The Mere Exposure Instruction Effect: Mere Exposure Instructions Influence Liking

Pieter Van Dessel[1]

Gaëtan Mertens[2]

Colin Tucker Smith[3]

Jan De Houwer[1]

[1]Ghent University

[2]Utrecht University

[3]University of Florida

Author note: Correspondence concerning this article should be addressed to Pieter Van Dessel, Department of Experimental-Clinical and Health Psychology, Ghent University, Dunantlaan 2, 9000 Ghent, Belgium. Phone: +32 3 9264 91 43. E-mail: Pieter.VanDessel@UGent.be.
Pieter Van Dessel is supported by a Postdoctoral fellowship of the Scientific Research Foundation, Flanders (FWO-Vlaanderen). Jan De Houwer is supported by Methusalem Grant BOF16/MET_V/002 of Ghent University and by the Interuniversity Attraction Poles Program initiated by the Belgian Science Policy Office (IUAPVII/33). The authors would like to thank Sean Hughes and Sascha Topolinski for their comments on an earlier draft of this manuscript.

Abstract

The mere exposure effect refers to the well-established finding that people evaluate a stimulus more positively after repeated exposure to that stimulus. We investigated whether a change in stimulus evaluation can occur also when participants are not repeatedly exposed to a stimulus, but are merely instructed that one stimulus will occur frequently and another stimulus will occur infrequently. We report seven experiments showing that (1) mere exposure instructions influence implicit stimulus evaluations as measured with an Implicit Association Test, personalized Implicit Association Test, or Affect Misattribution Procedure, but not with an Evaluative Priming Task, (2) mere exposure instructions influence explicit evaluations, and (3) the instruction effect depends on participants' memory of which stimulus will be presented more frequently. We discuss how these findings inform us about the boundary conditions of mere exposure instruction effects, as well as the mental processes that underlie mere exposure and mere exposure instruction effects.

Keywords: mere exposure, instructions, implicit evaluation, IAT, evaluative priming

The Mere Exposure Instruction Effect:

Mere Exposure Instructions Influence Liking

It has been recognized for several decades that many aspects of a person's behavior are determined largely by his or her likes and dislikes (Allport, 1935). Accordingly, understanding how preferences are formed and how they can be influenced is an important aim of psychological science. Research has shown that preferences can form as a result of direct experiences with a stimulus. These direct experiences typically involve regularities with regard to (a) the mere presence of the stimulus, (b) stimuli that are paired with the stimulus, or (c) actions that are paired with the stimulus (De Houwer, 2007, 2009a). With regard to (a), research on the mere exposure (ME) effect has shown that frequent exposure to a particular stimulus can lead to a more positive evaluation of that stimulus. This effect was reported early on in the history of psychology (Fechner, 1876; Maslow, 1937), and has received widespread attention since the pivotal publication of Zajonc (1968). A meta-analysis of 208 independent experiments established the ME effect as a reliable effect with a mean effect size of $d = 0.56$ (Bornstein, 1989). With regard to (b), evaluative conditioning (EC) research revealed that the evaluation of a stimulus can be influenced by pairing that stimulus with positive or negative stimuli (see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010, for a review). With regard to (c), studies have shown that the repeated performance of positively valenced actions such as approach movements or negatively valenced actions such as avoidance movements in response to a stimulus can lead to more positive or negative stimulus evaluations, respectively (Kawakami, Phills, Steele, & Dovidio, 2007; Woud, Maas, Becker, & Rinck, 2013).

Recent studies have established that some of these effects can occur also on the basis of mere instructions. In studies on instruction-based EC, it has been shown that changes in stimulus evaluations can arise when participants are informed about future stimulus pairings. For instance, De Houwer (2006) told participants that they would perform an experiment in which one novel stimulus (e.g., the letter string UDIBNON) would be paired with positive pictures and another novel stimulus (e.g., BAYRAM) would be paired with negative pictures. Despite the fact that participants never actually experienced the stimulus pairings, participants exhibited a preference for the former stimulus over the latter. Likewise, a recent study found effects similar to those of approach-avoidance training when participants did not actually perform approach-avoidance actions but were merely instructed that they would later have to perform these actions (Van Dessel, De Houwer, Gast, & Smith, 2015). For instance, participants who received instructions to approach one fictitious social group (e.g., Niffites) and avoid another fictitious social group (e.g., Luupites) showed a preference for the former group. These instruction-based effects were found to critically depend on participants' memory of the instructions such that only participants who correctly remembered the instructions showed these effects.

In accordance with experience-based procedures, it has been demonstrated that EC instructions and approach-avoidance instructions can influence not only explicit (i.e., non-automatic) but also implicit (i.e., automatic) stimulus evaluations (Gast & De Houwer, 2012; Van Dessel et al., 2015). These changes in implicit evaluations are not fully mediated by changes in explicit evaluations (Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016). The latter finding is important because it argues against an explanation of instruction effects in terms of mere demand compliance. Moreover, it challenges important models of implicit evaluation which assume that propositional information influences implicit evaluations only if this information is considered a valid basis for evaluation and, hence, is incorporated in explicit evaluations (Gawronski & Bodenhausen, 2006; see Van Dessel, De Houwer, Gast, et al., 2016, for a discussion).

The aim of the present study is to investigate whether changes in liking can also occur on the basis of ME instructions. This may seem implausible because the ME effect is often thought to result from mental processes that occur automatically when participants repeatedly experience stimulus presentations (Bornstein, 1989; Bornstein & D'Agostino, 1992; Topolinski, 2012). In line with this idea, some studies have provided evidence that the ME effect does not even require conscious recollection of the presentations of the stimuli (e.g., Bornstein & D'Agostino, 1992; Kunst-Wilson & Zajonc, 1980; Monahan, Murphy, & Zajonc, 2000). However, there are important reasons to believe that ME effects might (also) depend on processes that involve the conscious acquisition of propositional information. First, recent studies challenged the idea that ME effects are independent of stimulus awareness. In these studies, the ME effect occurred only when participants were aware that certain stimuli had been presented more often than others (Brooks & Watkins, 1989; de Zilva, Vu, Newell, & Pearson, 2013; Newell & Shanks, 2007; Stafford & Grimes, 2012). Second, Wang and Chang (2004) provided evidence that even the direction of the ME effect strongly depends on memory of stimulus presentations. They found that participants exhibited more positive evaluations of new stimuli that were judged to be old than of old stimuli that were judged to be new. These findings suggests that a person's propositional beliefs about the number of times stimuli are presented might be more important for the occurrence of a ME effect than the actual stimulus presentations.

In the current study, we adapted the instruction-based EC procedure of De Houwer (2006) in such a way that participants received instructions about a later phase in which they would encounter two novel words. One of the words would occur frequently and the other word would occur infrequently. After participants received these ME instructions, they were informed that they would first complete another task. This task was an implicit evaluation task that registered participants' implicit evaluations of the two words. We used this experimental set-up to test for ME instruction effects in three lab-based experiments and in four highly powered internet-based studies (see Zhou & Fishbach, 2016, for reasons why it can be important to establish new effects independently in both lab-based and internet-based studies). Note that our studies tested whether participants would show a preference for the instructed frequent word over the instructed infrequent word. This is different from ME studies in which the evaluation of a frequently presented stimulus is compared to the evaluation of a novel (i.e., never exposed) stimulus. We included an instructed infrequent word in our comparison to allow examination of the effect of instructions while keeping the number of exposures to the stimuli constant.

Although the main aim of this study was to examine whether ME instructions can influence stimulus evaluations, we also looked at two potential boundary conditions of these effects. First, we examined whether effects depend on the task that is used to measure evaluations. More specifically, across the different experiments, we used multiple tasks to measure implicit evaluations (Implicit Association Test, IAT, Greenwald, McGhee, & Schwartz, 1998, Experiments 1 and 4; personalized IAT, pIAT, Olson & Fazio, 2004, Experiment 2; Evaluative Priming Task, EPT, Fazio, Sanbonmatsu, Powell, & Kardes, 1986, Experiments 3, 5, and 6; Affect Misattribution Procedure, AMP, Payne, Cheng, Govorun, & Stewart, 2005, Experiment 7), as well as a measure of explicit evaluation (explicit rating task, Experiments 4-7). Each of the four implicit evaluation measures are sensitive to a number of factors other than the to-be-measured psychological construct of implicit evaluation (De Houwer, 2003; Gawronski & De Houwer, 2014). As a result, previous studies have found that effects do not necessarily converge on these measures. We included the IAT, EPT, and AMP because these are currently the most widely used implicit evaluation measures. These three measures differ in important ways (that might give rise to differences in observed effects). First, in the EPT and AMP, participants categorize target stimuli (i.e., valenced words or Chinese ideographs) as positive or negative. The evaluation stimuli (e.g., fevkani and lokanta) are included as primes that precede the presentation of the target stimuli and that might influence categorization responses. In contrast, in the IAT, participants perform two binary categorizations (i.e., the categorization of valenced words as positive or negative and the categorization of the evaluation stimuli on the basis of their identity). Because these categorizations are performed using the same response key, the (unintended) categorization of evaluation stimuli as positive or negative might influence task performance. It is therefore often assumed that IAT effects are based on fundamentally different processes than EPT and AMP effects (e.g., Payne & Gawronski, 2010). Second, whereas implicit evaluation scores in the IAT and EPT are calculated on the basis of response latencies, AMP scores are calculated on the basis of the number of positive and negative categorization responses. In addition to these three widely used implicit evaluation measures we also included the pIAT. This measure is assumed to minimize the impact of extra-personal knowledge in the IAT (Olson & Fazio, 2004), which is important because our aim was to measure personal liking of the stimuli. An overview of the most important procedural differences between the implicit evaluation measures is presented in Table 1.

Second, in Experiments 4-7 we also investigated whether ME instruction effects depend on participants' memory for the instructions. Based on the fact that effects of EC instructions and approach-avoidance instructions also critically depend on memory for the instructions, we expected that ME instructions would have a stronger influence on stimulus evaluations when participants have accurate memory of the instructions. If the acquisition of propositional information about the frequent or infrequent occurrence of a stimulus causes changes in participants' implicit stimulus evaluations, then this should depend on the extent to which participants register this information, which should be reflected in memory for the instructions. To gain information about the strength of evidence for the presence or absence of ME effects for participants who did or did not remember the instructions correctly, we supplemented traditional *t*-test analyses with Bayesian analyses. Bayesian analyses were performed with Cauchy prior width = .707, according to the procedures outlined by Rouder, Speckman, Sun, Morey, and Iverson (2009). They provide a Bayes Factor that gives 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). BFs between 1 and 3, between 3 and 10, and larger than 10, respectively designate ‘anecdotal evidence’, ‘substantial evidence’, and ‘strong evidence’ for either the null (BF_0) or the alternative hypothesis (BF_1) (Jeffreys, 1961).

Experiments 1, 2, and 3

Method

Participants. A total of 40 (33 women), 98 (67 women), and 39 (32 women) Dutch-speaking undergraduates participated in exchange for 4 euros in Experiments 1, 2, and 3, respectively. All participants had normal or corrected-to-normal vision. Participants were not given any information about the purpose of the experiment.

Apparatus and Materials. Two nonwords were used as evaluation stimuli, namely ‘LOKANTA’ and ‘FEVKANI’. These stimuli are often used in studies examining changes in (implicit) evaluations (e.g., Van Dessel, De Houwer, Roets, & Gast, 2016), including the seminal mere exposure studies of Zajonc (1968). Experiments were programmed and presented using the INQUISIT Millisecond Software package (Inquisit 3.0, 2011) on a Tori desktop PC with a 19-inch monitor.

Procedure. After participants had given informed consent, they were seated in front of the computer screen. Half of the participants read the following instructions (translated from Dutch):

In the following experiment you will see two words: The word LOKANTA will be presented often and the word FEVKANI will be presented rarely. It is very important that you remember this rule. You will need this information to complete the task *correctly. Before we present these words to you, you will complete a categorization task. This will take about 5 minutes of your time. Make sure that you do not forget how often you will see each word in the part of the study that follows the categorization task. Remember: The word LOKANTA will be presented often and the word FEVKANI will be presented rarely. Please press 'Continue' only when you are sure that you remember the instructions and are ready to begin the categorization task. Once you finish reading these instructions, this information will not be repeated!*

The other participants received identical instructions except that they were told that FEVKANI would be presented often (instructed frequent word) and LOKANTA would be presented rarely (instructed infrequent word). Whether the instructions presented the information about the instructed frequent word first and the information about the instructed infrequent word second, or vice versa, was counterbalanced across participants. After participants read the instructions, they could proceed to the next screen where the instructions for the implicit evaluation task were presented.

In the IAT of Experiment 1, participants categorized eight attribute words as ‘Positive’ (the Dutch words for HAPPY, HONEST, LOVE, and PEACE) or ‘Negative’ (the Dutch words for DEATH, SLIME, CANCER, and UGLY) and target words FEVKANI and LOKANTA as ‘Fevkani’ or ‘Lokanta’, by pressing a left (Q) or right (M) key on an AZERTY computer keyboard. All words were presented in uppercase letters in Arial Black font with font size 36. Participants began the IAT with 24 practice trials sorting the positive and negative attribute words and 20 practice trials sorting the target words. Next, participants completed two blocks of 48 trials in which FEVKANI and positive stimuli shared a single response key and LOKANTA and negative stimuli shared another response key (half of the participants completed the IAT in this way, whereas the other participants began by sorting LOKANTA and positive with the same key). IAT block order was orthogonally crossed with the between-subjects factor of instructed frequent word. Participants then practiced sorting target words with the response key assignment reversed for 24 trials and finally participants completed 2 blocks of 48 trials in which LOKANTA shared a response key with positive words and FEVKANI shared a response key with negative words (or vice versa). The order of the trials was determined randomly for each block and each participant separately. On each trial, a word 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 word was presented 400 ms later. If the response was incorrect, the word was replaced by a red “X” for 400 ms. The next word appeared 400 ms after the red “X” was removed from the screen.

The pIAT of Experiment 2 was identical to the IAT used in Experiment 1 with two exceptions. First, the category labels ‘positive’ and ‘negative’ were changed to “I like” and “I dislike”. Second, no error feedback was presented during the task (Olson & Fazio, 2004).

In the EPT of Experiment 3, participants categorized target words as either positive or negative using a left (A) and right (P) key on an AZERTY computer keyboard. In line with standard procedures (Spruyt, De Houwer, Hermans, & Eelen, 2007), a single trial consisted of the presentation of a fixation cross for 500 ms, a blank screen for 500 ms, a prime for 200 ms, a post-prime interval for 50 ms and the presentation of a target word for a maximum of 1500 ms. The inter-trial interval was set to vary randomly between 500 ms and 1500 ms. Whenever an incorrect response was made or participants did not respond prior to offset of the target stimulus, a red X was displayed in the center of the screen for 1000 ms before the next trial started. Participants were asked to respond as quickly as possible without making too many errors. Targets consisted of 10 positive words (e.g., the Dutch words for Love, Pleasure, Smile) and 10 negative words (e.g., the Dutch words for Hate, Pain, Sadness). The words FEVKANI and LOKANTA were used as prime stimuli. There were four different types of trials: trials with each of the two words as prime and a positive or negative word as target.

Participants first completed 15 practice trials and then completed 180 trials separated into three blocks of 60 trials, each containing 15 of the four types of trials, presented in random order.

Next, we assessed whether participants had correctly remembered the ME instructions. They were asked two questions: (1) "According to the instructions, how often will the word LOKANTA be presented?", and (2) "According to the instructions, how often will the word FEVKANI be presented?". Participants indicated their answer by selecting an option from a dropdown list (often/rarely/I don't remember). The order in which the questions were presented was determined randomly for each participant.

Finally, even though performance on this task was irrelevant for our hypotheses, participants performed 11 trials of a ME task in which they saw presentations of the frequent and infrequent instructed words. Participants were told that they would see a number of non-existing words presented on the screen and that they would not have to react to these words, but only to attentively watch the presentations. Each trial consisted of the presentation of the word FEVKANI or LOKANTA which remained on the screen for 1000 ms. The inter-trial interval was 1000 ms. The instructed frequent word was presented on nine occasions whereas the instructed infrequent word was presented on two occasions. This task was only included in order not to deceive participants in the earlier instructions and did not involve any measure of liking of the word stimuli.

Results

Data-preparation. In line with previous studies on the impact of instructions on stimulus evaluations (e.g., Van Dessel et al., 2015), we excluded the data of 12 (30%, Experiment 1), 29 (30%, Experiment 2), and 3 participants (8%, Experiment 3) because they did not correctly answer the memory questions. Including the data from all participants in the analyses reduced the magnitude of the instruction effects but did not change the statistical significance of any of the reported effects. We explored the role of memory for the instructions in ME instruction effects in Experiments 4-7, which were better suited to answer this question because they had higher statistical power. High statistical power is especially important for this question because the reliability of the necessary statistical tests depends on whether there are sufficient participants with different levels of memory for the instructions.

In Experiments 1 and 2, IAT and pIAT scores were calculated using the D4-algorithm (Greenwald, Nosek, & Banaji, 2003), such that higher scores indicate a stronger preference for the frequent instructed word. Spearman-Brown corrected split-half reliability of the IAT and pIAT scores, calculated on the basis of an odd-even split, was $r(26) = .62$ (Experiment 1), and $r(67) = .80$ (Experiment 2). [1]

In Experiment 3, data of EPT trials with an incorrect response were discarded (6.9%). In line with standard procedures at our lab (e.g., Spruyt, De Houwer, & Hermans, 2009), the impact of outliers was reduced by excluding trials with a response latency that deviated more than 2.5 standard deviations from an individual's mean latency for trials with the same prime and the same type of target (i.e., Lokanta-positive, Lokanta-negative, Fevkani-positive or Fevkani-negative trials; this resulted in an overall exclusion of 2.4% of the trials). EPT scores were created by (a) subtracting the mean latencies on trials with a positive target and the frequent instructed word prime from the mean latencies on trials with a negative target and the frequent instructed word prime, (b) subtracting the mean latencies on trials with a positive target and the infrequent instructed word prime from the mean latencies on trials with a negative target and the infrequent instructed word prime, and (c) subtracting the second difference score from the first difference score. Higher EPT scores indicate a stronger

[1] We analyzed IAT scores that were calculated using the D4 algorithm because these scores are most often used and are shown to bring improvement over other scores (e.g., they have better internal consistency, see Greenwald et al., 2003). For the sake of completeness, we also performed analyses on mean reaction times. Results of these analyses supported the reported conclusions and are reported in Electronic Supplementary Material 1.

preference for the frequent instructed word. Spearman-Brown corrected split-half reliability of the EPT score, calculated on the basis of an odd-even split, was $r(34) = .28$.

Data-analyses. We performed one-sample *t*-tests to examine whether participants' implicit evaluation scores in the three experiments were significantly larger than zero, indicating that participants preferred the instructed frequent word over the instructed infrequent word. We observed a significant effect of the ME instructions on IAT scores in Experiment 1 ($M = 0.23$, $SD = 0.42$), $t(27) = 2.84$, $p = .008$, $d_z = 0.54$, 95% confidence interval (CI) = [0.06, 0.39], $BF_1 = 5.32$, and on pIAT scores in Experiment 2 ($M = 0.16$, $SD = 0.48$), $t(68) = 2.72$, $p = .008$, $d_z = 0.33$, 95% CI = [0.04, 0.27], $BF_1 = 3.91$. We did not observe a significant ME instruction effect on EPT scores in Experiment 3 ($M = -3.34$, $SD = 46.15$), $t(35) = -0.43$, $p = .67$, $d_z = -0.07$, 95% CI = [-18.96, 12.28], $BF_0 = 5.06$ (Table 2).

Discussion

Experiments 1 and 2 showed that merely instructing participants that certain non-existing words would occur frequently or infrequently in a future task leads to more positive implicit evaluations of the instructed frequent word than of the instructed infrequent word. This effect is in line with typical ME effects, but it occurred even though participants were exposed to the stimuli only *after* their evaluations were probed. Hence, our results demonstrate a change in liking that is based only on instructions about stimulus frequency.

Results also revealed an important boundary condition of ME instruction effects. In Experiment 3, we did not observe a ME instruction effect on implicit evaluations as measured with an EPT. A number of explanations can be put forward for this result. First, ME instructions might influence IAT performance but not the actual (implicit) liking of the stimulus. For instance, Olson and Fazio (2004) proposed that IAT performance might not capture personal preferences but rather extra-personal knowledge (i.e., knowledge that a person has about societal views but regards as irrelevant for his or her own feelings about the attitude object). Participants might use this knowledge to improve their performance on the IAT, such that IAT scores reflect this knowledge rather than personal liking of the stimuli. However, this explanation is incompatible with the fact that we did observe a ME instruction effect on the pIAT which is specifically designed to minimize the impact of extra-personal knowledge. Another possibility is that both IAT and pIAT performance might be influenced by other non-attitudinal factors such as participants' recoding of IAT categories (Rothermund, Teige-Mocigemba, Gast, & Wentura, 2009). For instance, IAT categories might be recoded on the basis of asymmetries in the salience of IAT categories and this might facilitate IAT performance on certain blocks, producing IAT effects (Rothermund & Wentura, 2004). Importantly, processes that involve voluntary control may also contribute strongly to IAT performance (Fiedler & Bluemke, 2005), especially with novel stimuli (De Houwer, Beckers, & Moors, 2007). For instance, participants might consciously slow down responses in compatible or incompatible blocks, thereby influencing IAT effects. Thus, it is possible that ME instruction effects on the IAT reflect controlled changes in evaluation which are not registered with the EPT.

Second, the fact that we did not observe an effect on EPT scores may be due to a lack of power in this experiment. Indeed, a power analysis revealed that the power for revealing a small effect was very low (i.e., power = 0.32 for an effect size of d_z = 0.20). In combination with the fact that (1) we do not yet know whether ME instruction effects on implicit evaluation are typically of small or medium size and (2) evaluative priming procedures typically produce relatively small effect sizes and scores that are relatively low in reliability (Wittenbrink, 2007), it is therefore possible that the null effect observed in Experiment 3 constitutes a Type II error. Note, however, that the Bayes Factor indicated substantial evidence for the null hypothesis (reflecting the absence of a significant effect).

Experiments 4, 5, 6 and 7

In order to obtain more power for establishing whether ME instructions can cause changes in stimulus evaluations, we conducted four additional internet-based experiments that involved a large number of participants. We also investigated whether ME instruction effects are restricted to IAT-based implicit measures. More specifically, we examined effects of ME instructions on implicit evaluations as measured with either an IAT (Experiment 4), two different versions of an EPT (Experiments 5 and 6), and an AMP (Experiment 7). The addition of the AMP is important because this measure is based on a priming procedure (and thus fundamentally different from the IAT; see Gawronski & Ye, 2014) but is also known to result in more reliable effects than the EPT (e.g., Payne et al., 2005). Additionally, we examined whether ME instructions also influence explicit evaluations as measured with an explicit rating task (Experiments 4-7).

These experiments were also designed to investigate whether ME instruction effects depend on participants' memory of the instructions. Because an answer to this question requires sufficient statistical power (e.g., to detect a small effect size of d_z = 0.20 with sufficient power [power = 0.80] approximately 156 participants are needed in each memory condition), we decided to include a minimum of 700 participants in each of the experiments. Because we especially wanted sufficient statistical power in Experiments 5 and 6 that tested effects of ME instructions on EPT scores, we decided to include approximately 900 and 1200 participants in those studies. We were able to recruit this high number of participants by implementing our study on the Project Implicit research website (https://implicit.harvard.edu).

Method

Participants. Participants were 716 (Experiment 4), 898 (Experiment 5), 1259 (Experiment 6), and 724 (Experiment 7) visitors to the Project Implicit research website. Participation was restricted to United States citizens. In line with standard procedures of data-reduction for Project Implicit data (e.g., Smith, De Houwer, & Nosek, 2013), we excluded the data from participants who (1) did not complete all tasks (77 participants in Experiment 4: 10.75%; 116 participants in Experiment 5: 12.92%; 128 participants in Experiment 6: 10.17%; 120 participants in Experiment 7: 16.57%), (2) had error rates above 30% when considering all critical IAT test blocks or above 40% for any one of these blocks (10 participants in Experiment 4, 1.56%), (3) had error rates in the EPT that exceeded the population mean by more than 2.5 standard deviations (23 participants in Experiment 5: 2.94%, population mean = 7.23%, *SD* = 10.70%; 22 participants in Experiment 6: 1.95%, population mean = 4.29%, *SD* = 6.64%), or (4) responded either “positive” or “negative” to all AMP trials in any of the AMP blocks (55 participants in Experiment 7: 9.11%). The analyses were performed on the data of 629 participants (420 women, mean age = 35, *SD* = 14) in Experiment 4, 759 participants (446 women, mean age = 32, *SD* = 13) in Experiment 5, 1109 participants (676 women, mean age = 31, *SD* = 13) in Experiment 6, and 550 participants (326 women, mean age = 30, *SD* = 13) in Experiment 7.

Procedure. The procedure of Experiments 4-7 was identical to Experiments 1-3 (e.g., they involved the same evaluation stimuli and ME instructions) with the following exceptions. First, the IAT procedure of Experiment 4 differed from Experiment 1 in three ways. A first change involved the stimuli that were used in the IAT. Attribute stimuli consisted of ten positive words (e.g., Glorious, Success, Wonderful, Marvellous) and ten negative words (e.g., Agony, Unpleasant, Evil, Failure). Target words were FEVKANI and LOKANTA presented in eight different combinations of font types (Arial Black and Fixedsys), capitalizations (uppercase and lowercase), and size (16pt and 18pt), resulting in 8 different stimuli (for a similar procedure see Zanon, De Houwer, Gast, & Smith, 2014). This was done to avoid the possibility that target stimuli would be classified only on the basis of simple perceptual features. A second adaptation was that, in line with standard IAT procedures on the Project Implicit website, participants who made an error categorizing the stimuli saw a red “X” on the screen and had to correct their mistake in order to continue. Latencies were recorded until a correct response was made. A third adaptation was that participants now used the E and I keys of their computer keyboard to categorize the stimuli in the IAT.

Second, the EPT of Experiment 5 consisted of 120 instead of the 180 trials of Experiment 2, in accordance with previous studies on the Project Implicit website that used the EPT (van Dessel et al., 2015). Trials were separated into three blocks of 40 trials each. Prime stimuli were the eight different versions of the words FEVKANI and LOKANTA that were used as stimuli in the IAT of Experiment 4. Target stimuli for the EPT were 14 positive words (e.g., Attractive, Cheer, Beautiful) and 14 negative words (e.g., Annoying, Disaster, Loss). These target words were categorized by using the E and I keys of the computer keyboard. The EPT of Experiment 6 had a similar procedure with two exceptions. First, we implemented no more response deadline in the EPT (as is often the case in EPT procedures: e.g., Storbeck & Robinson, 2004). Second, the EPT consisted of only one block of 80 trials. This adaptation was made to counter the possibility that the ME instruction effect on the EPT might dissipate over time.

Third, in Experiment 7, implicit evaluations were registered with an AMP. In this task participants gave liking ratings for Chinese pictographs. Importantly, presentation of each pictograph was preceded by the presentation of a prime stimulus which consisted of the frequent instructed or infrequent instructed word. Participants were instructed to ignore the initial stimulus and rate whether the pictograph was either more pleasant or less pleasant than average. There were three different types of trials: trials in which the word FEVKANI, the word LOKANTA, or a grey rectangle served as prime. An individual trial began with the presentation of the prime for 100 ms. This was followed by a blank screen for 100 ms after which one of 72 different Chinese pictographs was presented for 100 ms. Finally, a mask image was presented (a black and white image) until the participant made a response by using the E or I key of their computer keyboard. Participants first completed three practice trials (with grey rectangle primes) and then completed 180 trials separated into three blocks of 60 trials, each containing 20 of each type of trial, presented in random order.

Finally, after the implicit evaluation task and before the memory questions and the ME task that ended the experiment, participants rated their liking of each of the words by answering two questions: “To what extent do you like FEVKANI?” and “To what extent do you like LOKANTA?”. Participants gave their evaluative ratings by selecting an option on a 9-point Likert scale (1= not liked at all; 9 = completely liked). The order of the two explicit evaluation questions was counterbalanced.

Results

Data-preparation. In Experiment 4, the IAT scores were calculated using the D2-algorithm (Greenwald et al., 2003). This procedure is similar to the D4 procedure used in Experiment 1 but it takes into account that the IAT procedure of Experiment 4 required participants to correct their mistakes before moving on to the next trial, in line with the standard IAT procedure on the Project Implicit website at the time of testing. Spearman-Brown corrected split-half reliability of the IAT score was $r(627) = .84$. In Experiments 5 and 6, EPT trials with an incorrect response were dropped (Experiment 5: 6.74%; Experiment 6: 3.63%) as were trials on which reaction times were at least 2.5 standard deviations removed from an individual’s mean for that type of trial (Experiment 5: 2.60%; Experiment 6: 2.93%). Spearman-Brown corrected split-half reliability of the EPT score was $r(757) = .37$ (Experiment 5), and $r(1126) = .34$ (Experiment 6). In Experiment 7, individual AMP scores were calculated by subtracting the proportion of “pleasant” responses on trials with the infrequent instructed word as prime from the proportion of “pleasant” responses on trials with the frequent instructed word as prime. Spearman-Brown corrected split-half reliability of the AMP score was $r(548) = .81$.

In all four experiments, an explicit rating score was calculated by subtracting the evaluative rating for the infrequent instructed word from the evaluative rating for the frequent instructed word. The explicit rating score correlated significantly with implicit evaluation scores in Experiment 4, $r(627) = .35$, $p < .001$, Experiment 5, $r(757) = .15$, $p < .001$, and Experiment 7 , $r(548) = .25$, $p < .001$, but not in Experiment 6, $r(1107) = .05$, $p = .11$.

To investigate the impact of memory for the instructions, we computed a memory variable on the basis of participants' answers to the questions related to how often the instructed frequent and instructed infrequent word would be presented. Participants' memory was coded as accurate when they selected the correct answer for both memory questions (Experiment 4: 459 participants, 73.93%; Experiment 5: 549 participants, 72.33%; Experiment 6: 731 participants, 65.92%; Experiment 7: 404 participants, 73.45%). Memory was coded as indiscriminate when only one of the questions was correctly answered or participants indicated 'I don't know' for one or both of the questions (Experiment 4: 105 participants, 16.69%; Experiment 5: 145 participants, 19.10%; Experiment 6: 247 participants, 22.27%; Experiment 7: 97 participants, 17.64%). Memory was coded as reversed when participants selected the incorrect option for both memory questions (Experiment 4: 59 participants, 9.38%; Experiment 5: 65 participants, 8.56%; Experiment 6: 131 participants, 11.81%; Experiment 7: 49 participants, 8.91%).

Data analysis of implicit evaluations. We subjected IAT scores in Experiment 4 to an analysis of variance (ANOVA) that contained Memory as a between-subjects factor. The effect of Memory on IAT scores was significant, $F(2, 626) = 5.36$, $p = .005$. Participants preferred the frequent instructed word when they had accurate memory for the instructions ($M = 0.12$, $SD = 0.48$), $t(458) = 5.42$, $p < .001$, $d_z = 0.25$, 95% CI = [0.08, 0.16], $BF_1 = 73208.54$, and when they had indiscriminate memory for the instructions ($M = 0.12$, $SD = 0.46$), $t(104) = 2.60$, $p = .011$, $d_z = 0.25$, 95% CI = [0.03, 0.20], $BF_1 = 2.63$, but not when they had reversed memory for the instructions (M = -0.09, SD = 0.48), $t(55) = -1.50$, $p = .14$, $d_z = -0.20$, 95% CI = [-0.22, 0.03], $BF_0 = 2.39$ (see Table 2 for an overview).

In Experiments 5 and 6, we did not observe an effect of Memory on EPT scores, Fs < 0.52, ps > .59. Participants did not exhibit a significant preference for the frequent instructed word when they had accurate memory for the instructions in Experiment 5 (M = 4.43, SD = 57.49), $t(548) = 0.58$, $p = .56$, $d_z = 0.02$, 95% CI = [-3.39, 6.25], $BF_0 = 17.60$, or in Experiment 6 (M = 1.98, SD = 103.59), $t(730) = 0.52$, $p = .61$, $d_z = 0.02$, 95% CI = [-5.54, 9.50], $BF_0 = 21.04$. We also did not observe a significant ME instruction effect for participants with indiscriminate or reversed memory in Experiments 5 or 6, ts < 0.64, ps > .53, $d_z s$ < 0.05, BF_0s > 7.10.

We observed a significant main effect of Memory on AMP scores in Experiment 7, $F(2, 547) = 3.89$, $p = .021$. Participants preferred the frequent instructed word when they had accurate memory of the instructions (M = 0.04, SD = 0.20), $t(403) = 4.01$, $p < .001$, $d_z = 0.20$, 95% CI = [0.02, 0.06], $BF_1 = 140.61$, but not when they had indiscriminate or reversed memory of the instructions, ts < 0.01, ps > .99, $d_z s$ < 0.01, BF_0s > 4.12.

Data analysis of explicit evaluations. We observed a significant main effect of Memory on explicit rating scores in Experiments 4, 5, and 6, Fs > 4.32, ps < .014, but not in Experiment 7, $F(2, 547) = 1.29$, $p = .28$. A significant preference for the frequent instructed word was observed for participants who correctly remembered the instructions in all experiments, ts > 2.11, ps < .037, d_zs [0.09 – 0.11], BF_1s [0.77 – 2.97]. Participants with indiscriminate memory also exhibited a significant instruction effect in Experiment 4 (M = 0.61, SD = 2.48), $t(104) = 2.52$, $p = .013$, $d_z = 0.25$, 95% CI = [0.13, 1.09], $BF_1 = 2.16$, and a marginally significant effect in Experiment 6 (M = 0.25, SD = 2.18), $t(246) = 1.78$, $p = .076$, $d_z = 0.11$, 95% CI = [-0.03, 0.52], $BF_0 = 2.94$, but no significant effect in Experiments 5 or 7, ts < 0.66, ps >.51, d_zs < 0.08, BF_0s > 7.26. Participants with reversed memory exhibited a significant contrast effect (i.e., a preference for the infrequent instructed word) in Experiment 6 (M = -0.41, SD = 2.31), t(130) = -2.04, p = .043, d_z = -0.18, 95% CI = [-0.81, -0.01], BF_0 = 1.39, a marginally significant contrast effect in Experiment 4 (M = -0.47, SD = 2.06), t(58) = -1.77, p = .082, d_z = -0.23, 95% CI = [-1.01, 0.06], BF_0 = 1.64, and no significant effects in Experiments 5 or 7, ts < 0.60, ps > .42, d_zs < 0.07, BF_0s > 4.70.

Combined analyses of Experiments 4-7. To compare the overall magnitude of the ME instruction effect on different implicit measures, we conducted an ANOVA on standardized implicit evaluation scores that included Measure (IAT, EPT, AMP) as a between-subjects factor. Scores were standardized by dividing the original implicit evaluation scores by their standard deviations. Data were included for all participants with accurate memory in the four experiments. The effect of Measure was significant, F(2, 2146) = 13.93, p < .001. The ME instruction effect was significantly smaller if participants performed the EPT than if they performed the IAT, t(1743) = 4.81, p < .001, BF_1 = 5324.94, or the AMP, t(1682) = 3.66, p < .001, BF_1 = 46.54. Participants who performed the IAT did not exhibit a larger instruction effect than participants who performed the AMP, t(567) = 0.65, p = .51, BF_0 = 10.66.

An ANOVA on standardized explicit evaluation scores in Experiments 4-7 with Memory and Previous Task (IAT, EPT, AMP) as between-subjects factors revealed only a main effect of Memory, F(2, 3038) = 7.59, p < .001. Participants preferred the frequent instructed word when they had accurate memory of the instructions (M = 0.10, SD = 0.98), t(2148) = 4.91, p < .001, d_z = 0.11, 95% CI = [0.06, 0.14], BF_1 = 3919.53. Participants who had indiscriminate memory of the instructions also exhibited a preference for the frequent instructed word (M = 0.12, SD = 1.03), t(593) = 2.79, p = .005, d_z = 0.11, 95% CI = [0.03, 0.20]. However, the Bayesian factor indicated that evidence for this effect was only anecdotal, BF_1 = 2.17. Participants who had reversed memory of the instructions exhibited a significant contrast effect (*M* = -0.13, *SD* = 1.08), $t(303) = -2.07$, $p = .040$, $d_z = -0.12$, 95% CI = [-0.25, -0.01]. The evidence for this effect, however, was also only anecdotal, $BF_1 = 1.03$.

Discussion

Experiments 4-7 provided corroborative evidence for ME instruction effects. ME instruction effects were not restricted to changes in IAT scores but also occurred in the AMP and an explicit rating task. We did not observe significant ME instruction effects on the EPT, even though we examined changes in EPT scores in two independent studies, with two different EPT procedures, and had sufficient power (power > .80) to detect even a small effect size of $d_z = 0.08$.[2]

Although we did not observe that ME instructions influence implicit evaluations as measured with an EPT, the fact that we did find effects on the IAT and AMP suggests that ME instructions influence actual implicit liking of the stimulus. ME instruction effects on the AMP cannot be explained by non-attitudinal factors that influence IAT performance (e.g., recoding of IAT categories). In line with the idea that ME instruction effects on both IAT and AMP are not entirely the result of controlled, non-automatic processes that involve the intentional use of the provided information, additional *t*-test analyses (reported in Electronic Supplementary Material 1) indicated that ME instructions produced larger changes in implicit evaluations (measured with the IAT and AMP) than explicit evaluations. This accords with other findings that instructions about stimulus events can have a direct influence on implicit evaluation (i.e., unmediated by changes in explicit evaluation; Van Dessel, De Houwer, Gast,

[2] To further investigate ME instruction effects on EPT performance, we also performed analyses with item-based linear mixed effects models (Bates, Mächler, Bolker, & Walker, 2014). This approach allowed us to investigate participants' raw reaction times (RTs) rather than an index of their performance as combined in one (unreliable) EPT score and to control for possible effects of counterbalancing factors such as the target words or prime words that were used. This approach has revealed more robust effects of approach-avoidance instructions in recent studies compared to traditional analyses (Van Dessel, Gawronski, Smith, & De Houwer, 2017). Importantly, however, the linear mixed effects regression analyses supported the conclusion of the main analyses that EPT performance was not influenced by ME instructions. We did not observe any main effects of Prime Type (instructed frequent or infrequent word) or interaction effects with Target Type (positive or negative target), χ^2s < 2.59, *p*s > .10.

et al., 2016). Moreover, statistical mediation analyses indicated that changes in explicit evaluations did not fully mediate the impact of ME instructions on implicit evaluation. In contrast, we observed full mediation of the ME instruction effect on explicit evaluation via implicit evaluation (see Electronic Supplementary Material 1) .[3]

The results of Experiments 4-7 also provided evidence that ME instruction effects depend on participants' memory of the instructions. When participants had accurate memory of the instructions, they preferred the instructed frequent word in all experiments (except for when evaluations were measured with the EPT). When participants had indiscriminate memory of the instructions, we observed a ME instruction effect on implicit and explicit measures of evaluation in Experiment 4, but not in any of the other experiments. In contrast, participants who had reversed memory of the instructions sometimes preferred the word that was instructed to occur infrequently (only when evaluations were registered with explicit measures). This contrast effect has been observed also for experience-based ME effects (Stafford & Grimes, 2012) and effects of approach-avoidance instructions (Van Dessel, De Houwer, & Gast, 2016). Importantly, however, Bayesian factors indicated that the evidence for ME instruction effects for participants with reversed and indiscriminate memory was only anecdotal, whereas strong evidence was obtained for ME instruction effects for participants with accurate memory for the instructions.

General Discussion

In seven experiments, we investigated whether ME instructions influence liking. Participants received instructions that one non-existing word would occur often in a subsequent phase and another non-existing word would occur rarely. After these instructions, but before participants actually experienced these regularities, evaluations of the two words

[3] Note that in the current studies the explicit evaluation task consisted of only a single item (which contrasts with the multi-trial procedure of implicit measures). This could contribute to the fact that we observed stronger effects of ME instructions on implicit than on explicit evaluations.

were assessed. Participants exhibited a preference for the word that was instructed to occur frequently on implicit measures of evaluation (IAT, pIAT, AMP, but not EPT) and on explicit measures of evaluation (explicit liking ratings). Robust ME instruction effects were observed only when participants were able to indicate correctly which word would occur more frequently.

Similar to the original demonstrations of instruction-based approach-avoidance and EC effects (De Houwer, 2006; Van Dessel et al., 2015), the current demonstration of a ME instruction effect is bound to raise a number of questions that will need to be addressed in future research. However, we believe that the current results already have a number of important theoretical and practical implications. In the remainder of this section, we discuss how the current results inform us about the boundary conditions of ME instruction effects and the mental processes that underlie ME instruction and ME experience effects.

Boundary conditions of ME instruction effects

The current studies provide evidence that ME instructions can produce robust changes in implicit and explicit stimulus evaluations. Similar to approach-avoidance instructions (see Van Dessel, De Houwer, Gast, et al., 2016), ME instructions might even produce stronger effects on implicit than on explicit measures of evaluation (see Electronic Supplementary Material 1). Importantly, however, ME instruction effects were not observed when evaluations were registered with an EPT. Though many previous studies have found that effects of manipulations on different measures of implicit evaluation do not necessarily converge, it often remains unclear why these dissociations arise (see De Houwer, 2003, for a discussion). In the current experiments, we found a dissociation between ME instruction effects on EPTs and effects on implicit measures that are also latency-based (IAT, pIAT) or priming-based (AMP). One possible reason for this dissociation is that the EPT was unable to capture the small effect sizes of ME instruction effects due to the lower reliability of the observed EPT scores (see also Wittenbrink, 2007). In line with this idea, correlational analyses revealed only small (Experiment 5) or non-significant (Experiments 6) correlations between EPT scores and explicit rating scores. Especially for novel stimuli such as unfamiliar non-existing words, stronger correlations between implicit and explicit evaluations are typically observed (Nosek, 2005).

Another possibility is that EPT procedures specifically hamper the observation of ME instruction effects. In line with this hypothesis, it has been observed that other instruction-based procedures (i.e., evaluative conditioning, approach-avoidance learning) do lead to reliable EPT effects (e.g., Van Dessel et al., 2015). In contrast, a recent study in which we investigated effects of actual stimulus presentations on stimulus evaluations found a ME effect on IAT scores and explicit ratings but not on EPT scores (Van Dessel, Mertens, Smith, & De Houwer, 2017).[4] Thus, whatever the reason for this moderation may be, it might be similar for ME instruction and ME experience effects. One possible reason for why reliable ME (instruction) effects are not observed on EPT performance is that ME (instructions) might increase the salience of the infrequent (instructed) stimulus compared to the frequent (instructed) stimulus (because presentation of the infrequent stimulus is more unexpected). This could influence EPT performance (e.g., because it draws attention away from evaluative stimulus features; see also De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009) and mask the influence of participants' liking of the stimulus on EPT performance (to a larger extent than IAT and AMP performance). A second possibility is that participants' preference for the (instructed) frequent word is reduced when participants encounter a number of presentations of the frequent and infrequent (instructed) word during the implicit evaluation task. Although this actual exposure might have reduced ME (instruction) effects on all implicit measures, it

[4] This study used the same non-words as stimuli, the same procedures for the implicit and explicit evaluation tasks, and the same online data-collection procedure as the current study. Similar to the ME instruction effects observed here, experience-based ME effects on implicit and explicit evaluations were also small to medium sized (IAT: d_zs [0.21-0.26]; EPT: d_z = 0.09; explicit ratings: d_zs [0.05-0.19]).

could have influenced EPT performance most strongly. For instance, compared to the EPT, the IAT might more easily pick up an initial preference because participants are required to link the stimuli with a certain valence at the beginning of the task (see Ebert, Steffens, von Stülpnagel, & Jelenec, 2009 for evidence that this can strengthen existing preferences). Note, however, that the explicit measures revealed a ME instruction effect despite the fact that the explicit measures always followed the implicit measures (which involved actual stimulus presentations). This indicates that performance of the implicit evaluation tasks does not totally cancel out the effects of ME instructions.

A third possible reason why we did not observe effects of ME instructions (and ME experience) on the EPT is that these learning procedures simply do not produce genuine changes in implicit evaluation. In this case, the observed effects on IAT scores might result from systematic construct-unrelated variance that is due to participants' recoding of IAT categories (Rothermund et al., 2009). More specifically, in the IAT, participants might categorize stimuli on the basis of features other than valence that are shared between two categories of stimuli. This recoding might simplify performance in blocks where these stimuli have to be categorized with the same key, which can lead to IAT effects (Rothermund & Wentura, 2004; Kinoshita & Peek-O'Leary, 2005). In our studies, recoding might occur in blocks where the instructed frequent word and positive stimuli are categorized with the same key. For instance, participants might categorize all stimuli on the basis of frequency because (1) positive words are typically contacted more frequently than negative words in daily life and (2) the instructed frequent word is said to occur more frequently than the instructed infrequent word. Recoding might also occur on the basis of stimulus salience because (1) negative stimuli are typically more salient than positive stimuli and (2) the instructed infrequent word might be more salient than the instructed frequent word because its frequent presentation in the IAT contrasts with the ME instructions. Such strategic recoding might explain the ME instruction effects on the IAT. Because of the task's structure, priming-based procedures such as the EPT (but also the AMP) are less susceptible to confounding recoding processes (Wentura & Degner, 2010; Gawronski & Ye, 2014). Hence, if the effects on the IAT reflect recoding rather than genuine changes in implicit liking, then effects should be absent on the EPT.

In the case of the AMP, ME instruction effects might reflect other systematic construct-unrelated variance such as variance that is due to strategic responding. It has been well established that AMP scores can be controlled intentionally (Teige-Mocigemba et al., 2016; Krieglmeyer & Sherman, 2012). In our study, participants might have controlled their evaluative responses in the AMP by emitting more positive evaluations of Chinese ideographs on trials with the frequent word as prime. It has been argued that EPT performance reflects more automatic processes than AMP performance (e.g., because categorization responses are emitted to clearly valenced stimuli rather than to ambiguous Chinese ideographs). Note however, that there is little evidence to support this claim (see De Houwer et al., 2009, for a review). In fact, EPT effects are also known to be susceptible to voluntary control (see Teige-Mocigemba & Klauer, 2013). Hence, it is always difficult to be absolutely sure that changes observed on implicit measures reflect genuine liking rather than strategic responding or other construct-unrelated variance, irrespective of the implicit measurement procedure that is used.

Nonetheless, there is some evidence that supports the idea that ME instructions did produce genuine changes in implicit evaluation. First, we found that changes in implicit evaluations were not fully mediated by changes in explicit evaluations (in fact, evidence was more in line with a reversed mediation model). If controlled responding fully accounts for changes in implicit evaluation then we should observe full mediation by changes in explicit evaluation because explicit evaluation measures are more vulnerable to intentional control (see also Van Dessel, De Houwer, Gast, et al., 2016). As a counterargument, one might point out that the explicit evaluation measure had less statistical power to capture these (controlled) effects because it only consisted of a single item (which contrasts with the multi-trial procedure of implicit measures). However, previous studies that used multiple items to measure explicit evaluations found high correlations between explicit evaluations measured with this item and other items (e.g., Van Dessel et al., 2015). Moreover, prior research showed that instruction-based manipulations even produced stronger effects on explicit evaluations that only incorporated this item compared to explicit evaluations that incorporated multiple items. Second, to examine the contribution of genuine implicit evaluation in the observed IAT effects, we applied the ReAL model to the IAT data of Experiment 1. This is a multinomial processing tree model that provides separate estimates for the influence of (automatic) evaluative processes and recoding processes in IAT performance (Meissner & Rothermund, 2013). Results provided evidence that evaluative processes significantly contributed to the IAT effect after controlling for recoding processes (see Electronic Supplementary Material 3).

In sum, although it is always difficult to be absolutely sure that changes observed on implicit measures reflect genuine changes in liking, we believe that our results provide sufficient grounds to support the conclusion that ME instructions influence implicit evaluations. To the extent that doubts remain about this conclusion, one should also doubt whether ME experience procedures produce genuine changes in (implicit) liking, given that both procedures led to similar changes in the various measures that we used. Future studies are required to test why ME instruction and ME experience effects are not (easily) registered with an EPT.

Another important observation is that ME instructions produced robust effects only for participants with accurate memory. This suggests that participants need to be consciously aware of the instructed information (enabling them to correctly report the information) in order to exhibit the changes in liking. We note, however, that the current study investigated whether ME instruction effects depend on participants' memory of which of two stimuli would occur most often (as measured with a single question that followed the evaluation tasks). It is possible that a third variable determines both ME instruction effects and participants' answer to this question such that more robust ME instruction effects are observed when participants have accurate memory. For instance, participants with accurate memory may be more attentive or engaged in the experiment and this could moderate both ME instruction effects and memory test performance.

Implications for mental process theories of ME and ME instruction effects

It is important to note that on a functional level (i.e., defining effects in terms of observable regularities in the environment) there is a clear distinction between standard ME effects and ME instruction effects. A ME effect is typically defined as a change in liking that is due to the frequent exposure to a certain stimulus. In our paradigm, there was no manipulation of the exposure frequency of stimuli. A preference for the instructed frequent word over the instructed infrequent word emerged even though the actual number of exposures to these words was identical. Hence, from this perspective, the results reported in this paper do not qualify as ME effects. Nonetheless, we do believe that ME instruction effects are relevant not only in their own right but also for the study of ME effects. ME (experience) research focuses on the impact of one type of events: the frequency of exposures. Because ME instructions provide verbal information about exactly those types of events, they can help elucidate how exposure frequency influences liking (see De Houwer & Hughes, 2016, for a similar argumentation in the context of EC instruction effects and EC). For instance, the finding that ME instructions produce changes in stimulus evaluations fits well with the idea that ME effects are mediated by conscious propositional beliefs about the frequency of occurrence of the stimuli (e.g., Stafford & Grimes, 2012; Wang & Chang, 2004). Based on this idea, one would indeed expect that liking changes irrespective of whether beliefs about frequency are formed on the basis of actual stimulus exposures or on the basis of instructions. Of course, neither our findings, nor previous evidence in support of belief-based models of ME, allow for definite conclusions regarding the mental mechanisms that mediate ME. For instance, it is also possible that multiple processes produce ME effects, some of which do not require the acquisition of propositional knowledge (e.g., ME effects with subliminal stimulus presentations, Monahan, Murphy, & Zajonc, 2000). Importantly, research on ME instruction effects can help shed light on these issues, for instance, by examining whether actual stimulus exposures add to the effects of mere instructions about stimulus exposures (see Raes, De Houwer, De Schryver, Brass, & Kalish, 2014, and Kurdi & Banaji, 2017, for a similar approach in the context of fear conditioning and EC).

In the remainder of this section we discuss possible explanations for ME instruction effects. A first process that might contribute to ME instruction effects is demand compliance (Sawyer, 1975). That is, participants might infer that the experimenter wants them to indicate a preference for the instructed frequent word and strategically adjust their responses according to this inference. We see three reasons why this explanation is unlikely to provide a full account of ME instruction effects. First, unlike what would be expected on the basis of a demand compliance account, changes in implicit evaluations were not fully mediated by changes in explicit evaluations (see Van Dessel, De Houwer, Gast, et al., 2016, for similar evidence in the context of approach-avoidance instructions). Second, as we described above, additional analyses suggest that ME instruction effects involved genuine changes in implicit liking (e.g., fitting the ReAL model on IAT data supported the contribution of automatic evaluative processes). Third, in a recent follow-up study that examined effects of ME instructions on implicit and explicit evaluations (Hughes, Van Dessel, & Smith, 2017), we included a question that asked participants on what basis they had emitted their explicit liking ratings. Only a small subset of participants (<5%) indicated that they had based their evaluations on what they expected that the experimenter wanted them to do. ME instruction effects were observed on implicit and explicit evaluations even when the data of these participants were excluded from the analyses. Of course, it is still possible that demand compliance contributed to ME instruction effects and participants just did not accurately report their demand compliance (e.g., because they believed that not reporting demand compliance would please the experimenter).

Another possible explanation for the ME instruction effect is that, when participants acquire knowledge about the frequency of a stimulus, they may elaborate on this information and infer that frequent stimuli are positive. People might make this inference because frequently occurring stimuli are typically safe and harmless (Zajonc, 2001).[5] Once this proposition is formed, this may influence both explicit and implicit stimulus evaluation (see De Houwer, 2014). This propositional explanation of ME (instruction) effects accords with recent propositional explanations of EC (De Houwer, 2009b) and approach-avoidance effects (Van Dessel, De Houwer, & Gast, 2016).

A third option is that ME instruction effects (partly) result from processing fluency, or the ease with which a stimulus can be processed. Processing fluency is often assumed to play an important role in ME experience effects (e.g., Fang, Singh, & Ahluwalia, 2007; Harmon-Jones & Allen, 2001). From this perspective, repeated exposure to a stimulus can result in facilitated processing of this stimulus when re-encountering it (Jacoby & Dallas, 1981). This may generate a feeling of fluency or ease, which (1) may evoke a positive feeling that can transfer to the stimulus (Reber, Winkielman, & Schwarz, 1998) or (2) can be misattributed to liking or to any other stimulus property that a participant is asked to rate (Bornstein &

[5] Note that people do not necessarily make such an inference in a controlled manner. Inferences often occur under certain conditions of automaticity (e.g., when the application of the underlying rule is well-practiced, see Kruglanski & Gigerenzer, 2011). Frequency of occurrence might be used as a heuristic such that, once people acquire information about the frequency of a stimulus, they spontaneously use this information for determining stimulus valence.

D'Agostino, 1992). Evidence for these ideas is provided by studies showing that other manipulations that induce a feeling of fluency (e.g., long exposure duration or high stimulus-background contrast) also produce positive stimulus evaluations (Reber et al., 1998; Reber, Schwarz, & Winkielman, 2004). One could argue that ME instructions can also influence processing fluency. For instance, learning that a new stimulus will occur frequently might facilitate processing of this stimulus through mental practice or rehearsal of the frequent stimulus (see Glenberg & Kaschak, 2002 for evidence that people engage in mental simulation of events that are described in instructions). This might result in more fluent processing of the instructed frequent word compared to the instructed infrequent word which could produce the changes in liking.

Another possible explanation is that ME instruction effects are actually specific instances of EC. EC refers to a change in liking that is due to the pairing of stimuli (De Houwer, 2007). One could argue that within the ME instructions, there is a (one-trial) pairing of a neutral stimulus (i.e., "Fevkani" or "Lokanta") and the word "often" or "rarely". The latter words might qualify as valenced stimuli, as indicated by a pilot study showing that participants exhibit more positive implicit evaluations of words related to the concept "frequent" than of words related to the concept "infrequent".[6] Hence, it is at least possible that the observed changes in stimulus evaluation resulted from the stimulus pairings within the ME instructions (e.g., Field, 2006). Nevertheless, we believe it is overly simplistic to assume that instructions have an effect because of their mere spatio-temporal properties. Instructions are structured symbolic messages that convey information about the properties of stimuli (see De Houwer & Hughes, 2016, for a discussion in the context of EC). It seems

[6] In a short study with 14 participants, we observed that participants had faster reaction times in IAT blocks in which they categorized frequent words (i.e. the Dutch words for OFTEN, FREQUENT, MANY, PERSISTENT) and positive words with the same response key and infrequent words (i.e. the Dutch words for SOMETIMES, RARELY, SCARCE, SPORADIC) and negative words with another response key than in IAT blocks with the response key assignment reversed, $t(13) = 7.75$, $p < .001$, $d_z = 2.07$.

highly probable that an instruction such as “Fevkani will be presented often” has an effect on liking not simply because words (e.g., “Fevkani” and “often”) co-occur in the same sentence but because the instruction specifies that these words are related in a specific manner (i.e., Fevkani will be presented often). In line with this idea, a recent study showed that effects of (approach-avoidance) instructions on implicit evaluation were moderated by the relational information specified in the instructions (Van Dessel, De Houwer, & Smith, 2017). Future research could address to what extent relational information also moderates ME instruction effects.

It is clear from this (non-exhaustive) overview that there are a number of possible mental process explanations for ME instruction effects. Whatever the mediating mental mechanisms might turn out to be, the observation of a ME instruction effect raises many novel questions, including fundamental questions about the unique contribution that actual stimulus presentations have on liking. In fact, the mere possibility that ME experience effects might in part depend on similar mechanisms as ME instruction effects might stimulate new research that is bound to increase our understanding of this important phenomenon. We believe that research on the effects of ME instructions can provide a new way to investigate the mechanisms underlying ME experience effects while also allowing us to gain a better understanding of instruction effects on (implicit) evaluations and evaluative learning in general (see Van Dessel, De Houwer, & Smith, 2017, for an example in the context of approach-avoidance instruction effects). For instance, testing whether ME instructions can produce other effects that are influenced by processing fluency (e.g., enhance attention to the infrequent stimulus, see Alter, Oppenheimer, Epley, & Eyre, 2007) or whether ME instruction effects depend on an EC mechanism might expand our knowledge on the mental processes underlying ME effects as well as instruction-based effects.

Concluding remarks

This set of studies provides strong evidence that ME instructions can influence both implicit and explicit evaluations and adds to recent work showing that evaluative learning effects which were traditionally assumed to rely entirely on automatic processes can also be obtained via mere instructions (e.g., EC: De Houwer, 2006; approach-avoidance training effects: Van Dessel et al., 2015). There are of course many questions that still need to be addressed, most prominently questions about the precise mechanism on which instructed ME effects rely and to what extent these mechanisms might also underlie experience-based ME effects. Nonetheless, the current studies already contribute substantially to the literature on ME and evaluative learning. Not only do they provide the first evidence ever for ME instruction effects, they also reveal both the generality (e.g., effects across various measures of liking) and boundary conditions of the effect (e.g., the absence of effects on the EPT and the impact of memory). Finally, the mere fact that the data raise many new questions is perhaps the most important contribution of our work simply because those questions are likely to stimulate new research. We therefore hope that the present studies will provide the basis for many important future discoveries.

Electronic Supplementary Material

ESM 1. Additional Analyses (ElectronicSupplementaryMaterial1.docx). This document reports *t*-test analyses and mediational analyses for ME instruction effects on implicit and explicit evaluation scores and on mean reaction times in implicit evaluation tasks.

ESM 2. Raw Data (RawData.zip). This zip-file contains raw data files of all the experiments reported in this manuscript.

ESM 3. ReAL Model Analyses (ElectronicSupplementaryMaterial3.docx). This document reports analyses with the ReAL Model for IAT scores in Experiment 1.

References

Allport, G. (1935). Attitudes. In Murchison, C. (Ed.). *A Handbook of Social Psychology* (pp. 789-843). Worcester, MA: Clark University Press.

Alter, A., Oppenheimer, D.M., Epley, N., & Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytical thought. *Journal of Experimental Psychology: General, 136*, 569–576. doi: 10.1037/0096-3445.136.4.569

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality and Social Psychology Bulletin, 38*, 1194-1208. doi: 10.1177/0146167212446835

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package v. 1.0-6. http://CRAN.R-project.org/package=lme4

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987. *Psychological Bulletin, 106*, 265–289. doi:10.1037/0033-2909.106.2.265

Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology*, *63*, 545–552. doi:10.1037/0022-3514.63.4.54

Brooks, J. O., & Watkins, M. J. (1989). Recognition memory and the mere exposure effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *15*, 968–976. doi:10.1037/0278-7393.15.5.968

De Houwer, J. (2003). A structural analysis of indirect measures of attitudes. In J. Musch & K.C. Klauer (Eds.), *The Psychology of Evaluation: Affective Processes in Cognition and Emotion* (pp. 219-244). Mahwah, NJ: Lawrence Erlbaum.

De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation*, *37*, 176–187. doi:10.1016/j.lmot.2005.12.002

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. The Spanish Journal of Psychology, 10, 230-241. doi:10.1017/S1138741600006491

De Houwer, J. (2009a). How do people evaluate objects? A brief review. *Social and Personality Psychology Compass, 3,* 36-48. doi:10.1111/j.1751-9004.2008.00162.x

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

De Houwer, J. (2014). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass, 8*, 342-353. doi:10.1111/spc3.12111

De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology, 43*, 972-978. doi:10.1016/j.jesp.2006.10.007

De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition, 34*, 480–494. doi:10.1521/soco.2016.34.5.480

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*, 347-368. doi:10.1037/a0014211

de Zilva, D., Vu, L., Newell, B. R., & Pearson, J. (2013). Exposure Is Not Enough: Suppressing Stimuli from Awareness Can Abolish the Mere Exposure Effect. *PLoS ONE*, *8*(10), e77726. doi:10.1371/journal.pone.0077726

Ebert, I. D., Steffens, M. C., von Stülpnagel, R., Jelenec, P. (2009). How to like yourself better, or chocolate less: Changing implicit attitudes with one IAT task. *Journal of Experimental Social Psychology, 45*, 1098–1104. doi:10.1016/j.jesp.2009.06.008

Fang, X., Singh, S., & Ahluwalia, R. (2007). An Examination of Different Explanations for the Mere Exposure Effect. *Journal of Consumer Research*, *34*(1), 97–103. doi:10.1086/513050

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238. doi: 10.1037/0022-3514.50.2.229.

Fechner, G. T. (1876). Vorschule der Aesthetik. Leipzig, Germany: Breitkopf and Härtel

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. Basic and Applied Social Psychology, 27, 307–316.

Field, A. P. (2006). Is conditioning a useful framework for understanding the development and treatment of phobias? *Clinical Psychology Review*, *26*, 857-875. doi: 10.1016/j.cpr.2005.05.010

Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *Quarterly Journal of Experimental Psychology*, *65*, 1657-1674. doi:10.1080/17470218.2012.665061

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731. doi:10.1037/0033-2909.132.5.692

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd edition, pp. 283–310). New York: Cambridge University Press.

Gawronski, B., & Ye, Y. (2014). What drives priming effects in the affect misattribution procedure? *Personality and Social Psychology Bulletin*, *40*, 3-15. doi:10.1177/0146167213502548

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic*

Bulletin & Review, 9, 558–565. doi:10.3758/BF03196313

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*, 1464–80. 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

Harmon-Jones, E., & Allen, J. J. B. (2001). The Role of Affect in the Mere Exposure Effect: evidence from Psycho-physiological and Individual Differences Approach. *Personality and Social Psychology Bulletin*, *27*, 889–898. doi:10.1177/0146167201277011

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

Hughes, S., Van Dessel, P., & Smith, C. T. (2017). A comparative analysis of different evaluative learning procedures. Manuscript in preparation.

Jacoby, L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306-340. doi:10.1037/0096-3445.110.3.306.

Jeffreys, H. (1961). Theory of Probability. Oxford: Oxford University Press.

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

Kinoshita, S., & Peek-O'Leary, M. (2006). Two bases of the compatibility effect in the Implicit Association Test (IAT). *The Quarterly Journal of Experimental Psychology,* *59*, 2102–2120. doi:10.1080/17470210500451141

Krieglmeyer, R., & Sherman, J.W. (2012). Disentangling stereotype activation and stereotype application in the Stereotype Misperception Task. *Journal of Personality and Social Psychology, 103,* 205-224.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberative judgements are based on common principles. *Psychological Review, 118,* 97–109.

Kunst-Wilson, W., & Zajonc, R. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, *207*(4430), 557–558. doi:10.1126/science.7352271

Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes? *Journal of Experimental Psychology: General*, *146*, 194–213. doi: 10.1037/xge0000239

Maslow, A. H. (1937). The influence of familiarization on preference. *Journal of Experimental Psychology, 21*, 162-180. doi:10.1037/h0053692

Meissner, F. & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology, 104*,45-69. doi:10.1037/a0030734

Monahan, J. L., Murphy, S. T., & Zajonc, R. B. (2000). Subliminal Mere Exposure: Specific, General, and Diffuse Effects. *Psychological Science, 11*, 462–466. doi:10.1111/1467-9280.00289

Newell, B. R., & Shanks, D. R. (2007). Recognising what you like: Examining the relation between the mere-exposure effect and recognition. *European Journal of Cognitive Psychology*, *19*, 103–118. doi:10.1080/09541440500487454

Nosek, B.A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565–584. doi:10.1037/0096-3445.134.4.565

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653-667. doi:10.1037/0022-3514.86.5.653

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277-293. doi:10.1037/0022-3514.89.3.277

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1-15). New York, NY: Guilford.

Raes, A. K., De Houwer, J., De Schryver, M., Brass, M., & Kalisch, R. (2014). Do CS-US pairings actually matter? A within-subject comparison of instructed fear conditioning with and without actual CS-US pairings. *PLoS ONE, 9(1):* e84888. doi:10.1371/journal.pone.0084888

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*, 364–82. doi:10.1207/s15327957pspr0804_3

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science*, *9,* 45–48. doi:10.1111/1467-9280.00008

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Minimizing the influence of recoding in the Implicit Association Test: The Recoding-Free Implicit Association Test (IAT-RF). *The Quarterly Journal of Experimental Psychology, 62,* 84–98. doi:10.1080/17470210701822975

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139–165. doi:10.1037/00963445.133.2.139

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

Sawyer, A. G. (1975). Detecting Demand Characteristics in Laboratory Experiments in Consumer Research: the Case of Repetition-Affect Research. In: Schlinger, M. J. (Ed.). *Advances in Consumer Research*, 2, Ann Abor, MI : Association for Consumer Research, 713-724.

Smith, C. T., De Houwer, J., & Nosek, B. (2013). Consider the source: persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin*, 39, 193-205. doi:10.1177/0146167212472374

Spruyt, A., De Houwer, J., Hermans, D., & Eelen, P. (2007). Affective priming of non-affective semantic categorization responses. *Experimental Psychology, 54*, 44–53. doi:10.1027/1618-3169.54.1.44

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

Stafford, T., & Grimes, A. (2012). Memory Enhances the Mere Exposure Effect. *Psychology & Marketing*, *29*(12), 995–1003. doi:10.1002/mar.20581

Storbeck, J., & Robinson, M.D. (2004). When preferences need inferences: A direct comparison of the automaticity of cognitive versus affective priming. *Personality and Social Psychology Bulletin*, *30*, 81-93.

Teige-Mocigemba, S. & Klauer, K.C. (2013). On the controllability of evaluative-priming effects: Some limits that are none. *Cognition & Emotion, 27,* 632-657.

Teige-Mocigemba, S., Penzl, B., Becker, M., Henn, L., & Klauer, K. C. (2016). Controlling the "Uncontrollable": Faking effects on the Affect Misattribution Procedure. *Cognition & Emotion*, *30*, 1470-1484.

Topolinski, S. (2012). The sensorimotor contributions to implicit memory, familiarity, and recollection. *Journal of Experimental Psychology: General*, *141*(2), 260–281. doi:10.1037/a002565

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 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. Manuscript under review.

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., Mertens, G., Smith, C.T., & De Houwer, J. (2017). Mere Exposure Effects on Implicit Evaluation: The moderating role of Evaluation Task, Number of Stimulus Presentations and Memory of Stimulus Presentations. Manuscript in preparation

Wang, M.-Y., & Chang, H.-C. (2004). The mere exposure effect and recognition memory. *Cognition & Emotion, 18,* 1055–1078. doi:10.1080/02699930341000374

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95-116). New York, NY: Guilford Press.

Wittenbrink, B. (2007). Measuring attitudes through priming. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 17-58). New York: Guilford Press.

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

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monographs, 9,* 1-27. doi: 10.1037/h0025848

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science*, *10*, 224–228. doi:10.1111/1467-8721.0015428

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

Zhou, H., & Fishbach, A. (2016). The Pitfall of Experimenting on the Web: How Unattended

Selective Attrition Leads to Surprising (Yet False) Research Conclusions. *Journal of Personality and Social Psychology, 11*, 493-504. doi: 10.1037/ pspa0000056

Tables

Table 1. *Main procedural differences between the four implicit evaluation measures used in Experiments 1-7.*

Implicit evaluation measure	Target of categori-zation	Type of categorization	Basis for computation of the evaluation score
IAT: Experiments 1,4	Valenced words + evaluation stimuli	Categorization in 2 categories: positive/negative and Fevkani/Lokanta	Latencies
pIAT: Experiment 2	Valenced words + evaluation stimuli	Categorization in 2 categories: I like/ I dislike and Fevkani/Lokanta	Latencies
EPT: Experiments 3,5,6	Valenced words	Categorization in 1 category: positive/negative	Latencies
AMP: Experiment 7	Chinese ideographs	Categorization in 1 category: positive/negative	The number of positive and negative responses

Table 2. *Mean evaluation scores indicating a preference for the frequent instructed words and ME instruction effects for participants with accurate memory of the instructions in Experiments 1-7.*

Experiment	N	Mean (SD)	Test statistic	Effect size d	Bayesian *t*-test
Experiment 1: IAT score	28	0.23 (0.42)	$t(27) = 2.84$, $p = .008$	0.54	$BF_1 = 5.32$, Substantial *(H_1)*
Experiment 2: pIAT score	69	0.16 (0.48)	$t(68) = 2.72$, $p = .008$	0.33	$BF_1 = 3.91$, Substantial *(H_1)*
Experiment 3: EPT score	35	-3.34 (46.15)	$t(34) = -0.43$, $p = .67$	-0.07	$BF_0 = 5.06$, Substantial *(H_0)*
Experiment 4: IAT score	459	0.12 (0.48)	$t(458) = 5.42$, $p < .001$	0.25	$BF_1 = 73208.54$, Strong *(H_1)*
Experiment 5: EPT score	549	4.43 (57.49)	$t(548) = 0.58$, $p = .56$	0.02	$BF_0 = 17.60$, Strong *(H_0)*
Experiment 6: EPT score	731	1.98 (103.59)	$t(730) = 0.52$, $p = .61$	0.02	$BF_0 = 21.04$, Strong *(H_0)*
Experiment 7: AMP score	404	0.04 (0.20)	$t(403) = 4.01$, $p < .001$	0.20	$BF_1 = 140.61$, Strong *(H_1)*
Experiments 4-7: Explicit score	2149	0.10 (0.98)	$t(2148) = 4.91$, $p < .001$	0.11	$BF_1 = 3919.53$, Strong *(H_1)*