Tracking hand movements captures the response dynamics of the evaluative priming effect

Naoaki Kawakami & Emi Miura

To cite this article: Naoaki Kawakami & Emi Miura (2018): Tracking hand movements captures the response dynamics of the evaluative priming effect, Cognition and Emotion, DOI: 10.1080/02699931.2018.1483897

To link to this article: https://doi.org/10.1080/02699931.2018.1483897
Tracking hand movements captures the response dynamics of the evaluative priming effect

Naoaki Kawakami\textsuperscript{a} and Emi Miura\textsuperscript{b}

\textsuperscript{a}Faculty of Human Sciences, Shimane University, Matsue, Japan; \textsuperscript{b}Graduate School of Education, Shimane University, Matsue, Japan

ABSTRACT
We tested the response dynamics of the evaluative priming effect (i.e. facilitation of target responses following evaluatively congruent compared with evaluatively incongruent primes) using a mouse tracking procedure that records hand movements during the execution of categorisation tasks. In Experiment 1, when participants performed the evaluative categorisation task but not the non-evaluative semantic categorisation task, their mouse trajectories for evaluatively incongruent trials curved more toward the opposite response than those for evaluatively congruent trials, indicating the emergence of evaluative priming effects based on response competition. In Experiment 2, implementing a task-switching procedure in which evaluative and non-evaluative categorisation tasks were intermixed, we obtained reliable evaluative priming effects in the non-evaluative semantic categorisation task as well as in the evaluative categorisation task when participants assigned attention to the evaluative stimulus dimension. Analyses of hand movements revealed that the evaluative priming effects in the evaluative categorisation task were reflected in the mouse trajectories, while evaluative priming effects in the non-evaluative categorisation tasks were reflected in initiation times (i.e. the time elapsed between target onset and first mouse movement). Based on these findings, we discuss the methodological benefits of the mouse tracking procedure and the underlying processes of evaluative priming effects.

It has been assumed that people spontaneously evaluate all incoming stimulus information as pleasant or unpleasant, liked or disliked, good or bad. In accordance with this hypothesis, Fazio, Sanbonmatsu, Powell, and Kardes (1986) showed that participants need less time to judge the evaluative connotation of a target stimulus (e.g. friend) after the presentation of an evaluatively congruent prime stimulus (e.g. love) than after the presentation of an evaluatively incongruent prime stimulus (e.g. pain). Over the last three decades, it has been demonstrated that this evaluative priming effect is based on fast-acting processes and depends neither on the conscious identification of the primes, nor on the presence of ample processing resources or an explicit evaluative processing goal (for reviews, see Fazio, 2001; Klauer & Musch, 2003). This suggests that humans are indeed endowed with an evaluative decision mechanism that allows them to automatically and unconsciously evaluate all incoming stimuli (e.g. Spruyt, Hermans, De Houwer, & Eelen, 2002).

Although evaluative priming research has made progress in describing and documenting the occurrence of the evaluative priming effect, there is disagreement concerning the nature of the underlying mechanism that is responsible for the observed effects. Several researchers have claimed that evaluative priming effects may be produced by processes that operate at an encoding level (e.g. Bargh, Chaiken, Raymond, & Hymes, 1996; Hermans, De Houwer, & Eelen, 1994). According to this encoding account of the evaluative priming effect, affectively...
polarised prime stimuli pre-activate the memory representations of evaluatively related targets. Therefore, when a target is preceded by a prime with the same valence, its semantic encoding is facilitated, resulting in less time needed to judge the evaluative connotation of the target.

Other researchers have argued that it is possible to explain evaluative priming effects with response competition that operates at a response selection stage rather than – or in addition to – the encoding level (e.g. De Houwer, Hermans, Rothermund, & Wentura, 2002; Klauer, Roßnagel, & Musch, 1997). In a typical evaluative priming study (e.g. Fazio et al., 1986), participants are instructed to categorise the targets based on their valence (i.e. the evaluative categorisation task). According to the response perspective, Stroop-like response interference processes could underlie the evaluative priming effect in this task. That is, even though people do not have to respond to an affectively polarised prime stimulus, the prime automatically activates a particular response (e.g. “good”) that is either congruent or incongruent with the response activated by the target (see Klauer & Musch, 2003). Thus, people tend to respond faster when the activated responses to the prime and target are congruent, and slower when these responses are incongruent.

In sum, evaluative priming effects can be explained by an encoding perspective, which posits that the prime facilitates encoding of the target, and by a response perspective based on Stroop-like response interference, which posits that the prime facilitates or inhibits the response made to the target. In both perspectives, it is hypothesised that the reaction time in an evaluatively congruent trial would be faster than the reaction time in an evaluatively incongruent trial. The important question that remains is the cause of the reaction time differences. Does the congruity between the prime and target influence how easily the target can be encoded and/or how quickly a response to the target can be executed (or other processes that cannot be fully explained by both mechanisms)?

To solve this theoretical discrepancy concerning the nature of the underlying mechanism of evaluative priming effects, researchers have adopted other variations of the sequential evaluative priming paradigm in which participants are not required to make an evaluative response to the target (e.g. the pronunciation task, semantic categorisation task). In a pronunciation task, for example, participants are simply required to pronounce the name of the target (e.g. friend) as quickly as possible rather than evaluate it as “good” (e.g. Bargh et al., 1996). As there is no confound between a response evoked by the prime and a response required by the target, researchers are able to rule out the response competition perspective if evaluative congruency effects are observed. In addition, the non-evaluative task most often used to examine evaluative priming effects is the semantic categorisation task (e.g. De Houwer et al., 2002; Klinger, Burton, & Pitts, 2000). In this task, for example, when both primes and targets are positive or negative nouns, participants are required to indicate whether the target refers to a person (e.g. friend, snob) or an animal (e.g. butterfly, cockroach). If the encoding facilitation mechanism operates in priming tasks, because primes should also facilitate non-evaluative semantic processing of evaluatively congruent targets, it is expected that significant evaluative priming will also occur in the semantic categorisation task. However, according to a meta-analysis that covered a quarter century of evaluative priming experiments (Herring et al., 2013), reliable evaluative priming effects have only been found in the evaluative categorisation task. Although significant evaluative priming effects also occur in the pronunciation tasks, the effect size is smaller than those in the evaluative categorisation task. On the contrary, researchers often failed to detect effects (e.g. Klauer & Musch, 2001; Spruyt, Hermans, Pandelaere, De Houwer, & Eelen, 2004). As for the semantic categorisation tasks, the meta-analysis revealed no significant evaluative priming effect. Thus, it seems possible that the evaluative priming effect is primarily driven by processes that operate at a response selection stage rather than by processes that operate at an encoding stage, although these mechanisms are not mutually exclusive.

It should however be noted that it is insufficient to switch from the evaluative categorisation task to a semantic categorisation task to distinguish encoding from response mechanisms in an evaluative priming effect. When researchers switch from the evaluative categorisation task to a semantic evaluative categorisation task, this change also changes the focus of attention induced by the tasks. While the typical evaluative categorisation task focuses participants’ attention on evaluation by requiring them to categorise the targets based on the stimuli’s valence, the semantic categorisation task does not require participants to pay attention to the stimuli’s evaluation to give
the correct response (Spruyt, De Houwer, Hermans, & Eelen, 2007). Thus, one route to address this problem is to disentangle the two processes while keeping the evaluation task. However, distinguishing encoding from response mechanisms is difficult with outcome-based measures, such as reaction times. This is because reaction times represent the combined output of both these processes (and others). Therefore, these can be limited in their ability to make inferences about what sort of cognitive processing is occurring across time, as only ultimate outputs (e.g. key-presses) of completed decisions are available. As such, these tell us little about the ongoing decision-making processes underlying categorisation that lead to faster or slower responding. Thus, several researchers have employed measurements and methods other than reaction times in an attempt to address this limitation. For instance, Eder, Leuthold, Rothermund, and Schweinberger (2012) recorded the lateralised readiness-potential (LRP) in the human brain as an online-index of activation of the left and right response hand. Results showed that the incorrect response hand was activated more strongly in evaluatively incongruent trials than in evaluatively congruent trials, suggesting a response competition in incongruent trials. They also demonstrated that the amplitude of the N400 was greater in evaluatively incongruent trials than in evaluatively congruent trials, suggesting processing of evaluative mismatches between primes and targets. Nevertheless, a regression analysis revealed that the LRP effect was a better predictor of the evaluative priming effect (reaction time difference) than the N400 effect. Moreover, Voss, Rothermund, Gast, and Wentura (2013) adopted a diffusion model data analysis which helps to separately estimate the mediating effects of the two processes, suggesting that response competition plays an important role in the emergence of evaluative priming effects. In this way, using measures other than reaction times has contributed to better understanding of the mechanisms involved in evaluative priming effects. Nevertheless, given the complexity of findings from the electrophysiological and diffusion model approaches, it seems worthwhile to demonstrate another approach which does not depend on outcome-based measures. In the present study, we will show the response dynamics of an evaluative priming effect with continuous measures of performance, allowing for enhanced insight into the underlying processes of evaluative priming effects.

To investigate the dynamic aspect of the evaluative priming effect, the present study utilised a mouse tracking procedure (Freeman & Ambady, 2010) that records and analyses hand movements during the execution of categorisation tasks. This methodology provides an online measure of the spontaneous behaviours performed by the participant and opens a window to the possible underlying psychological processes (Freeman & Ambady, 2009). In a typical study using a mouse tracking procedure (e.g. Freeman, Ambady, Rule, & Johnson, 2008), participants are required to categorise a stimulus (e.g. a human face) by moving the mouse from the bottom centre of the screen to specific response locations on either the top-left or top-right corners on the screen as quickly as possible (see Figure 1). A number of early studies have repeatedly demonstrated that mouse trajectories reflect the degree of response competition between evaluations (e.g. Freeman & Ambady, 2010). Such an approach has not been used in evaluative priming research, but has been successfully applied in other domains. For instance, when participants were presented with sex-atypical male and female faces (with a mixture of masculine and feminine cues), mouse trajectories showed a continuous attraction toward the opposite sex category, as measured by the Area Under the Curve (AUC), despite participants ultimately selecting the correct category. This indicates a dynamic aspect of continuous competition between femininity and masculinity, which operates at a response selection stage. Therefore, if the evaluative priming effect can be at least partially explained by response competition that operates at a response selection stage, then the evaluative congruency effects should be detectable in the mouse trajectory. More specifically, for evaluatively incongruent trials, this response competition can be indexed by a greater attraction of mouse trajectories toward the opposite valence of response compared to evaluatively congruent trials. Such a finding can add to the evaluative priming literature, showing that the response dynamics of the evaluative priming effect can be captured by tracking hand movements.

Experiment 1

In Experiment 1, we examined the standard evaluative priming paradigm in an evaluative categorisation task and a semantic categorisation task using a mouse tracking procedure. In addition to our primary measure of mouse trajectories, we also assessed two
kinds of reaction time measures: initiation time (IT) and ultimate reaction time (URT). IT is defined as the time elapsed between target onset and the first mouse movement, and URT is defined as the time elapsed between presentation of target and clicking the correct response location. Measuring these reaction times helped us explore possible consistency with previous research.

**Methods**

**Participants**

Sixty-two Japanese undergraduate students (34 female and 28 male, mean age = 20.42 years, range = 18–24 years) participated in this study in exchange for course credit. We recruited as many participants as possible until the end of the predefined time period for the study, and the data of all participants were included in the analysis. All participants were randomly assigned to one of the two tasks: evaluative categorisation (n = 31) or semantic categorisation (n = 31). This sample size allowed detection of an evaluation congruency effect of $dz = 0.46$ with power $1 – \beta = .80$ (alpha = .05, one-tailed). This effect size roughly corresponds to what Herring et al. (2013) reported as the mean effect size ($dz = .45$) for evaluative congruency effects using the evaluation task. Power analyses were done using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

**Materials**

Stimuli were selected based on a preliminary study in which 70 participants rated 1034 Japanese nouns on a nine-point rating scale (1 = very negative to 9 = very positive). Targets were 60 nouns (30 positive, 30 negative). Within each valence category, 15 targets referred to persons (e.g. baby, thief), whereas 15 targets referred to objects (e.g. treasure, weapon). The mean rating was 6.89 ($SD = 0.45$) for the positive person targets, 2.36 ($SD = 0.39$) for the negative person targets, 6.73 ($SD = 0.45$) for the positive object targets, and 2.44 ($SD = 0.49$) for the negative object targets. Primes were 60 nouns (30 positive, 30 negative) that do not refer to persons or objects (e.g. peace, false). The mean rating was 6.85 ($SD = 0.50$) for the positive primes and 2.35 ($SD = 0.42$) for the negative primes. A complete list of stimuli is provided in the Appendix.

To track mouse movements during the categorisation task, we used Mouse Tracker software (Freeman & Ambady, 2010). Mouse Tracker records real time $x$ and $y$ mouse coordinates at an approximate 60–75 Hz sampling rate. Stimuli were displayed on a 24-inch LCD monitor connected to a laptop, running at a 1920 x 1080 resolution. The task was performed using a standard computer mouse, with the cursor speed set at the 6/11 default mode in Windows 10.

**Procedure**

Participants were tested individually in a soundproof room. For both the evaluative and the semantic categorisation tasks, there were 120 trials that were divided into two blocks of 60 trials. For each block, the 60 target words were randomly assigned to one of the 60 prime words such that the number of evaluatively congruent and evaluatively incongruent trials would be balanced. Participants were asked to categorise targets according to their valence (i.e. positive vs. negative) in the evaluative categorisation task, and according to their semantic categories (i.e. person vs. object) in the semantic categorisation task.

Each trial started when participants clicked with the mouse on the START button located at the bottom of the screen (see Figure 1). After a 500 ms fixation point presentation (slightly lower than the centre of the screen), the prime was presented for 200 ms. Following the prime, a blank screen was presented for 50 ms, after which the target appeared and participants were instructed to immediately move the mouse and click the appropriate response button located either at the top-left or top-right of the computer screen. The location of the positive versus negative (person vs. object) response button at the top-left or top-right of the computer screen was counterbalanced across participants. Each target word was presented until participants had provided their response or 2,000 ms elapsed. In the case of errors, a red cross was displayed for 500 ms. To ensure that mouse trajectories reflected online processing, participants were encouraged to begin their movements as quickly and accurately as possible and were warned if movement was initiated later than 700 ms following target word presentation. This instruction is customarily included in mouse tracking studies so that trajectories reflect the dynamics of a decision process rather than simply reflecting the kinematics of a response choice after the choice has already been made (Freeman & Ambady, 2009). After having clicked a response button, participants were instructed to move back to the START button and click it, so that the following trial could start. Before the experimental phase, there were 12 practice trials.
in which 6 person target words (3 positive, 3 negative) and 6 object target words (3 positive, 3 negative) that were not used in the experimental phase, were randomly assigned to one of the 12 prime words (6 positive, 6 negative) that were also not used in the experimental phase.

Results

Participants completed a total of 7,440 trials (3,720 per categorisation task). Of these, there were 31 trials with incorrect responses (0.42%), 18 trials with ITs greater than 700 ms (0.24%), 116 trials with ultimate decision times exceeding 2.5 SD (1.56%), and 78 trials (1.05%) with aberrant mouse movements (i.e. erratic output producing clockwise/anticlockwise looping that could not be interpreted), which were discarded. Overall, 97% of the trials were included in the following analyses.

Trajectory analyses

X-coordinates over time. We computed the AUC as an index for response competition in the evaluative priming effect. AUC is defined as the geometric area between the observed mouse trajectory and an idealised straight-line trajectory (a straight line between the trajectory’s start and endpoints). That is, a greater AUC is interpreted as greater response competition. All trajectories were rescaled into a standard coordinate space (top left: x, y = [−1, 1.5]; bottom right: x, y = [1, 0]). The AUC values were then averaged for each participant and for each condition (congruent vs. incongruent). An analysis of variance (ANOVA) revealed a significant interaction between categorisation task (evaluative categorisation vs. semantic categorisation) and evaluative congruence (evaluatively congruent vs. evaluatively incongruent), $F(1, 60) = 98.27, p < .000$, $h_{p}^2 = .62$ (see Table 1). For the evaluative categorisation task, participants were more likely to demonstrate a greater AUC in response to evaluatively incongruent trials when compared to evaluatively congruent trials, $F(1, 30) = 190.50, p < .000$, $h_{p}^2 = .86$, indicating that trajectories for incongruent trials curved more toward the opposite response relative to those for congruent targets (see Figure 2).

Table 1. Mean (SD) of performance measures for mouse movements in Experiment 1.

<table>
<thead>
<tr>
<th>Task</th>
<th>Ultimate reaction time (ms)</th>
<th>X-coordinates over time (AUC)</th>
<th>Initiation time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Evaluative congruency</td>
<td>Evaluative congruency</td>
<td>Evaluative congruency</td>
</tr>
<tr>
<td></td>
<td>Congruent</td>
<td>Incongruent</td>
<td>PE</td>
</tr>
<tr>
<td>Evaluative categorisation</td>
<td>784 (107.95)</td>
<td>825 (92.81)</td>
<td>41**</td>
</tr>
<tr>
<td>Semantic categorisation</td>
<td>831 (132.36)</td>
<td>834 (118.94)</td>
<td>3</td>
</tr>
</tbody>
</table>

Note: PE = priming effect (evaluatively incongruent – evaluatively congruent).

**p < .01.
suggests an effect of evaluative congruence based on response competition. Conversely, there was no significant difference between evaluatively congruent and evaluatively incongruent trials in the semantic categorisation task, $F(1, 30) = 1.32, p = .26, \eta^2_p = .04$.

**Distributional analysis.** The trajectory data above showed clear response competition in the evaluative categorisation task, but not in the semantic categorisation task. These results provide a good example of mutual validation of the most prevalent accounts of evaluative priming and the mouse tracking procedure. However, an alternative explanation could also explain the data. It is possible that the continuous attraction we see in Figure 2 is simply the result of averaging across fundamentally different types of trials (Freeman et al., 2008). For example, whereas some trials showed movement straight to the correct answer, other trials showed movement initially directed straight at the incorrect answer, followed by a sudden reanalysis and corrective movement redirecting the trajectory straight to the correct category. In this case, the distribution of the AUC values would be bimodal. To test against this possibility, we analysed the distribution of AUC values for indications of bimodality. Specifically, we computed a bimodality coefficient (SAS Institute, 1989) for the distribution of AUC z-scores for evaluatively incongruent trials in the evaluative categorisation task. The calculated bimodality coefficient was 0.492, which is less than the minimum value of 0.555 that would represent a bimodal distribution.

**Reaction time analyses**
Two different indices were considered: URT and IT. The URT is the interval between the onset of the target word and the click on the response button. Conversely, IT is the time elapsed between the target onset and the first mouse movement.

In the analysis of URT, an ANOVA revealed a significant interaction between categorisation task and evaluative congruence, $F(1, 60) = 7.26, p = .009, \eta^2_p = .11$ (see Table 1). For the evaluative categorisation task, participants were faster in reaching the response locations in the case of congruent trials than in the case of incongruent trials, $F(1, 30) = 34.57, p = .000, \eta^2_p = .54$. For the semantic categorisation task, there was no significant difference between evaluatively congruent and evaluatively incongruent trials, $F(1, 30) = 0.04, p = .85, \eta^2_p = .00$. Thus, a congruency effect was found only for the evaluative categorisation task, which supports previous research. This suggests that URT reflects evaluative priming effects as an outcome-based measure.

As for the IT, an ANOVA revealed no significant interaction, $F(1, 60) = 0.23, p = .63, \eta^2_p = .00$ (see Table 1).

**Discussion**
Results showed that congruency effects were found in mouse trajectories, as assessed by AUC, only in the evaluative categorisation task. This asymmetrical result between two tasks is consistent with the general findings of an outcome-focused measure (Herring et al., 2013). Since the mouse trajectories are sensitive to the response competition between evaluations (Freeman & Ambady, 2010), the main result of Experiment 1 provides compelling evidence that the evaluative priming effect can be at least partially explained by response competition that operates at a response selection stage.
These results demonstrated the methodological benefits of the mouse tracking procedure to examine the response dynamics of evaluative priming effects. However, it is important to note the possible concerns with the evaluative categorisation task in Experiment 1. These concerns include the extremely large effect size of the AUC ($dz = 2.48$), and ITs that were of relatively short duration. Because participants were encouraged to begin their movements as quickly as possible, it is possible that they began movement prior to identifying a target stimulus based on intentional processing of the valence of the prime. Within the context of the current experiment, this prime-triggered strategy will result in a fast initiation (as they were asked for) and the correct trajectory in 50% of the trials. In the remaining 50% trials, this will lead to a fast initiation and enough time to turn the trajectory after starting their movements. This strategy may be the cause of the large AUC effects and short ITs observed. To overcome these concerns and gain more insight into the underlying processes of evaluative priming effects, we performed a second experiment with a task-switching procedure introduced by Spruyt et al. (2007).

**Experiment 2**

The results of Experiment 1 support a response perspective which posits that the prime facilitates or inhibits the response made to the target. However, Spruyt et al. (2007; see also Spruyt, De Houwer, & Hermans, 2009; Spruyt, De Houwer, Everaert, & Hermans, 2012) found that evaluative priming effects can also be obtained in a non-evaluative categorisation task when participants are encouraged to assign attention to the evaluative stimulus dimension. According to Spruyt and colleagues, the evaluative categorisation task and other non-evaluative tasks do not only differ with regard to the confounding of evaluative congruency and response compatibility. Another important difference is the focus of attention induced by the tasks. While the typical evaluative categorisation task focuses participants’ attention on evaluation by requiring them to categorise the targets based on the stimuli’s valence, the non-evaluative tasks, such as a semantic categorisation task, do not require participants to pay attention to the stimuli’s evaluation to give the correct response. Based on this consideration, they claim that failures to replicate the evaluative priming effect in non-evaluative tasks should be attributed to reduced evaluative (prime) processing. To test this hypothesis, Spruyt et al. (2007) manipulated participants’ attention to stimuli evaluation by implementing a task-switching procedure. Participants were presented with target pictures that portrayed positively and negatively valenced examples of the categories “animals” and “objects” (primes were also positively and negatively valenced pictures, but none of them portrayed an animal or an object). For response assignments, participants were asked to switch between two tasks, depending on a cue that was presented at target onset: participants were required to respond depending on either the valence of the targets (i.e. evaluative categorisation trials) or the semantic category (i.e. non-evaluative semantic categorisation trials). Crucially, evaluative categorisation responses were required on the majority of the trials (75%) in one group of participants (i.e. the remaining 25% trials were non-evaluative semantic categorisation responses), whereas the proportion of evaluative categorisation trials was restricted to just 25% in a second group (i.e. the majority of the trials were non-evaluative semantic categorisation responses). The results showed that when participants were asked to respond on the basis of the valence of the target stimuli on the majority of trials, reliable evaluative priming effects were found with both evaluative categorisation trials and non-evaluative semantic categorisation trials. However, when participants were required to respond to non-evaluative semantic categorisation responses on the majority of trials, evaluative priming effects occurred on neither the evaluative categorisation trials nor the non-evaluative semantic categorisation trials. These results suggest two important points. First, selective attention to the evaluative stimulus dimension should be a sufficient condition for the emergence of evaluative priming effects. Second, processes other than response competition should be involved in the evaluative priming effect because the non-evaluative semantic categorisation trials do not contain dimensional overlap between the prime and response.

We can go a step further in the examination of methodological and theoretical standpoints by combining this task-switching paradigm and the mouse tracking procedure. First, in the task-switching paradigm, because participants do not know which task should be followed before target onset, they cannot adopt the prime-triggered strategy described...
above. This will further contribute to probing the validity of the mouse tracking procedure. Second, it was expected that the results of Spruyt et al. (2007) would be replicated for URTs of the mouse tracking procedure as the outcome-focused measure. Specifically, the URTs of the congruent trials for both evaluative categorisation responses and non-evaluative semantic categorisation responses would be faster than the URTs of the incongruent trials when attention was directed at the evaluative stimulus dimension. If this was the case, whether response competition causes congruency effects could be assessed by examining mouse trajectories. To this end, we adopted a mouse tracking procedure with four choice options (e.g. Freeman, Nakayama, & Ambady, 2013).

### Methods

#### Participants

Sixty-five Japanese undergraduate students (35 female and 30 male; mean age = 20.58 years, range = 18–24 years) participated in exchange for course credit. We recruited as many participants as possible until the end of the predefined time period for the study, and the data of all participants were included in the analysis. All participants were randomly assigned to one of the two conditions: the 75% evaluation condition (n = 33) or the 25% evaluation condition (n = 32). This sample size allowed detection of a between-participants interaction effect of $d = 0.71$ with power $1-\text{Beta} = .80$ (alpha = .05). The interaction effect found by Spruyt and colleagues (2007) was $d = 0.74$.

#### Materials

Targets were 60 nouns (30 positive, 30 negative). Within each valence category, 15 targets referred to persons, whereas 15 targets referred to objects. Primes were 60 nouns (30 positive, 30 negative) that did not refer to persons or objects. These stimuli and the apparatuses used in this experiment were identical to that used in Experiment 1.

#### Procedure

In the 75% evaluation condition, to promote selective attention for the evaluative stimulus dimension, participants were required to categorise target words according to their valence on the majority of all trials (75%): the remaining 25% trials required non-evaluative semantic categorisation responses. Conversely, in the 25% evaluation condition, the proportion of evaluative categorisation trials was restricted to just 25%: the remaining 75% trials required non-evaluative semantic categorisation responses.

Participants were instructed to immediately move the mouse and click the appropriate response button on the screen: POSITIVE, NEGATIVE, PERSON, or OBJECT. These labels were located at the top-left, top-right, bottom-left, or bottom-right of the screen (see Figure 3). The assignment of these labels to the response buttons was counterbalanced across participants, but the pair of evaluative categories (positive vs. negative) and pair of semantic categories (person vs. object) was always located on the same line (top or bottom, respectively).

In the 75% evaluation condition, participants were asked to categorise the words according to their valence, unless the words appeared with a black circle. In that case, participants were asked to

![Figure 3](image-url)
categorise the targets according to their semantic categories (i.e. person vs. object). Importantly, the targets with a black circle were presented in only 25% of the trials. In the 25% evaluation condition, participants were asked to categorise the target as persons or objects, unless the words appeared with a black circle. These words were to be categorised based on their valence. For both conditions, 120 trials were divided into two blocks of 60. For each block, 60 target words were randomly assigned to one of the 60 prime words, such that the number of evocatively congruent and evocatively incongruent trials would be balanced. In both groups, 25% of the targets appeared with a black circle. The temporal details on any given trial were identical to that used in Experiment 1.

Prior to the experimental phase, there were 12 practice trials in which 6 person target words (3 positive, 3 negative) and 6 object target words (3 positive, 3 negative), which were not used in the experimental phase, were randomly assigned to one of the 12 prime words (6 positive, 6 negative) that were also not used in the experimental phase. In 25% of these trials, the target appeared with a black circle so that participants were required to categorise the targets according to their valence on exactly 9 trials in the 75% evaluation condition and 3 trials in the 25% evaluation condition.

Results

Participants completed a total of 7,800 trials (3,900 per condition). Of these, there were 136 trials with incorrect responses (1.7%), 65 trials with ITs greater than 700 ms (0.83%), 122 trials with ultimate decision times exceeding 2.5 SD (1.56%), and 170 trials (2.17%) with aberrant mouse movements, which were discarded. Overall, 95% of the trials were included in the following analyses.

URTs

First, to investigate the evaluative priming effects with an outcome-based measure, we analysed URTs (see Table 2). A 2 (condition: 75% evaluation vs. 25% evaluation) × 2 (categorisation response: evaluative categorisation vs. semantic categorisation) × 2 (evaluative congruence: evocatively congruent vs. evocatively incongruent) ANOVA revealed a significant interaction between condition and categorisation response, $F(1, 63) = 31.28, p < .001$, $\eta^2_p = .33$. When participants performed the evaluative categorisation trials in the 75% evaluation condition, they reached the correct response locations faster than in the semantic categorisation trials, $F(1, 32) = 9.42, p = .004$, $\eta^2_p = .23$. In the 25% evaluation condition, participants were faster in reaching the response locations in the case of semantic categorisation trials than in the case of evaluative categorisation trials, $F(1, 31) = 28.53, p < .001$, $\eta^2_p = .48$.

More importantly, the ANOVA also revealed a significant interaction between evaluative congruence and condition, $F(1, 63) = 16.80, p < .001$, $\eta^2_p = .21$. Further analyses showed that the effect of evaluative congruence was significant in the 75% evaluation condition, $F(1, 32) = 48.87, p < .001$, $\eta^2_p = .60$, but not in the 25% evaluation condition, $F < 1$. Importantly, both in the 75% evaluation condition and the 25% evaluation condition, the interaction between evaluative congruence and categorisation response was not significant, $Fs < 1$. That is, the effect of evaluative congruence reached significance in the 75% evaluation condition, irrespective of whether the targets were classified according to their valence, or according to their semantic category membership. Conversely, a significant effect of evaluative congruence could be detected neither with the evaluative categorisation trials nor with the non-evaluative semantic categorisation trials in the 25% evaluation condition. Thus, in line with Spruyt et al. (2007), the evaluative priming effect was found only under the 75% evaluation condition with an outcome-focused measure.

Trajectory analyses

We computed AUC as an index for response competition (see Table 2). An ANOVA revealed a significant three-way interaction, $F(1, 63) = 16.39, p < .001$, $\eta^2_p = .21$. To further investigate this interaction, we performed separate 2 (categorisation response: evaluative categorisation vs. semantic categorisation) × 2 (evaluative congruence: evocatively congruent vs. evocatively incongruent) ANOVAs, one for the 75% evaluation condition and one for the 25% evaluation condition. The ANOVA for the 75% evaluation condition yielded a significant interaction, $F(1, 32) = 45.31, p < .001$, $\eta^2_p = .58$. Follow-up analyses revealed that when participants performed the evaluative categorisation trials, the AUC in response to evocatively incongruent trials was greater compared to evocatively congruent trials, $F(1, 32) = 59.31, p < .001$, $\eta^2_p = .65$ (see Figure 4). Distributional analysis also showed that the distribution of AUC values for evocatively incongruent trials ($b = .518$) was within the $b < .555$ bimodality-free zone. In contrast, when
semantic categorisation trials were performed, there was no significant AUC difference between evaluatively congruent and evaluatively incongruent trials, $F(1, 32) < 0.01$, $p = .98$, $\eta^2_p < .01$. However, in the 25% evaluation condition, there was no significant interaction, $F(1, 31) = 0.30$, $p = .59$, $\eta^2_p < .01$. Yet, in the 25% evaluation condition, there was no significant interaction, $F(1, 31) = 0.30$, $p = .59$, $\eta^2_p < .01$. **Discussion**

Overall, the results of Experiment 2 replicated the findings of Spruyt et al. (2007, 2009), which demonstrated that evaluative priming effects of non-evaluative semantic categorisation responses can be obtained if attention is directed toward the evaluative stimulus dimension. Specifically, in the 25% evaluation condition, there were no effects for any of the 3 measures examined (URT, IT, and AUC), regardless of the types of categorisation responses. However, in the 75% evaluation condition, significant evaluative priming effects could be detected with an outcome-based measure (URT), both in the evaluative categorisation trials and in the non-evaluative semantic categorisation trials. Interestingly, the processes that caused the URT differences between congruent trials and incongruent trials were dependent upon the categorisation responses. When the participants classified the targets according to their valence, the AUC in

**Figure 4.** Mean mouse trajectories for congruent and incongruent trials are illustrated separately for evaluative and semantic responses. (A) Shows average trajectories taken from the 75% evaluation condition. (B) Shows average trajectories taken from the 25% evaluation condition.
response to evaluatively incongruent trials was greater compared to evaluatively congruent trials ($dz = 1.34$), leaving the IT unaffected. In contrast, when the participants classified the targets according to the semantic category, a shorter IT was found on evaluatively congruent trials compared to on evaluatively incongruent trials, leaving the AUC unaffected. These results might suggest that the underlying mechanisms of evaluative priming effects in evaluative and semantic categorisation tasks are different when attention is directed at the evaluative stimulus dimension. In the evaluative categorisation task, evaluative congruency affected the mouse trajectory, showing that response competition caused the evaluative priming effect. Conversely, in the semantic categorisation task, the evaluative congruency affected not the mouse trajectory, but IT. This indicates that the evaluative priming effects observed in the semantic categorisation task were caused by a mechanism other than response competition, suggesting task-dependency in the mechanism of the effects.

Moreover, Experiment 2 succeeded in ruling out the prime-triggered strategy suggested in Experiment 1. In the current experiment, because participants did not know which task would need to be performed prior to target onset, they could not adopt this strategy. In addition, ITs were longer than in Experiment 1, and the ITs of the 75% condition and 25% condition were at the same level. These results confirm the validity of the mouse tracking procedure.

**General discussion**

The purpose of the present study was to use a mouse tracking procedure to show the response dynamics of an evaluative priming effect, in an effort to gain more insight into the underlying processes of evaluative priming effects in general. These effects have been explained from an encoding and/or response perspective. The encoding perspective posits that the prime facilitates encoding of the target. Alternatively, the response perspective supposes that the prime facilitates or inhibits the response to the target. Many researchers have tested these two accounts using variations of the sequential evaluative priming paradigm (e.g. evaluative categorisation, semantic categorisation, and pronunciation tasks). However, it is difficult to determine how much the encoding and response mechanisms contribute to the evaluative priming effect since researchers have typically only used outcome-focused measures (e.g. reaction times). This is because reaction time differences do not inform researchers about the preceding stages of the evaluation process that lead to faster or slower responding, as only the ultimate output of completed decisions are available.

To overcome this limitation, we adopted a mouse tracking procedure that allows for analysis of hand movements during the execution of categorisation tasks. Recently, many researchers have argued that evaluative priming effects can be explained by response competition that operates at a response selection stage on the basis of the observation that reliable priming effects are found only in evaluative categorisation tasks (as opposed to semantic categorisation tasks; Herring et al., 2013). In Experiment 1, we obtained this asymmetrical result between two tasks with an outcome-based measure (URT) in the same manner as the previous studies. Moreover, we also obtained the same result with mouse trajectory (as assessed via AUC). These results can be a mutual validation of the most prevalent account of evaluative priming effects and a mouse tracking procedure as a means to measure response competition. Overall, the main result of Experiment 1 provides compelling evidence that the evaluative priming effect can be explained by response competition that operates at a response selection stage.

It should be noted that there is also a large volume of evidence to support the hypothesis that evaluative priming effects are at least partly due to a response competition mechanism (e.g. De Houwer et al., 2002; Klauer et al., 1997; Wentura, 1999). However, one could argue that the evidence is rather indirect, in that it is primarily based on the observation that reliable reaction time differences between evaluatively congruent and incongruent trials emerge in evaluative categorisation tasks, but not in other non-evaluative tasks in which there are no fixed matches between target valence and response (e.g. semantic categorisation task). This logic and result can help to rule out the encoding facilitation mechanism in the non-evaluative task, but cannot rule out the possibility of the effect of encoding facilitation in the evaluative categorisation task because response competition and encoding facilitation act in the same direction in a typical evaluative categorisation task. In fact, several researchers have employed measurements and methods other than reaction times in an attempt to disentangle the two processes in the evaluative categorisation task (Eder et al., 2012; Voss et al., 2013), indicating...
that response competition plays a major role in causing evaluative priming effects. In that sense, our result for the analysis of hand movement dynamics also provides additional direct evidence that evaluative priming effects in the evaluative categorisation task can mostly be explained by response competition although the two processes are not mutually exclusive.

The result of Experiment 1 seems to suggest that the occurrence of the evaluative priming effect depends upon whether dimensional overlap between the prime and response is included in a task. However, several studies have succeeded in detecting evaluative priming effects, in spite of the absence of dimensional overlap in tasks (e.g. Bargh et al., 1996; Hermans et al., 1994; Schmitz & Wentura, 2012; Spruyt, Hermans, De Houwer, Vandromme, & Eelen, 2002, 2007, 2009, 2012; Wentura & Frings, 2008). This suggests that processes other than response competition could be involved in the evaluative priming effect. In particular, Spruyt et al. (2007, 2009, 2012) found that when participants focused their attention on evaluation, evaluative priming effects emerged not only in the evaluative categorisation task, but also in the non-evaluative semantic categorisation task. Since the semantic categorisation task cannot be based on response conflict, it seems plausible that reaction time differences between evaluatively congruent and incongruent trials in the semantic categorisation task are explained by encoding facilitation. In our second experiment, using a task-switching paradigm (Spruyt et al., 2007, 2009, 2012) with a mouse tracking procedure, we succeeded at replicating these previous findings with URTs as the primary outcome measure. Interestingly, in analysing continuous hand movement, while the evaluative congruency affected AUCs in the case of the evaluative categorisation task, it affected ITs in the case of the semantic categorisation task. What do these results mean? First, since the evaluative congruency effects in the semantic categorisation task could not be found in AUCs and the evaluative congruency effects in the evaluative categorisation task occurred quite clearly with AUCs in Experiment 1, it is reasonable to assume that the evaluative priming effects in each task are based on different processes in a task-switching paradigm. Second, since mouse trajectories are sensitive to response competition (Freeman & Ambady, 2010) and the evaluative categorisation task contains dimensional overlap between the prime and response, it is plausible to interpret that the evaluative congruency effects with AUCs reflect a response competition mechanism. Finally, given that the evaluative priming effects in the semantic and evaluative categorisation tasks are based on different processes, and given that AUC reflects response competition, what do the IT differences reflect? Although it is too early to conclude the mechanism, ITs might reflect encoding facilitation, based on the previous research.

Nevertheless, we acknowledge that the interpretation of the ITs is still controversial and the degree to which the ITs relate to encoding facilitation is an open question for future research. For example, one could argue the possibility that some kind of interference effect might exert an influence on ITs rather than reflecting encoding facilitation (Werner & Rothermund, 2013). In Experiment 2, the participants could anticipate the majority categorisation task in most cases. Thus, there might be a default for what participants will expect and try to execute. With the appearance of the target, they might try to execute the majority categorisation first, regardless of whether this is the task actually demanded. For the 75% evaluation condition, since the majority (default) task was evaluative categorisation, execution of the semantic categorisation task might elicit a response conflict. Thus, a combination of this default conflict and evaluative incongruency might delay the ITs of the evaluatively incongruent trials in the semantic categorisation task. Overall, although the interpretation of the ITs is currently tentative, we took a first step in assessing the evaluative priming effects by use of a mouse tracking procedure (see Supplementary material for another cautionary indicator for the sensitivity of evaluative priming paradigms). We believe that tracking hand movements adds to the methodological repertoire that researchers can draw upon to further unravel the dynamic and multifaceted nature of evaluative priming effects.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Data statement**

The data supporting this publication can be accessed on the Open Science Framework: doi:10.17605/OSF.IO/8ZY7T.
References


Appendix

Stimuli used in Experiment 1 and 2

*Positive prime words*: 平和 (peace), 栄光 (honour), 希望 (hope), 楽園 (paradise), 朝日 (sunrise), 仏教 (Buddhism), 世界 (world), 科学 (science), 文化 (culture), 奇跡 (miracle), 音楽 (music), 芸術 (art), 芸術 (art), 光明 (light).

*Negative prime words*: 不幸 (misfortune), 災難 (disaster), 絶望 (despair), 戦争 (war), 悪臭 (bad smell), 改造 (reconstruction), 偽り (lie), 汚染 (pollution), 拷問 (torture), 極寒 (cold), 事故 (accident), 失敗 (failure), 水洪 (flood), 衝突 (collision), 騒音 (noise), 退治 (retribution), 素晴拉 (wonderful), 堕落 (corruption), 破産 (bankruptcy), 水破 (break), 腐敗 (corrupt), 退治 (retribution), 堕落 (corruption), 塩分 (salt), 花嫁 (bride), 兄弟 (brother), 友人 (friend), 勝者 (winner).
Negative person target words: 浮気者 (cheater), 暗殺者 (assassin), 極悪人 (fiend), 死者 (dead person), 侵入者 (invader), 奴隷 (slave), 泥棒 (thief), テロリスト (terrorist), 変質者 (eccentric person), 狂人 (madman), 犠牲者 (victim), 腰抜け (traitor), 罪人 (offender), 犯人 (murderer), 人質 (hostage).

Positive object words: 宝物 (treasure), 花束 (flower), 贈り物 (gift), 勲章 (decoration), 金 (gold), ダイヤモンド (diamond), ちみつ (honey), ケーキ (cake), 食物 (food), チョコ (chocolate), フルーツ (fruit), ドレス (dress), 真珠 (pearl), 王冠 (crown), 宝石 (jewellery).

Negative object words: 泥 (mud), ゴミ (garbage), 疾病 (tumour), 爆弾 (bomb), 病原菌 (bacterium), 棺桶 (coffin), 賄賂 (bribe), 拳銃 (gun), かび (mould), 毒物 (poison), ギロチン (guillotine), 埃 (dust), 注射 (injection), 武器 (weapon), 骸骨 (skeleton).