Running head: Automatic Evaluation despite Faking

Strategic Modification of the Evaluative Priming Effect Does Not Reduce Its Sensitivity to Uncontrolled Evaluations

Yoav Bar-Anan
Ben-Gurion University of the Negev, Beer-Sheva, Israel

Contact details: Yoav Bar-Anan, baranany@bgu.ac.il, phone: +972-8-6472030

Author's Note: Correspondence concerning this article should be sent to Yoav Bar-Anan, Department of Psychology, Ben-Gurion University of the Negev, Beer Sheva, Israel, 84105. Electronic mail may be sent to baranany@bgu.ac.il

Abstract

In the evaluative priming procedure the processing of a target stimulus is facilitated when preceded by a prime of the same valence. This procedure is used to investigate and measure the unintentional and uncontrolled influence of attitudes. Consistent with previous findings, in this research, when participants knew that primes are more likely to precede targets of opposite valence the typical priming effect was reversed. This may suggest that non-evaluative processes can eliminate the effect of unintentional evaluation. However, in five studies, success in reversing the priming effect was still related to people's evaluation of the primes. This suggests that unintentional evaluation affects performance in the evaluative priming procedure even when people successfully counteract the priming effect. Although behaviors that are sensitive to evaluative processes may be altered by rival processes, the rival processes do not necessarily decrease the absolute influence of the evaluative processes on those behaviors.

Key-words: Evaluative priming, Affective priming, Implicit measures, Faking, Attentional control, Automaticity, Strategic effects, Automatic evaluations

Strategic Modification of the Evaluative Priming Effect Does Not Reduce Its Sensitivity to Uncontrolled Evaluations

In the evaluative priming (EP) procedure, the primary task of the participants requires the processing of target evaluative stimuli (e.g., classify word adjectives as pleasant or unpleasant). Each target is preceded by a prime stimulus (e.g., a smiling face) unrelated to the primary task. Numerous studies using this paradigm found an EP effect: people were faster and more accurate to process good words after positive primes, and bad words after negative primes, than to process good words after negative primes and bad words after positive primes (Fazio, 2001; Klauer & Musch, 2003). The EP effect is considered an unintentional and uncontrolled effect because it happens very quickly (the prime precedes the target by less than 300 ms; Hermans, De Houwer & Eelen, 2001), and because the effect sometimes reflects an evaluation that participants are motivated to hide (Fazio et al., 1995). Because of that, the EP effect is of the main sources of evidence that evaluation can influence behavior with no need for conscious decision to evaluate, and it is a main tool for measurement and investigation of unintentional evaluation (Bargh, 1994; Bargh et al., 1996; Duckworth et al., 2002; Fazio et al., 1986; Fazio, 1986, 2007).

However, recent studies found that instructions can decrease (or increase) the priming effect (Degner, 2008; Klauer & Teige-Mocigemba, 2007; Teige-Mocigemba & Klauer, 2008). For instance, German participants in a study conducted by Teige-Mocigemba and Klauer (TMK; 2008, Study 1) completed an EP procedure in which some prime-target pairs (Arab primes before positive targets and celebrities primes before negative targets) were presented more often than the other pairs (*Arab-bad*, *celebrity-good*). Participants who were not informed about this imbalance showed the expected EP effect: faster responses in *celebrity-good* and *Arab-bad* trials than in *celebrity-bad* and *Arab-good* trials. Participants who were informed about the specific frequency—and therefore expected good words after Arab primes and bad words after celebrity primes—did not show the EP effect.

TMK's findings suggest that participants' knowledge about imbalanced prime-target frequencies can alter the priming effect. One account for this effect is that the knowledge eliminated the automatic effect of the primes' evaluation. This would entail that it is possible to directly turn off the evaluation effect. However, the alteration of the overall priming effect

does not indicate that the *evaluative* priming effect was altered. Another possibility is that the knowledge about the imbalanced frequencies influenced the priming effect in one direction, while the evaluation still influenced the priming effect in the opposite direction. Put differently, perhaps people can decrease the relative influence of evaluations on the priming effect by activating processes that also influence the priming effect, but they cannot decrease the absolute influence of the evaluations on the priming. In that case, the sensitivity of the priming effect to variations in the evaluations of the primes should remain – only the overall priming effect would shift. The present research investigated that possibility.

Five studies tested whether the priming effect in a *stated imbalanced EP* (an imbalanced EP when participants are informed about the frequencies) was related to evaluations of the primes measured by other measures. The other measures were evaluative priming, self-report, and the Implicit Association Test (IAT, Greenwald, McGhee & Schwartz, 1998). If the priming effect would be related to people's evaluations of the primes, then it will suggest that this form of control on the priming effect does not eliminate the sensitivity of the priming effect to evaluations.

Overview of the Studies

In all studies, participants completed a few measures of their racial (Studies 1, 3 and 5) or political (Studies 2 and 4) attitudes. Studies 1-2 started with an *unstated imbalanced EP* (i.e., participants were not informed about the imbalanced prime-target proportions). Next, participants were informed about the prime-target proportions before completing another imbalanced EP. Studies 3-4 were similar, but half of the participants first completed the standard EP (balanced evenly with equal prime-target proportions) instead of the unstated imbalanced EP. In Study 5, participants completed an EP procedure and an IAT. The EP was either the stated imbalanced EP or a standard EP. In all studies, participants also explicitly reported about their attitudes. The method of Study 1 is described first, followed by the modifications in the rest of the studies, and the rationale for each modification.

Methods

Participants. Volunteers at the Project Implicit research website (https://implicit.harvard.edu; see Nosek, 2005 for more information) were randomly assigned to the study from a large pool of available studies. The details about the number of participants are presented in Table 1. The analyses did not include participants who did not have above-

chance success rate (51%) in all tasks, or did not have at least one trial in each of the conditions of the relevant task (e.g., the four prime-target conditions in EP).

Table 1
Number of participants, demographics, and dropout rates

Group	Started (% women, mean age, SD age)	Completed (% of started)	Removed from analyses (% of completed)
Study 1	281 (65%, 28, 12)	223 (79%)	30 (13%)
Study 2	243 (65%, 27, 12)	194 (80%)	25 (13%)
Study 3: Standard EP	163 (75%, 25, 11)	137 (84%)	22 (16%)
Study 3: Imbalanced EP	166 (76%, 27, 12)	127 (77%)	13 (10%)
Study 4: Standard EP	149 (70%, 29, 14)	119 (80%)	10 (8%)
Study 4: Imbalanced EP	156 (69%, 26, 11)	128 (82%)	16 (13%)
Study 5: Standard EP	156 (65%, 27, 12)	125 (80%)	3 (2%)
Study 5: Imbalanced EP	210 (61%, 27, 12)	152 (72%)	9 (6%)

Notes: (a) In Studies 3-4, in the imbalanced EP condition, the first EP had imbalanced prime-target frequencies (the same as the second EP); whereas in the standard EP, the first EP had equal frequencies. In the imbalanced EP condition in Study 5, participants performed the imbalanced EP and were informed about the frequencies beforehand. (b) Participants were removed from the analyses if their success-rate in one of the EPs was not above chance (less the 51%) or if they did not respond with at least one correct response for each of the four prime-target conditions. (c) The difference in dropout rate between the conditions in each study was never significant.

Procedure and Materials

Stimuli. The attitude-object stimuli were face images of 12 of Black and 12 White men (the young men stimuli from Gawronski et al., in press). The target words in the EP were 14 positive and 14 negative nouns and adjectives.

EP. In each trial, the prime stimulus was presented for 275 ms, followed immediately by a target word which remained on the screen until 800 ms had passed or a response was given by pressing one of two keyboard keys (these durations were used in TMK's procedure). After an incorrect response, a red X appeared for 275 ms. The intertrial interval was 250 ms. Each EP procedure consisted of three 60-trial blocks.

Prime-target pairs that were inconsistent with the common preference in Project Implicit's participant pool (*Black men-good* and *White men-bad*) appeared more often. Each of the two inconsistent pairs appeared 20, 19 and 18 times in blocks 1-3, respectively; and each consistent pair appeared 10, 11, and 12 times in blocks 1-3.

Participants first completed three blocks of this task with the following instructions: "Images and words will appear one after another. Ignore the images and categorize the words

as good or bad." Before completing another three blocks, participants were informed about the imbalanced prime-target frequencies: "When you see an image of a **Black man**, it is more likely that a **positive** word will appear next. When you see an image of a **White man**, it is more likely that a **negative** word will appear next." [Bold in original].

Self-report. A thermometer rating probed feelings toward Black and White men on a scale from 0, *the coldest* to 10, *the warmest*. The explicit attitude was the difference score.

Design. The presentation of the self-report questionnaire (before or after the EPs) was counterbalanced between participants.

Modifications in Study 2. The prime stimuli were American politicians: six Democrats and six Republicans. Because self-reported political attitudes are strongly related to indirectly-measured political attitudes (Nosek, 2005) explicit attitude in this study should be more helpful in detecting evaluative influence in the stated imbalanced EP. Because most participants in the pool identify as Liberals, the more frequent prime-target pairs were *Republican-good* and *Democrats-bad*.

Modifications in Study 3. The primes were 7 Black men and 7 White women. For half of the participants, the first EP was the standard EP with 15 trials for each prime-target pair in each block. The other half started with the unstated imbalanced EP, like in Studies 1-2. The objective was to examine whether the stated imbalanced EP would be related to a standard EP.

Modifications in Study 4. This was a combination of Studies 2 and 3: politicians were the primes, and half of the participants completed a standard EP before the stated imbalanced EP.

Modifications in Study 5. The study compared the relationship between an IAT and the standard EP to the relationship between the IAT and the stated imbalanced EP. The stimuli were the same as in Study 3. The IAT used the same face stimuli and the categories *Black people*, *White people*, *Good* and *Bad*. The IAT was the standard 7-block IAT (Greenwald et. al, 2001), and was scored after removing latencies slower than 10000ms or faster than 400ms, and including error latencies (Greenwald, Nosek & Banaji, 2003). The proportions in the imbalanced EP were 22-8, 20-10, 20-10 in blocks 1-3, respectively. Participants were randomly assigned to one of four conditions in a 2 (EP: standard, unbalanced) X 2 (Measures-order: IAT, self-report, EP or EP, IAT, self-report) design. The IAT was used to add evidence

that the relationship between the different EPs in Studies 1-4 was due to the primes' evaluation and not due to method-related non-evaluative factors.

Results and Discussion

In all attitude measures, positive scores indicated preference for White people or preference for Democrats. Measures-order manipulations did not moderate the results of any of the following tests.

Self-report and IAT measures. The mean scores of the self-report and the IAT are detailed in Table 2. Participants reported preference for White over Black people in Studies 1, 3 and 5, with effect sizes of *ds* = .18, .11, .38, respectively. Participants reported preference for Democrats over Republicans, *ds* = .46, .62, in Studies 2 and 4, respectively. The IAT in Study 5 also indicated preference for White women over Black men, *d* = .77.

Table 2
Self-report and IAT scores

Measure	Mean (SD)	T-test: difference from zero
Study 1: Self-report (White-Black men)	0.32 (1.74)	$t(192) = 2.53$, $p = .01$
Study 2: Self-report (Democrats-Republicans)	1.97 (4.25)	$t(168) = 6.02$, $p < .0001$
Study 3: Self-report (White women – Black Men)	0.21 (1.86)	$t(228) = 1.74$, $p = .08$
Study 4: Self-report (Democrats-Republicans)	2.58 (4.01)	$t(220) = 9.56$, $p < .0001$
Study 5: Self-report (White women – Black Men)	0.86 (2.13)	$t(264) = 6.56$, $p < .0001$
Study 5: IAT (White women – Black Men)	0.30 (0.39)	$t(264) = 12.44$, $p < .0001$

Notes: All self-report measures are the difference scores between two thermometer ratings on a 10-point scale; The IAT score is a D score. In all measures, zero indicates no preference.

EP. Before computing the priming effect score, error responses, responses faster than 300ms, and responses with latency more than 2.5 SDs away from the participant's average latency were discarded. Then, the latencies of conditions compatible with the common preference (*White women-good*; *Black men-bad* in Study 1) were subtracted from the latencies of the incompatible conditions (*White men-bad*; *Black men-good*). The EP Analyses were based on log-transformed response latencies. For clarity, the reported means are non-transformed.

The means of all the priming effects and their significance are detailed in Table 3. The results across all studies are very consistent: The standard EP always produced the expected priming effect (Studies 3-5). The unstated imbalanced EP never produced a significant priming effect (Studies 1-4). Most important, in all five studies, the stated imbalanced EP produced a significant reverse priming effect of more than 50 ms for the race EPs and between 10 to 20 ms

in the politics EPs. As detailed in Table 3, these reverse priming effects were always significantly different from the priming effect in the other EP procedures. The reverse priming effects were still significant even after a very stringent cleaning of outlier observations (see supplementary materials for details). Therefore, TMK's findings were replicated: knowing about the imbalanced frequencies of the prime-target pairs had a strong influence on the priming effect.

Table 3
Priming effects

Group	Mean (SD)	T-test: Difference from zero	T-test: Difference from the stated imbalanced EP
Study 1			
Unstated imbalanced EP	0 (38)	$t(192) = 0.16$, $p = .87$	$t(192) = 11.07$, $p < .0001$
Stated imbalanced EP	-53 (66)	$t(192) = 11.09$, $p < .0001$	
Study 2			
Unstated imbalanced EP	0 (40)	$t(168) = .02$, $p = .99$	$t(168) = 6.23$, $p < .0001$
Stated imbalanced EP	-21 (47)	$t(168) = 5.77$, $p < .0001$	
Study 3			
Standard EP	12 (33)	$t(114) = 3.12$, $p = .0001$	$t(114) = 9.53$, $p < .0001$
Unstated imbalanced EP	-3 (40)	$t(113) = 0.89$, $p = .38$	$t(113) = 7.92$, $p < .0001$
Stated imbalanced EP	-52 (69)	$t(228) = 11.55$, $p < .0001$	
Study 4			
Standard EP	9 (33)	$t(108) = 2.85$, $p = .005$	$t(108) = 3.90$, $p = .0002$
Unstated imbalanced EP	3 (35)	$t(111) = 0.80$, $p = .42$	$t(111) = 2.17$, $p = .03$
Stated imbalanced EP	-10 (48)	$t(221) = 2.09$, $p = .003$	
Study 5			
Standard EP	11 (39)	$t(121) = 3.31$, $p = .001$	$t(263) = 7.99$, $p < .0001$
Stated imbalanced EP	-54 (86)	$t(142) = 7.50$, $p < .0001$	

Note: The means are in millisecond, but the statistical analyses used log transformed values.

Imbalanced prime-target frequencies reduced the priming even when people had no knowledge of these frequencies (consistent with findings reviewed by Klauer & Musch, 2003, pp. 18-19). In Studies 3 and 4, one group of participants completed the unstated imbalanced EP at the beginning of the study, whereas the other group completed the standard EP. The priming effect in the unstated imbalanced EP was significantly smaller than in the standard EP, $t(234, 278) = 3.64, 2.00$, $ps = .0003, .05$, $ds = .43, .21$, in Studies 3 and 4, respectively. This may raise the possibility that the reverse priming effect in the stated imbalanced EP in Studies 1-4 was due to implicit learning during the previous unstated imbalanced EP task, and not conscious knowledge of the imbalanced frequencies. However, there was no evidence for implicit learning across blocks in the unstated imbalanced EP; and, in Studies 3 and 4, the

stated imbalanced EP showed the same strong reverse priming effects even when it was preceded by the standard EP (see supplementary materials for full details).

Relationship with other measures. The main question in this research was whether the priming effect in the stated imbalanced EP would be related to the evaluations of the primes. As detailed in Table 4, across the five studies, the stated imbalanced EP significantly correlated with 5 out of 6 of the other EP procedures, 3 out of 5 of the explicit scores, and with the IAT. Without bivariate outliers (the values in the parentheses in Table 4), 9 out of these 12 correlations were significant, another 2 were marginally significant, and only one (with the unstated imbalanced EP in Study 4) was far from significant. Finally, the correlations of the stated imbalanced EP with explicit attitudes and the IAT were never significantly smaller than the correlations of the standard and the unstated imbalanced EPs with the same self-report and IAT measures (see the supplementary materials for more details).

Table 4
Correlations between the attitude measures

Study 1	Explicit	Unstated Imbalanced EP	
Stated imbalanced EP	.16* (.14#)	.26*** (.27***)	
Unstated Imbalanced EP	.19** (.23**)		
Study 2	Explicit	Unstated Imbalanced EP	
Stated imbalanced EP	.46*** (.44***)	.50*** (.50***)	
Unstated Imbalanced EP	.42*** (.38***)		
Study 3	Explicit	Unstated Imbalanced EP	Standard EP
Stated imbalanced EP	.12# (.13*)	.15 (.17#)	.19* (.18*)
Unstated Imbalanced EP	.25** (.23*)		
Standard EP	.02 (.01)		
Study 4	Explicit	Unstated Imbalanced EP	Standard EP
Stated imbalanced EP	.27*** (.29***)	.09 (.10)	.20* (.27**)
Unstated Imbalanced EP	.31*** (.32**)		
Standard EP	.35*** (.40***)		
Study 5	Explicit	IAT	
Stated imbalanced EP	.09 (.26**)	.17* (.16#)	
Standard EP	.03 (.06)	.29** (.28**)	
IAT	.32*** (.36***)		

Notes. In parentheses: the correlation without bivariate outliers. These were individual scores with a Cook's D value above the threshold of the relevant sample size, or with an absolute studentized residual larger than 2; # $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$;

Conclusions

Consistent with previous findings, knowing about the imbalanced prime-target frequencies altered the priming effect and reversed it. However, even when reversed, the priming effect was still related to people's evaluation of the primes, not less than the priming effect in the standard EP and the unstated imbalanced EP. This suggests that when the priming effect is altered by expecting certain prime-target contingencies, it does not lose its sensitivity to the evaluation of the primes.

Findings about the sensitivity of indirect attitude measures to "faking" manipulations may seem of little significance because researchers do not use these manipulations when they aim to measure attitudes. However, some participants might spontaneously use strategies similar to those induced by these manipulations. Additionally, to comprehend the importance of a test, one must consider the implications of the opposite results: had TMK found that conscious expectancies cannot alter the priming effect this would have been very strong evidence that this effect is highly resistant to non-evaluative processes.

Similar rationale explains the significance of the present research. First, it is true that if some participants spontaneously use strategies that alter the priming effect, the quality of EP as a measurement of evaluations would be impaired because the variations in the decision to use these strategies would add noise. But, such strategies would have impaired the measurement much worse had they been able to eliminate or decrease the influence of evaluations on the priming. The present results suggest that this impairment is not as damaging as it could have been. Second, if conscious expectancy about the prime-target frequencies had reduced the sensitivity of the priming effect to evaluation, this would have suggested that top-down processes may turn the spontaneous influence of evaluations off. Yet, the present results tell a different story about the robustness of spontaneous evaluations: even when knowledge about the prime-target frequencies has a very strong effect on performance, it still does not reduce the sensitivity of the performance to primes' evaluation. This is novel evidence about the stability of spontaneous evaluative processes in face of opposing processes.

The present research also conveys a broader idea: In research about factors that may influence unintentional and uncontrollable processes, examining the effect of these factors on the measure that assesses the automatic process is not the only informative test. It is also informative to examine whether these factors influence the sensitivity of the measure to the

automatic process. If the sensitivity is not moderated by the investigated factors, then this probably suggests that these factors influence the measurement and not the actual automatic process in question.

References

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition (Vol. 1*, pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology, 32,* 104-128.

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science, 13*, 513-519.

Degner, J. (2008). On the (un)controllability of affective priming: Strategic manipulation is feasible but can possibly be prevented. *Cognition and Emotion, 28*, 327-354.

Fazio, R. H. (1986). How do attitudes guide behavior? In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 204-243). New York: Guilford Press.

Fazio, R. H. (2001). On the automatic activation of associated evaluations : An overview. *Cognition and Emotion, 15*, 115-141.

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

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.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.

Greenwald, A. G.,McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

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.

Gawronski, B., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (in press). Attentional influences on affective priming: Does categorization influence spontaneous evaluations of multiply categorizable objects? *Cognition and Emotion*.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition and Emotion, 15*, 143–165

Klauer, K. C. , & Musch, J. (2003). Affective priming: Findings and theories . In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7-49). Mahwah, NJ: Lawrence Erlbaum.

Klauer, K. C., & Teige-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology, 43*, 648-655.

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

Teige-Mocigemba, S., & Klauer, K. C. (2008). "Automatic" evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology, 44*, 1414-1417.

Supplementary Online Materials for "Intentional Modification of the Evaluative Priming Effect Does Not Reduce Its Sensitivity to Uncontrolled Evaluations"

Reverse Priming Effects after a Stringent Cleaning of Outliers

As described in the main text, in all five studies the information about the imbalanced prime-target frequencies caused a reverse priming effect, such that people were faster to respond to the prime-target pairs that were inconsistent with the common racial or political attitudes in Project Implicit's participant pool. Because these reverse priming effects were stronger than the common priming effect, additional analyses examined these effects without outlier scores. The outliers were removed recursively: participants with scores 2 SDs away from the mean score were removed. Then, the mean score was computed without those participants, and again participants with scores 2 SDs away from the new mean score were removed. This method was repeated until none of the scores of the remaining participants were 2 SDs away from the mean of those participants. This removed about 20%-40% of the participants in the five different studies. After these outlier observations were removed, the reverse priming effects were reduced by about 10-20ms. However, as detailed in Table 1S, even with this extremely stringent removal of outliers, all the effects were still significantly smaller than zero, and all were still significantly smaller than the priming effects in the other EPs. In conclusion, the reverse priming effects were robust and not the result of outlier observations.

Table S1

Evaluative Priming Effect in the disclosed imbalanced EP (stringent outliers cleaning)

Group	Mean (SD)	T-test: Difference from zero	T-test: Difference from the unstated imbalanced EP
Study 1	-38 (46)	$t(175) = 11.13$, $p < .0001$	$t(175) = 10.56$, $p < .0001$
Study 2	-10 (23)	$t(129) = 5.11$, $p < .0001$	$t(129) = 4.21$, $p < .0001$
Study 3	-30 (38)	$t(194) = 10.92$, $p < .0001$	$t(101) = 8.14$, $p < .0001$
Study 4	-4 (26)	$t(189) = 2.05$, $p = .04$	$t(115) = 2.07$, $p = .04$
			T-test: Difference from the standard EP
Study 5	-40 (47)	$t(125) = 9.85$, $p < .0001$	$t(227) = 10.15$, $p < .0001$

Implicit Learning of the Imbalanced Frequencies

In Studies 1 and 2, participants first performed three blocks of EP with imbalanced prime-target frequencies without receiving the explicit information that the prime-target

frequencies are imbalanced. Only after these three blocks, they were informed about the imbalanced frequencies and then completed another three blocks of the same procedure. In Studies 3 and 4, one group of participants experienced the same sequence of events: they completed six blocks of the imbalanced EP but were informed about the imbalance only after the third block. The other group of participants started with three blocks of a balanced EP. Like in the other conditions, after the third block they were informed that in the next blocks the prime-target frequencies will be imbalanced, and they then completed the three blocks with the imbalanced frequencies.

As detailed in Table 3 of the main text, in all four studies, the priming effect in the unstated imbalanced EP (i.e., the first three blocks of the imbalanced EP) was not significant. In Studies 3 and 4, the balanced EP produced significant priming effects, and it was significantly stronger than the priming effect in the unstated imbalanced EP, $t(234, 278) = 3.64, 2.00$, $ps = .0003, .05$, $ds = .43, .21$, in Studies 3 and 4, respectively. This suggests that the imbalanced frequencies influenced the priming effect even when participants were not explicitly informed about them. Further, this may suggest that the reverse priming effects observed in the stated imbalanced EPs were the result of learning during the first three blocks and not the result of the conscious knowledge about the imbalance.

Yet, an inspection of the mean priming effects in each of the six blocks suggests that learning without explicit information about the imbalanced frequencies had a small effect on performance (Table S2). In all studies, there was no significant decrease in the priming effect from the first block of the unstated imbalanced EP to the second block of that procedure; and there was no significant decrease in the priming effect from the second block to the third. In all studies, the fourth block, right after the participants were informed about the imbalance, was the one that showed the larger decrease in the priming effect, usually to a very strong reverse priming effect.

Additionally, the priming effect in block 4 (the first block of the states imbalanced EP) was not significantly smaller when it was preceded by three imbalanced blocks (Study 3: $M = -71$, $SD = 93$; Study 4: $M = -18$, $SD = 72$), in comparison to when it was preceded by balanced blocks (Study 3: $M = -49$, $SD = 103$; Study 4: $M = -10$, $SD = 60$), $ts(217, 222) = 1.64, 1.22$, $ps = .10, .22$, in Studies 3 and 4, respectively. Finally, there was no difference between the overall priming effect in the stated imbalanced EP when it was preceded by the imbalanced EP

(Study 3: M = -56, SD = 68; Study 4: M = -9, SD = 45), in comparison to when it was preceded by the balanced EP (Study 3: M = -49, SD = 69; Study 4: M = -10, SD = 49), ts < 1. All these suggest that explicit information about imbalanced frequencies was the main reason for the reverse priming effect observed in the stated imbalanced EP.

Table S2

The priming effect by blocks in Studies 1-4 (in parentheses, the standard error)

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
Study 1 (all blocks imbalanced)	8 (5)	-3 (4)	-3 (3)	-64* (8)	-44* (5)	-37 (5)
Study 2 (all blocks imbalanced)	4 (6)	-2 (4)	-1 (4)	-20* (6)	-21 (5)	-10 (4)
Study 3 (all blocks imbalanced)	5 (4)	-16 (5)	-5 (5)	-71* (9)	-51 (8)	-42 (7)
Study 3 (first three blocks balanced)	19 (6)	14 (4)	9 (4)	-49* (10)	-46 (8)	-37 (7)
Study 4 (all blocks imbalanced)	-3 (7)	-4 (5)	8 (5)	-6 (6)	-9 (6)	-10 (6)
Study 4 (first three blocks balanced)	0 (6)	3 (5)	19* (5)	-18* (7)	-7 (5)	-3 (6)

Notes. (a) Before block 4, participants were always informed that the next three blocks will have imbalanced prime-target frequencies. (b) * indicates that the mean is significantly (< .05) different than the mean in the previous block. (c) Notice that the values in the parentheses are standard errors and not standard deviations.

Comparisons between the correlations of the EP tasks and other measures

The correlations between the different attitude measures in the five studies are detailed in Table 4 in the main text. In each of the studies, it was possible to test whether the stated imbalanced EP had significantly lower correlations with the explicit measure (and the IAT in Study 5) in comparison to the other EPs. The largest difference was in Study 3, between the correlation of the explicit attitude with the stated imbalanced EP, r(118) = .10, p = .29, and with the unstated imbalanced EP r(118) = .25, p = .008. A Williams test of equality of dependent correlations found that the two correlations were not significantly different, t(115) = 1.27, p = .21. The second largest difference was found in Study 5, after removing bivariate outliers, when the correlation between the standard EP and the IAT, r(118) = .28, p = .001, appeared bigger than the correlation between the stated balanced EP and the IAT, r(135) = .16, p = .07. Again, this difference (comparing independent correlations) was far from significant, z = 1.08, p = .31. In conclusion, the correlation of the stated imbalanced EP with non-EP attitude measures was never inferior to the correlations of the other EPs with the same non-EP measures.