one of the two valid keys. If the response was correct, the word disappeared and the next trial started 400 ms later. If the response was incorrect, the word was replaced by a red “X” for 400 ms. The next trial started 400 ms after the red “X” was removed from the screen.

After the IAT, we assessed participants’ explicit evaluations of Black and White people. First, participants completed liking ratings and thermometer ratings of self-reported feelings of warmth towards Black and White people on two 9-point Likert scales (1 = not liked/warm at all; 9 = completely liked/warm). Second, participants completed a ten-item version of Pettigrew and Meertens’ (1995) questionnaire assessing subtle and blatant racial prejudice, adapted to a Belgian context (Van Hiel & Mervielde, 2005).

Subsequently, a funnel debriefing procedure was used to question participants for awareness of the stimuli, stimulus-action contingencies and research hypotheses. Participants indicated: (1) what the purpose of the experiment and of the joystick task was, (2) what the relationship was between the joystick task and the categorization task, (3) whether they had noticed anything suspicious about the background in the joystick task, and (4) what the specific content of the background flashes was. We also asked participants to indicate what the stimulus-action contingencies were, but only if they had correctly identified the nature of the stimuli.
Finally, participants performed a perceptibility task in which they categorized the photographs of Black and White persons’ faces masked under the same conditions as in the AA training task. The task consisted of two blocks of 48 trials. Trials were identical to the AA training trials with the exception that (1) target words were replaced with strings of ‘XXXX’ and (2) participants did not perform AA actions but responded to the photographs that were presented before these strings. They responded by pushing the ‘Z’ key when the photograph depicted the face of a Black person and the ‘B’ key when the photograph depicted the face of a White person. Participants did not receive feedback about the accuracy of their responses.

Results

Awareness and Perceptibility. Thirteen participants (21%) indicated that photographs of faces had been presented during the AA training task. Seven participants (11%), 6 of which were in the approach Blacks training condition, indicated that these were faces of Black and White people. Of these participants, six (10%) expressed suspicion that the purpose of performing the AA training task was related to the purpose of performing the IAT and four (6%), all in the approach Blacks training condition, indicated that the purpose of the AA training task was to target racial prejudice by approaching faces of Black people.

The overall accuracy in the perceptibility task was 54% ($SD = 6\%$). We computed the signal detection sensitivity measure $d'$ with hits defined as ‘Black’ responses on trials with photographs of Blacks, and false alarms as ‘Black’ responses on trials with photographs of Whites. The $d'$ score indicated that participants’ detection performance was (marginally significantly) above chance level ($M = 0.17, SD = 0.67$), $t(61) = 1.98, p = .053$. Twenty-two participants (35%), 12 in the approach Blacks training condition, including 6 of the participants who indicated they had seen faces of Black and White people, had an individual $d'$ score above
the 95% confidence interval \((d' > 0.34)\), which indicates a potential capability to see the masked prime (Macmillan & Creelman, 2005). Performance of two of these participants was very high \((d' > 1)\) even though they had not reported awareness of the stimuli in the debriefing questions.

**Implicit evaluation.** IAT D4 scores were calculated following the procedure by Greenwald, Nosek and Banaji (2003). Positive scores reflect a preference for White people over Black people. Similar to Kawakami et al. (2007), a one-sample t-test revealed that participants displayed a strong implicit preference for White people \((M = 0.23, SD = 0.40), t(61) = 4.47, p < .001, d = 0.58, BF_1 = 595.87\). However, contrasting Kawakami et al.’s findings, a between-subjects t-test did not reveal a significant difference in IAT scores for participants who approached Blacks and avoided Whites \((M = 0.24, SD = 0.38)\) compared to participants in the left/right control condition \((M = 0.22, SD = 0.42), t(60) = -0.21, p = .83, d = 0.05\). The BF score provided substantial evidence in favor of the null hypothesis \((BF_0 = 4.49)\). We performed an additional t-test excluding the data of the 13 participants who expressed awareness of the presentation of subliminal presentation of faces. Again, we observed no subliminal AA training effect, \(t(47) = 0.29, p = .78, d = 0.07, BF_0 = 3.01\). We also did not observe significantly reduced IAT scores for those participants in the approach Blacks training condition who had reported seeing photographs of Black and White people, \(t(5) = 0.83, p = .78, BF_0 = 1.70\), or who had an individual \(d'\) score above the 95% confidence interval, \(t(11) = 0.47, p = .68, BF_0 = 2.46\). Note that the latter analyses revealed only anecdotal evidence in favour of the null hypothesis, which is not surprising given that they included only a very small sample of participants.

**Explicit evaluation.** We calculated liking rating scores \((M = 0.19, SD = 0.33)\) and warmth rating scores \((M = 0.15, SD = 0.29)\) by subtracting each participants’ liking and warmth ratings for Black people from the corresponding ratings for White people. A score for subtle racism \((M\)
= 3.77, $SD = 0.86$, Cronbach’s $\alpha = 0.51$) and a score for blatant racism ($M = 1.80$, $SD = 0.69$, Cronbach’s $\alpha = 0.73$) was calculated by summing the ratings for the items in the racism scales and dividing this by the number of items. The four resulting explicit evaluation scores correlated significantly (blatant racism – subtle racism: $r[60] = 0.57$; blatant racism – liking rating: $r[60] = 0.32$; blatant racism – warmth rating: $r[60] = 0.47$; subtle racism – liking rating: $r[60] = 0.39$; subtle racism – warmth rating: $r[60] = 0.32$; liking rating – warmth rating: $r[60] = 0.70$), $ps < .012$. The IAT score correlated significantly with the liking rating score, $r(60) = 0.29$, $p = .020$, and warmth rating score, $r(60) = 0.30$, $p = .016$, but not with the blatant racism scale score, $r(60) = 0.08$, $p = .54$, or the subtle racism scale score, $r(60) = 0.18$, $p = .17$. Similar to the IAT score, the explicit rating scores indicated a significant preference for White people over Black people. Participants’ liking and warmth ratings were higher for White people (liking: $M = 6.85$, $SD = 1.34$; warmth: $M = 6.61$, $SD = 1.40$) than for Black people (liking: $M = 6.11$, $SD = 1.48$; warmth: $M = 6.02$, $SD = 1.59$), $t$s $> 4.03$, $ps < .001$, $BF_{1s} > 153.60$. Most importantly, however, none of the explicit evaluation scores revealed significant differences between the approach Blacks training group and the control group, $t$s $< 0.32$, $ps > .74$, $ds < 0.01$, $BF_{0s} > 2.99$.

**Discussion**

Despite using the same subliminal AA training procedure as Kawakami et al. (2007), we were unable to replicate the effect on implicit evaluations (or provide evidence for an additional effect on explicit evaluations). To account for these discrepant results, a number of explanations can be considered. First, though power calculations were used to determine the sample size and ascertain that we had sufficient power to detect the effect size observed by Kawakami et al., we may have lacked the power to detect an effect of smaller effect size (i.e., for a medium effect: power = 0.62; for a small effect: power = 0.25). However, Bayesian analyses indicated that our
findings already provide substantial evidence that subliminal AA training does not allow for changes in participants’ implicit or explicit evaluations. Second, as suggested to us by one of the authors of Kawakami et al., the discrepancy between our results and the original results may relate to important differences in participants’ baseline levels of racial prejudice due to cross-cultural differences. Our participant sample consisted of Belgian undergraduates, whereas participants in Kawakami et al.’s study were undergraduates from North America. Prejudice towards Black people may be more relevant (and more robust) for the latter population (e.g., due to a long history of conflict between these two racial groups; see Perlmutter, 1999). Because previous evidence suggests that AA training effects are observed only when the training is inconsistent with participants’ racial bias (e.g., Phills et al., 2011), a smaller overall effect of AA training may be observed if a smaller subset of participants have this bias in our sample, which may explain why we did not find any subliminal AA training effects. Nevertheless, it should be noted that in our study, we did observe a strong preference for White people over Black people both on implicit and explicit measures of evaluation. To corroborate this, we examined the IAT scores of volunteers with Belgian or US nationality who completed a race IAT on the Project Implicit research website (https://implicit.harvard.edu) in the years 2002-2012 (Xu, Nosek, & Greenwald, 2014). We observed a significant preference for White people over Black people for both Belgian ($M = 0.44, SD = 0.39), t(71) = 9.46, p < .001, d = 1.13, and US participants ($M = 0.37, SD = 0.43), t(30343) = 152.19, p < .001, d = 0.86. Importantly, The IAT prejudice score was even slightly lower for US participants than for Belgian participants. Hence, the lack of substantial AA effects in this study could not be attributed to lower baseline levels of racial prejudice toward Black people in the present sample.
Alternatively, the effect that was observed by Kawakami et al. (2007) may have simply been a Type I error. In line with this idea, Bayesian analyses indicated that the data obtained by Kawakami et al. provided only anecdotal evidence in favor of the alternative hypothesis (i.e., that IAT scores differed significantly between approach Blacks training and control conditions). In contrast, our data provided substantial evidence for the null hypothesis. Another possible explanation is that in the experiment reported by Kawakami et al., some participants identified the stimulus-action contingencies and, though they failed to report this in the awareness questions, fuelled the effect. In line with this idea, our experiment provided evidence that (1) with the reported presentation times, participants are sometimes able to discriminate the race of the person presented in the photograph, and (2) participants may not always report this in the debriefing questions (e.g., the two participants with the best performance in the stimulus discriminability task did not report stimulus awareness). Moreover, we found that participants who performed the subliminal approach Blacks training (but not participants in the control group) sometimes identified the purpose of performing the AA training task. This may reduce participants’ implicit prejudice if, for instance, participants adapt to the demands of the interpersonal context (Castelli & Tomelleri, 2008; Richeson & Ambady, 2003). In our study, however, this did not cause any significant differences in evaluations between participants in the approach Blacks training group and the control group. More specifically, we did not even find AA training effects for participants who were able to discriminate the race of the presented persons or reported seeing pictures of Black or White persons.

**Experiment 2**

Experiment 1 failed to provide support for the idea that training to approach respectively avoid subliminally presented photographs of Black and White people causes changes in the
evaluations of these well-known social groups. In Experiments 2 and 3, we shifted our focus to the investigation of subliminal AA training effects for novel stimuli\(^3\). Subliminal AA training effects may be more robust for these stimuli because AA effects are more easily established for neutral as compared with initially valenced stimuli (e.g., Priester, Cacioppo, & Petty, 1996; Woud, Becker, Lange, & Rinck, 2013; Van Dessel, De Houwer, Gast, & Smith, 2015). This finding is related to the general idea that changing preferences is more difficult than establishing novel preferences (e.g., Gregg, Seibt, & Banaji, 2006). The stimuli we used in Experiment 2 were unfamiliar faces. Previous AA training studies have provided robust effects for these stimuli (Woud et al., 2008; 2013; Van Dessel, De Houwer, & Gast, 2015; but see: Vandenbosch & De Houwer, 2011) and evidence suggests that AA effects are larger for face stimuli compared to other pictorial stimuli (Laham, Kashima, Dix, & Wheeler, in press).

The purpose of Experiment 2 was twofold: (1) to establish whether changes in implicit and explicit evaluations of novel face stimuli can arise as the result of subliminal AA training, and (2) to compare effects of subliminal AA training with effects that result from instructions about stimulus-action contingencies. To this end, participants were informed about the contingencies between the four face stimuli and the AA actions in the subliminal AA training task before they performed the task. Unbeknownst to participants, the contingency information was incompatible with the AA training for half of the stimuli. Consequently, each face was assigned to one of the following four conditions: (1) approach training and approach instructions, (2) approach training and avoid instructions, (3) avoid training and approach instructions, or (4) avoid training and avoid instructions. Participants’ evaluations of the faces were registered with an explicit evaluative ratings and an implicit measure of evaluation (the evaluative priming task; Fazio, Sanbonmatsu, Powell, & Kardes, 1986).
This experimental set-up allowed us to test specific predictions derived from associative and propositional accounts of AA training effects. If AA training effects arise as the result of the gradual automatic formation of associations, then changes in evaluations should reflect the contingencies in the subliminal AA training task. These effects may be observed more strongly on implicit evaluations because implicit evaluations are considered to reflect the automatic activation of associations in memory (Gawronski & Bodenhausen, 2011). If, on the other hand, AA training effects critically depend on the acquisition of propositional information, then changes in evaluations should always be in line with the provided contingency information. Acquiring this information may influence both explicit and implicit stimulus evaluations (see De Houwer, 2014).

Method

Participants. Seventy-six native Dutch-speaking undergraduates (61 women) participated in exchange for a monetary reward of 5 euros. None of the participants had previously participated in Experiment 1.

Apparatus and Materials. Four photographs of female faces, selected from the set of materials used by Van Dessel, De Houwer, and Gast (2015), served as stimuli for the AA training task. For the evaluative priming task, seven positive words (the Dutch words for happy, pleasant, sweet, kind, friendly, sympathetic, and fun), and seven negative words (the Dutch words for unfriendly, irritating, hostile, bad, moody, unpleasant, and mean) were selected to serve as target stimuli and the four photographs of faces were used as primes.

The experiment was programmed in C-language and presented using the C-library Tscope package (Tscope 1.0.171.) on a Tori PC with a 19-inch monitor (80 Hz refresh rate), a keyboard and a joystick (Wingman Attack 2) attached to it.
Procedure. After participants had given informed consent, they were seated in front of a computer screen. Half of the participants read the following instructions (translated from Dutch):

*In this experiment, you will see photographs of different faces. You will approach a specific number of these faces by pulling the joystick towards you. You will avoid other faces by pushing the joystick away from you.*

*The photographs of faces will be presented in such a way that you will not be able to perceive them consciously. For this reason, we will present a word immediately following the presentation of a specific face. This word will indicate whether a face was presented that is to be approached or avoided.*

*Please respond by moving the joystick forward when the word ‘APPROACH’ is presented and respond by moving the joystick backward when the word ‘AVOID’ is presented.*

*Before we start the task we will teach you which faces you will approach and which faces you will avoid.*

These participants thus received instructions to approach by pulling a joystick towards them (an arm flexion movement) and to avoid by pushing the joystick away from them (an arm extension movement). The other half of the participants received identical instructions except that they were instructed to approach by moving themselves towards the screen with the joystick (an arm extension movement) and to avoid by moving themselves away from the screen with the joystick (an arm flexion movement).  

Subsequently, participants were shown the four faces that would be presented during the AA training task. Above two faces approach instructions were presented: “These are the faces you will have to approach” (‘approach instruction faces’) and avoid instructions were presented above the other two faces: “These are the faces you will have to avoid” (‘avoid instruction faces’).
Participants were asked to make sure that they would not forget which action belonged with each face.

The AA training task consisted of two blocks of 160 trials. During each training block, each of the four faces was presented on 40 occasions and was always presented with either the word ‘approach’ or the word ‘avoid’. For each participant, one of the approach instruction and one of the avoid instruction faces was always paired with approach (‘approach training faces’) and the other faces were always paired with avoid (‘avoid training faces’). To avoid biases, we randomized the assignment of faces to the AA training action and AA instruction action. Similar to Experiment 1, on each trial of the AA training task a moonscape mask was presented for 300 ms, followed by a photograph presented for 23 ms and a backward moonscape mask presented for 33 ms. Then the word approach or avoid was presented until participants responded correctly with the joystick by performing a vertical movement towards the screen or away from the screen. After 200 ms the next trial started.

For half the participants, the AA training task was immediately followed by an evaluative priming task. The other half of the participants first completed explicit ratings. In the evaluative priming task, participants categorized target words as either positive or negative using the E and I keys of a computer keyboard. The assignment of the response keys to either the positive or negative category was counterbalanced across participants and across task order and action framing conditions. Participants were instructed to perform this categorization task as quickly as possible, while making as few mistakes as possible. Participants were further told that they would see photographs of faces presented before the words and that they could look at these photographs, but that their task was simply to respond to the positive and negative words. In line with standard procedures at our lab (Spruyt, De Houwer, Hermans, & Eelen, 2007), a single trial
consisted of a fixation cross presented for 500 ms, a blank screen for 500 ms, a prime for 200 ms, a post-prime pause for 50 ms and the target word in white font for 1500 ms or until the participant had given a response. Error feedback was presented on the screen (i.e., the Dutch word for ‘wrong’ presented in red font) for 250 ms if participants made an error. The inter-trial interval was set to vary randomly between 500 ms and 1500 ms. Participants completed 224 trials separated into two blocks, each containing 14 trials with each of the faces as prime and a positive or negative word as target, presented in random order.

The explicit ratings consisted of two questions for each of the faces. Participants indicated whether they liked the person in the photograph and whether they thought the person in the photograph was friendly on two eight-point Likert scales (0 = not liked at all/not friendly at all; 7 = liked a lot/very friendly). For each face, we collapsed these score ratings into one explicit rating score by averaging the respective ratings. The internal consistency of this score was good (mean Cronbach’s $\alpha = 0.76$, $SD = 0.04$).

After the implicit and explicit evaluation tasks, participants completed questions assessing their memory for the instructed stimulus-action contingencies. Each of the faces was presented in a random order and participants were asked to indicate what action they had performed in response to this face according to the instructions. Participants answered by selecting an option on a dropdown menu with “Approach”, “Avoid”, and “I don’t remember” as possible answers.

Subsequently, participants indicated whether they had ever seen any face during the joystick task by choosing from three options (i.e., “Yes, clearly”, “Yes, but I couldn’t identify which face”, and “No, never”). Then each of the faces was presented again and participants reported on how many occasions they had seen this face during the subliminal training task by entering a number between 0 and 320.
Finally, participants performed a perceptibility task which consisted of 80 trials. Trials were identical to the subliminal AA training trials with the exception that (1) the words approach and avoid were replaced with strings of ‘XXXX’, and (2) participants indicated which of the four faces was presented on each trial by pushing one of four keys on the numeric keypad (1-4).

Results

**Awareness and Perceptibility.** On average, participants selected the correct instructed action for the faces 82% of the time (SD = 32%). The number of correct instructed action identifications was identical for faces with compatible instruction and training (e.g., approach instructions and approach training; M = 82%, SD = 34%) and faces with incompatible instruction and training (e.g., approach instructions and avoid training; M = 82%, SD = 34%).

In response to the question whether they had ever seen any face during the joystick task, one participant (1%) indicated ‘yes, clearly’, eleven participants (15%) indicated ‘yes, but I couldn’t identify which face’, and 64 participants (84%) indicated they had not seen any of the faces.

Participants’ overall detection accuracy in the perceptibility task was 26.17% (SD = 4.85%), which was significantly above the chance level of 25%, t(75) = 2.10, p = .039. Individual accuracy scores of 27 participants (36%) were above the 95% confidence interval (> 27.28%).

**Implicit Evaluation.** For the analysis of the evaluative priming task reaction time data, trials with an incorrect response were dropped (4.6%) as well as trials in which reaction times (RTs) were at least 2.5 standard deviations removed from an individual’s mean (2.9%) (Spruyt, De Houwer, Hermans, & Eelen, 2007). RTs were subjected to a mixed analysis of variance
(ANOVA) that contained 3 within-subject factors: Prime AA Instructions (approach, avoidance), Prime AA Training (approach, avoidance), Target Valence (positive, negative), and 2 between-subject factors: Order (evaluative priming task first, explicit rating task first) and Action Framing (approach by pulling, approach by moving towards the screen). We observed a main effect of Target Valence, $F(1,72) = 56.64, p < .001, \eta^2_p = 0.44$, indicating that participants were faster to detect a positive target ($M = 569, SD = 82$) than to detect a negative target ($M = 590, SD = 85$).

We also observed a main effect of Action Framing, $F(1,72) = 6.14, p = .016, \eta^2_p = 0.08$, indicating that, overall, participants were faster if they had approached faces by pulling the joystick in the subliminal AA training task. More importantly, we observed an interaction effect of Target Valence and Prime AA Instructions, $F(1,72) = 5.68, p = .020, \eta^2_p = 0.07$ (Table 2).

Participants were faster when the valence of the target word was compatible with the valence of the AA instruction for the face prime (i.e., positive target and approach instruction face [$M = 565, SD = 88$], or negative target and avoid instruction face [$M = 588, SD = 83$]), than when they were incompatible (i.e., positive target and avoid instruction face [$M = 572, SD = 77$], or negative target and approach instruction face [$M = 593, SD = 87$]). In contrast, the interaction of Target Valence and Prime AA Training was not significant, $F(1,72) = 0.03, p = .87, \eta^2_p < 0.01$, nor were any other main or interaction effects, $Fs < 1.44., ps > .23.$

To further examine participants’ implicit preference for approach training and approach instruction faces, we calculated two indices of AA effects. An index of the AA training effect was calculated by subtracting evaluative priming RTs for trials where target valence was compatible with the valence of the AA training for the face prime (i.e., trials with approach training face and positive target or avoid training face and negative target) from RTs for trials where target valence was incompatible with the valence of the AA training for the face prime.
SUBLIMINAL APPROACH-AVOIDANCE TRAINING

(i.e., trials with avoid training face and positive target or approach training face and negative target). An index of the AA instruction effect was calculated in the same manner for trials where target valence and valence of the AA instructions for the face prime were compatible or incompatible. One-sample t-tests indicated a significant effect of AA instructions ($M = 23, SD = 83$), $t(75) = 2.40, p = .019, d = 0.28, BF_1 = 4.58$, but no effect of AA training ($M = -2, SD = 94$), $t(75) = -0.17, p = .87, d = 0.02, BF_0 = 6.49$.

We performed a linear regression of participants’ detection scores in the perceptibility task on the index of the AA training effect. We did not observe a significant positive intercept, $b_1 = -23.27, t(74) = -0.39, p = .70$, indicating that participants whose detection performance was at chance level did not exhibit a significant AA training effect. The slope was also not significant, $b_2 = 81.98, t(74) = 0.37, p = .72$, which indicates that the AA training effect was not a function of stimulus visibility.

Explicit evaluation. The explicit rating scores were subjected to a mixed ANOVA that included AA Instructions and AA Training as within-subjects factors, and Order and Action Framing as between-subjects factors. The analysis revealed a main effect of AA Instructions, $F(1,72) = 16.19, p < .001, \eta^2_p = 0.19$ (Table 3). Participants preferred faces they were instructed to approach ($M = 3.88, SD = 1.50$) over faces they were instructed to avoid ($M = 2.93, SD = 1.44$). The main effect of Action Framing was also significant, indicating that participants who approached by pulling and avoided by pushing evaluated the faces more positively compared to participants where the framing was reversed, $F(1,72) = 7.18, p = .009, \eta^2_p = 0.09$. No other effects reached significance, $Fs < 1.94, ps >.16$.

We calculated indices of the AA training and AA instruction effect by subtracting the explicit rating scores for avoid training faces from explicit rating scores for approach training
faces and explicit rating scores for avoid instruction faces from explicit rating scores for approach instruction faces. These indices correlated significantly with the corresponding indices of AA effects on implicit evaluations, $r_s > 0.39$, $p_s < .001$. Similar to the results for implicit evaluations, we observed a significant effect of AA instructions ($M = 1.90$, $SD = 4.06$), $t(74) = 4.09$, $p < .001$, $d = 0.47$, $BF_1 = 412.97$, but no significant effect of AA training ($M = -0.07$, $SD = 2.26$), $t(74) = -0.28$, $p = .78$, $d = 0.03$, $BF_0 = 7.00$.

**Discussion**

In Experiment 2, we obtained no evidence that subliminal training to approach or avoid face stimuli causes changes in implicit or explicit evaluations of these stimuli. In contrast, we observed changes in both implicit and explicit evaluations as the result of instructions about the stimulus-action contingencies. This corroborates previous findings of AA instruction effects (Van Dessel, De Houwer, Gast, & Smith, 2015), and extends these findings by showing that AA instruction effects occur even if (subliminal) AA training is provided that is incompatible with the provided information. These findings are more consistent with a propositional account of AA training effects than with associative accounts which suggest that AA training effects result from the gradual and automatic formation of associations.

Some aspects of our procedure, however, may have impeded the detection of subliminal AA training effects. First, we may have lacked the power to detect a small effect (power = .53 to detect an effect-size of $d = 0.20$). However, Bayesian analyses indicated that our data provide substantial evidence that subliminal AA training does not cause changes in evaluations. In contrast, these analyses indicated strong evidence in favor of AA instruction effects. Second, the mere presence of AA instructions might have somehow interfered with the subliminal training effect. For instance, subliminal AA training may not work in the presence of opposite AA
instructions because (1) these instructions allow the stimulus to acquire a specific valence (Van Dessel, De Houwer, Gast, & Smith, 2015), and (2) recently established (implicit) evaluations may not be easily undone (Gregg et al., 2006). Third, we used Kawakami et al. (2007)’s procedure for the subliminal presentation of the face stimuli. Though this procedure may allow participants to identify the race of the depicted person, it may be more difficult to register the identity of the face stimulus. Arguing against this explanation, however, research with event related brain potentials (ERPs) suggests that participants can process the identity of masked face stimuli that are presented for these short durations (e.g., Lee, Lim, Lee, & Choi, 2009). Moreover, performance in the detection task indicated that participants were able to discriminate the face stimuli better than chance level.

**Experiment 3**

In Experiment 3, we changed our set-up to create more optimal circumstances for subliminal AA training effects to occur. First, novel non-words were used as stimuli. Previous research suggests that AA effects on evaluations are weaker for associatively rich stimuli than for associatively-impoverished stimuli such as non-words (e.g., Priester et al., 1996). Second, we tried to maximize the possibility that participants would be able to register the identity of the stimuli during the subliminal training task by modeling the subliminal presentation procedure after studies that established subliminal priming effects on the basis of the identity of non-words (Holcomb & Grainger, 2006; Beyersmann, Castles, & Coltheart, 2009). Third, we tested a larger sample of participants (N = 96) to have sufficient statistical power (power = 0.80) to detect even a small effect size. Finally, we included a manipulation that might facilitate association formation under conditions of subliminal stimulus presentations (Custers & Aarts, 2011).
Custers and Aarts (2011) proposed that association formation strongly depends on attentional processes. When people experience specific contingencies in the environment this may allow for the formation of associations in memory if their attentional system is prepared to process these contingencies. Under these circumstances, associative learning effects may arise even in the absence of awareness of the contingencies or in the absence of stimulus awareness (see also Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Dijksterhuis & Aarts, 2010). They further argued that the existence of a contingency relationship in the environment may tune participants’ attention to detect other such relationships. Thus, when participants experience specific contingencies before a learning phase with subliminal stimulus presentations, this may facilitate effects of associative learning. They tested this in three experiments and obtained evidence that the learning of predictive relations in a subliminal priming task was facilitated when participants engaged in predicting targets on the basis of cues in an earlier, unrelated task.

If attention can be tuned to process contingency relations and if this facilitates association formation, then participants who experienced specific contingencies in an earlier AA training task may show stronger effects of subliminal AA training. To test this, we designed an experiment where participants performed an AA training task with supraliminal stimulus presentations prior to performing the subliminal AA training task. For half the participants the supraliminal AA training task involved specific stimulus-action contingencies (attention-tuning condition), for the other participants the identity of the stimulus did not predict the action: 50% contingency (control condition).

Method

Participants. Ninety-six native Dutch-speaking undergraduates (78 women) participated in exchange for a monetary reward of 5 euros. None of the participants had previously
participated in Experiments 1 or 2. Participants were randomly assigned to the attention-tuning condition or to the control condition.

**Apparatus and Materials.** Four non-words were selected as evaluation stimuli from previous studies that registered Dutch-speaking participants’ evaluations of non-words (Zanon, De Houwer, & Gast, 2012). The non-words for the supraliminal AA training task were UDIBNON and SARICIK. For the subliminal AA training task the non-words were LOKANTA and FEVKANI. The attribute stimuli used in the IAT were four positive words (the Dutch words for happy, pleasant, fun, and nice) and four negative words (the Dutch words for mean, unpleasant, irritating, and bad).

**Procedure.** Upon entering, participants were seated at a desk and were informed that they would perform a series of tasks. They were asked to remember that some of these tasks involved reacting to two non-existing words, specifically LOKANTA and FEVKANI (the non-words for the subliminal training task). Participants were then instructed that in the first task they would respond to the words “approach” and “avoid” by making approach or avoidance movements with the joystick. The same action framing instructions were used as in Experiment 2.

The supraliminal AA training task consisted of 40 trials. The non-words SARICIK and UDIBNON were each presented on 20 occasions. For participants in the attention-tuning condition, one of the non-words was always presented with the word approach and the other non-word was always paired with the word avoid. Assignment of the non-words SARICIK and UDIBNON to the approach or avoidance action was counterbalanced across participants. For participants in the control condition each non-word was equally often followed by the words approach and avoid. Each trial started with the presentation of a white fixation cross presented in the center of the screen for 500 ms, followed by a forward mask of # symbols presented for 500
ms. Then, one of the non-words was displayed for 500 ms after which the word approach or avoid was presented until participants responded correctly with the joystick by performing a vertical movement towards the screen or away from the screen. The inter-trial interval was 200 ms.

Subsequently, participants were informed that they would now perform another joystick task. Trials in the subliminal training task were identical to the supraliminal training task trials with the following exceptions. First, there were 2 blocks of 50 trials and in each of the blocks the non-word FEVKANI and the non-word LOKANTA were each presented on 25 occasions. Second, for participants in both conditions one of the non-words was always presented with the word approach and the other non-word was always presented with the word avoid. Assignment of the non-words FEVKANI and LOKANTA to the approach or avoidance action was counterbalanced across participants and across conditions. Third, the presentation time of the non-words was limited to 40 ms.

In the IAT, participants categorized attribute words as ‘positive’ or ‘negative’ and target words FEVKANI and LOKANTA as ‘Fevkani’ or ‘Lokanta’ by using the E and I keys of a computer keyboard. To avoid that stimuli were classified only on the basis of simple perceptual features, each target stimulus was presented in four different fonts (lower case Arial Black, upper case Arial Black, lower case Fixedsys, and upper case Fixedsys), resulting in 8 different target stimuli. All other procedural details of the IAT were identical to Experiment 1.

Participants then completed liking ratings and thermometer ratings of self-reported warmth feelings towards each of the four non-words on two 8-point Likert scales (0 = not liked/warm at all; 7 = completely liked/warm). For each non-word, we collapsed these ratings
into one explicit rating score by averaging the respective ratings. The internal consistency of this measure was good (mean Cronbach’s \(\alpha = 0.76, SD = 0.04\)).

Next, participants answered questions about the supraliminal AA training task: They indicated (1) whether they had noticed any regularities in the presentations of the words approach or avoid and the non-words UDIBNON and SARICI K, and (2) what action word had been presented most often with these non-words by choosing from four options (i.e., ‘approach’, ‘avoid’, ‘both an equal number of times’ or ‘I don’t remember’). For the subliminal training task, participants reported whether they had ever seen a stimulus appear before the words approach or avoid (yes/no). Then, participants were informed that the non-words FEVKANI and LOKANTA had been presented during this task and they indicated whether they had ever seen these words and what action word they thought had been presented most often with these non-words.

Finally, participants performed a d’ perceptibility task which consisted of 50 trials. Trials were identical to trials in the subliminal AA training task with the exception that (1) the words approach and avoid were replaced with strings of ‘XXXX’, and (2) participants indicated whether FEVKANI or LOKANTA had been presented by pushing key 1 or 2 on the numeric keypad.

**Results**

**Awareness and Perceptibility.** First, we investigated participants’ awareness of the contingencies in the supraliminal AA training task. As expected, participants in the attention-tuning condition indicated more often that they had noticed regularities in the presentation of action words and non-words (52% of the time) than participants in the control condition (25% of the time), \(t(94) = 2.81, p = .006\). Participants in the attention-tuning condition selected the correct action for the non-words (\(M = 44\%, SD = 48\%) more often than the incorrect action (\(M = 5\%, SD = 21\%\)), \(t(47) = 4.79, p < .001\). Participants indicated that they did not know the correct
action for the nonwords 43% of the time ($SD = 48\%$) and that they had performed both actions an equal number of times 8% of the time ($SD = 24\%$).

For the subliminal AA training task, 20 participants (21%) indicated that they had noticed that a stimulus was sometimes presented before the approach/avoid word and 16 participants (17%) indicated that the non-words FEVKANI and LOKANTA were presented. Participants did not select the correct action for the non-words ($M = 6\%$, $SD = 23\%$) significantly more often than the incorrect action ($M = 5\%$, $SD = 21\%$), $t(95) = 0.32$, $p = .75$.

Participants’ overall accuracy in the perceptibility task was 65% ($SD = 11\%$). The signal detection sensitivity measure $d'$ was computed with hits defined as LOKANTA responses on trials where LOKANTA was presented, and false alarms as LOKANTA responses on trials where FEVKANI was presented. The $d'$ score indicated that participants’ detection performance was above chance level ($M = 0.86$, $SD = 0.71$), $t(95) = 11.81$, $p < .001$.

**Supraliminal AA training.** We examined whether supraliminal AA training caused changes in explicit rating scores for contingency aware and contingency unaware non-words. Non-words were classified as contingency aware if participants had indicated the correct action for the non-word (44%) and as contingency unaware if they had indicated the incorrect action or if they indicated that they didn’t know the correct action or that both actions had been performed an equal number of times (56%). Analyses were restricted to the data of participants in the attention-tuning condition because only these participants experienced specific stimulus-action contingencies. In line with Van Dessel, De Houwer, and Gast (2015), analyses were performed with item-based linear mixed effects models as implemented in R package lme-4 (Bates, Maechler, Bolker, & Walker, 2014). We tested for an effect of supraliminal AA training by fitting a model that included Participant as a random factor and the fixed factors of Action (approach,
avoid), Non-word (Saricik, Udibnon), and Contingency Awareness (contingency aware, contingency unaware). We observed a main effect of Action, $\chi^2(1) = 3.84, p = .050$, which was qualified by a marginally significant interaction effect of Action and Contingency Awareness, $\chi^2(1) = 2.88, p = .090$. A preference for approached non-words ($M = 3.48, SD = 1.45$) over avoided non-words ($M = 2.50, SD = 1.11$) was observed for contingency aware non-words, $\chi^2(1) = 5.80, p = .016, d = 0.67, BF_1 = 5.81$, but not for contingency unaware non-words, $\chi^2(1) = 0.03, p = .87, d = 0.07, BF_0 = 3.01$.

**Subliminal AA Training.** IAT D4 scores were calculated such that higher scores indicate a preference for the approached non-word over the avoided non-word. Importantly, we did not observe a significant preference for the approached non-word ($M = -0.03, SD = 0.51$), $t(95) = -0.52, p = .60, BF_0 = 9.14$. An ANOVA on IAT score with Non-word (Lokanta, Fevkani), Condition (attention-tuning, control), and Action Framing (approach by pulling, approach by moving towards the screen) as between-subjects factors only revealed a main effect of Non-word, $F(1,88) = 15.68, p < .001, BF_1 = 209.41$, indicating that participants preferred LOKANTA over FEVKANI. The main effect of Condition was not significant, $F(1,88) = 0.41, p = .52, BF_0 = 3.95$. We did not observe a significant preference for the approached non-word for participants in the attention-tuning or control condition, $t_s < 0.07, p_s > .95, BF_{0s} > 4.44$. No other main or interaction effects were significant, $F_s < 0.98, p_s > .32$. We performed a linear regression of $d'$ scores on IAT scores. The intercept, $b_1 = -0.06, t(94) = -0.73, p = .47$, and slope were not significant, $b_2 = 0.04, t(94) = 0.51, p = .61$.

An explicit preference score for the approached non-word was calculated by subtracting participants’ explicit rating score for the avoided non-word, from the explicit rating score for the approached non-word ($M = -0.39, SD = 4.16$). This score correlated significantly with the IAT.
score, $r(94) = 0.59, p < .001$. Importantly, we did not observe a significant preference for the approached non-word ($M = 6.94, SD = 2.45$) over the avoided non-word ($M = 7.32, SD = 2.43$), $r(95) = -0.91, p = .37, BF_0 = 11.38$. An ANOVA on the explicit rating score revealed only the main effect of Non-Word, $F(1,88) = 28.00, p < .001, BF_1 = 11261.91$, but no main effect of Condition, $F(1,88) = 0.18, p = .67, BF_0 = 4.39$, or any interaction effect, $Fs < 1.35, ps > .24$. We did not observe a preference for the approached non-word in the attention-tuning or control condition, $ts < 0.01, ps > .99, BF_0s > 6.03$.

**Discussion**

Despite more optimal conditions for subliminal perception and an increased power for detecting small effects we did not obtain evidence for the conclusion that subliminal AA training caused changes in evaluations of novel non-words. We also did not find more robust subliminal AA training effects when participants experienced specific stimulus-action contingencies in an earlier supraliminal AA training task. This might indicate that the attention-tuning manipulation did not enhance participants attention to contingency relations (but note that participants in the attention-tuning condition did indicate awareness of the contingencies in the supraliminal AA training task) or that tuning participants attention to process these relations does not facilitate associative learning (but see Custers & Aarts, 2011). Alternatively, the findings by Custers and Aarts (2011) may simply not extend to AA training effects because awareness of the specific contingencies is crucial for AA training effects rather than attention to contingency relations in general. In line with this idea, we observed effects of *supraliminal* AA training only when participants indicated awareness of the stimulus-action contingencies, corroborating previous findings (Van Dessel, De Houwer, & Gast, 2015).

**General Discussion**
In three experiments with different stimuli and procedures, we tested whether subliminal AA training changes stimulus evaluations. Participants repeatedly performed AA movements in response to approach or avoidance cues that were preceded by subliminally presented stimuli: faces of Black or White people (Experiment 1), unfamiliar faces (Experiment 2), or unfamiliar non-words (Experiment 3). In contrast to previous findings (Kawakami et al., 2007; Jones et al., 2013), we did not observe changes in implicit or explicit evaluations of these stimuli that could be attributed to the subliminal AA training. We did, however, observe AA effects when participants were provided with correct or incorrect information about the stimulus-action contingencies in the subliminal AA training task and when participants performed *supraliminal* AA training and were aware of the stimulus-action contingencies.

**How can these results be reconciled with previous findings?**

Kawakami et al. (2007) and Jones et al. (2013) reported effects of subliminal AA training in two and three experiments, respectively; we report three experiments that do not show these effects. To explain these discrepant results two options can be considered. On the one hand, it is possible that AA training cannot cause changes in stimulus evaluations when the stimuli are presented subliminally during the training phase. Under that assumption, we might look at the original findings and suggest that the reported effects are simply Type I error, that is, a false rejection of the null hypothesis. In line with this idea, Bayesian analyses indicated that these original results provided only anecdotal evidence for subliminal AA training effects while our results consistently provided more substantial evidence favoring the idea that subliminal AA training does not allow for changes in evaluations. A Bayes Factor meta-analysis including the data of the successful subliminal AA training studies by Kawakami et al. and Jones et al., as well as our studies indicates that the available evidence favors the null hypothesis that subliminal AA
training does not influence implicit evaluations, \( BF_0 = 3.64 \), or explicit evaluations, \( BF_0 = 6.51 \) (Table 1). However, some caution is warranted when interpreting these results, most importantly because the subliminal training studies all differed substantially in their methodology.

Alternatively, it is possible that the effects reported by Kawakami et al. and Jones et al. occurred because stimuli were on some occasions presented above the perceptual threshold. By using \( d' \) measures of stimulus perceptibility, we found evidence that, with the subliminal AA training procedure used by Kawakami et al., some participants’ are able to detect the presented stimuli. Although we did not observe that this caused changes in evaluations in our experiments, this may have contributed to the original reports of subliminal AA training. One could object that in three of the experiments reported by Kawakami et al. and Jones et al. effects were observed even though participants did not report awareness of the stimuli when probed. As previously contended, however, the accuracy of these reports may be limited because (1) awareness was assessed only after the conditioning and evaluation phase, and (2) participants do not always report having been aware of a stimulus despite having some subjective experience of it (Cleeremans, 2001).

On the other hand, it is possible that subliminal AA training effects can be reliably observed, but we failed to do so. The detection of these effects, then, may require certain methodological idiosyncrasies that we failed to incorporate in our experiments. First, for some reason, our procedures might have interfered with the perception of the subliminal stimuli (Bargh & Morsella, 2008). It is unlikely that this was a problem in our experiments. Not only did we model our subliminal presentation procedures after Kawakami et al.’s study and other studies that report robust subliminal priming effects, also did the \( d' \) perceptibility scores indicate that some participants were even able to consciously identify the identity of the presented stimuli. Note,
however, that d’ perceptibility scores may overestimate awareness to the stimuli (Vermeiren & Cleeremans, 2012; but see Amihai, 2012). For instance, directing participants’ attention to the subliminal stimuli during the d’ perceptibility task may lead to higher visibility of these stimuli. Second, AA training may cause changes in evaluations only when the training involves a sufficiently large number of training trials (e.g., because associative learning is a slow and gradual process; Rydell & McConnell, 2006). Again we see no reason why this would be a problem in our studies. We matched the number of training trials to that in Kawakami et al. (Experiment 1) or to the number of training trials recommended by Woud, Becker, and Rinck (2011) for AA training with novel stimuli. Third, subliminal AA training effects may be limited to well-known stimuli such as spiders and well-known social groups. For instance, visual attention may be more strongly directed toward a well-known stimulus because the stimulus’ related evaluation is particularly accessible (Roskos-Ewoldsen and Fazio, 1992; Young & Fazio, 2013). However, in both our experiments with novel stimuli, we presented the stimuli in the instructions directly preceding the subliminal training task to ensure that a representation of the stimulus would recently have been made accessible. Moreover, participants were asked to remember these stimuli (or stimulus-action contingencies), which may even facilitate selective visual attention (Soto, Humphreys, & Heinke, 2006; Soto, Heinke, Humphreys, & Blanco, 2005) and thus learning (e.g., Mackintosh, 1975).

In sum, although we realized different variations of an AA training task that are representative of AA training tasks in the literature, we did not replicate or establish effects on subliminally presented stimuli in any of these variations. Of course such a set of results can never exclude the possibility that an effect on subliminally presented stimuli might occur under a different set of conditions. It does, however, indicate that these effects, if they exist, depend on
boundary conditions that are not yet identified. Further research is necessary to determine if and under what circumstances subliminal AA training effects can be reliably established. In the absence of this research, one should be cautious in drawing strong theoretical conclusions on the basis of previous reports of subliminal AA training effects. This also means that evidence for subliminal AA training is currently not reliable enough to be treated as conclusive evidence in the debate between association formation versus propositional models of associative learning in general and AA training specifically.

**Contingency awareness and AA training**

The current results fit nicely with the idea that awareness of the stimulus-action contingencies mediates AA training effects (Van Dessel, De Houwer, & Gast, 2015). If contingency awareness is necessary to obtain AA training effects, then subliminal AA training effects should not occur. Of course, the reversed inference does not hold: The fact that we did not observe AA training effects in the absence of awareness of the stimuli does not prove that AA training cannot cause changes in evaluations in the absence of contingency awareness. However, some aspects of our data do support a causal role of contingency information in AA training effects. First, in Experiment 2, we obtained evidence that aware contingency knowledge can lead to AA effects without actual practice of the AA contingencies. When participants were informed about the stimulus-action contingencies that would be presented in the subliminal AA training task, they exhibited changes in implicit and explicit evaluations in line with this information (also see Van Dessel, De Houwer, Gast, & Smith, 2015). This effect occurred even when the contingency information contrasted with the contingencies that were actually presented in the training task. Second, the results of Experiment 3 provide evidence that contingency awareness is also a necessary condition for AA training effects. When participants were provided with
supraliminal AA training, a preference for approached non-word stimuli was observed only when participants were able to consciously report the relation between non-word and action. This corroborates previous findings that AA training influences evaluations (of novel faces) only in the presence of contingency awareness (Van Dessel, De Houwer, & Gast).

**Theoretical implications**

The current results have important implications for accounts of AA training effects. Most importantly, they pose a challenge to theories that assume that the repeated paring of stimuli and AA actions allows for the automatic formation of associations that underlie (implicit) evaluations (Woud et al., 2013; Phills et al., 2011). For our findings to be reconciled with these accounts it seems necessary to assume that specific boundary conditions determine when these automatic effects may arise. Current association formation theories, however, remain unclear on the necessary environmental conditions (Shanks, 2007). In Experiment 3, we explored one possible boundary condition of unconscious associative learning (i.e., whether attention is tuned to process contingency relations; see Custers & Aarts, 2011). However, we failed to find evidence that our attention-tuning manipulation moderated subliminal training effects. In the absence of a clear description of boundary conditions for association formation, the current results may be more easily explained by an alternative, propositional account of AA training effects. Because a propositional account suggests that the acquisition of information about stimulus-action contingencies directly causes changes in implicit and explicit stimulus evaluations, this fits well with the observation that contingency awareness plays an important role in establishing AA training effects.

Our findings may also advance the debate about the validity of dual-process and single-process propositional models of evaluative learning. Dual-process models assume two routes of
evaluative learning: learning on the basis of the automatic formation of associations and learning on the basis of conscious propositional reasoning (Smith & DeCoster, 2000; Strack & Deutsch, 2004). In contrast, single-process propositional theories postulate that all learning necessarily involves the conscious acquisition of propositional information (e.g., De Houwer, 2009; Mitchell, De Houwer, & Lovibond, 2009). Proponents of these theories argue that unambiguous evidence in favor of unconscious associative learning is scarce and the existence of an automatic association-formation mechanism can therefore be questioned (see Mitchell et al., 2009, for a review). In the evaluative conditioning (EC) literature, for example, investigations which used more fine-grained methodologies often failed to find evidence for subliminal EC and for EC without contingency awareness (Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Gast, De Houwer, & De Schryver, 2012; but see Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012; see Sweldens, Corneille, & Yzerbyt, 2014, for a recent review).

In the area of AA training effects, Kawakami et al. (2007) and Jones et al. (2013) report evidence for subliminal effects. As we noted in the introduction, these findings have the potential to provide strong support for the hypothesis that preferences can be learned unconsciously, and hence support the idea that an automatic association formation mechanism exists. This is not only relevant for understanding AA training effects, but also for the more general debate about associative, propositional, and dual-process models. Considering the relevance of such findings the reports of unconscious AA training effects, however, have not been scrutinized sufficiently, unlike reports of unaware learning in other research areas. In the current paper we both evaluate the previous evidence for subliminal AA training effects and report three additional studies that test it. Across our analyses we consistently failed to find reliable evidence for the existence of subliminal AA training effects. We nevertheless hope that future studies will further try to
establish the empirical validity of findings that, if valid, may provide strong evidence for the existence of an association-formation mechanism.

Our results also have important implications for theories of evaluation. Evaluation is often explained in terms of dual-process models in which implicit and explicit evaluation depend on different systems or processes (Gawronski & Bodenhausen, 2006, 2011, 2014; Rydell & McConnell, 2006; Rydell, McConnell, Mackie, & Strain, 2006; Rydell, McConnell, Strain, Claypool, & Hugenberg, 2007). Implicit evaluations are typically thought to emerge from associative mental processes that operate through the spreading activation of associations in memory. Because associations are assumed to be formed gradually and automatically, implicit evaluations may not be easily changed (Gregg et al., 2006). Explicit evaluations are typically assumed to be governed by propositional processes that operate on the basis of propositional reasoning, and therefore may more strongly and immediately reflect new information. In line with this idea, Rydell and McConnell (2006) showed that verbal information about a novel stimulus caused changes in the explicit, but not implicit evaluation of this stimulus, whereas the repeated pairing of the same stimulus with valenced primes, presented subliminally, changed only implicit evaluations. In Experiment 2, we provided evidence that directly contrasts these results. Both implicit and explicit evaluations changed as the result of verbal information about the contingencies between a stimulus and AA action, but not as the result of the repeated pairing of the stimulus with AA actions. This accords with recent findings that both implicit evaluations and explicit evaluations can be changed even by a single piece of verbal information (Cone & Ferguson, 2015; De Houwer, 2006; Gast & De Houwer, 2012; Peters & Gawronski, 2011; Van Dessel et al., 2015; Zanon et al., 2014) and fits more easily with theories which assume that propositional information influences both implicit and explicit evaluations (De Houwer, 2014).
Again, further research is required to distinguish between the various accounts. A first step in this endeavour can be to try and replicate any studies reporting evidence that clearly favors the idea that implicit and explicit evaluative change relies on separate processes.

Concluding Remarks

On the basis of the available evidence, we believe that there is insufficient evidence to conclude that AA training causes changes in evaluations in the absence of awareness. If subliminal AA training effects do exist, they must be subjected to very stringent boundary conditions. In contrast, evidence seems to be accumulating that conscious propositional knowledge about stimulus-action contingencies plays an important role in AA training effects. Nevertheless, given the important theoretical implications of subliminal AA training effects, we hope that future studies will further explore these effects and attempt to replicate studies reporting these effects.

Acknowledgments

The authors would like to thank Kerry Kawakami and Curtis Phillis for their help in establishing the procedure for Experiment 1, and Maarten De Schryver for his statistical support in the analyses of this paper.

APA copyright notice: This article may not exactly replicate the final version published in the APA journal. It is not the copy of record. http://www.apa.org/pubs/journals/psp/
References


Footnotes

1 In contrast to the procedure by Kawakami et al. (2007), we decided not to include a condition where participants received training to avoid photographs of Black people and approach photographs of White people. This was done out of ethical concerns and because participants in the control condition and avoid Blacks condition typically do not exhibit significant changes in their evaluations (see Kawakami et al.; Phills et al., 2011).

2 We ensured by careful pre-testing that presentation times of photographs and backward masks, as registered by the Direct RT program, were exactly 23.52 ms and 35.29 ms on each presentation.

3 In both experiments the novel stimuli were introduced shortly before participants performed the AA training task to ensure that participants would not have to create a representation of the stimulus under conditions of subliminal presentation.

4 Some theories have argued that AA effects depend on whether the AA action consists of arm flexion or arm extension (e.g., Cacioppo, Priester, & Berntson, 1993), or on the specific instructions that are used to frame these actions as approach or avoidance (e.g., Laham et al., 2014). Because Kawakami et al. (2007) and Jones et al. (2013) used different action framing instructions (and incongruent mappings of movement to AA action label), we decided to include both action framing instructions in our experiment.

5 Including these trials did not result in any shift in significance for any of the reported effects.
### Subliminal Approach-Avoidance Training Experiment Results

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Evaluation object</th>
<th>Evaluation task</th>
<th>N</th>
<th>Test statistic</th>
<th>Bayesian t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kawakami et al. (2007, Exp. 2)</td>
<td>Black people</td>
<td>IAT</td>
<td>50</td>
<td>(t(47) = 2.28, p = .027)</td>
<td>BF(_1) = 2.31, Anecdotal ((H_1))</td>
</tr>
<tr>
<td>Jones et al. (2011, Exp. 1)</td>
<td>Insects</td>
<td>Rating task</td>
<td>42</td>
<td>(t(40) = 2.08, p = .044)</td>
<td>BF(_1) = 1.64, Anecdotal ((H_1))</td>
</tr>
<tr>
<td>Jones et al. (2011, Exp. 2)</td>
<td>Spiders</td>
<td>Personalized IAT</td>
<td>118</td>
<td>(t(116) = 2.04, p = .044)</td>
<td>BF(_1) = 1.26, Anecdotal ((H_1))</td>
</tr>
<tr>
<td>All previous experiments</td>
<td></td>
<td></td>
<td>210</td>
<td></td>
<td>BF(_1) = 46.46, Strong ((H_1))</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>Black people</td>
<td>IAT</td>
<td>62</td>
<td>(t(60) = -0.21, p = .83)</td>
<td>BF(_0) = 4.49, Substantial ((H_0))</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rating task</td>
<td></td>
<td>(t(60) = 0.19, p = .85)</td>
<td>BF(_0) = 3.34, Substantial ((H_0))</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>Unfamiliar faces</td>
<td>Evaluative priming task</td>
<td>76</td>
<td>(t(74) = -0.17, p = .87)</td>
<td>BF(_0) = 6.49, Substantial ((H_0))</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rating task</td>
<td></td>
<td>(t(74) = -0.28, p = .78)</td>
<td>BF(_0) = 7.00, Substantial ((H_0))</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>Novel non-words</td>
<td>IAT</td>
<td>96</td>
<td>(t(95) = -0.52, p = .60)</td>
<td>BF(_0) = 9.14, Substantial ((H_0))</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rating task</td>
<td></td>
<td>(t(95) = -0.91, p = .37)</td>
<td>BF(_0) = 11.38, Strong ((H_0))</td>
</tr>
<tr>
<td>All current experiments</td>
<td></td>
<td>Explicit evaluation tasks</td>
<td>234</td>
<td></td>
<td>BF(_0) = 10.34, Strong ((H_0))</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implicit evaluation tasks</td>
<td></td>
<td></td>
<td>BF(_0) = 10.82, Strong ((H_0))</td>
</tr>
<tr>
<td>All subliminal AA training experiments</td>
<td></td>
<td>Explicit evaluation tasks</td>
<td>352</td>
<td></td>
<td>BF(_0) = 3.64, Substantial ((H_0))</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implicit evaluation tasks</td>
<td>326</td>
<td></td>
<td>BF(_0) = 6.51, Substantial ((H_0))</td>
</tr>
</tbody>
</table>
Table 2.

*Mean RTs in the evaluative priming task in Experiment 2 as a function of Prime AA Instructions, Prime AA Training, and Target Valence.*

<table>
<thead>
<tr>
<th></th>
<th>Approach Instructions</th>
<th>Avoid Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Approach training</td>
<td>Avoidance training</td>
</tr>
<tr>
<td>Positive Target</td>
<td>565 (83)</td>
<td>565 (93)</td>
</tr>
<tr>
<td>Negative Target</td>
<td>593 (84)</td>
<td>593 (90)</td>
</tr>
</tbody>
</table>

|                  | Approach Training     | Avoidance training |
| Positive Target  | 571 (77)              | 572 (78)           |
| Negative Target  | 586 (83)              | 589 (84)           |

*Note.* Standard deviations are in parentheses.
Table 3.

*Mean explicit rating scores in Experiment 2 as a function of AA Instructions and AA Training.*

<table>
<thead>
<tr>
<th>Approach Instructions</th>
<th>Avoid Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approach training</td>
<td>Avoidance training</td>
</tr>
<tr>
<td>3.91 (1.62)</td>
<td>3.86 (1.38)</td>
</tr>
<tr>
<td>Approach Training</td>
<td>Avoidance training</td>
</tr>
<tr>
<td>2.87 (1.41)</td>
<td>2.99 (1.43)</td>
</tr>
</tbody>
</table>

*Note.* Standard deviations are in parentheses.