**Data Preparation**

214 participants (108 females, *Mage =* 30.65*, SD =* 6.08) took part to the study on Prolific. We excluded data from 15 participants who did not complete the entire session. The data of participants who had IAT error rates above 30% across the entire task, or above 40% for any one of the four critical blocks, or for participants who responded faster than 400ms on more than 10% of IAT trials (*N* = 20) were also excluded from the analyses. This led to a final sample of 179 participants.

**Results**

We first looked at the effect of CS-US color matching on explicit attitude. We calculated a difference score by subtracting explicit ratings for CS2 from CS1. The one-way ANOVA revealed a main effect for color matching, *F*(1,178) = 58.61, *p* < .001, η2partial **=** 0.25, suggesting that the CSs acquired the valence of the US when the two shared colors that have been symbolically related to one another during the color training phase. We also found a significant effect of color matching on the IAT score, *F*(1, 178) = 14.39, *p* < .001, η2partial=0.08: participants that saw CS1 appearing in the color that was matched with the color of positive words showed higher IAT scores (revealing a preference for CS1 over CS2). Then, we checked whether the color matching manipulation also resulted in a significant effect on intentions towards the two CSs. We looked at the difference in the proportion of participants who intended to purchase CS1 and CS2 in the two conditions. The proportion of responses in favor of CS1 was higher when the color of CS1 matched the color of the positive USs (0.30) than when it matched the negative ones (.13, *z* = 2.84, *p* = 0.005). The opposite pattern emerged for the proportion of choices in favor of CS2 (.15 vs. .31, *z* = -2.57, *p* = 0.01).

*CS-US contingency awareness.* We found that 75% of participants (*N* = 134) responded correctly to the two questions about the CS-US contingencies (i.e., they selected “One word always had a positive meaning and the other a negative meaning” when presented with the following question: “In the second part of the experiment we presented CS1 [CS2] along with two other words. Did those two other words always have a positive meaning, have a negative meaning, or did one word always have a positive meaning and the other a negative meaning?”). Of the remaining 25% (n = 45), nine (5%) responded to one of the CSs correctly and the other incorrectly. The other 36 participants responded incorrectly on both questions, either indicating that they could not remember (n = 3) or answering both questions incorrectly. Coding participants as either having passed or failed the contingency awareness test, and adding this as a factor in a one-way ANOVA together with Color Matching did not reveal any interactions for explicit, *F*(1, 177) = 0.79, *p* = .38 or implicit scores, *F*(1, 177) = 1.34, *p* = .25.

*CS-US Color contingency awareness*. Participants were asked which colors were associated with the CSs and USs (i.e “In the second part of the experiment, when CS1 [CS2] appeared on the screen with two other words, what color was CS1 [CS2] presented in:” or “In the second part of the experiment, when CS1 or CS2 appeared on the screen, what color were the POSITIVE WORDS presented in:”). 82 participants (46%) correctly matched all the colors with their associated CSs and USs. Of the remaining 97 participants, 50% responded correctly to at least two questions and 34 of them had no correct responses. 11% of all participants indicated not remembering any of the associations. If we code individuals as having passed or failed this test, and add this factor to a one-way ANOVA together with Color Matching, we did not see a significant interaction effect for explicit, *F*(1, 177) = 0.20, *p* = .65, and implicit evaluations, *F*(1, 177) = 0.002, *p* = 0.96.

*Color contingency awareness*. Participants were asked which colors were associated with one another (i.e. “Think back to the first part of the experiment where you learned about the relationship between colors. What color was BLUE related to:”). A color contingency score that ranged from 0 to 4 was calculated. The majority of participants (N = 142) was able to link all four colors (blue, yellow, green and purple) correctly to one another (79%). Only 13 participants scored 0 and were therefore unable to correctly identify the relationship between colors.

*Hypothesis (color) and influence Awareness.* We looked at participants’ response to the color awareness question (i.e., “*Think back to the first part of the experiment. During that part of the study, we trained you to relate Blue to Yellow and Green to Purple. In the second part, we presented one of the words on the left in blue (or green) and the words on the right in yellow or purple. Did you notice this during the study?*”). We found that 84% of participants did notice this whereas 16% did not. We re-analyzed the data considering hypothesis awareness as a factor in a one-way ANOVA together with Color Matching. We found a significant interaction effect on explicit evaluations, *F*(1,177) = 7.51, *p* = 0.007, η2partial **=**0.04**,** such that explicit attitudes towards CS1 [CS2] were stronger for participants that noticed how the color relationship from the first part related to the color arrangement in the second part (*M* = 2.73, *SD* = 4.20), compared to participants that did not notice this procedure (*M* = 0.44, *SD* = 2.89). In case of the implicit measure, we did not find such an interaction effect, *F*(1,177) = 0.17, *p* = 0.68. Among participants who noticed this particular procedure, we found that only 51% of them reported that the color matching influenced the way they evaluated the CSs, whereas the remaining 49% said it did not. Eighty-two participants explicitly reported being influenced by the color matching. As was the case with hypothesis awareness, influence awareness shaped the impact of our manipulation on explicit attitudes, as demonstrated by the significant interaction effect with color matching, *F*(1, 177) = 26.63, *p* < 0.001, η2partial **=**0.13. This suggests that participants who believed this color matching influenced their subsequent attitude towards CS1 [CS2], in fact showed the strongest explicit EC effect (*M* = 4.00, *SD* = 4.58) compared to participants who did not believe it had an influence (*M* = 0.98, *SD* = 3.03). In case of the implicit attitudes, no such interaction effect was found, *F*(1, 177) = 0.68, *p* = 0.41.

*Demand.* We had 19% of demand compliant participants for the explicit measures and 14% for the IAT. The exclusion of these participants did not affect the magnitude of implicit or explicit attitude.

*Reactance.* We had 15% of reactant participants for the explicit measures and 8% for the IAT. The exclusion of reactant participants did not affect the magnitude of implicit or explicit attitude.

*Writing down manipulation check.* We had 4% of participants that indicated writing down what happened in order to help them figure out what was going on.