

Running Head: INSTRUCTIONS VS. EXPERIENCE QUAD MODELING

The Impact of Instruction- and Experience-Based Evaluative Learning on IAT Performance: A
Quad Model Perspective

Colin Tucker Smith*

University of Florida

Jimmy Calanchini*

University of Freiburg

University of California Riverside

Sean Hughes, Pieter Van Dessel, and Jan De Houwer

Ghent University

Authors' Note

*The first two authors contributed equally to the preparation of this manuscript. CTS, University of Florida, Gainesville, FL, USA; JC, University of Freiburg, Freiburg, Germany, University of California Riverside, Riverside, CA, USA; SH, PVD, JDH, Ghent University, Ghent, Belgium. This research was conducted with the support of a postdoctoral research fellowship from the Alexander von Humboldt Foundation to JC, a post-doctoral research fellowship from the Fellowship of the Research Foundation – Flanders (FWO) to PVD, and Grant BOF16/MET_V/002 of Ghent University to JDH. Correspondence concerning this article should be sent to colinsmith@ufl.edu.

Word count: 10139

Abstract

Learning procedures such as mere exposure, evaluative conditioning, and approach/avoidance training have been used to establish evaluative responses as measured by the Implicit Association Test (IAT). In this paper, we used the Quad model to disentangle the processes driving IAT responses instantiated by these evaluative learning procedures. Half of the participants experienced one of these three procedures whereas the other half only received instructions about how the procedure would work. Across three experiments (total $n = 4231$), we examined the extent to which instruction-based versus experience-based evaluative learning impacted Quad estimates of the Activation of evaluative information in IAT responses. Relative to a control condition, both instruction- and experience-based evaluative learning procedures influenced Activation. Moreover, and contrary to what prevailing models of implicit evaluations would predict, in no instance did experience-based procedures influence (positive or negative) Activation more strongly than instruction-based procedures. This was true for analyses which combined procedures and also when testing all three procedures individually. Implications for the processes that mediate evaluative learning effects and the conditions under which those processes operate are discussed.

Keywords: Attitudes, IAT, Quad modeling, Learning, Implicit evaluations

Abstract word count: 177

The Impact of Instruction- and Experience-Based Evaluative Learning on IAT Performance: A Quad Model Perspective

In the years since the introduction of the Implicit Association Test (IAT: Greenwald, McGhee, & Schwartz, 1998), a number of evaluative learning procedures have been found to influence responses on that task. Three such procedures are the repeated presentation of stimuli (as in mere exposure studies), the pairing of stimuli (as in evaluative conditioning studies), and contingencies between stimuli and approach or avoidance responses (as in approach-avoidance training studies). Mere exposure effects refer to a change in liking due to repeated stimulus presentations (e.g., viewing one consumer product frequently leads people to like it more than a second product viewed less frequently: Zajonc, 1986). Evaluative conditioning refers to a change in liking due to the pairing of stimuli (e.g., pairing a person with positive words leads that person to be liked more than another who was paired with negative words: Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). Finally, approach-avoidance effects refer to changes in liking due to a particular relation between stimuli and responses (e.g., repeatedly approaching one beverage and avoiding another leads people to like the former more than the latter: Wiers, Eberl, Rinck, Becker, & Lindenmeyer, 2011).

Recent research has demonstrated that each of these three evaluative learning procedures can lead to reliable differences in IAT *D*-scores (mere exposure: Van Dessel, Mertens, Smith, & De Houwer, in press; evaluative conditioning: Gregg, Seibt, & Banaji, 2006; Mitchell, Anderson, & Lovibond, 2003; Olson & Fazio, 2001; approach-avoidance: Van Dessel, De Houwer, Gast, Smith, & De Schryver, 2016). *D*-scores are based on differences in response latencies and errors in one set of IAT trials (e.g., in which insect and pleasant concepts share a response key) versus another (e.g., in which insect and unpleasant concepts share a response key), and are typically

interpreted as evidence of evaluative response strength and/or assumed to reflect the automatic activation of associations between representations in memory. However, subsequent research has revealed that IAT *D*-scores are not “process pure” indices of automatic associations or evaluative response strength but, instead, reflect the operation of multiple processes (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) To the extent that multiple processes contribute to responses on the IAT, any interpretation of evaluative learning effects on implicit evaluations based on IAT *D*-scores lacks precision because these effects may be mediated by relatively automatic processes, relatively controlled processes, or a combination of both. With this limitation in mind, we set out to disentangle the contribution of multiple processes to IAT responses resulting from evaluative learning procedures using the quadruple process model (Quad model: Conrey et al., 2005).

The Quad Model

The Quad model has been implemented as a multinomial model (see Batchelder & Riefer, 1999) designed to estimate the independent contributions of multiple processes from responses on implicit measures (see Sherman, 2006; Sherman et al., 2008). According to the model, performance on implicit measures such as the IAT is influenced by four qualitatively distinct processes. The *Activation* parameter¹ refers to the degree to which evaluative information² (e.g., information connecting a stimulus with positive valence) is activated by the

¹ In previous research, the Activation parameter (commonly abbreviated as AC) has been almost exclusively referred to as reflecting the “activation of associations” (e.g., Conrey et al., 2005). However, the Quad model does not speak to the representational nature of what is activated, only that something has been activated. Moreover, past work has never conclusively demonstrated that associations rather than other cognitive representations are captured by this parameter. Consequently, in the present manuscript we describe the Activation parameter using language that does not rely on *a priori* assumptions about the underlying representational structure. This more conservative approach corresponds more closely to how the Activation parameter is described by Sherman et al. (2008): “the activation of an impulsive response tendency” (p. 316).

² The Activation parameter does not necessarily have to reflect evaluative information. Instead, it can reflect relationships between concepts and attributes such as stereotypes (e.g., Calanchini, Sherman, Klauer, & Lai, 2014).

presence of a stimulus on a given IAT trial. The more accessible the information (e.g., the stronger the association is between “insects” and “unpleasant”, or the stronger the propositional belief that “insects are unpleasant”), the more likely that information is to be activated and produce a response tendency in a direction consistent with that information. The *Detection* parameter reflects the likelihood that the participant can discriminate between correct and incorrect responses, according to task requirements. Sometimes, activated evaluative information conflicts with the detected correct response. For example, on an IAT trial in which a picture of a disliked insect appears, and insects and pleasant stimuli share a response key (i.e., a so-called “incompatible” trial), the response tendency activated by the negatively evaluated insect (i.e., to press the button labeled “unpleasant”) conflicts with the detected correct response (i.e., to press the button labeled “pleasant”). In this case, the Quad model proposes that an *Overcoming Bias* process resolves the conflict. This parameter refers to an inhibitory process that prevents activated evaluative information from influencing behavior when this information conflicts with detected correct responses. Finally, the *Guessing* parameter reflects any other processes that guide responses in the absence of influence from the other three parameters. The construct validity of the Quad model has been extensively demonstrated in previous research (see Calanchini & Sherman, 2013; Sherman et al., 2008).

Applications of the Quad Model to the IAT

The Quad model has previously been used to examine the cognitive processes that underpin different interventions designed to influence implicit evaluations. For example, exposure to liked Black people and disliked White people (e.g., Dasgupta & Greenwald, 2001) leads to a significantly lower Black-unpleasant Activation estimate compared to a control condition, but does not influence other model parameters (Gonsalkorale, Allen, Sherman, &

Klauer, 2010). In contrast, depicting members of social groups in positive versus negative contexts (e.g., Wittenbrink et al., 2003) does not influence Activation but does influence Overcoming Bias (Allen, Sherman, & Klauer, 2010). A third example is perhaps most relevant to the present research: participants completed a version of an evaluative conditioning task in which they viewed pictures of Black and White people paired with positive and negative images (e.g., Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000). Rather than passively observing the pairings - as in typical studies using this task - in one condition, participants responded “yes” to stimulus pairings that are counter to typical prejudices (i.e., Black-pleasant; White-unpleasant), whereas, in another condition, participants responded “yes” to stimulus pairings that are consistent with typical prejudices (i.e., Black-unpleasant; White-pleasant). Thereafter, participants completed an IAT measuring evaluations of Black and White people. When the Quad model was applied to these data, participants in the counter-prejudicial training condition demonstrated lower Black-unpleasant and White-pleasant Activation estimates than did participants in either the prejudice-consistent training or a control condition, as well as an increase in Detection (Calanchini, Gonsalkorale, Sherman, & Klauer, 2013). In each of these examples, Quad modeling had important implications for theories about the mental processes underlying observed effects that would be overlooked by a reliance on only *D*-scores.

The Present Research

In the current work, we applied Quad modeling to IAT responses produced by mere exposure, evaluative conditioning, and approach-avoidance in order to examine the mental processes that mediate evaluative learning effects. Specifically, ‘experience’ in the form of repeated presentations of a single stimulus, pairing of stimuli, or the pairing of stimuli with responses has long been argued to lead to the installation of mental associations in memory

(Rydell & McConnell, 2006; Smith & DeCoster, 2001). However, a number of researchers have recently proposed an alternative idea: that mere exposure, evaluative conditioning, and approach-avoidance effects are instead mediated by processes that operate on the basis of propositions (De Houwer, 2009, 2018; Mitchell, De Houwer, & Lovibond, 2009). Unlike associations, which are simple links between concepts (e.g., “insects + unpleasant), propositions can include more information in that they can specify how concepts are related (e.g., “Some insects eat other insects”).

Relatedly, a number of studies now show that implicit evaluations can be learned not only via experience, but also via instructions. For instance, mere instructions about the repeated presence of a single stimulus, pairing of stimuli, or relationship between stimuli and actions can produce IAT *D*-scores that are similar to, if not stronger than, those produced via experience (De Houwer, 2006; Kurdi & Banaji, 2017; Van Dessel, Mertens, Smith, & De Houwer, 2017; Van Dessel, Gawronski, Smith, & De Houwer, 2017). Given that participants who only receive instructions describing evaluative learning paradigms never directly experience the procedures that are assumed to create associations in memory, instruction-based evaluative learning effects are assumed to be mediated by non-associative processes, such as propositional processes (but see Fazio, 2007; Gawronski & Bodenhausen, 2006).

Overview of Experiments

We examined the process-level effects of evaluative learning on IAT responses in two on-line (Experiments 1 and 3) and one laboratory-based study (Experiment 2). Participants were exposed to an experience- or instruction-based variation of one of three evaluative learning procedures (mere exposure, evaluative conditioning, approach-avoidance), followed by an IAT.³

³ The data from Experiment 1 are a subset of data from a larger study (Hughes, Van Dessel, Smith, & De Houwer, 2019) which included additional explicit measures and experimental conditions not relevant for the current purposes.

We then applied the Quad model to the IAT data, in order to examine the processes underpinning the effects of instruction- and experience-based procedures on IAT responses. If IAT effects are mediated by associative rather than propositional processes, and if associations are formed via experience and not (or more weakly) via instructions, then we should observe higher estimates of Activation in the experience versus instruction conditions. In contrast, if evaluative learning influences IAT responses through propositional processes, then we should observe similar Activation estimates in the experience and instruction conditions.

Experiment 1

Method

Participants. Participants were 1495 volunteers at the Project Implicit website (<https://implicit.harvard.edu>) randomly assigned to this study from a pool of approximately 10 studies. The mean age was 36.7 years, ($SD = 15.1$) and the majority (62.7%) were women. Participants were citizens of 67 different countries, with 60% being from the United States, 9% from the United Kingdom, 6% from Canada; all other countries <2%.

Materials

Stimuli. Two nonsense words (Vekte and Empeya) served as brand names during the evaluative learning and measurement phases. These brands were selected on the basis of a pre-rating study in which a separate set of Project Implicit participants ($n = 634$) rated twenty fictitious brand name and logo compounds on a scale ranging from 1 to 7 with 4 as a neutral point. Vekte ($M = 4.02$, $SD = 0.96$) and Empeya ($M = 4.03$, $SD = 1.04$) were rated most

The full design of that study is available on the Open Science Framework (see Supplement 1: osf.io/v7y4s). Participants included in the current analyses are those who were in conditions for which the sole manipulation was the evaluative learning procedures described here (e.g., participants were not included who also received counter-attitudinal information before completing the IAT).

neutrally. Whether Vekte or Empeya was paired with positive or negative stimuli in the following procedures was counterbalanced across participants.

Evaluative learning. Participants were randomly assigned to one of six evaluative learning conditions: an experienced or instructed version of mere exposure, evaluative conditioning, or approach-avoidance.⁴ See Appendix A for full text of all of the instruction conditions.

Mere Exposure: Experience. Participants were told that they would see images of different brands. They were then presented with the logo of one brand ten times and the logo of another brand once. Each image was presented on the screen for 500ms with a 1000ms inter-trial interval. Order of stimulus presentation was randomized.

Mere Exposure: Instructions. Participants were instructed that, later on in the study one brand would be presented frequently and another brand would be presented infrequently.

Evaluative Conditioning: Experience. Participants were first told that they would see images of brands that would be paired with a second image. The task proceeded automatically for 30 trials. In 15 trials, one brand was paired with one of five positive images; in the other 15 trials, the other brand was paired with one of five negative images. Each pair of images was on the screen for 2000ms with 750ms between trials.

Evaluative Conditioning: Instructions. Participants were instructed that, later on in the study, they would see positive and negative images paired with the two brand names.

Specifically, they were told that whenever one brand name was presented a positive image would also appear, and whenever another brand name was presented a negative image would appear.

⁴ In all instruction-based conditions, an attention check was used to ensure that participants could accurately recall the instructions, and progression to the next part of the experiment was contingent on a correct response. Subsequent data were included from all participants; those who initially answered this item incorrectly were asked to respond again until a correct response was made.

Approach-Avoid: Experience. Participants were told to use the up and down keys of the computer keyboard to control a stick figure depicted onscreen. They were also told that brand logos would appear in the center of the screen with green or blue borders around them, and they were instructed to make the stick figure approach the brand with a green border and avoid the brand with a blue border. Throughout the task, one brand always had a green border around it and the other brand always had a blue border around it, such that participants consistently approached the former brand and avoided the latter. At the beginning of each trial, the stick figure randomly appeared either above or below the logo. In this way, half of the “approach” movements required participants to use the up arrow and the other half the down arrow. When the appropriate arrow was pressed, the stick figure moved toward or away from the brand; no movement occurred if the incorrect arrow was pressed. Participants began the approach-avoid task with four practice trials, one each for approaching or avoiding the brands with an upward and downward motion. They then continued the task for an additional 56 trials.

Approach-Avoid: Instructions. Participants were instructed that, later on in the study, they would see two brand names that they would need to approach and avoid. Specifically, whenever they would see one brand they would need to approach it and whenever they would see the other brand they would need to avoid it.

Implicit Association Test (IAT). Participants completed an IAT (Greenwald et al., 1998) designed to assess automatic evaluative responding towards the two brands. In this task, participants assigned stimuli to the following categories: ‘Vekte’, ‘Empeya’, ‘Good’, and ‘Bad’. Stimuli for the Empeya and Vekte categories consisted of four different versions of the brand name and logo (see Appendix A). Evaluative stimuli were positive (e.g., Happy) and negative (e.g., Sad) adjectives. Stimuli from one of these four categories were presented sequentially on

the participant's computer screen against a white background. Empeya and Vekte were presented in green and the evaluative categories were presented in blue; labels appeared in the upper-left and upper-right of the screen. Participants used the "E" key and the "I" key to sort stimuli to the left and right, respectively, and were instructed to sort stimuli quickly while making as few errors as possible.

The IAT was constructed following the recommendations of Nosek and colleagues (Nosek, Greenwald, & Banaji, 2005). Participants began by sorting Vekte and Empeya stimuli (20 trials) and then sorted evaluative stimuli (20 trials). Next, participants completed 60 trials in which stimuli related to Vekte and positive shared a single response key and stimuli related to Empeya and negative shared a single response key. Participants then practiced sorting stimuli related to Vekte and Empeya with the reversed response mapping for 20 trials before completing a second set of 60 trials in which Vekte stimuli shared a response key with negative and Empeya stimuli shared a response key with positive. If the participant made an error, a red "X" appeared on the screen; participants had to make a correct response in order to continue. Whether participants completed the task with Vekte first paired with positive or negative was counterbalanced.

Procedure

After providing informed consent, participants were told that they would encounter two brands (Empeya and Vekte) that might be introduced into supermarkets in the USA and elsewhere around the world. They completed one of the six evaluative learning procedures and then the IAT. Finally, they received feedback about their IAT score as well as information about the study aims.

Results

Data Preparation

In keeping with standard treatment of data collected at Project Implicit (Smith, De Houwer, & Nosek, 2012), IAT data were removed for participants who (a) had error rates above 30% when considering all IAT blocks or above 40% for any one of the critical IAT test blocks, and/or (b) responded faster than 400ms on more than 10% of the IAT trials (49 participants; 3.3%). This left us with 1446 participants for analyses.

Parameter Estimation

The structure of the Quad model is depicted as a processing tree in Figure 1. In order to estimate the parameters specified in the Quad model, we employed the Bayesian approach proposed by Klauer (2006, 2010) to fit a multilevel extension of the model that treats participants and items as random factors for each model parameter (Judd, Westfall, & Kenny, 2012), as implemented by the TreeBUGS R package (Heck, Arnold, & Arnold, 2018). In this Bayesian approach, the T_1 statistic summarizes how well the model accounts for the pattern of observed assignment frequencies aggregated across participants within each condition (Klauer, 2010), corresponding to the goodness-of-fit statistic chi-square used in traditional modeling approaches (Batchelder & Riefer, 1999). The T_2 statistic summarizes how well the model accounts for the variances and correlations of these frequencies computed across participants, which thereby quantifies how well the model accounts for individual differences between participants in the individual response frequencies (Klauer, 2010).

For each participant, we calculated two Activation parameter estimates, and one estimate each for Detection, Overcoming Bias, and Guessing.⁵ One Activation parameter reflected the extent to which positive information is activated in response to the brand that was the target of positive attitude induction, and the other Activation parameter reflected the extent to which negative information is activated in response to the brand that was the target of negative attitude induction. The Guessing parameter was coded so that higher scores represented a bias toward responding with the “good” key. Across all conditions, participants made 5.29% errors. At the individual level, the median p -value for T_1 was $p = .464$. At the group level, the observed versus predicted values for T_1 were 0.1441 and 0.0063, respectively, $p < .001$, and the observed versus predicted values for T_2 were 4.4610 and 0.4364, respectively, $p < .001$. The non-significant p value for the individual-level statistic suggests that the Quad model provides good fit to these data, but the significant p values for the group-level statistics suggest that the observed outcomes differed significantly from the predicted outcomes. However, the individual-level test is arguably underpowered to detect misfit, and given our large sample, the group-level tests are highly powered to detect even small amounts of misfit. There is no agreed-upon method to quantify model fit for the analyses used here that controls for sample size. Consequently, we include graphs of the observed versus predicted frequencies and covariances for all experiments in Appendix C. Visual inspection of these graphs indicates that differences between observed and predicted outcomes are minimal, which suggest that the Quad model provides good fit to these data. We report all parameter estimates for each experimental condition in Table 1.

⁵ Though IAT research often focuses on D -scores (Greenwald, Nosek, & Banaji, 2003), the current manuscript is focused on analysis of Quad model parameters. As such, for the sake of space and clarity, we do not include D -scores in the main text, but report them for all three experiments in Appendix B.

Hypothesis Testing

We conducted a series of planned contrasts for each Quad parameter in order to compare the process-level effects of instruction- versus experience-based evaluative learning. We did so by subtracting the distributions for all posterior samples of a given parameter for the three experience-based procedures from the distributions for all posterior samples of the same parameter for the three instruction-based procedures. In the resulting distribution of credible mean differences, the effects of experience- versus instruction-based learning can be interpreted as being credibly different from one another if the 95% Highest Density Interval (HDI, which corresponds to a Confidence Interval in traditional analyses) does not contain zero, with positive estimates reflecting a stronger effect of instruction-based evaluative learning and negative estimates reflecting a stronger effect of experience-based evaluative learning. We summarize below the results of these contrasts, and report in Table 2 the results of all contrasts between the instruction- versus experience-based version of each evaluative learning paradigm. The full set of all possible contrasts is available at the project page (Supplement 2: osf.io/v7y4s).

Comparing instructions versus experience across learning procedures. Collapsing across the three evaluative learning procedures, instruction-based learning had a stronger effect on the Activation of positive evaluative information than did experience-based learning, 0.0039, 95% HDI [0.0008, 0.0085]. None of the other parameters differed between instruction- and experience-based conditions.

Comparing instructions versus experience within learning procedures. The above analyses examined the effects of instruction- versus experience-based evaluative learning paradigms on Quad parameters, collapsed across learning paradigm. Though this approach increases statistical power by focusing on the key variable of interest in the present research (i.e.,

the distinction between instruction- versus experience-based evaluative learning), this approach is limited in that it may obscure or overlook differences within individual learning paradigms. However, there is little evidence of this - among 15 possible within-procedure comparisons, only three were significant. Within approach-avoidance procedures, instructions had a stronger effect on the Activation of positive evaluative information more than did experience, 0.0057, 95% HDI [0.0004, 0.0138]. Within the evaluative conditioning procedures, instructions had a stronger effect on Detection than did experience, 0.0237, 95% HDI [0.0077, 0.0400], whereas experience had a stronger effect on Overcoming Bias parameter than did instructions, -0.8928, 95% HDI [-1.000, -0.1082]. See Supplement 2 (osf.io/v7y4s) for tests of all possible comparisons between Quad parameters in this and subsequent studies.

Discussion

Experiment 1 revealed stronger effects of instruction- versus experience-based evaluative learning on the Activation of positive evaluative information. However, instruction- and experience-based learning had equivalent effects on Detection, Overcoming Bias, Guessing, and Activation of negative evaluative information. These results suggest that the process-level effects of instruction-based learning are largely - but not entirely - equivalent to those of experience-based learning. Though the limitations of interpreting null results are well-known, we nevertheless believe these findings are meaningful because of their implications for the dominant theoretical perspective that evaluative learning effects on implicit evaluations are associative in nature. The evaluative learning procedures used in Experiment 1 have long been presumed to operate via association formation processes (see Hofmann et al., 2010; Van Dessel, Hughes, & De Houwer, 2018; Zajonc, Markus, & Wilson, 1974), and the IAT is presumed to be a measure of association strengths – as indicated by its very name – and the Activation parameter of the

Quad model is generally described as an index of association activation (e.g., Conrey et al., 2005). From this perspective, the pattern of results reported in Experiment 1 is certainly noteworthy; mere instructions resulted in automatically activated response tendencies to a similar or greater extent than did experience, and in no case did direct experience with an evaluative learning procedure allow for a stronger activation of evaluative information as compared to instructions about the same procedure.

Experiments 2-3

The purpose of Experiment 1 was to examine the relative influences of instruction- versus experience-based learning on implicit evaluations. However, Experiment 1 did not include a control condition which limits the conclusions we can draw from these data. One interpretation of the results of Experiment 1 is that both instructions and experience had equally-strong effects on (most of) the Quad parameters, but an alternate interpretation is that neither instructions nor experience had any effect on (most of) the Quad parameters. Therefore, in Experiments 2 and 3, we included a control condition in which an additional group of participants completed an IAT measuring implicit evaluations of Empeya and Vekte in the absence of any prior evaluative learning. Comparing the effects of experience- and instruction-based evaluative learning against a condition in which no evaluative learning took place will help to determine whether both forms of learning had equivalent effects or no effects. Given that the designs of Experiments 2 and 3 are largely identical, we report them together for the sake of convenience.

Method

Participants. Participants in Experiment 2 were 486 students at the University of Florida who completed the study in partial fulfillment of a course requirement. Mean age was 19.3 years

($SD = 2.6$); 70.8% of the participants were women. We ran the study until we had 50 participants per cell. Participants in Experiment 3 were 2250 volunteers at the Project Implicit website. The mean age was 38.1 years, ($SD=14.6$) and a slight majority (53.6%) were women. Participants were citizens of 67 different countries, with 62% being from the United States, 8% from the United Kingdom, 7% from Canada; all other countries <3%. Participants who had participated in Experiment 1 (or any other Project Implicit studies using Empeya/Vekte stimuli) were not eligible to participate in the current study.

Materials and Procedure

Experiment 2 differed from Experiment 1 in two ways. The first difference was that participants in a control condition were asked to imagine that two new brands (Empeya and Vekte) would be introduced in the United States and elsewhere around the world and that they would have to complete a speeded categorization task related to those brands. They then proceeded directly to the IAT. The second difference was that stimulus identity was held constant such that, in the attitude induction procedures, Empeya was always subject to positive and Vekte to negative attitude induction. Experiment 3 also included a control condition and held stimulus identity constant. Additionally, whereas task order⁶ was counterbalanced in Experiments 1 and 2, it was fixed in Experiment 3 such that the IAT was always completed directly after the learning procedure.

⁶ In all three experiments, participants answered five questions (e.g., explicit attitude, confidence) which are not relevant for the current purposes and which we did not analyze. The full text of these items is available at the OSF project page (see Supplement 3: osf.io/v7y4s).

Results

Parameter Estimation

Quad model parameters were estimated as in Experiment 1. Across all conditions, participants in Experiment 2 made 6.76% errors and participants in Experiment 3 made 4.92% errors. At the individual level, the median p -value for T_1 was $p = .418$ in Experiment 2, and the median p -value for T_1 was $p = .454$ in Experiment 3. At the group level, the observed versus predicted values for T_1 were 0.3223 and 0.0192, $p < .001$, and the observed versus predicted values for T_2 were 6.2688 and 1.4166 in Experiment 2, and the observed versus predicted values for T_1 were 0.2540 and 0.0040, $p < .001$, and the observed versus predicted values for T_2 were 8.2376 and 0.2798 in Experiment 3. Visual inspection of graphs of the observed versus predicted frequencies and covariances (see Appendix C) suggest that the Quad model provides good fit to these data. We report all parameter estimates for each experimental condition in Table 1. Sample size, analyses, and hypotheses for Experiment 3 were pre-registered (osf.io/v7y4s).

Hypothesis Testing

In addition to the analyses reported in Experiment 1, in which we compared the effects of instruction- versus experience-based learning, in Experiments 2 and 3 we also conducted a series of planned contrasts to examine the effects of instruction- and experience-based learning against the control condition. To do so, we subtracted the distributions for all posterior samples of a given parameter for the control condition from the distributions for all posterior samples of the same parameter for the three instruction-based procedures and, separately, subtracted the distributions for all posterior samples of a given parameter for the control condition from the distributions for all posterior samples of the same parameter for the three experience-based procedures. In the resulting distribution of credible mean differences, the effects of experience-

and instruction-based learning can be interpreted as being credibly different from control if the 95% HDI does not contain zero, with positive estimates reflecting a stronger effect of evaluative learning relative to control. We summarize below the results of these contrasts, and report in Table 2 the results of all contrasts between the instruction- versus experience-based version of each evaluative learning paradigm. The full set of all possible contrasts is also available (osf.io/v7y4s).

Comparing instructions versus experience across learning paradigms. In Experiment 2, positive Activation estimates were larger in both the instruction-based, 0.0067, 95% HDI [0.0005, 0.0198], and experience-based evaluative learning conditions, 0.0070, 95% HDI [0.0006, 0.0208] relative to the control condition. Detection estimates in the instruction-based evaluative learning conditions were also larger than in the control condition, .0453, 95% HDI [.0152, .0778], whereas Detection estimates in the experience-based evaluative learning conditions were not different from the control condition, .0284, 95% HDI [-.0034, .0616]. The negative Activation, Guessing and Overcoming Bias parameters did not differ from control in either the instruction- or experience-based conditions. No Quad parameters differed between the instruction- and experience-based conditions.

In Experiment 3, positive Activation estimates were larger in both the instruction-based, 0.0032, 95% HDI [0.0009, 0.0068], and experience-based evaluative learning conditions, 0.0026, 95% HDI [0.0005, 0.0057], relative to the control condition. Negative Activation estimates were also larger in both the instruction-based, 0.0048, 95% HDI [0.0017, 0.0101] and experience-based evaluative learning conditions, 0.0028, 95% HDI [0.0008, 0.0062], relative to the control condition. Detection estimates in the instruction-based evaluative learning conditions were larger than in the control condition, 0.0106, 95% HDI [0.0005, 0.0214], whereas Detection estimates in

the experience-based evaluative learning procedures were not different from the control condition, 0.0059, 95% HDI [-0.0046, 0.0168]. The Guessing and Overcoming Bias parameters did not differ from control in either the instruction- or experience-based conditions. No Quad parameters differed between the instruction and experience-based conditions.

Comparing parameters within learning procedure. In Experiment 2, instructed mere exposure increased Detection more than did experienced mere exposure, 0.0416, 95% HDI [0.0075, 0.0774]. No other Quad parameters differed within procedures between instruction- and experience-based conditions. In Experiment 3, instructed evaluative conditioning increased Detection more than did experienced evaluative conditioning, 0.0132, 95% HDI [0.0002, 0.0259] and experienced evaluative conditioning influenced Guessing in the direction of negative responses more than did instructed evaluative conditioning, -0.0807, 95% HDI [-0.1209, -0.0404]. No other Quad parameters differed by condition.

Discussion

Experiments 2 and 3 provide further evidence that instructed and experienced evaluative learning procedures have highly similar effects on the Quad parameters. Both types of procedures resulted in stronger Activation parameters relative to control. However, instructed versus experienced learning are not identical: in both experiments, instruction- but not experience-based procedures increased Detection relative to control, though the effects of the two types of procedures on Detection were not different from one another. Additionally, Experiments 2 and 3 rule out the alternative explanation for the findings of Experiment 1: instruction- and experience-based evaluative learning procedures largely have equivalent, rather than null, effects on Quad parameters. Put simply, the processes driving IAT responses do not

rely solely on experience with a learning procedure, but are also affected by simple instructions about the procedure.

General Discussion

Previous research has demonstrated that both experience with and instructions about evaluative learning procedures, such as repeated stimulus presentations (as in mere exposure studies), stimulus pairings (as in evaluative conditioning research), and contingencies between stimuli and responses (as in approach-avoidance training) can produce evaluative responses as captured by tasks such as the IAT. Building upon these findings, in the present research we employed Quad modeling to investigate the processes that mediate IAT responses established by these procedures. Whereas many argue that the effects produced by these procedures, including those reflected in the IAT, are mediated by the automatic activation of associations in memory (Hofmann et al., 2010; Phills, Kawakami, Tabi, Nadolny, & Inzlicht, M., 2011; Zajonc et al., 1974), others have increasingly argued that those same effects are mediated by propositions (De Houwer, 2018; Van Dessel et al., 2019; Van Dessel, Mertens et al., 2017). In support of propositional accounts, previous research has shown that instructions about stimulus presentations, pairings, or stimulus-behavior relations can influence performance on the IAT even when those procedures are never actually administered (e.g., De Houwer, 2006; Kurdi & Banaji, 2017, Van Dessel et al., 2015, Van Dessel, Mertens et al., 2017). The present research extends these findings, examining the processes underlying instruction- versus experience-based evaluative learning effect on implicit evaluations using Quad modeling. If evaluative learning effects on the IAT are mediated by associations, and if associations are formed only through direct experience, then experience-based procedures should produce stronger Activation estimates than instruction-based procedures. However, if evaluative learning influences implicit

evaluations via propositional processes, then both instruction- and experience-based procedures should produce similar Activation estimates. The three experiments reported herein provide evidence in support of the latter perspective.

Comparing Instructions to Experience

In the present research, instruction- and experience-based evaluative learning procedures consistently influenced the Activation parameters of the Quad model. Both instructions and experience influenced the Activation of positive information to a similar extent in Experiments 2 and 3, and also influenced the Activation of negative information to a similar extent in Experiment 3. However, instructions had stronger effects on the positive Activation parameter than did experience in Experiment 1. Instructions, but not experience, also influenced the Detection parameter in Experiments 2 and 3. In many ways, these results are consistent with previous theory and research. For example, the purpose of evaluative learning is to create evaluations towards novel targets, and the Activation parameter is assumed to reflect activation of learned evaluative information related to the presented stimuli, so it makes sense that the most consistent effects of evaluative learning procedures were on the Activation parameter. Additionally, Activation estimates in the control conditions of Experiments 2 and 3 were not different from zero, which is what should be expected of neutral, novel targets, and suggests that the control condition functioned as an appropriate baseline. Taken together, these findings reveal a degree of consistency in the influence of instruction- and experience-based evaluative learning procedures on Quad parameters, but also a degree of inconsistency. The inconsistencies between the two types of learning may reveal mechanisms underlying both approaches. For example, previous research indicates that the Detection parameter is reduced when cognitive capacity is constrained (Conrey et al., 2005). In light of this, one possible explanation for the Detection

effects observed in the present research is that both instructions and experience increase Detection, but the act of experiencing an evaluative learning procedure also depletes the cognitive resources that Detection depends on to a greater degree than does simply reading instructions. Consequently, experience-based evaluative learning procedures may have countervailing effects on the kind of accuracy-oriented cognitive process reflected in the Detection parameter that instruction-based learning procedures do not seem to have. That said, the effects of instruction- versus experience-based learning procedures on Detection did not differ when they were compared directly with one another; this difference only appeared when each form of learning was compared to control. Nevertheless, this example highlights ways in which the current approach opens new avenues of investigation.

Comparing Instructions to Experience within Learning Procedures

Though the focus of the present research was on the effects of instructed versus experienced evaluative learning on the processes underlying implicit evaluations, the interested reader may reasonably wonder whether these effects differed across the three evaluative learning paradigms employed in the present research. The short answer is no. Out of twenty possible comparisons across all three experiments, only five comparisons revealed significant differences between the instructed and experienced version of a specific paradigm. Importantly, these differences were not concentrated within a certain paradigm or Quad parameter. Moreover, none of these significant differences replicated across all three experiments, and only one (i.e., stronger effects of instructed than experienced Evaluative Conditioning on Detection) replicated across two experiments. Taken together, paradigm-level effects were inconsistent and unreliable, especially in comparison to the main analyses collapsed across instruction- and experience-based conditions, which showed reliable effects of both forms of learning. In turn, these within-

paradigm analyses indicate that the effects of instructed and experienced evaluative learning are relatively consistent across learning paradigms.

Relevance for Cognitive Theories of Evaluation

The consistency of these findings also reveals the limitations of extant theory and methods. For example, participants who experienced evaluative learning procedures undeniably engaged in a qualitatively different procedure from those who were merely instructed about those procedures, and those procedural differences can be expected to result in each form of learning to be represented differently in memory. However, as the present research indicates, the Quad model cannot distinguish between evaluations formed through instructions versus experience. (Importantly, the *D*-score suffers the same limitation: De Houwer, 2006; Kurdi & Banaji, 2017, Van Dessel et al., 2015, Van Dessel, Mertens et al., 2017). These findings call into question the utility of analytic approaches such as Quad modeling that are grounded in dual-process models of cognition that emphasize the importance of associative processes.

The field of social cognition has been dominated by dual-process perspectives arguing that explicit evaluations are mediated by propositional (i.e., belief-based) processes, but that implicit evaluations are mediated by the automatic activation of associative links in memory (between the representation of a target stimulus and a valenced representation; e.g., Strack & Deutsch, 2004; Wilson, Lindsey, & Schooler, 2000). According to such models, to the extent that Activation parameters in the Quad model reflect evaluative information, then repeated presentation of a single stimulus, pairing of stimuli, or responses to stimuli should have led to stronger Activation parameters than only instructions about those relationships. Of note, some associative models (e.g., Gawronski & Bodenhausen, 2011; see also Fazio, 2007) assert that the associations presumed by the model to underlie implicit evaluations can be impacted indirectly

via the type of propositional reasoning likely to occur during the instruction-based procedures (see also Van Dessel, Gawronski et al., 2017). However, given that these models argue that the experience-based procedures would have direct effects on the associations, both types of procedures should influence the Activation parameter (for example), but experience-based procedures should result in larger parameter estimates. Yet this was not the case.

Alternately, a perhaps more parsimonious explanation for the present findings is that the effects of both types of procedures resulted in associative representations, but through different mechanisms. For example, whereas experience-based procedures could create associations in memory via the direct co-activation of representations in memory (e.g., Hebb, 1949), instruction-based procedures might produce associations via an indirect co-activation of representations in memory when participants mentally simulate the events described in the instructions. Although such an account is perhaps technically feasible, it would effectively undermine the popular idea that association formation is a low-level process specifically directed at capturing regularities in the actual environment (e.g., McConnell & Rydell, 2014). Instead, association formation would become a second way of encoding the content of higher-order cognitive processes, next to propositional representations that can capture the full relational complexity of higher-order cognition. One may wonder what the benefits are of having such a second memory system, especially when taking into account that propositional representations can have automatic effects on behavior (De Houwer, 2014). Nevertheless, as the present research indicates, extant methods, analyses, and theory are ill-suited to draw definite conclusions about the representational nature of the processes that mediate implicit evaluations, including implicit evaluations that are based solely on instructions about learning procedures.

Evaluative Learning and Automaticity

The present research adds to our understanding of the automaticity of evaluative learning effects in the context of the IAT. A cognitive process can be considered automatic if it operates quickly, outside of conscious awareness, is minimally dependent on cognitive resources, does not require deliberate intent, or cannot be stopped once started (Bargh, 1994; Moors & De Houwer, 2006). Importantly, a wealth of work indicates that the various conditions of automaticity do not perfectly co-vary: if a cognitive process is fast, there is no guarantee that it is also efficient or uncontrollable (e.g., Sherman, Krieglmeier, & Calanchini, 2014).

Previous research has examined the conditions under which the mental processes specified by the Quad model operate in an attempt to reveal the extent to which each process possesses features of automaticity. For instance, Detection and Overcoming Bias are reduced when a response deadline is implemented on an IAT, but Activation and the collection of processes reflected in the Guessing parameter⁷ are not influenced by such time constraints, suggesting that the former are relatively slow and perhaps resource-dependent processes and the latter processes (or collection of processes, in the case of Guessing) operate in a relatively fast and efficient manner (Conrey et al., 2005). Overcoming Bias is also lower among older people than younger people (Gonsalkorale, Sherman, & Klauer, 2009; 2014), whereas Activation, Detection, and Guessing do not vary by age. Age-related deficits in higher-order cognitive functioning are well-documented (e.g., Connelly, Hasher, & Zacks, 1991; Hasher & Zacks, 1988), especially in the context of inhibition (e.g., Kane, Hasher, Stoltzfus, Zacks, & Connelly, 1994). To the extent that Overcoming Bias is an inhibitory process related to such higher-order

⁷ Guessing is operationalized in the Quad model as the tendency to select ‘pleasant’ versus ‘unpleasant’ responses. However, Guessing should not be interpreted as a specific cognitive process but, rather, as reflecting any processes that influence responses in addition to Activation, Detection, and Overcoming Bias.

cognitive functions, Gonsalkorale and colleagues' (2009; 2014) findings thus suggest that Overcoming Bias is a relatively resource-dependent process, but that Activation, Detection, and the collection of processes reflected in the Guessing parameter are relatively more efficient. Finally, Activation, Detection, and Overcoming Bias are all influenced by implementation intentions to respond on the IAT in an egalitarian manner but Guessing is not (Calanchini, Lai, & Klauer, 2019), suggesting that the former processes are susceptible to deliberate intent, whereas the collection of processes reflected in the Guessing parameter is not.

The present research leverages previous research on the operating conditions of the Quad parameters to reveal a relatively nuanced picture of the automaticity of evaluative learning effects in the IAT. Both instructed and experienced evaluative learning procedures influenced Activation, which possesses some features of automaticity but not others: it is relatively fast and efficient, but susceptible to deliberate intent. Additionally, neither experience- or instruction-based evaluative learning influenced Overcoming Bias, which possesses features of control: it is relatively slow, resource-dependent, and susceptible to deliberate intent.⁸ Instruction-based procedures influenced Detection, which possess some features of control but not others: it is relatively slow and susceptible to deliberate intent, but is efficient. Finally, neither form of learning influenced the collection of processes reflected in the Guessing parameter, which operates under conditions associated with automaticity: it is relatively fast, efficient, and not susceptible to deliberate intent. This type of specific profile of effects would be overlooked by analytic approaches that do not account for the influence of multiple cognitive processes (e.g.,

⁸ The Quad model is structured such that the Overcoming Bias parameter influences responses to target (i.e., Empeya, Vetke) but not attribute (i.e., pleasant, unpleasant) stimuli, and only in the incompatible blocks of the IAT. In contrast, Activation, Detection, and Guessing influence responses to both type of stimuli in both blocks of the IAT. Consequently, the Overcoming Bias parameter is estimated from fewer trials and, thus, less reliably than the other three parameters. As such, the present research should not be interpreted as strong evidence that evaluative learning has absolutely no effect on Overcoming Bias but, instead, that any effects are too small to be reliably detected given the present samples.

the IAT *D*-score), or by theoretical perspectives that conceptualize automaticity and control dichotomously rather than as consisting of multiple facets.

Not only does the present research speak to the automaticity versus controllability of the processes underlying evaluative learning effects in implicit social cognition, but it also highlights the relative contributions of qualitatively different types of processes. The Quad parameters are estimated on a likelihood scale, such that 0 reflects no contribution and 1 reflects consistent contribution.⁹ As Table 1 indicates, Detection estimates were generally very high (median $D=0.9245$ across experiments and conditions), Overcoming Bias estimates were moderate (median $OB=0.4356$ across experiments and conditions), and Activation estimates were generally very low (median $AC=0.0045$ across experiments and conditions). Thus, on any given IAT trial in the experiments reported here, Detection is highly likely to have influenced responses, Overcoming Bias is moderately likely to have influenced responses¹⁰, and Activation of evaluations is the least likely to have influenced responses. To the extent that Detection and Overcoming Bias are control-oriented processes (in that they operate to constrain the expression of activated evaluations), this pattern of results challenges assumptions that responses on the IAT minimize the influence of control-oriented processes. Moreover, the magnitude of Evaluation estimates calls into question, perhaps ironically, the relative contributions of evaluations to evaluative learning effects in implicit cognition. To be clear, the impact of evaluative learning procedures on Activation estimates was consistently observable. It was simply small and,

⁹ This is not true of the Guessing parameter, which is anchored at .5 rather than 0. Guessing estimates $> .5$ reflect a tendency to respond with the “good” key, estimates $< .5$ reflect a tendency to respond with the “bad” key, and estimates $= .5$ reflect no evaluative response bias. Consequently, Guessing parameters cannot be interpreted in the same way as the other Quad parameters.

¹⁰ The Activation, Detection, and Guessing parameters are specified in the Quad model to influence responses to both target and attribute stimuli in both blocks of the IAT. In contrast, the Overcoming Bias parameter is specified to only influence responses to target stimuli and only in the incompatible blocks of the IAT.

perhaps, smaller than one would have expected. Future research should continue to investigate the role of evaluations in other evaluative learning paradigms and using other implicit measures.

Alternative Process-Level Perspectives

Although the Quad model is well-validated, it is not the only analytic method available to gain process-level insight into implicit cognition. For example, Payne's (2001) process dissociation (PD) model has also been applied to a wide variety of implicit measures. However, it is unclear whether the PD model would reveal anything in the present experiments that the Quad model overlooked. The standard version of the PD model includes two parameters, one representing Automatic processes and another representing Controlled processes, which conceptually map onto the Activation and Detection parameters of the Quad model (Payne & Bishara, 2009). A variant of the PD model includes a Guessing parameter (e.g., Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012), which maps directly onto the Guessing parameter in the Quad model. The key difference between the Quad and PD models is that the former includes the Overcoming Bias parameter. Given that we found no evidence that evaluative learning influences Overcoming Bias in the present research, or that instruction- versus experience-based evaluative learning differentially affect Overcoming Bias, the PD model is unlikely to have changed the pattern of results reported here.

The ReAL model (Meissner & Rothermund, 2013) also accounts for the contributions of multiple processes to IAT responses. However, it has only been validated on a modified IAT procedure: in contrast to the standard IAT procedure, which is what we used in the present research, Meissner and Rothermund's (2013) modified IAT procedure includes an extended block and trial structure, task-switch as well as task-repeat trials, and a response deadline. Given

that the ReAL model has not been validated on a standard IAT, it would not be appropriate for use in the present research.

Hütter and De Houwer (2017) also applied a process dissociation model to effects of evaluative learning, and more specifically evaluative conditioning instructions. Their study came to a similar conclusion as the current study: that instruction-based learning procedures can also influence more automatic parameters of evaluative learning. That study, however, modeled memory influences on evaluative ratings rather than on IAT responses.

Conclusion

The current work is the most comprehensive investigation to date directly comparing instructed and experienced evaluative learning procedures on implicit measures of evaluations. We employ a single design to compare the effectiveness of three different procedures either in instructed or experienced forms. Instructed learning led to activation of positive and negative information during the IAT, and these effects were as strong as, or stronger than, experienced learning. In our view, these findings pose a serious challenge for existing dual-process theories of evaluation which emphasize the importance of associations to implicit evaluations. The present research suggests that instruction- and experience-based evaluative learning procedures largely have similar effects at a process level.

References

- Allen, T. J., Sherman, J. W., & Klauer, K. C. (2010). Social context and the self-regulation of implicit bias. *Group Processes & Intergroup Relations, 13*, 137-149.
- Bargh, John A. "The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition." *Handbook of social cognition 1* (1994): 1-40.
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review, 6*, 57-86.
- Calanchini, J., Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2013). Counter-prejudicial training reduces activation of biased associations and enhances response monitoring. *European Journal of Social Psychology, 43*, 321-325.
- Calanchini, J., Lai, C. K., & Klauer, K. C. (2019). A process-level meta-analysis of implicit bias-reduction interventions. *Unpublished manuscript*.
- Calanchini, J., & Sherman, J. W. (2013). Implicit attitudes reflect associative, non-associative, and non-attitudinal processes. *Social and Personality Psychology Compass, 7*, 654-667.
- Calanchini, J., Sherman, J. W., Klauer, K. C., & Lai, C. K. (2014). Attitudinal and non-attitudinal components of IAT performance. *Personality and Social Psychology Bulletin, 40*, 1285-1296.
- Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology, 89*, 469-487.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800-814.

- De Houwer, J. (2006). Using the Implicit Association Test does not rule out an impact of conscious propositional knowledge on evaluative conditioning. *Learning and Motivation, 37*, 176-187.
- De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*, 1–20.
- De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass, 8*, 342-353.
- De Houwer, J. (2018). Propositional Models of Evaluative Conditioning. *Social Psychological Bulletin, 13*(3): e28046.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731.
- Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. In *Advances in experimental social psychology* (Vol. 44, pp. 59-127). Academic Press.
- Gonsalkorale, K., Allen, T. J., Sherman, J. W., & Klauer, K. C. (2010). Mechanisms of group membership and exemplar exposure effects on implicit attitudes. *Social Psychology, 41*, 158-168.
- Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology, 45*, 410-414.

- Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2014). Measures of implicit attitudes may conceal differences in implicit associations: The case of antiaging bias. *Social Psychological and Personality Science*, *5*, 271-278.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464-1480.
- 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*(2), 197-216.
- Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*, 1-20.
- Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*, 264-284.
- Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*, 390-421.
- Hughes, S., Van Dessel, P., Smith, C. T., & De Houwer, J. (2019). A comparative analysis of different evaluative learning procedures. Manuscript in preparation.
- Hütter, M., & De Houwer, J. (2017). Examining the contributions of memory-dependent and memory-independent components to evaluative conditioning via instructions. *Journal of Experimental Social Psychology*, *71*, 49-58.

- Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General*, *141*, 539-557.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54-69.
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*, 871-888.
- Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *71*, 7-31.
- Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*, 70-98.
- Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes? *Journal of Experimental Psychology: General*, *146*, 194-213.
- McConnell, A. R., & Rydell, R. J. (2014). The Systems of Evaluation Model: A dual-systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 204-217). New York: Guilford.
- Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology*, *104*, 45-69.

- Mitchell, C. J., Anderson, N. E., & Lovibond, P. F. (2003). Measuring evaluative conditioning using the Implicit Association Test. *Learning and Motivation, 34*, 203-217.
- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *The Behavioral and Brain Sciences, 32*, 183–198.
- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297-326.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166-180.
- Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413-417.
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81*, 181-192.
- Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology, 20*, 272-314.
- Phills, C. E., Kawakami, K., Tabi, E., Nadolny, D., & Inzlicht, M. (2011). Mind the gap: Increasing associations between the self and Blacks with approach behaviors. *Journal of Personality and Social Psychology, 100*, 197–210.
- Sherman, J. W. (2006). Clearing up some misconceptions about the Quad model. *Psychological Inquiry, 17*, 269-276.

- Sherman, J. W. (2008). Controlled influences on implicit measures: Confronting the myth of process-purity and taming the cognitive monster. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 391-426). New York, NY, US: Psychology Press.
- Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review, 115*, 314-335.
- Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Handbook of implicit social cognition: Measurement, theory, and applications.
- Sherman, J. W., Krieglmeier, R., & Calanchini, J. (2014). Process models require process measures. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 121-138). New York: Guilford.
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108–131.
- Smith, C. T., De Houwer, J., & Nosek, B. A. (2013). Consider the source: Persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin, 39*, 193-205.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*, 220-247.
- Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. T. (2015). Instruction-based approach–avoidance Effects: changing stimulus evaluation via the mere instruction to approach or avoid stimuli. *Experimental Psychology, 62*, 161-169.

- Van Dessel, P., De Houwer, J., Gast, A., Smith, C. T., & De Schryver, M. (2016). Instructing implicit processes: When instructions to approach or avoid influence implicit but not explicit evaluation. *Journal of Experimental Social Psychology, 63*, 1-9.
- Van Dessel, P., Hughes, S., & De Houwer, J. (2019). How Do Actions Influence Attitudes? An Inferential Account of the Impact of Action Performance on Stimulus Evaluation. *In press at Personality and Social Psychology Review*.
- Van Dessel, P., Gawronski, B., Smith, C. T., & De Houwer, J. (2017). Mechanisms underlying approach-avoidance instruction effects on implicit evaluation: Results of a preregistered adversarial collaboration. *Journal of Experimental Social Psychology, 69*, 23-32.
- Van Dessel, P., Mertens, G., Smith, C. T., & De Houwer, J. (2017). The mere exposure instruction effect: Mere exposure instructions influence liking. *Experimental Psychology, 64*, 299-314.
- Van Dessel, P., Mertens, G., Smith, C. T., & De Houwer, J. (2019). Mere exposure effects on implicit stimulus evaluation: The moderating role of evaluation task, number of stimulus presentations, and memory for presentation frequency. *In press at Personality and Social Psychology Bulletin*.
- Wiers, R. W., Eberl, C., Rinck, M., Becker, E. S., & Lindenmeyer, J. (2011). Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science, 22*, 490-497.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101-126.
- Zajonc, R. B., Markus, H., & Wilson, W. R. (1974). Exposure effects and associative learning. *Journal of Experimental Psychology, 10*, 248-263.

Tables

	Experiment 1		Experiment 2		Experiment 3	
	Estimate	95% HDI	Estimate	95% HDI	Estimate	95% HDI
AC Negative ME-I	0.0020	[0.0005, 0.0050]	0.0017	[0.0000, 0.0078]	0.0017	[0.0002, 0.0045]
AC Negative EC-I	0.0074	[0.0027, 0.0149]	0.0127	[0.0011, 0.0422]	0.0076	[0.0025, 0.0162]
AC Negative AA-I	0.0060	[0.0020, 0.0124]	0.0032	[0.0001, 0.0129]	0.0063	[0.0020, 0.0137]
AC Negative ME-E	0.0017	[0.0003, 0.0041]	0.0065	[0.0002, 0.0253]	0.0012	[0.0001, 0.0032]
AC Negative EC-E	0.0050	[0.0015, 0.0109]	0.0061	[0.0003, 0.0222]	0.0049	[0.0014, 0.0111]
AC Negative AA-E	0.0052	[0.0017, 0.0110]	0.0066	[0.0003, 0.0245]	0.0034	[0.0008, 0.0083]
AC Negative Control	N/A	N/A	0.0019	[0.0000, 0.0090]	0.0004	[0.0000, 0.0013]
AC Positive ME-I	0.0036	[0.0010, 0.0077]	0.0058	[0.0002, 0.0217]	0.0013	[0.0002, 0.0032]
AC Positive EC-I	0.0087	[0.0035, 0.0169]	0.0141	[0.0016, 0.0439]	0.0067	[0.0024, 0.0136]
AC Positive AA-I	0.0092	[0.0035, 0.0177]	0.0029	[0.0001, 0.0111]	0.0043	[0.0014, 0.0088]
AC Positive ME-E	0.0025	[0.0005, 0.0059]	0.0049	[0.0001, 0.0178]	0.0018	[0.0004, 0.0043]
AC Positive EC-E	0.0039	[0.0011, 0.0086]	0.0089	[0.0006, 0.0296]	0.0037	[0.0011, 0.0083]
AC Positive AA-E	0.0034	[0.0010, 0.0074]	0.0102	[0.0008, 0.0338]	0.0047	[0.0014, 0.0104]
AC Positive Control	N/A	N/A	0.0010	[0.0000, 0.0047]	0.0008	[0.0000, 0.0024]
D ME-I	0.9284	[0.9170, 0.9383]	0.9069	[0.8831, 0.9276]	0.9287	[0.9190, 0.9375]
D EC-I	0.9389	[0.9285, 0.9484]	0.9112	[0.8877, 0.9315]	0.9398	[0.9310, 0.9479]
D AA-I	0.9333	[0.9227, 0.9430]	0.9004	[0.8773, 0.9213]	0.9412	[0.9330, 0.9488]
D ME-E	0.9217	[0.9103, 0.9323]	0.8653	[0.8359, 0.8915]	0.9230	[0.9130, 0.9325]
D EC-E	0.9152	[0.9020, 0.9272]	0.8982	[0.8763, 0.9177]	0.9267	[0.9166, 0.9362]
D AA-E	0.9391	[0.9290, 0.9480]	0.9040	[0.8785, 0.9262]	0.9458	[0.9366, 0.9542]
D Control	N/A	N/A	0.8608	[0.8306, 0.8882]	0.9260	[0.9163, 0.9350]
G ME-I	0.5203	[0.4872, 0.5543]	0.5159	[0.4542, 0.5782]	0.5691	[0.5414, 0.5970]
G EC-I	0.5564	[0.5208, 0.5921]	0.4733	[0.4118, 0.5350]	0.4854	[0.4542, 0.5166]
G AA-I	0.5386	[0.5033, 0.5748]	0.5195	[0.4638, 0.5755]	0.5509	[0.5219, 0.5800]
G ME-E	0.5138	[0.4814, 0.5470]	0.5227	[0.4685, 0.5751]	0.5377	[0.5095, 0.5659]
G EC-E	0.5323	[0.4995, 0.5657]	0.4893	[0.4363, 0.5421]	0.5661	[0.5375, 0.5957]
G AA-E	0.5143	[0.4787, 0.5511]	0.5313	[0.4680, 0.5930]	0.5176	[0.4828, 0.5530]
G Control	N/A	N/A	0.4920	[0.4374, 0.5463]	0.5203	[0.4929, 0.5480]
OB ME-I	0.3684	[0.0000, 1.0000]	0.7569	[0.0016, 1.0000]	0.5659	[0.0038, 0.9999]
OB EC-I	0.0389	[0.0000, 0.4672]	0.8138	[0.0431, 1.0000]	0.4057	[0.0003, 0.9922]
OB AA-I	0.7803	[0.0237, 1.0000]	0.3108	[0.0000, 0.9965]	0.4327	[0.0007, 0.9965]
OB ME-E	0.3656	[0.0000, 1.0000]	0.5466	[0.0000, 1.0000]	0.3756	[0.0000, 0.9971]
OB EC-E	0.9317	[0.2162, 1.0000]	0.8101	[0.0357, 1.0000]	0.6298	[0.0113, 0.9999]
OB AA-E	0.1545	[0.0000, 0.9460]	0.7714	[0.0094, 1.0000]	0.3434	[0.0000, 0.9836]
OB Control	N/A	N/A	0.4386	[0.0000, 1.0000]	0.2939	[0.0000, 0.9788]

Table 1. Parameter estimates for all experimental conditions for all experiments. AC Negative = Activation of negative evaluations. AC Positive = Activation of positive evaluations. D = detection. G = guessing. OB = overcoming bias. ME-I = mere exposure instructions. EC-I = evaluative conditioning instructions. AA-I = approach-avoidance instructions. ME-E = mere exposure experience. EC-E = evaluative conditioning experience. AA-E = approach-avoidance experience. Control = control condition.

	Positive Activation	Negative Activation	Detection	Overcoming Bias	Guessing
Experiment 1					
<i>Instruct vs. Exp</i>	0.0039 [0.0008, 0.0085]	0.0012 [-0.0015, 0.0046]	0.0081 [-0.0006, 0.0170]	-0.0881 [-0.5792, 0.3311]	0.0183 [-0.0075, 0.0439]
Experiment 2					
<i>Instruct vs. Exp</i>	-0.0004 [-0.0100, 0.0089]	-0.0005 [-0.0091, 0.0079]	0.0170 [-0.0017, 0.0357]	-0.0822 [-0.5703, 0.3547]	-0.0115 [-0.0570, 0.0348]
<i>Instruct vs. Control</i>	0.0067 [0.0005, 0.0198]	0.0039 [-0.0029, 0.0158]	0.0453 [0.0152, 0.0778]	0.1886 [-0.6164, 0.7927]	0.0109 [-0.0505, 0.0729]
<i>Exp vs. Control</i>	0.0070 [0.0006, 0.0208]	0.0045 [-0.0025, 0.0171]	0.0284 [-0.0034, 0.0616]	0.2708 [-0.5186, 0.9999]	0.0225 [-0.0371, 0.0820]
Experiment 3					
<i>Instruct vs. Exp</i>	0.0007 [-0.0016, 0.0033]	0.0020 [-0.0003, 0.0056]	0.0047 [-0.0025, 0.0118]	0.0185 [-0.4158, 0.4660]	-0.0053 [-0.0283, 0.0178]
<i>Instruct vs. Control</i>	0.0032 [0.0009, 0.0068]	0.0048 [0.0017, 0.0101]	0.0106 [0.0005, 0.0214]	0.1742 [-0.5094, 0.8409]	0.0148 [-0.0144, 0.0437]
<i>Exp vs. Control</i>	0.0026 [0.0005, 0.0057]	0.0028 [0.0008, 0.0062]	0.0059 [-0.0046, 0.0168]	0.1557 [-0.5295, 0.7525]	0.0201 [-0.0091, 0.0497]

Table 2. Planned contrasts between Instruction- and Experience-based evaluative learning

(Experiments 1, 2, and 3); between Instruction-based learning and control, and between

Experience-based learning and control (Experiments 2 and 3). Values reported here are coded

such that positive sign reflects a larger effect in the former versus latter condition as listed in the

first column of the table. Values in brackets reflect 95% HDIs.

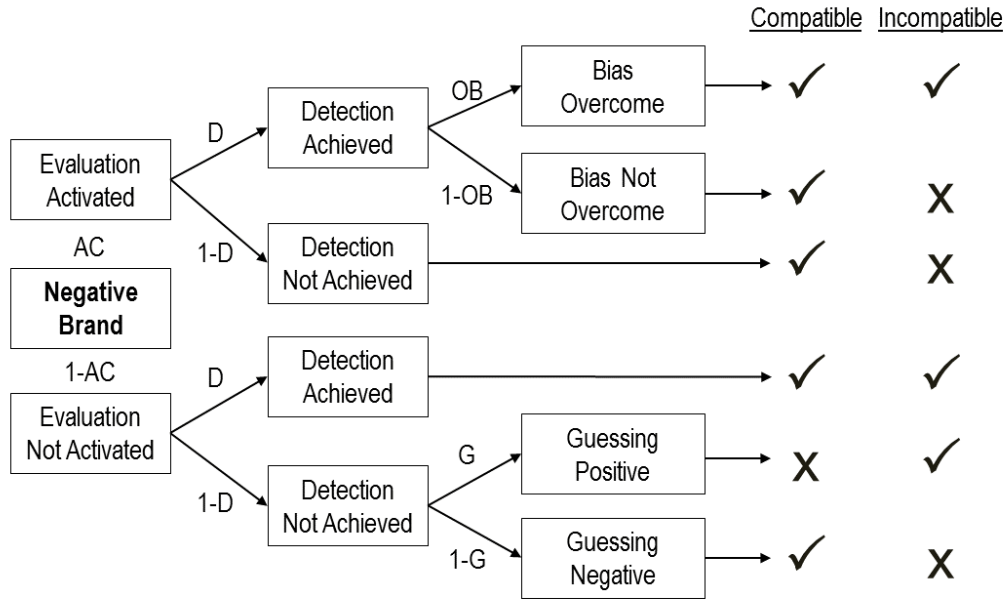


Figure 1.


A portion of the Quadruple Process Model (Quad Model) depicting possible outcomes when a stimulus appears representing the negatively-evaluated brand. Each square represents a parameter and each path represents a likelihood. All parameters are conditional upon all preceding paths. The table on the right side of the figure depicts correct (✓) and incorrect (✗) responses as a function of process pattern.

Appendix A

Screen shots of Instruction Conditions (All Experiments)

Mere Exposure Instructions

Later on in this study you will see two new brand names (Empeya and Vekte) that may be introduced to supermarkets in the USA and elsewhere around the world. Remember that later on in this experiment:

The brand name  will be presented to you **OFTEN**

The brand name  will be presented to you **RARELY**

It is VERY important that you remember how often these two brand names will be presented later in the experiment. You will need this information to complete the tasks correctly. This information will NOT be presented later in the experiment - ONLY NOW - so please read it carefully.

Continue

Evaluative Conditioning Instructions

Later on in this study, you will see a positive image (e.g., a happy person) or a negative image (e.g., a disgusted person). Each image will be paired with a brand name (either Empeya or Vekte) that may be introduced to supermarkets in the USA and elsewhere in the world. Specifically, later in this experiment:

Whenever you see the brand name  , a **POSITIVE IMAGE** will also appear.

Whenever you see the brand name  , a **NEGATIVE IMAGE** will also appear.


It is VERY important to remember which brand name will be paired with positive or negative words in the second part of the experiment. You will need this information to complete the tasks successfully. This information will NOT be presented later in the experiment - ONLY NOW - so please read it carefully.

Continue

Approach Avoid Instructions

Later on in this study you will see two brand names (Empeya and Vekte) that may be introduced to supermarkets in the USA and elsewhere around the world. You will have to make a particular action every time you see one of these brand names. Specifically, later in this experiment:

Whenever you see the brand name  you will need to **APPROACH** it.

Whenever you see the brand name  you will need to **AVOID** it.

It is VERY important to remember what brand name you will have to approach or avoid later in the experiment. You will need this information to successfully complete the study. This information will NOT be presented later in the experiment - ONLY NOW - so please read it carefully.

Continue

IAT Stimuli (All Experiments)

Empeya



Vekte



Appendix B**IAT *D*-Scores by Condition and Experiment**

	Experiment 1		Experiment 2		Experiment 3	
	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>	<i>n</i>	<i>M (SD)</i>
AA-E	190	0.22 (0.42)	47	0.20 (0.44)	201	0.17 (0.46)
AA-I	205	0.33 (0.43)	64	0.28 (0.36)	329	0.32 (0.42)
EC-E	195	0.36 (0.43)	78	0.40 (0.37)	254	0.38 (0.46)
EC-I	200	0.47 (0.39)	53	0.35 (0.42)	267	0.42 (0.44)
ME-E	222	0.07 (0.48)	58	0.12 (0.39)	270	0.09 (0.46)
ME-I	209	0.13 (0.42)	49	0.12 (0.44)	287	0.07 (0.43)
Control			50	0.01 (0.49)	290	0.00 (0.41)

Note. Positive IAT *D*-scores reflect a pro-Empeya bias (in line with the direction of the evaluative learning procedure); AA-E = approach-avoidance experience; AA-I = approach-avoidance instructions; EC-E = evaluative conditioning experience; EC-I = evaluative conditioning instructions; ME-E = mere exposure experience; ME-I = mere exposure instructions

Appendix C

Graphs of Observed versus Predicted Response Frequencies and Covariances as indices of model fit.

The x-axis of all graphs are labeled as follows:

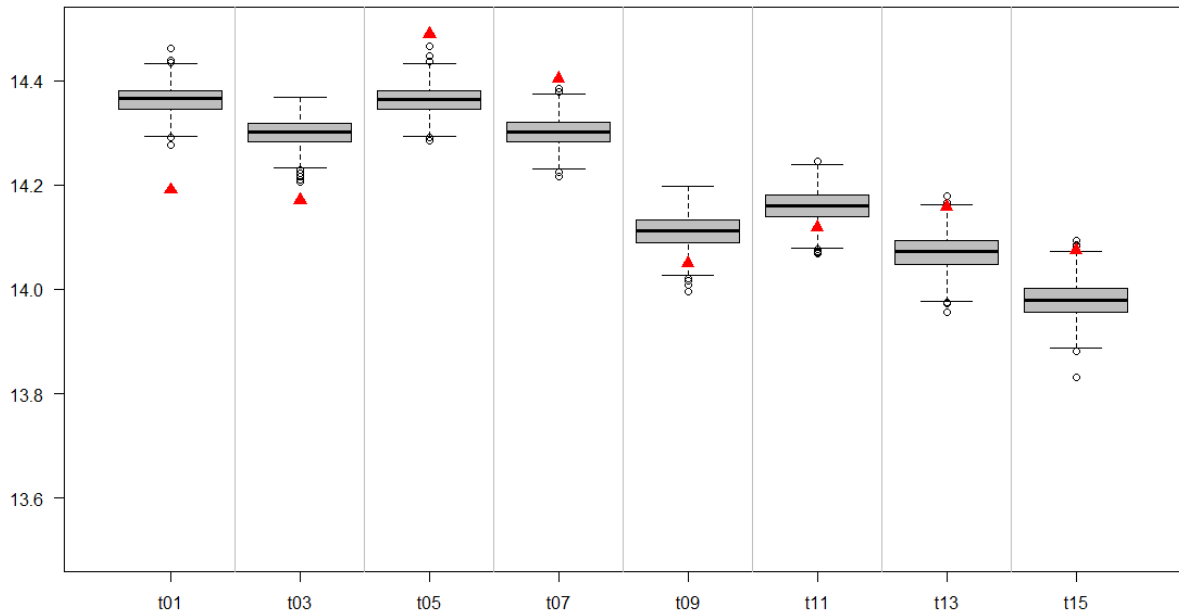
- t01: correct responses to stimuli representing the positive brand when the positive brand / good words share a response key
- t03: correct responses to stimuli representing the negative brand when the negative brand / bad words share a response key
- t05: correct responses to good words when the positive brand / good words share a response key
- t07: correct responses to bad words when the negative brand / bad words share a response key
- t09: correct responses to stimuli representing the positive brand when the positive brand / bad words share a response key
- t11: correct responses to stimuli representing the negative brand when the negative brand / good words share a response key
- t13: correct responses to good words when the negative brand / good words share a response key
- t15: correct responses to bad words when the positive brand / bad words share a response key

The y-axis of the response frequency graphs represents number of correct responses for each response category, which can range from 0 (no correct responses) to 15 (all correct responses).

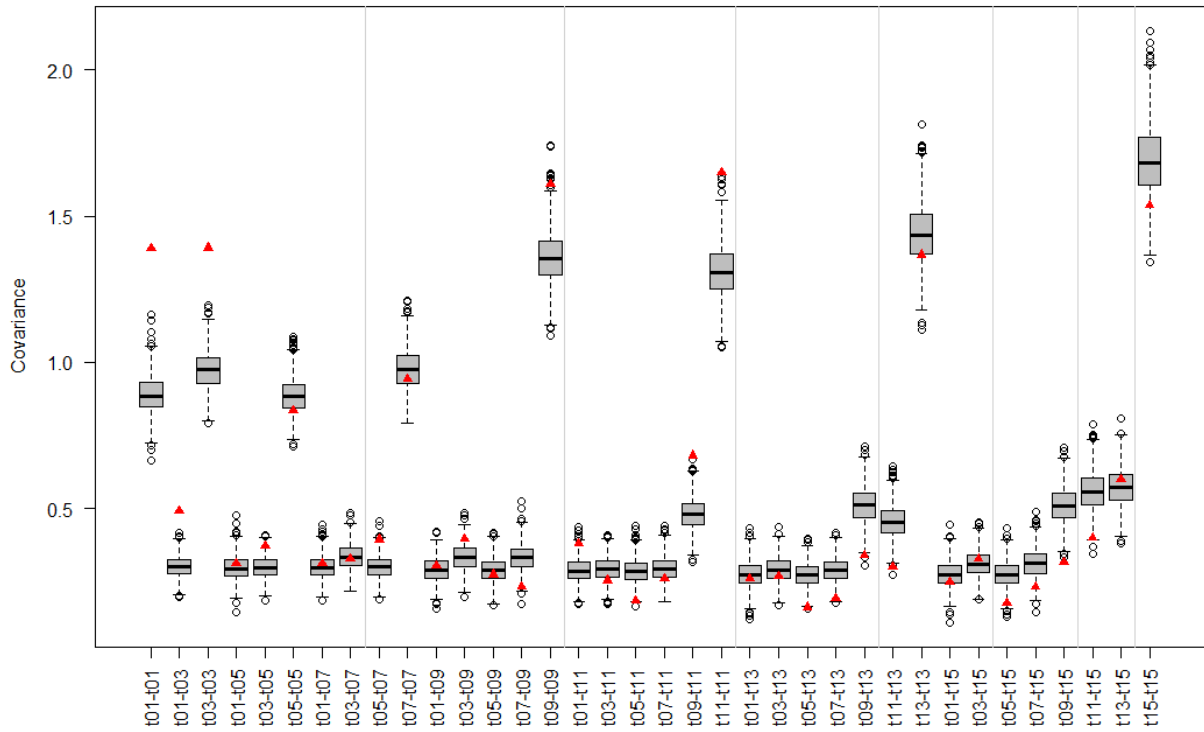
The y-axis of the covariance graphs represents the covariances between each response category.

Experiment 1:

Observed (red) and predicted (boxplot) mean frequencies

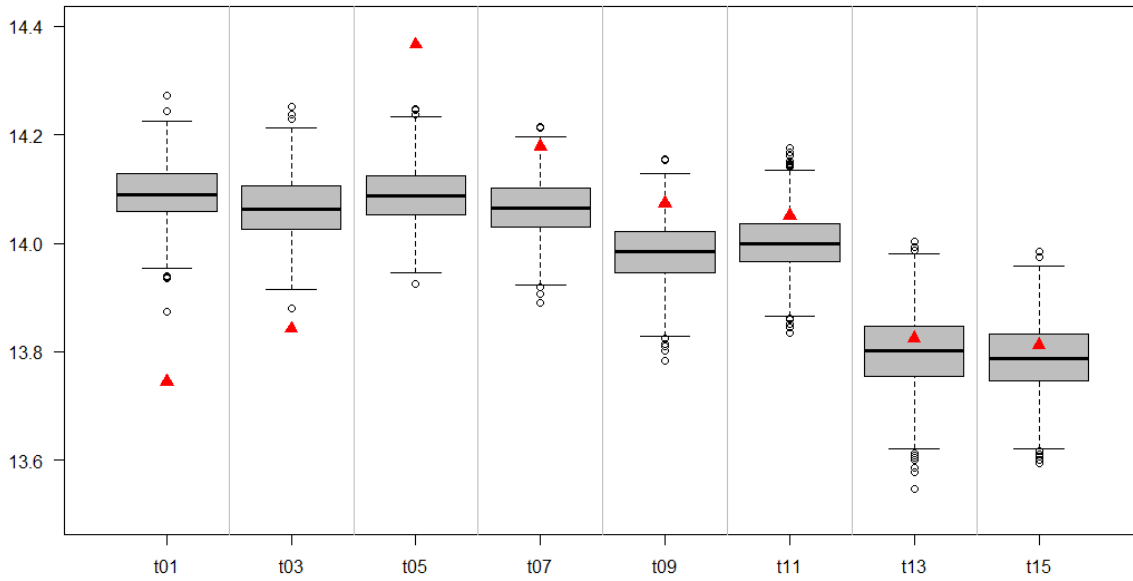


Observed (red) and predicted (gray) covariances

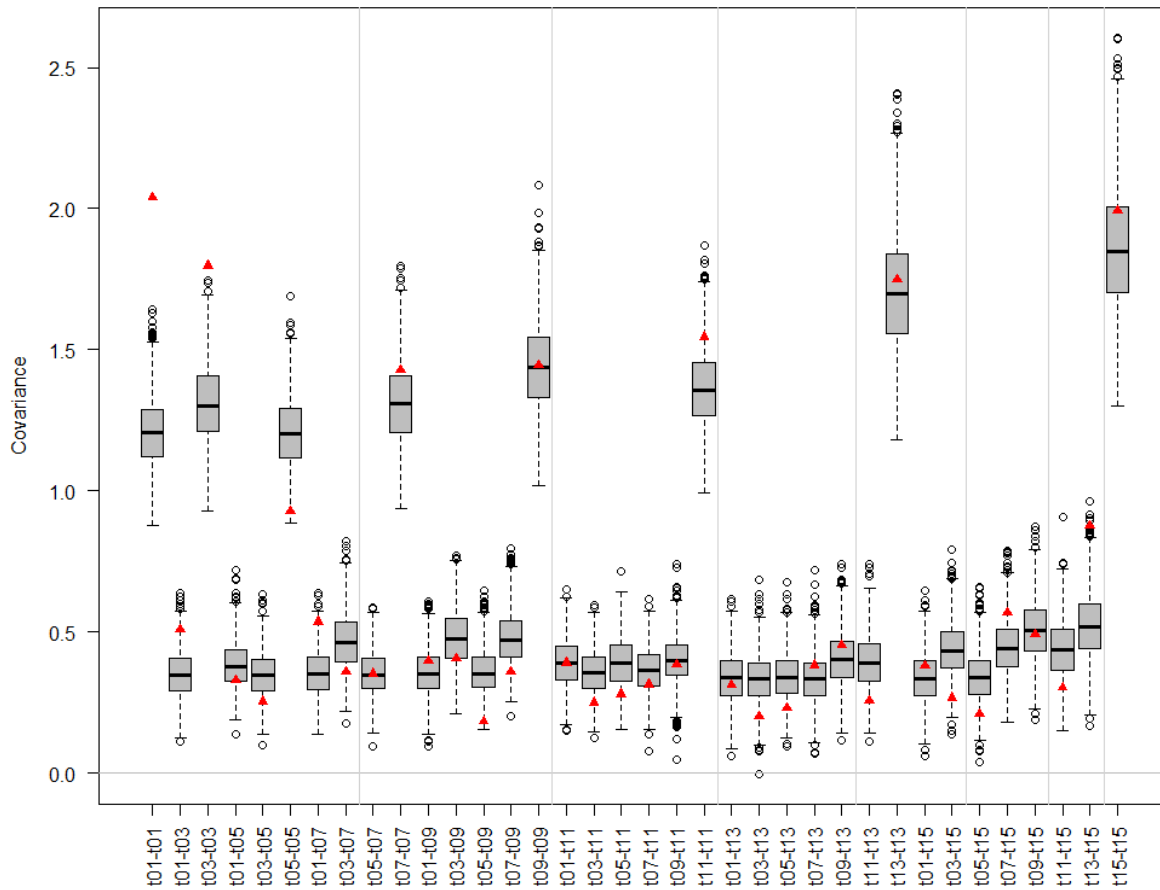


Experiment 2:

Observed (red) and predicted (boxplot) mean frequencies

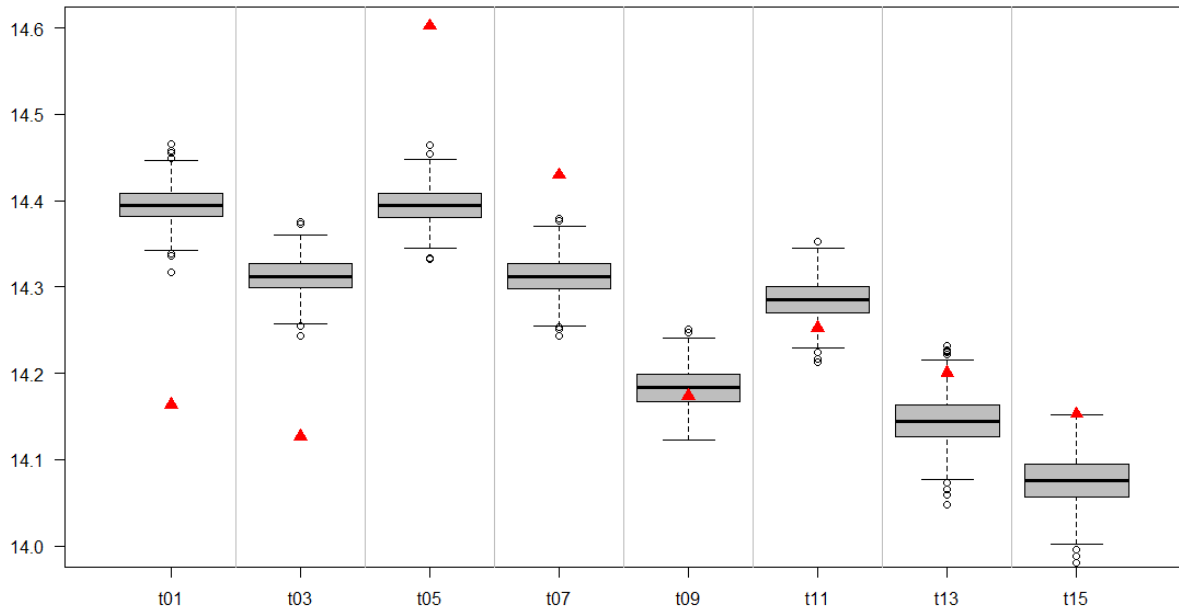


Observed (red) and predicted (gray) covariances



Experiment 3:

Observed (red) and predicted (boxplot) mean frequencies



Observed (red) and predicted (gray) covariances

