Primed to Like Yourself - Can Self-evaluative Cognition Be Changed by Learning Contingencies of Self-evaluative Statements and Truth-values?

Jusepeitis, Adrian; Rothermund, Klaus

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-95432-7

Personality Psychology

Primed to Like Yourself – Can Self-evaluative Cognition Be Changed by Learning Contingencies of Self-evaluative Statements and Truth-values?

Adrian Jusepeitis[1] [a], Klaus Rothermund[1]

[1] Department of Psychology, Friedrich-Schiller-University Jena, Germany

Collabra: Psychology

Measures of automatic propositional self-evaluation have been shown to predict adverse outcomes above and beyond measures of deliberate self-evaluation, thereby suggesting an independent source of automatic self-evaluation that might also provide a pathway to change self-esteem and its correlates. Based on theoretical models of automatic, proposition-based evaluative cognition, we hypothesize that automatic self-evaluation can be changed by raising the accessibility of specific truth-values in the presence of self-positive and self-negative statements. To test this hypothesis, we exposed $N = 160$ participants to a learning procedure based on the Propositional Evaluation Paradigm on three consecutive days. This procedure implemented contingencies between self-positive statements and truth in one condition and between self-positive statements and falsity in the other condition. Investigating the performance of the participants in the learning procedure itself, we found evidence for short-term effects of the contingencies as well as cumulative effects across days. However, the learning procedure had no effect on external criteria such as questionnaires of affect and self-esteem as well as the preference for one's own initials. Implications and suggestions for future research on the malleability of automatic propositional self-evaluation are discussed.

Self-esteem differs from many other stable individual traits in that it is not conceptualized as a capability, a behavioral disposition, or an affective disposition (cf. general intelligence, extraversion, or neuroticism) but a specific evaluative disposition, namely a global evaluative disposition concerning the self (Rosenberg et al., 1995). This unique position allows us to make sense of self-esteem from the background of the rich literature on evaluative cognition, which provides a framework to describe and understand self-evaluative processes their properties, their development, and their potential malleability.

The predominant current perspective on (self-)evaluative cognition is heavily shaped by dual-process models. These models differentiate controlled, deliberate self-evaluation (historically named "explicit self-esteem") and automatic self-evaluation (historically named "implicit self-esteem"). Multiple models imply an interdependence of both (self-)evaluative processes, but still suggest that they rely on different learning experiences and influence different outcomes (Asendorpf et al., 2002; Fazio, 1990; Gawronski & Bodenhausen, 2007; Strack & Deutsch, 2004). Different models diverge with regard to their assumptions about the characteristics of both types of evaluation (for an overview, see Strack & Deutsch, 2015), however, one frequently proposed duality is the duality of associative and propositional processes. Associative processes are thought to operate in a network that connects conceptual/perceptual/behavioral nodes by unqualified links. The strength of these links and therefore the degree to which one node is activated when a linked node is activated ("spread of activation") is determined by the frequency and recency of the combined activation of both nodes. On the other hand, propositional processes consist in syllogistic reasoning validating propositions (statements about the world) in light of the currently accessible information. Applied to self-evaluation, associative self-evaluation is thought of as relying crucially on the associative strength of the concept "I" and "good" (or "bad") and propositional self-evaluation is conceptualized as the outcome of syllogistic processes operating on self-relevant knowledge assigning truth-values to self-evaluative statements ("I am good").

---

a Correspondence concerning this article should be addressed to Adrian Jusepeitis, Am Steiger 3, Institut für Psychologie, 07743 Jena. Email: adrian.jusepeitis@uni-jena.de

Dual process models differ in how far they equate "associative" with "automatic" (i.e., showing one or more features of automaticity: goal independent, unconscious, efficient, fast; Moors et al., 2010) and "propositional" with "controlled" (i.e., not showing features of automaticity). For example, the Reflective-Impulsive Model (Strack & Deutsch, 2004) assumes associative processes to be fast and efficient (i.e., automatic) and propositional processes to be slow and resource intensive (i.e., controlled), while the Associative-Propositional Evaluation (APE, Gawronski & Bodenhausen, 2007) model explicitly conceptualizes "associative vs. propositional" and "automatic vs. controlled" as theoretically orthogonal dimensions (operating principles vs. operating conditions), thereby theoretically allowing for automatic propositional processes. However, one central driving force of the equation of "automatic" and "associative" and thereby an impediment to the investigation of automatic propositional processes was and is the notion of "implicitness".

Across different research streams and traditions, different researchers used and use the "implicit" terminology to mean different things. Corneille and Hütter (2020) identified three predominant connotations of this term: Implicit-as-Indirect, Implicit-as-Automatic, and Implicit-as-Associative, each of which emphasizes a different meaning of a measurement instrument, a measurement outcome, or a mental process being "implicit". They criticize this conflation of meanings in one term not only on the grounds of clear communication and concise theory but also on empirical grounds. Among other important findings, previous work has demonstrated that indirect measures do not exclusively capture evaluative associations but also non-evaluative processes (Ito et al., 2015; Rothermund et al., 2009) and that they can easily be influenced by propositional information (Kurdi & Banaji, 2017; Van Dessel et al., 2018). Moreover, there is little evidence for the automatic acquisition of mental associations (Corneille & Stahl, 2019) and even directly reported evaluations can be influenced by processes that are not strategically controlled (Hütter & Sweldens, 2018). For a comprehensive review of the relevant findings, see Corneille and Hütter (2020).

Despite the diversity in conceptualizations in dual process models as well as cumulating evidence against the concordance of automaticity and associative processes on the one hand and controlled and propositional processes on the other hand, automatic (self-)evaluation has long been investigated only in terms of associative (self-)evaluation. Case in point, measures of automatic self-evaluation were almost exclusively designed to not contain propositional information (e.g., relational qualifiers or full sentences) but to present isolated self-relevant stimuli (portraits, pronouns, names, letters, etc.) and to indirectly assess evaluative tendencies that are caused by these stimuli, presumably mediated via evaluative associations.

Automatic Propositional Self-evaluation

Starting with the Implicit Relational Assessment Procedure (IRAP; Power et al., 2009), however, multiple measures have been proposed that do not follow these equations of "controlled = propositional" and "automatic = associative". These measures A) try to establish conditions of automaticity by requiring fast responding, B) are indirect insofar as they do not ask respondents to self-evaluate, but C) present propositional information and require true/false responses. With this pattern of features showing properties of both conventionally "implicit" and "explicit" measures, they clash with conventional research practices by specifically investigating automatic propositional processes. Both in self-esteem research as well as in other attitude research, said measures have begun to show some promising results even outperforming the traditional "implicit" measures in some cases.

For example, in a self-esteem IRAP (Remue et al., 2013) respondents were instructed to react to combinations of sample stimuli (e.g., "I AM") and target stimuli (positive / negative adjectives) using "true" / "false" responses. In a compatible block, they were asked to press "true" ["false"] for stimulus combinations that indicated self-positivity [self-negativity] and in an incompatible block this instruction was reversed. The difference in performance between the blocks was interpreted as an indirect measure of self-evaluation and was shown to be significantly predicted by the level of depressive symptoms (low vs. high; Remue et al., 2013, 2014), which the scores of a self-esteem IAT were not.

Dentale et al. (2020) designed a self-esteem Relational Responding Task (RRT; De Houwer et al., 2015) that confronted respondents with self-evaluative statements in two blocks. In one block, participants were instructed to respond with "true" when presented with self-positive statements and with "false" after self-negative statements. In the other block, the instruction was reversed. The performance difference between both blocks was interpreted as a measure of the automatic agreement with self-evaluative statements. Besides being correlated with a conventional self-esteem questionnaire and measures of self-esteem correlates, this measure showed incremental validity in predicting the severity of depressive symptoms beyond the conventional self-esteem questionnaire. Jusepeitis and Rothermund (2022a) were able to replicate this result using the Propositional Evaluation Paradigm (Müller & Rothermund, 2019). In this study, the speed of true/false responses primed by self-evaluative statements was informative concerning negative affect and depressive symptoms beyond a conventional self-esteem questionnaire. Assuming that automatic propositional self-evaluation is indeed a causal antecedent of these adverse outcomes, an important question is how they develop and how they can be changed in a way to ultimately have beneficial consequences.

Changing Automatic Self-evaluation

As automatic self-evaluation so far was mostly conceptualized in terms of associations, learning procedures attempting to change it have often relied on establishing or strengthening associations between the self and a positive evaluation, that is, evaluative conditioning. For example, Baccus et al. (2004) paired self-relevant and non-self-relevant stimuli with smiling, neutral, or frowning faces in

a standard evaluative conditioning procedure. For participants in the experimental group, self-relevant stimuli were always paired with smiling faces. In the control group, no such contingency existed. Post-training composite scores of a self-esteem Implicit Association Test (Greenwald & Farnham, 2000) and a Name Letter Task (Kitayama & Karasawa, 1997; Koole et al., 2001; Nuttin, 1985) showed significant differences between the groups. A similar procedure using subliminal presentations of stimuli produced varying results in multiple studies (Dijksterhuis, 2004; Fleming & Burns, 2017; Grumm et al., 2009; Versluis et al., 2018).[1] Furthermore, Maricuțoiu et al. (2019) used a variant of the Self-Referencing Task (Perkins & Forehand, 2012), in which respondents were instructed to categorize self-related and other-related words as well as pleasant and neutral images to condition self-positivity. A self-positive group that used one common response for classifying stimuli as self-related or positive and another response for both other categories consistently showed increased self-reported self-esteem compared to a control group that only categorized positive and neutral images.[2] However, no group difference was observed for the Name Letter Task (here: Initial Preference Task).

Overall, the evidence for the effectiveness of evaluative conditioning in changing self-evaluation seems ambiguous, but a recent meta-analysis attested these procedures to have small but significant effects on average ($d = .25$; Niveau et al., 2021). Due to partially incomplete designs (see footnotes 1 and 2), however, it is unclear what portion of these effects is actually driven by factors other than evaluative conditioning, in the narrow sense of contingency learning.

Concerning the longevity of possible effects of these procedures and their beneficial consequences going beyond self-evaluation less research has been conducted. To the best of our knowledge, only Espinosa et al. (2018), Fleming and Burns (2017), and Maricuțoiu et al. (2019) investigated whether self-evaluative conditioning procedures have effects on outcomes that do not reflect measures of self-esteem. In their "proof-of-concept" study ($N = 28$), Espinosa et al. (2018) found inconclusive results concerning the amelioration of subclinical psychotic symptoms and paranoia ideation brought about by the learning procedure used by Baccus et al. (2004). Fleming and Burns (2017) found no effect of self-evaluative conditioning on negativity toward homosexuality in homosexual men. Maricuțoiu et al. (2019) found increased self-reported well-being (Study 2) but no change in mental health (Study 3) in their experimental groups.

Although recent accounts of evaluative conditioning explicitly acknowledge the influence of propositional processes (De Houwer, 2018), to our knowledge there is no research yet specifically investigating the malleability of automatic propositional self-evaluation. In general, the APE model proposes that propositional evaluations are specifically affected by introducing novel propositions or raising the salience of self-relevant propositions (e.g., by plainly presenting them or by creating experiences that facilitate their deduction) that conflict with previously endorsed propositions. A variety of interventions has been designed based on similar frameworks, most notably Cognitive Behavioral Therapy which has been found to be the most studied self-esteem intervention and produces moderate effects (Niveau et al., 2021). Importantly, cognitive therapy and similar interventions rely on slow and deliberate processing of self-relevant information in changing self-esteem and studies regarding their effectiveness mostly rely on self-reported (i.e., potentially controlled) dependent variables. Accordingly, those studies are not directly informative concerning automatic propositional self-evaluation.

Which interventions then, can be assumed to specifically affect automatic propositional self-evaluation? According to the APE model, propositional reasoning can vary in complexity (i.e., how many different propositions are considered). If more propositions are considered, the probability of encountering propositions inconsistent with associative evaluations increases. This is why a dissociation of associative and propositional evaluations becomes increasingly likely (although not necessarily) with more complex propositional reasoning. Approaching measures that attempt to assess automatic propositional evaluations from this background, while not measuring mere evaluative associations, they can be assumed to measure propositional evaluations with very low elaboration and complexity. Within a PEP, IRAP, or RRT trial, there is neither enough time nor incentive to consider many relevant propositions. Hence, we assume that these measures assess the most accessible truth-value for a given prime statement regardless of its consistency with other possibly relevant propositions. Just as with evaluative associations, it is reasonable to assume that the accessibility of truth-values depends on the frequency and recency of their co-occurrence with the according proposition or similar propositions in the individual's learning history. This view is also consistent with the behavior analytic Relational Frame Theory (Hayes et al., 2001) underlying the rationale of the IRAP and more specifically the Relational Elaboration and Coherence model (Hughes

---

1 Note, however, that their designs did not control for the overall frequency of positive stimuli across conditions and therefore leave some room for alternative explanations of their respective findings. Also, see Replicability-Index (2016) for an critical evaluation of the replicability of Dijksterhuis (2004).

2 Importantly, this design presents self-relevant and non-self-relevant words only to the experimental group. Thus, besides evaluative conditioning (association formation between the self and positivity), the mere exposure to and categorization of self- and non-self-related words must be considered as an alternative explanatory factor (e.g., familiarity, fluency; Reber et al., 1998) of the group differences.

et al., 2012) nested within RFT. Very briefly put, while avoiding references to any cognitive or mentalistic constructs (e.g., "beliefs"), the REC model posits that fast responses to stimuli rely on relational responding with low levels of derivation and low complexity. Here, "low levels of derivation" indicates that there has been an extensive learning history concerning the relation of two stimuli (e.g., "I" and "good"), excluding that the relational response is not derived from stimulus relations only during the assessment procedure. "Low complexity" means that responses rely on few and simple stimulus relations (e.g., "I = good") instead of a vast network of many stimuli and their manifold relations.

Hence, we hypothesize that automatic propositional evaluations can be modified analogously to automatic associative evaluations by increasing the availability of certain truth-values in the presence of self-evaluative propositions, that is, by providing a learning history of simple contingencies between certain self-evaluative propositions and certain truth values.

Summary and Outline of the Study

In sum, automatic propositional self-evaluation, despite being implied by influential models of evaluative cognition, has only started to be investigated recently. In this line of research, indirect measures of self-evaluation presenting propositional information have proven useful in incrementally predicting adverse outcomes and thus suggest novel pathways in changing self-evaluation and ultimately said outcomes. In line with propositional models of evaluative learning (e.g., De Houwer, 2018) and RFT, we hypothesize that indirect relational measures of self-esteem capture fairly unelaborate propositional self-evaluations that do not involve many other relevant propositions or vast relational networks but simply rely on the most available truth-value when presented with a self-evaluative statement. To try to increase the availability of certain truth-values in the presence of certain statements we developed an active learning procedure based on the procedure of the Propositional Evaluation Paradigm (Müller & Rothermund, 2019). This learning procedure consists in a priming paradigm with stimulus-target contingencies, the stimuli being self-evaluative statements and the targets being truth-values. In a between-group design, we exposed one group to contingencies that are in line with a positive self-evaluation and the other group to contingencies that are in line with a negative self-evaluation. We investigated whether this was able to influence A) the availability of truth-values in the presence of self-evaluative statements as estimated by the performance on the priming procedure itself and in a questionnaire with speeded responses, B) deliberate self-evaluation in a self-esteem questionnaire, C) current affect measured in a self-report questionnaire, and D) a conventional indirect measure of self-evaluation based on the idea of associative self-evaluation. To increase the impact of the learning procedures and to capture the progress of change across time, the procedure was repeated in three sessions across three consecutive days and dependent measures were assessed at various times during the study.

Methods

The preregistration (https://osf.io/m5kea), as well as study materials, raw data, and R syntaxes (https://osf.io/zx2aq/) can be accessed in the Open Science Framework (Jusepeitis & Rothermund, 2022b). The procedure of this study was approved by the Ethical Commission of the Faculty of Social and Behavioral Sciences of the University of Jena (reference number FSV 22/024).

We conducted a pre-registered pilot study in which highly similar results were obtained as in the present study, which was conducted as a follow-up to the pilot study. The main study had slightly more power and a more extensive learning procedure. Hence, we only give a brief report of the pilot study and its results in Supplementary Materials. All data, syntaxes, and the preregistration of the pilot (https://osf.io/tgkzh) study are also available on OSF (under the same link).

Sample

For this online study spanning three sessions on three consecutive days, German-speaking participants were recruited via Prolific (www.prolific.co) to complete a series of experiments that were created using the PsychoPy Builder interface (Peirce et al., 2019), output to PsychoJS experiments (Bridges et al., 2020), and hosted on Pavlovia (www.pavlovia.org). Participants received monetary compensation for taking part.

The data collection ran until $N = 160$ complete datasets were collected. As pre-registered, this sample size was chosen to ensure a power of at least .80 to detect small ($f = .1$) interactions in a mixed-design ANOVA assuming a pre-post correlation of .60 for the dependent measures. This power analysis was based on the effect of the learning procedure on the external criteria, which were assessed before and after the manipulation. To ensure adequate power we assumed a small effect and a pre-post correlation of .60, which can be thought of as a lower bound of plausible correlations within self-report measures across an interval of three days. Hence, our power analysis is based on conservative assumptions: Larger effects and higher pre-post-correlations would only imply a higher power for the given same sample size. Power calculations were run using G*Power 3.1 (Faul et al., 2009).

To ensure comparable sample sizes in the four (2 experimental conditions x 2 questionnaire scale poles, see below) conditions, participants were allocated to the conditions in a semi-random way. That is, participants were randomly assigned to one of the conditions with the lowest number of completed and active participants when starting the first session. Once allocated, participants remained in one condition across all three sessions. $n = 190$ complete datasets were collected for session 1 (dropout rate = 16%). The dropout rates from session 1 to session 3 did not differ between the self-positive and self-negative condition, $\chi^2 = 0.019$, $df = 1$, $p = .89$. The final sample (80 female, Age: $M = 33.59$, $SD = 11.09$) was distributed evenly ($n = 80$) across both experimental conditions. The two samples coincidentally differed with regard to the distributions of sex (more

males in the self-negative condition), $\chi^2 = 4.225$, $df = 1$, $p = .040$, but not age, *Welch t* = 0.263, *df* = 156.19, *p* = .793. Because of this we later tested whether any of the effects of the condition could be better explained as effects of sex. However, in no analysis did sex explain variance in the outcome variables let alone explain more variance than the condition factor.

Procedure

In session 1, participants first completed a pre-training measurement of two self-esteem questionnaires. After being introduced to the PEP and completing a practice phase, they completed the self-evaluation training. At the end of the session, their current affect was measured with a questionnaire. Session 2 repeated the same procedure without the self-esteem questionnaires and the practice phase of the PEP. In session 3, participants again completed the self-evaluation training and affect measures. Afterward, a post-training measurement of the self-esteem questionnaires was taken. Finally, self-esteem items had to be rated again in a speeded self-report and lastly, the Name Letter Task was completed.

Materials

Self-esteem questionnaires. A German version of the Rosenberg Self-Esteem Scale (RSES; Collani & Herzberg, 2003; Rosenberg, 1965) as well as a subset of the items of the Core Beliefs Questionnaire (CBQ; Wong et al., 2017), which we translated to German and amended with inverted items (see Tables S1 and S2 for all items), were used to assess self-esteem before and after the training. RSES items were rated on a 7-point Likert scale and CBQ items were rated on a 7-point Likert scale. Mean scores for all items were calculated for both questionnaires after inverting ratings of negatively phrased items. The resulting scores will be abbreviated as RSES and CBQ in the following.

The assignment of poles to the left and right end of the self-esteem questionnaires, as well as the affect questionnaire and the Name Letter Task described below, was counterbalanced across participants. That is, for some participants the lowest point on the scale signified agreement and for others it signified disagreement. This was done to prevent possible learning effects in the scales to arise from spatial contingencies (e.g., self-positive – right) instead of semantic contingencies (e.g., self-positive – true) in the learning procedure.

Learning procedure. The learning procedure consisted in a modified Propositional Evaluation Paradigm (PEP; Müller & Rothermund, 2019). The PEP is an indirect priming-based measure of truth-evaluations of statements. In a PEP trial, respondents are primed with a sentence in rapid serial visual presentation (for temporal details, see Müller

& Rothermund, 2019). After the sentence, a target word is presented. If this word is "TRUE" participants are instructed to respond with "true", if the target is "FALSE" respondents are instructed to respond with "false" (for a graphical representation of PEP trials, see Figure 1). In the present version of the PEP, responses were given by moving the mouse cursor into the upper right ("true") or upper left ("false") corner of the screen, starting from a starting zone in the bottom center of the screen (see Cummins & De Houwer, 2021). Analogously to an affective priming task, the prime statement is thought to facilitate compatible responses and inhibit incompatible responses. That is, if the statement "Berlin is the capital of Germany" is presented, responses to the target "TRUE" are more accurate and faster than to the target "FALSE". In contrast to an affective priming task, the PEP also contains catch trials besides the probe trials just explained. In catch trials, the target "?????" is presented and respondents are instructed to give their subjective truth-evaluation of the preceding sentence. These trials are necessary to create an "evaluative mindset" and prevent participants from ignoring the prime sentences (Wiswede et al., 2013).

When the PEP is used as a dependent measure, each statement is presented as often followed by the target "TRUE" as followed by the target "FALSE" in the probe trials. This means that there is no contingency between the statement and the target.[3] A compatibility effect can then be calculated as the difference in performance (here: the time it takes the participant to move the cursor to the correct corner of the screen after the target is presented) of both of these trial-types. Usually, this compatibility effect is not calculated for a single statement but for a collection of statements thought to measure the same underlying construct, just as in a questionnaire.

In our training PEP, the usual procedure was changed to establish contingencies between the type of sentence and the target. As prime statements, CBQ items were used that were either self-positive (e.g., "I am strong.") or self-negative (e.g., "I am weak."). Participants were assigned either to a self-negative condition or a self-positive condition. In the self-positive condition, self-positive sentences were followed by the target "TRUE" in 80% of probe trials and by the target "FALSE" in 20% of probe trials. In accordance, self-negative sentences were followed by the target "FALSE" in 80% of probe trials and by the target "TRUE" in 20% of probe trials. For participants in the self-negative condition, these contingencies were exactly reversed. Hence, participants were presented with 80% self-positive trials (see 1 and 2 in Figure 1) in the self-positive condition and with 80% self-negative trials (see 3 and 4 in Figure 1) in the self-negative condition. These percentages were chosen to render the sentences highly predictive of the following target, while still allowing for the computation of

---

[3] Depending on the version of the PEP, this is only true regarding the critical trials (probe trials), which nonetheless are most important since they are the only trials entering into the computation of the PEP effect.
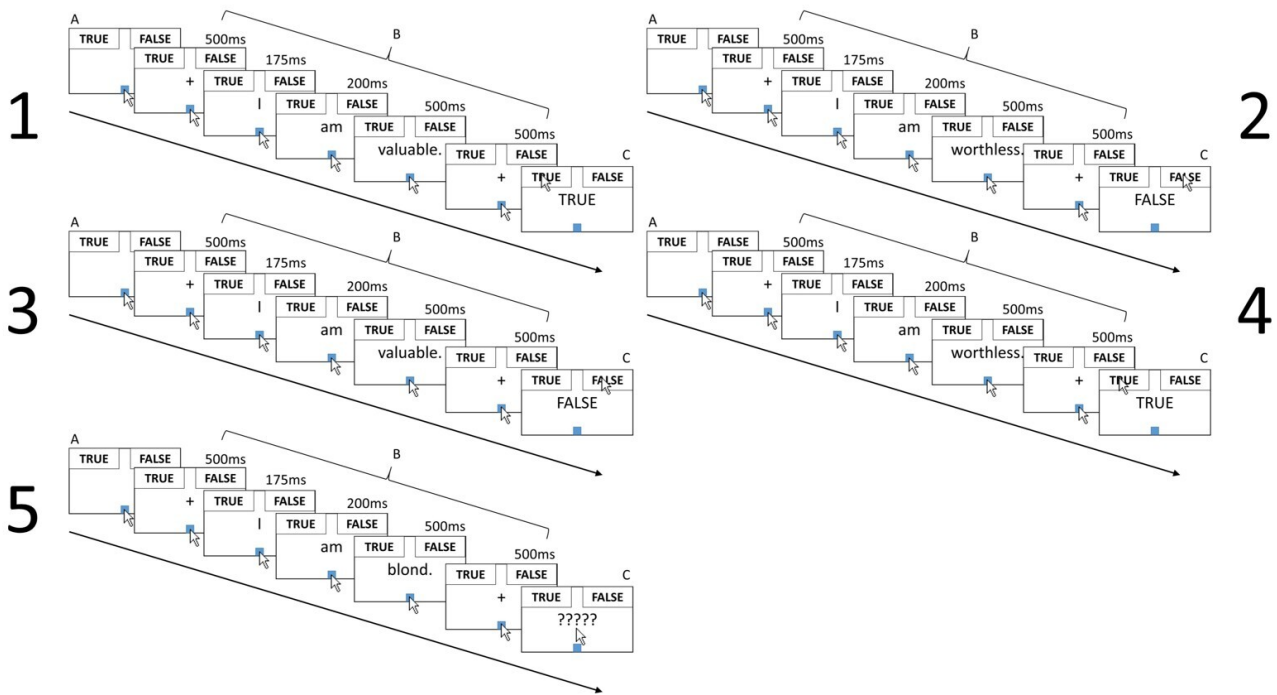
Figure 1. Sequence of events in exemplary SE-PEP trials

Notes. Stimuli are not drawn to scale. A = starting screen, clicking the bottom center area started the trial. B = RSVP of prime sentences and fixation crosses. C = target, presented until the cursor reached a response box. 1 and 2 are self-positive trials (self-positive statement followed by "TRUE" target and self-negative statement followed by "FALSE" target), 3 and 4 are self-negative trials (self-positive statement followed by "FALSE" target and self-negative statement followed by "TRUE" target), 5 are catch trials (self-neutral statement followed by "?????"). Filler trials are not depicted here.

a reasonably reliable compatibility effect to investigate the effects of the training in the PEP performance itself. To not change these contingencies by the implementation of catch trials, we used only neutral statements (e.g., "I am dark-haired", full set in Table S3) to be presented before the target "?????". To hide the fact, that only neutral statements were ever to be directly evaluated, we also implemented a small number of trials where a neutral statement was followed by "TRUE" or "FALSE". We will call the according trials "filler trials" in the following. The compatibility effect was calculated as the difference of reaction times in self-negative and self-positive trials divided by the standard deviation of reaction times in both trial types. It will be abbreviated as $PEP_{\Delta RT}$ in the following. High self-esteem implies a poor performance in self-negative trials and a good performance in self-positive trials, meaning that high self-esteem should imply high values of $PEP_{\Delta RT}$.

The PEP procedure in each session began with a test block of 46 trials (28 probe, 14 catch, 4 filler) without contingencies between statement type (prime) and target to assess a baseline performance that was not affected by short-term retrieval of stimulus-response bindings (Rothermund et al., 2005; for a review see Frings et al., 2020). Afterward, the learning procedure in the strict sense began, consisting of three blocks of 51 trials each (30 probe, 15 catch, 6 filler). The statements to be presented were sampled from the CBQ items in a semi-random way, ensuring that a single statement was not presented disproportionally often. The order of trials was also chosen in a semi-random way that prevented the same primes or targets to be presented more than twice in a row. The trial lists in the

test blocks were identical for all participants in both conditions. were identical for each participant in a condition but differed across the blocks and days. That means that across the study we presented 1 (test block) x 3 (days) + 3 (learning blocks) x 3 (days) x 2 (conditions) = 21 different trial lists. The PEP procedure in the first session was preceded by detailed instructions and a practice block with only neutral sentences (15 trials).

Affect questionnaire. Self-evaluation is linked to affect to the degree that discrimination of mood and state self-esteem has been an important issue in designing self-esteem questionnaires (Heatherton & Polivy, 1991). Hence, we used a German version of the Positive Affect Negative Affect Schedule (Breyer & Bluemke, 2016; Watson et al., 1988) as one of the external criteria. Participants had to rate ten negative and ten positive affective adjectives on a five-point Likert scale with regard to how strongly they experienced this affect in the present moment. Mean ratings for negative items were subtracted from mean ratings for positive items to yield a score that expresses the dominant affective valence and controls for unspecific affective excitement and response biases. This score will be abbreviated as AFF in the following.

Speeded SE questionnaire. Besides the rating of CBQ items in the questionnaire format, they were also answered in a speeded and dichotomized format. To this end, CBQ items were presented just as the catch trials in the PEP and "true"/ "false" responses had to be given as fast as possible by moving the mouse to the respective corner of the screen. The dependent variable from this task was the rate of self-positive responses (responding "true" to a self-pos-

itive statement or "false" to a self-negative statement) and will be abbreviated as CBQ_{speed}.

Name Letter Task. The Name Letter Task (NLT; Koole et al., 2001; Nuttin, 1985) was implemented as an additional indirect criterion measure after the registration of the study. In the NLT, participants are asked to rate the letters of the (German) alphabet on a 5-point scale from "not at all beautiful" to "extremely beautiful". Afterward, they were asked to report their initials. A score indicating their preference for their initials was calculated using the I-algorithm suggested by LeBel and Gawronski (2009) and will be abbreviated as NLT in the following.

Assessment of contingency awareness. To assess whether participants were aware of the prime-target contingencies in the PEP, they were asked to give two estimates after completing all other relevant measures. Specifically, they should estimate what percentage of self-positive statements was followed by the target word "TRUE" as well as what percentage of self-negative statements was followed by the target word "TRUE" across all blocks of all three sessions. A score representing the perceived contingency between prime and target was calculated as the difference between these two estimates. Because of a technical error, three participants skipped this portion of the experiment and the score could not be computed for them.

Results

Exclusion of data

As preregistered, PEP trials with reaction times smaller than 200ms and greater than 4000ms as well as subject-, day-, and trial-type specific Tukey outliers and inaccurate responses were excluded before calculating $PEP_{\Delta RT}$. No participant reached the preregistered exclusion criterion of more than 20% of excluded probe trials in this way.

NLT data for ten participants who did not report their initials or rated every letter as being equally beautiful were excluded from the analysis concerning the NLT.

Psychometric properties of measures and descriptive statistics

Investigating the internal consistency of both self-esteem questionnaires across both conditions we found good Cronbachs α's for the self-esteem questionnaires both pre- and post-training, min α = .94. Across days, negative affect items showed slightly lower internal consistency, min α = .83, than positive affect items, min α = .90. NLT scores calculated independently for the first and last initial showed a modest split-half reliability of .54. The split-half reliabili-

ties of PEP test blocks calculated based on two $PEP_{\Delta RT}$ estimates for odd and even trials ranged from .48 to .55 across days.

Development of $PEP_{\Delta RT}$ across the training

Figure 2 depicts the course of average $PEP_{\Delta RT}$ across the blocks and days of the training in both conditions.

To model this course, a mixed-effects model with fixed effects for block * day * condition and random (i.e., subject-specific) effects for days was fitted to the $PEP_{\Delta RT}$ data using maximum likelihood estimation (Bates et al., 2015).[4] This model accounted for R^2 = .399 of the observed variance, with R^2 = .096 accounted for by fixed effects only (both calculated according to Arel-Bundock, 2022; Nakagawa et al., 2017). To test our hypotheses concerning the course of $PEP_{\Delta RT}$ across the training, we estimated the model implied marginal means and investigated the simple contrast between the conditions as well as its linear change across the levels of "block" and "day" (i.e., "trends of contrasts" or "interaction contrasts"; Lenth, 2022).

Investigating the effect of the condition factor (averaged across blocks and days), we found a significant difference between the conditions, *contrast* = 0.307, *t* = 6.285, *df* = 162[5], *p* < .0001, indicating more positive self-evaluations in the self-positive compared to the self-negative condition. Averaged across days, this contrast grew significantly across blocks, *linear interaction contrast* = 1.165, *t* = 6.285, *df* = 1458, *p* <.0001. That is, as expected the average $PEP_{\Delta RT}$ of both groups grew significantly apart from block to block. However, averaged across blocks, the group difference in $PEP_{\Delta RT}$ did not grow across days, *linear interaction contrast* = 0.020, *t* = 0.286, *df* = 162, *p* = .775. That is, the average $PEP_{\Delta RT}$ in both groups did not differ more from day to day but the difference remained stable. The growth of the condition contrast across blocks did non-significantly decrease across days, *linear interaction contrast* = -0.855, *t* = -1.883, *df* = 1458, *p* = .060 (two-tailed). Hence, the rate at which the contingencies influenced the performance of participants more and more in each block did not speed up from day to day but rather showed a trend towards slowing down.

All previous models included test as well as learning blocks. Therefore, the results describe the immediate effect of the contingencies that respondents were subjected to in the learning blocks. Most importantly, however, in the neutral test block, the condition contrast also grew across days, *linear interaction contrast* = 0.194, *t* = 1.723, *df* = 844, *p* = .043 (one-tailed), meaning that from day to day both groups differed more strongly in their average $PEP_{\Delta RT}$ in the test

---

4 As preregistered, we compared multiple potential models with regard to their fit using χ^2 tests. All competing models contained fixed effects for block * day * condition but differed in terms of the random effects modeled. Model 1 contained only random intercepts, model 2 contained random effects for day, model 3 contained random effects for block and model 4 contained random effects for day and block. The presented model (model 2) fit the data best.

5 Since there are multiple methods to estimate degrees of freedom in mixed models, we compared both approaches available in the software we used (Kenward-Roger and Satterthwaite). As both approaches converged with regard to our results, we stuck to estimating degrees of freedom using the Kenward-Roger method throughout our analyses.
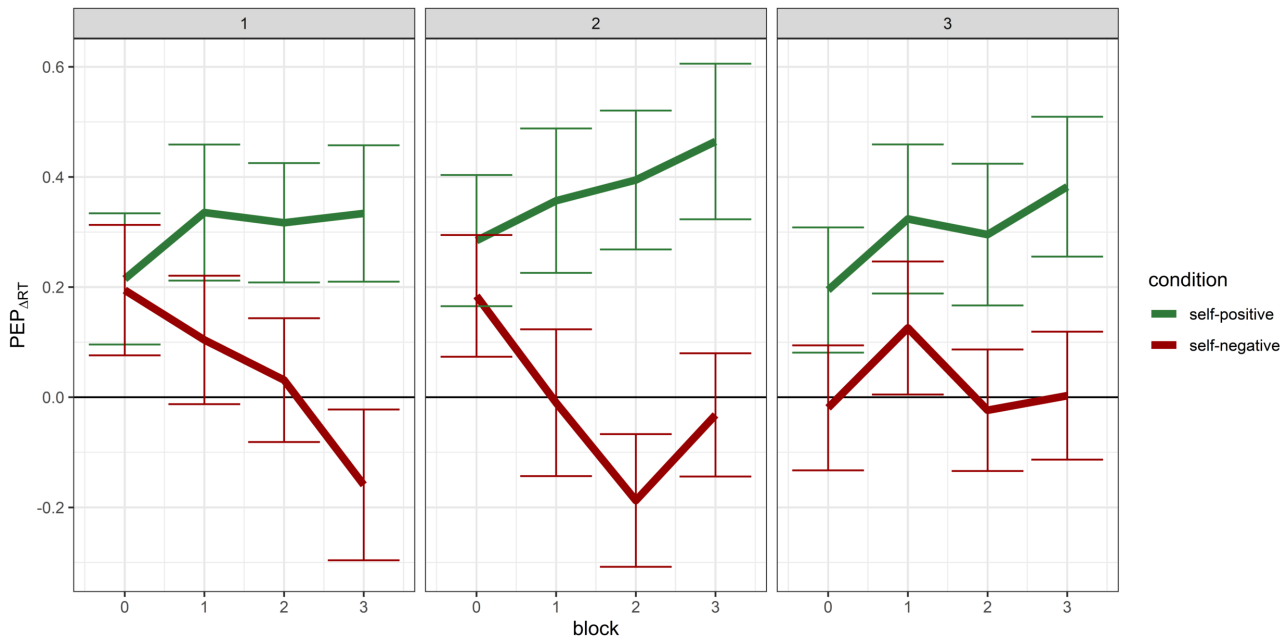
Figure 2. Course of average PEP$_{\Delta RT}$ per condition across blocks and days of the training

Notes. PEP$_{\Delta RT}$ = Average standardized mean reaction time difference between self-positive and self-negative trials. Error bars indicate 95% confidence intervals. Block 0 is the test block without prime-target contingencies (i.e., the contingency is zero for these blocks), blocks 1 to 3 are learning blocks implementing the condition-specific contingencies.

block in the beginning of the session, where they were not exposed to the training contingencies of their respective condition.

Effect on external criteria

Effects of the learning procedure on the self-esteem questionnaires were investigated in terms of the interaction contrast of the within-subject factor "time" (pre-training, post-training) and the between-subjects factor "condition" in mixed effects models with fixed effects for time * condition and random (i.e., subject-specific) intercepts. Neither for the RSES, *interaction contrast* = -.211, t = -1.536, df = 162, p = .937, nor for the CBQ, *interaction contrast* = 0.058, t = 0.499, df = 162, p = .619, did our analyses indicate an increase of the group difference in the expected direction across time.

Effects on AFF were tested similarly, in a mixed effects model with the within-subject factor "day" (1, 2, 3) and the between-subjects factor "condition". There was neither a significant group difference when averaged across days, *contrast* = 0.033, t = 0.229, df = 162, p = .410, nor was there a significant growth of the group contrast across days, *interaction contrast* = 0.063, t = 0.573, df = 324, p = .284.

T-tests comparing CBQ$_{speed}$, *Welch t* = -0.945, df = 155.2, p = .826, and NLT, *Welch t* = 0.030, df = 146.54, p = .976, between the two training groups did not yield significant results either.

Contingency awareness (not preregistered)

The estimations of prime-target contingency given by the participants at the end of the third day were on average much closer to 0 than the actual contingencies (46% / -46%

including the test blocks), self-positive condition: M = 7.64%, self-negative condition: M = -25.13%. Nonetheless, on average, the participants were aware of the contingency of primes and targets, as the contingency estimates differed significantly between the groups, *Welch t* = -8.095, df = 153.9, p < .001. Within both groups, they also differed significantly from 0, max p = .006. Also, estimations differed less strongly from 0 in the self-positive condition than in the self-negative condition, *Welch t* = 4.318, df = 153.9, p < .001, indicating that a high frequency of self-positive trials was less obvious to participants than a high frequency of self-negative trials.

Since many learning psychologists are interested in disentangling conscious influences (i.e., expectations that can be verbalized explicitly by participants and that can be used for strategic responding) from more automatic influences (i.e., those that cannot be explicitly verbalized but are based on changes in spontaneous responding that are due to – in the case of the PEP – actual changes in the subjective plausibility of the presented statements) on learning, we explored to what degree contingency awareness predicted the learning effects in the PEP. To this end, we regressed standardized PEP$_{\Delta RT}$ scores in the test blocks of sessions 2 and 3 onto three predictors: A) standardized PEP$_{\Delta RT}$ in the test block of day 1 (= baseline), B) the group factor, and C) the contingency awareness score. For both days, contingency awareness predicted the change from baseline, max p = .001, while the group factor did not, min p = .463. Due to the calculation of the contingency awareness as a difference with 0 meaning "no awareness", this finding indicates that for respondents who were not aware of the contingencies, there was no divergence between the groups in PEP$_{\Delta RT}$ from day 1 to days 2 and 3.

Discussion

In the present study, we investigated the malleability of automatic propositional self-evaluations and their effect on deliberate self-evaluation and current affect. To this end, we exposed participants to contingencies between self-evaluative statements and truth-values in a learning procedure consisting in a modified Propositional Evaluation Paradigm (PEP). In three sessions on three consecutive days, participants completed a PEP that associated self-positive statements with either the truth-value "true" (self-positive condition) or "false" (self-negative condition) and self-negative statement with the truth-value "false" (self-positive condition) or "true" (self-negative condition).

Our analyses of the compatibility effects evident in the reaction times of participants in the learning procedure indicated that participants on average indeed learned to respond in line with contingencies across the blocks of the training on each day. Participants in the self-positive condition on average performed better and better in self-positive trials than in self-negative trials while participants in the self-negative group performed better and better in self-negative trials compared to self-positive trials. Although the procedure did not consistently reverse the self-positivity bias at baseline, it weakened it across the training. This difference between the groups grew across the learning blocks within the training sessions. Most importantly, from day to day, the two groups also diverged with regard to their compatibility effects in a neutral test block preceding the learning phase on each day, indicating a transfer from the previous training sessions and an effect of the contingencies that went beyond their presence. An exploratory analysis of the effects of contingency awareness preliminarily hinted at the importance of contingency awareness for effects that accumulated across sessions and went beyond the immediate presence of the contingencies.

Importantly, while the procedure was effective in influencing performance in the task that was used to manipulate the contingencies (i.e., in the PEP) itself, it did not have any influence on external criteria, such as directly instructed deliberate and speeded self-evaluations, affect, or preference for one's initials (NLT). Pre-post differences in self-esteem questionnaires, the self-reported affective valence at the end of each session as well as speeded self-evaluation and scores in the Name Letter Task at post-training did not differ between the groups.

Taken together, these findings suggest that the learning procedure was indeed able to temporarily raise the accessibility of specific truth-values in the presence of positive and negative self-evaluative statements. However, our participants did not rely on the primed truth-values when directly instructed to self-evaluate regardless of the speed of the instructed self-evaluation, independently of whether the presented statement were identical to the ones in the learning procedure or not.

A comparison of these findings with previous evidence for the effect of evaluative conditioning on self-reported self-esteem does seem to imply the superiority of associative learning procedures in influencing self-esteem. However, as there is significant heterogeneity in findings across different conditioning procedures (e.g., compare Maricuțoiu et al., 2019, and Grumm et al., 2009) as well as confounding of the group factor with the presentation frequency of either self-related stimuli (Maricuțoiu et al., 2019) or positive stimuli (Dijksterhuis, 2004; Fleming & Burns, 2017; Grumm et al., 2009; Versluis et al., 2018) more systematic research is needed before a fair and informative comparison is possible.

The null effect of our procedure concerning self-reported self-evaluation is, on a first glance, in line with the assumption of the APE model that suggests that deliberate propositional self-evaluation should depend on more than just the availability of a truth-value of a single proposition. Instead, it should depend on syllogistic inferences concerning a host of self-relevant propositions. However, considering the temporal characteristics of the different self-evaluative judgments, the absence of a transfer from the learning procedure to the questionnaires might still be surprising. Our subjects took 0.539 seconds on average (*SD* = 0.212) to judge self-evaluative statements (CBQ items) in the speeded and dichotomized questionnaire and 2.253 (CBQ, *SD* = 0.999) and 3.471 (RSES, *SD* = 1.250) seconds to answer the items of the regular questionnaires. While dual process models are of course not formalized enough as to suggest specific temporal cut-offs for different processes, we think it is still uncontroversial to suggest that these response latencies are very small for elaborate propositional reasoning. Thus, we have to conclude that increasing the availability of specific truth-values in the presence of certain self-evaluative statements does not influence even reasonably fast and accordingly rather unelaborate instructed propositional judgments.

One possible explanation for this finding suggests itself when taking the course of the average compatibility effect in the learning procedure (see Figure 2) into consideration. Both groups started the procedure with relatively more available self-positive judgments (i.e., significantly positive compatibility effects). Although the learning procedure was able to weaken the effect in the self-negative group, their average compatibility effect was only significantly negative (not adjusting for multiple testing) in three of the eleven blocks that were potentially influenced by the learning procedure. At the end of the procedure on the last day, shortly before the measurement of the criteria, the average compatibility effect in the self-negative condition was practically zero, suggesting an equal availability of self-negative and self-positive evaluations. Hence, it could be that the procedure was powerful enough to gradually change the availability of truth-values, but not powerful enough to fully reverse a pre-existing disposition. Thus, when forced into a categorical endpoint, the results of the self-evaluative processes remained unchanged.

Limitations

Two important limitations concern the design of the training procedure. To make sure that participants process the whole statement and its meaning we included catch trials, in which respondents were instructed to report their subjective truth-value of the presented statement. How-

ever, since almost all statements began with the words "I am" (German: "Ich bin") followed by a positive, negative, or neutral adjective it is still possible that only the adjectives and their contingencies with the following target drove the learning effect in the probe trials. That is, respondents might not have learned based on contingencies of "I am good" and "true" ("false"), but only of "good" and "true" ("false"). Moreover, the target ("TRUE" or "FALSE") was perfectly confounded with the direction of the necessary response. Accordingly, the relevant contingencies that produced the effects in the learning procedure could be merely spatial (e.g., "self-positive – right"). This could be another reason why the training did have an effect on the performance in the training itself but did not transfer to deliberate self-evaluations. To produce more unambiguous findings, future studies could use more varied statement formats or control sentences like "He is god / bad" followed by "TRUE" in 50% percent of trials regardless of the condition and switch responses assigned to "TRUE" and "FALSE" within the learning procedure. This would enforce the learning of contingencies between statement-meaning and truth-values as no single word alone would have the same predictive power concerning the target as the whole sentence and the target would on average not be informative with regard to the necessary response.

Another possible methodological explanation of the lack of transfer between the learning procedure and the external criteria is the fact that the pre- and post-measurements used the same items to directly assess self-esteem and were only about 48 hours apart. Accordingly, it is possible that participants simply remembered how they responded before and, in an attempt to remain consistent, gave the same answers again after the learning procedure. In future studies, this could be prevented by using counterbalanced parallel tests for the pre- and post-measurement. Furthermore, speeded judgments of self-evaluative statements could also be presented at baseline and varied with regard to their position before or after the regular questionnaire. This would exclude the possible order effect of speeded self-evaluation being influenced systematically by the deliberate evaluation preceding it.

Moreover, the results concerning the importance of contingency awareness for the temporal transfer of effects are limited by the fact that contingency awareness was only assessed once at the end of the training. Therefore, the causal direction of contingency estimation and PEP effects is unknown. A design that would allow for stronger conclusions could assess contingency estimates after each block under the pretense of a change in contingencies. However, it needs to be considered that this assessment in itself may change the relationship between contingency estimation and learning effects.

Lastly, the critical dependent variable in the PEP itself – the effect in the test blocks – demonstrated a rather low split-half reliability. While apparently still being sensitive to group differences and effects of contingency awareness, it might be fruitful to include multiple test blocks thereby increasing the sensitivity to smaller effects, for example, incremental effects of the condition factor when controlling for contingency awareness. Another possibility for testing training effects would be to include another indirect relational measure of self-esteem (e.g., an SE-RRT; Dentale et al., 2020). However, for such a design to be informative, it would need to be preceded by further research establishing the convergence of different indirect relational measures of self-esteem.

Conclusion

Investigating the malleability of automatic propositional self-evaluation by increasing the accessibility of certain self-evaluative judgments, we found that a modified version of the Propositional Evaluation Paradigm was able to temporarily affect the performance of subjects in the learning procedure itself. Although some open questions and alternative explanations need to be addressed in future work, this study provides first evidence that the performance in indirect measures of propositional self-evaluation can be manipulated. However, the manipulation had no effect on criteria external to the learning procedure like affect and directly assessed self-esteem. Therefore, as of now, it has to be regarded as a tool in basic research and is no candidate for changing "implicit beliefs" with far-reaching consequences. In this, our study converges with many previous attempts to adopt basic learning procedures like evaluative conditioning to change attitudes and psychopathology.

<div align="center">..</div>

Competing interests

We have no competing interest to declare.

Data accessibility statement

The preregistration (https://osf.io/m5kea) as well as study materials, raw data, and R syntaxes (https://osf.io/zx2aq/?view_only=21f5ebcce09044ee9b6ce11ce56e63f1) can be accessed in the Open Science Framework.

References

Arel-Bundock, V. (2022). modelsummary: Data and model summaries in R. *Journal of Statistical Software*, *103*(1), 1–23. https://doi.org/10.18637/jss.v103.i01

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology*, *83*(2), 380–393. https://doi.org/10.1037/0022-3514.83.2.380

Baccus, J. R., Baldwin, M. W., & Packer, D. J. (2004). Increasing implicit self-esteem through classical conditioning. *Psychological Science*, *15*(7), 498–502. https://doi.org/10.1111/j.0956-7976.2004.00708.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Breyer, B., & Bluemke, M. (2016). *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel)*. https://doi.org/10.6102/ZIS242

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Collani, G. von, & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, *24*(1), 3–7. https://doi.org/10.1024//0170-1789.24.1.3

Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review*, *24*(3), 212–232. https://doi.org/10.1177/1088868320911325

Corneille, O., & Stahl, C. (2019). Associative Attitude Learning: A Closer Look at Evidence and How It Relates to Attitude Models. *Personality and Social Psychology Review*, *23*(2), 161–189. https://doi.org/10.1177/1088868318763261

Cummins, J., & De Houwer, J. (2021). The shape of belief: Developing a mousetracking-based relational implicit measure. *Social Psychological and Personality Science*, *12*(8), 1517–1526. https://doi.org/10.1177/1948550620978019

De Houwer, J. (2018). Propositional Models of Evaluative Conditioning. *Social Psychological Bulletin*, *13*(3). https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J., Heider, N., Spruyt, A., Roets, A., & Hughes, S. (2015). The Relational Responding Task: Toward a new implicit measure of beliefs. *Frontiers in Psychology*, *6*, 319. https://doi.org/10.3389/fpsyg.2015.00319

Dentale, F., Vecchione, M., Ghezzi, V., Spagnolo, G., Szemenyei, E., & Barbaranelli, C. (2020). Beyond an associative conception of automatic self-evaluations: Applying the Relational Responding Task to measure self-esteem. *The Psychological Record*, *70*(2), 227–242. https://doi.org/10.1007/s40732-020-00392-4

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology*, *86*(2), 345–355. https://doi.org/10.1037/0022-3514.86.2.345

Espinosa, R., Valiente, C., Varese, F., & Bentall, R. P. (2018). Can we ameliorate psychotic symptoms by improving implicit self-esteem? A proof-of-concept experience sampling study of an evaluative classical conditioning intervention. *Journal of Nervous & Mental Disease*, *206*(9), 699–704. https://doi.org/10.1097/nmd.0000000000000858

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In *Advances in Experimental Social Psychology* (Vol. 23, pp. 75–109). Elsevier. https://doi.org/10.1016/s0065-2601(08)60318-4

Fleming, J. B., & Burns, M. N. (2017). Online evaluative conditioning did not alter internalized homonegativity or self-esteem in gay men. *Journal of Clinical Psychology*, *73*(9), 1013–1026. https://doi.org/10.1002/jclp.22388

Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Kiesel, A., Kunde, W., Mayr, S., Moeller, B., Möller, M., Pfister, R., & Philipp, A. (2020). Binding and Retrieval in Action Control (BRAC). *Trends in Cognitive Sciences*, *24*(5), 375–387. https://doi.org/10.1016/j.tics.2020.02.004

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition*, *25*(5), 687–717. https://doi.org/10.1521/soco.2007.25.5.687

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*(6), 1022–1038. https://doi.org/10.1037/0022-3514.79.6.1022

Grumm, M., Nestler, S., & Collani, G. von. (2009). Changing explicit and implicit attitudes: The case of self-esteem. *Journal of Experimental Social Psychology*, *45*(2), 327–335. https://doi.org/10.1016/j.jesp.2008.10.006

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/b108413

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, *60*(6), 895–910. https://doi.org/10.1037/0022-3514.60.6.895

Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioral Science*, *1*(1–2), 17–38. https://doi.org/10.1016/j.jcbs.2012.09.003

Hütter, M., & Sweldens, S. (2018). Dissociating Controllable and Uncontrollable Effects of Affective Stimuli on Attitudes and Consumption. *Journal of Consumer Research*, *45*(2), 320–349. https://doi.org/10.1093/jcr/ucx124

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology*, *108*(2), 187–218. https://doi.org/10.1037/a0038557

Jusepeitis, A., & Rothermund, K. (2022a). *Is there something you don't tell me? Evidence for the incremental predictive validity of indirectly assessed self-esteem using the Propositional Evaluation Paradigm* [Manuscript submitted for publication. Institut für Psychologie, Friedrich-Schiller-Universität Jena].

Jusepeitis, A., & Rothermund, K. (2022b). *Materials for "Primed to like yourself – Can self-evaluative cognition be changed by learning contingencies of self-evaluative statements and truth-values?"* https://osf.io/zx2aq/

Kitayama, S., & Karasawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin*, *23*(7), 736–742. https://doi.org/10.1177/0146167297237006

Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, *80*(4), 669–685. https://doi.org/10.1037/0022-3514.80.4.669

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*(2), 194–213. https://doi.org/10.1037/xge0000239

LeBel, E. P., & Gawronski, B. (2009). How to find what's in a name: Scrutinizing the optimality of five scoring algorithms for the name-letter task. *European Journal of Personality*, *23*(2), 85–106. https://doi.org/10.1002/per.705

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans

Maricuțoiu, L. P., Payne, B. K., & Iliescu, D. (2019). It works both ways. Enhancing explicit self-esteem using the self-reference task. *Journal of Experimental Social Psychology*, *80*, 8–16. https://doi.org/10.1016/j.jesp.2018.08.017

Moors, A., Spruyt, A., & De Houwer, J. (2010). In search of a measure that qualifies as implicit: Recommendations based on a decompositional view of automaticity. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. Guilford Press.

Müller, F., & Rothermund, K. (2019). The propositional evaluation paradigm: Indirect assessment of personal beliefs and attitudes. *Frontiers in Psychology*, *10*, 2385. https://doi.org/10.3389/fpsyg.2019.02385

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society, Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Niveau, N., New, B., & Beaudoin, M. (2021). Self-esteem interventions in adults – A systematic review and meta-analysis. *Journal of Research in Personality*, *94*, 104131. https://doi.org/10.1016/j.jrp.2021.104131

Nuttin, J. M. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology*, *15*(3), 353–361. https://doi.org/10.1002/ejsp.2420150309

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perkins, A. W., & Forehand, M. R. (2012). Implicit Self-Referencing: The Effect of Nonvolitional Self-Association on Brand and Product Attitude. *Journal of Consumer Research*, *39*(1), 142–156. https://doi.org/10.1086/662069

Power, P., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) as a measure of implicit relative preferences: A first study. *The Psychological Record*, *59*(4), 621–640. https://doi.org/10.1007/bf03395684

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, *9*(1), 45–48. https://doi.org/10.1111/1467-9280.00008

Remue, J., De Houwer, J., Barnes-Holmes, D., Vanderhasselt, M.-A., & De Raedt, R. (2013). Self-esteem revisited: Performance on the Implicit Relational Assessment Procedure as a measure of self- versus ideal self-related cognitions in dysphoria. *Cognition & Emotion*, *27*(8), 1441–1449. https://doi.org/10.1080/02699931.2013.786681

Remue, J., Hughes, S., De Houwer, J., & De Raedt, R. (2014). To be or want to be: Disentangling the role of actual versus ideal self in implicit self-esteem. *PloS One*, *9*(9), e108837. https://doi.org/10.1371/journal.pone.0108837

Replicability-Index. (2016). *A replicability analysis of "I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning"*. https://replicationindex.com/2016/12/03/a-replicability-analysis-ofi-like-myself-but-i-dont-know-why-enhancing-implicit-self-esteem-by-subliminal-evaluative-conditioning/

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press. https://doi.org/10.1515/9781400876136

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review*, *60*(1), 141. https://doi.org/10.2307/2096350

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). *Quarterly Journal of Experimental Psychology*, *62*(1), 84–98. https://doi.org/10.1080/17470210701822975

Rothermund, K., Wentura, D., & Houwer, J. de. (2005). Retrieval of incidental stimulus-response associations as a source of negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 482–495. https://doi.org/10.1037/0278-7393.31.3.482

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Strack, F., & Deutsch, R. (2015). The duality of everyday life: Dual-process and dual system models in social psychology. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, Volume 1: Attitudes and social cognition.* (pp. 891–927). American Psychological Association. https://doi.org/10.1037/14341-028

Van Dessel, P., De Houwer, J., & Smith, C. T. (2018). Relational information moderates approach-avoidance instruction effects on implicit evaluation. *Acta Psychologica*, *184*, 137–143. https://doi.org/10.1016/j.actpsy.2017.03.016

Versluis, A., Verkuil, B., & Brosschot, J. F. (2018). Converging evidence that subliminal evaluative conditioning does not affect self-esteem or cardiovascular activity. *Stress and Health*, *34*(2), 235–246. https://doi.org/10.1002/smi.2777

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wiswede, D., Koranyi, N., Müller, F., Langner, O., & Rothermund, K. (2013). Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience*, *8*(6), 647–653. https://doi.org/10.1093/scan/nss042

Wong, Q. J. J., Gregory, B., Gaston, J. E., Rapee, R. M., Wilson, J. K., & Abbott, M. J. (2017). Development and validation of the Core Beliefs Questionnaire in a sample of individuals with social anxiety disorder. *Journal of Affective Disorders*, *207*, 121–127. https://doi.org/10.1016/j.jad.2016.09.020

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/74788-primed-to-like-yourself-can-self-evaluative-cognition-be-changed-by-learning-contingencies-of-self-evaluative-statements-and-truth-values/attachment/157537.docx?auth_token=nra8KbXwMVYQ7Q91aFQV

Supplemental Material

Download: https://collabra.scholasticahq.com/article/74788-primed-to-like-yourself-can-self-evaluative-cognition-be-changed-by-learning-contingencies-of-self-evaluative-statements-and-truth-values/attachment/157540.docx?auth_token=nra8KbXwMVYQ7Q91aFQV