

Diffusion Model-based Understanding of Unconscious Affective Priming in Continuous Flash Suppression

Minchul Kim

Korea Advanced Institute of Science and Technology

Jeeyeon Kim

Korea Advanced Institute of Science and Technology

Jaejoong Kim

Korea Advanced Institute of Science and Technology

Bumseok Jeong (✉ bs.jeong@kaist.ac.kr)

Korea Advanced Institute of Science and Technology

Research Article

Keywords: Continuous flash suppression (CFS), binocular rivalry, Nonetheless, influences subsequent decision making

Posted Date: January 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-152941/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Diffusion model-based understanding of unconscious affective priming in continuous
flash suppression**

Author names: Minchul Kim, MD¹, Jeeyeon Kim¹,
Jaejoong Kim, MD, PhD¹, Bumseok Jeong, MD, PhD¹

Author affiliations:

¹ Graduate School of Medical Science and Engineering (GSMSE), Korea Advanced Institute of Science and Technology (KAIST)

291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea

Corresponding author:

Bumseok Jeong, MD, PhD

Graduate School of Medical Science and Engineering (GSMSE), Korea Advanced Institute of Science and Technology (KAIST)

291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea

Tel: +82-42-350-4285

E-mail: bs.jeong@kaist.ac.kr

Abstract

Affective states influence our decisions even when processed unconsciously. Continuous flash suppression (CFS) is a new variant of binocular rivalry that can be used to render the prime invisible and thus unconscious. Nonetheless, it is unclear how prior information from emotional faces provided by CFS influences subsequent decision making. Here, we employed the CFS priming task to examine the effect of nonconscious information on the evaluation of target words as either positive or negative. The hierarchical diffusion model was used to investigate the underlying mechanisms. Two experiments were performed to investigate the effects of facial identity and facial expression. As a result, a significant affective priming effect on response time was observed only for angry faces but not happy and neutral faces. The results of diffusion model analyses revealed that both the drift rate and nondecisional process are accountable for the 'positive bias' - the processing advantage of positive over negative stimuli. Priming effects of facial identity were mapped onto the drift rate and eliminated 'positive bias'. Meanwhile, positive emotional faces increased the nondecision time with negative target words. The model-based analysis implies that both facial identity and emotion are processed under CFS.

Introduction

Emotions can be attributed to people's attitudes toward unrelated objects, as suggested by "affect-as-information" ^{1,2}. This emotional influence on people's decision making has been revealed by the affective priming task ³, which examines the implicit affective association between an emotion prime and target words. In a typical affective priming task, the prime and the target are presented sequentially, and participants are instructed to indicate the valence of the target as quickly as possible. If participants categorize the target faster when it is valence-congruent with the prime (e.g., prime and target are both positive) than in the valence-incongruent case (e.g., prime is positive while the target is negative), the reaction time difference between the congruent and incongruent conditions is called the 'affective priming effect'.

Previous studies have shown that people evaluate targets based on primed affective information, and this emotion-induced bias was particularly prominent when the prime was rendered invisible because it escapes regulation by conscious awareness ^{1,4,5}. Traditionally, briefly presented primes with forward/backward noise patterns are used to mask the prime from conscious awareness ⁶. Continuous flash suppression (CFS), a variant of binocular rivalry and flash suppression, is a relatively new method with several advantages in making a prime unconscious. One notable difference between traditional masking and CFS is the duration of unconscious presentation that can be sustained: masking can render primes invisible for tens of milliseconds, whereas CFS can suppress primes from being perceived for seconds ^{7,8}. A previous report comparing traditional backward masking and CFS discovered that nonconscious affective priming under CFS is more prominent negatively valenced facial expressions ^{9,10} and can even guide eye movements ¹¹. However, how nonconscious information from CFS is processed and biases people's judgment remains unclear. There are

two main types of information in emotional faces: facial identity and expression¹². It has been suggested that low-level (i.e., facial identity) but not high-level information (i.e., emotional information from facial expressions) can be processed under interocular suppression or that very limited prime-related information may undergo processing under CFS^{5,9}. In this article, we conducted two priming experiments to investigate unconscious facial priming effects under CFS. Specifically, in Experiment 1, we used primes of neutral, inverted and scrambled faces to further investigate the priming effect of facial identity in a valence categorization task. In Experiment 2, we explored the priming effects of facial expression using positive and negative emotional faces in the same task.

We used the diffusion model to investigate the underlying mechanisms of how priming influences people's decisions. The diffusion model, initially proposed by Roger Ratcliff in 1978, is a well-developed cognitive model that accounts for the time course of human decision making in two-choice tasks¹³. The diffusion model conceptualizes a decision between two choices based on the accumulation of evidence toward one of the decision alternatives¹⁴. When a participant is asked to categorize the target, the evidence from the target is accumulated over time until it hits an upper or lower boundary. This drift process is characterized by four parameters (Figure 1A): the initial bias toward one of the alternatives (β), the total time devoted to general, nondecision-related processes (τ , which includes perceptual encoding and motor preparation), the rate of evidence accumulation (v), and the distance between decision boundaries (α)¹⁵. If a participant's reaction time (RT) is high in an experimental condition, we can determine why this is so based on these parameters. Additionally, when the RTs are similar, diffusion model analysis can reveal what different mechanisms contributed to the results (Figure 1B and C).

From previous studies, we hypothesized two candidate parameters that would be responsible for the nonconscious facial information from CFS. The first is the drift rate parameter. The associative network model, a model to explain the affective priming effect, suggests that the activation of the semantic representation of the prime increases the activation level of the semantic representation of the target^{3,16}. Voss *et al.* conducted a series of experiments and used the diffusion model to show that the associative (semantic) priming effects were mapped onto the drift rate parameter, indicating that semantic associations facilitate information uptake¹⁷. In addition, as valence categorization tasks are performed, the processing advantage of positive over negative stimuli – ‘positive bias’ – is expected to be observed¹⁸. This bias may be captured by the drift rate¹⁹. Another possible candidate is the nondecision time. One study found that affective priming with words moderated nondecision time¹⁷. Additionally, Yap *et al.* reported that when stimuli were degraded, the priming effect mapped onto both the drift rate and nondecision time²⁰. From these studies, we identified two parameters of interest; in particular, we hypothesized that the facial priming effect of CFS would be mapped onto the drift rate parameter and/or the nondecisional time.

Materials and methods

Experiment 1

Participants

To determine the sample size, we looked up several relevant studies that had tested congruent priming effects under CFS^{5,21}. Almeida *et al.* reported average standardized effect sizes of $d_z = 0.58$ and 0.51 ²². Using G*Power 3.1²³, we then determined that for the given effect sizes and type I error probability of $\alpha = 0.05$, a sample size of 19 to 25 was required to

achieve a power of 0.80 (t test for matched pairs, one-sided).

A total of twenty-one individuals (23.2 ± 3.5 years old; 15 males) were recruited via a web-based message board. All participants had normal or corrected-to-normal vision and reported having no ophthalmic disease or convulsions, which could impair the preciseness of the results. All participants provided written consent and received monetary compensation for their participation. Participants were informed not to consume caffeine drinks or cigarettes from one hour before the experiment. All participants provided written informed consent to participate in the experiment. The study was approved by the Korean Advanced Institute of Science and Technology Institutional Review Boards in accordance with the Declaration of Helsinki.

Stimuli and design

Instructions and stimuli were displayed on a Qnix 24-inch LCD monitor (75 Hz). The participants viewed the stimuli through a geoscope mirror stereoscope at a distance of approximately 50 cm with their head position fixed. The participants performed the tasks in a dimly lit soundproof chamber. Facial priming stimuli consisted of seventy neutral face images (half male) from the Karolinska Directed Emotional Faces dataset²⁴. All images were transformed into ellipse-cut grayscale images with low contrast and luminance by MATLAB using the ‘*imadjust*’ function with the following parameters: [0 1], [0.2 0.5]. As we wanted to reveal a facial identity effect in the first experiment, scrambled and inverted face images were created and used from the processed grayscale images and the ‘*randblock*’ function in MATLAB (the code to make images is available at <https://osf.io/zqnb/>). CFS stimuli were Mondrian pattern masks refreshed at a rate of 10 Hz generated by MATLAB (Figure 2B; http://martin-hebart.de/code/make_mondrian_masks.m). All stimuli were presented

surrounded by a black frame to facilitate binocular fusion, and two stimuli subtended approximately 9.1 degrees of visual angle.

In the valence categorization task, eighty personality adjectives (half positive, half negative) were used as target words. The targets were three or four syllable Korean words selected from a pool of ninety words based on a survey conducted in a separate group ($N=30$). The level of valence was balanced between positive and negative words that were matched ($t = 0.30, p = 0.76$, mean valence rate of positive and negative words = 1.78 and 1.76, respectively).

Procedure

The eye dominance of each participant was measured using the hole-in-the-card method. The participants were instructed to fix the eye gaze at the center of the stimuli. A central cross was presented to each eye except when the face stimuli and target words were displayed.

Figure 2A shows the experimental design of the nonconscious affective priming task (NAPT). In a given trial, the participants were presented with a 500-ms fixation cross surrounded by black square frames to both eyes. Following fixation, high-contrast colorful grids refreshed at a rate of 10 Hz were presented to the participant's dominant eye for 1000 ms to keep the face prime invisible. Concurrently, a low-luminance, low-contrast grayscale face prime was presented to the participant's nondominant eye for 600 ms. To enhance suppression by CFS, the face prime was presented 200 ms after the onset of colorful grids and was removed 200 ms before the offset of the colorful grids. Following the priming step, a fixation point was presented to both eyes for 500 ms. The target word was then presented at the center of each visual field to both eyes. The participants were instructed to categorize the word as either

positive or negative as quickly and accurately as possible. Speed and accuracy were equally emphasized. It was also emphasized to not ponder the meaning of the word. The participants responded by pressing one of two buttons using their dominant hand. The two button keys were the left and right direction keys on the keyboard, in which the left and right direction keys indicated ‘negative’ and ‘positive’, respectively.

To check whether participants perceived the suppressed faces, they were asked to indicate whether they saw any perceptible image other than flashing colorful grids. As soon as the word categorization response was registered, the text “Have you seen any image other than the flashing grids?” was presented with the no/yes option below. The participants responded by pressing either the left or right direction button keys indicating ‘no’ or ‘yes’, respectively.

Two hundred forty trials, divided into three blocks, were presented to each participant, yielding 40 trials per condition (3 (scrambled, neutral, and inverted faces) by 2 (positive and negative words) within-subject design). Each block contained forty positive and negative words to yield eighty trials. One block took 4 to 5 minutes depending on the participant. To ensure that the results were not affected by familiarity with the task, the participants had 20 practice trials before starting the main task and answered if they got used enough. In between the blocks, the participants were asked to take at least two minutes of rest to reduce the possibility of CFS breaking. The participants were provided with disposable artificial tears to prevent eye drying. The total experiment time for the NAPT, including the break time, was within ~30 minutes.

To check each participant’s knowledge about the target words, the participants completed an evaluation survey with the target words based on two criteria. First, they indicated how much they understood the meaning of the word on a 4-point scale from ‘very well’ (4) to ‘don’t know at all’ (1). Second, they categorized the word into one of the three categories:

positive, negative and cannot judge.

Experiment 2

Participants

A total of thirty-one individuals were recruited (21.5 ± 2.5 years old; 15 males) with the same inclusion and exclusion criteria used in Experiment 1. No one participated in both experiments.

Stimuli, design and procedure

The procedure in Experiment 2 was similar to that used in Experiment 1, with the exception of two aspects. First, the facial priming used in the second experiment was different since the purpose of the experiment was to reveal the priming effects of facial expressions. Affective stimuli consisted of seventy face images (half male) with happy, angry and neutral facial expressions from the Karolinska Directed Emotional Faces dataset²⁴. Happy and angry emotions were chosen from among the available other facial expressions since these two emotions were considered adequate to examine the pure effect of valence, and angry faces were found to modulate decisions under CFS in a previous study⁹. Second, we used a different CFS mask. In Experiment 2, we used a pattern mask with the same-sized squares generated using the Psychopy toolbox (https://perso.univ-lyon2.fr/~brogniar/notes/psychopy-continuous-flash/#head.flash_colors.exp). Three hundred and twenty trials, divided into four blocks, were presented to each participant, yielding 40 trials per condition (4-by-2 within-subject design). Each block contained twenty happy, angry, neutral, and scrambled face stimuli and forty positive and negative words to yield eighty trials.

Data analysis for Experiments 1 and 2

Preanalysis data exclusion

Trials in which participants responded to perceived images (aware trials) were excluded from the analysis^{4,5}. Next, we excluded participants who did not complete a sufficient number of within-subject trials (at least 30 trials per condition). To identify outliers, trials with RTs that were beyond two standard deviations from the mean were excluded for both unaware and aware trial analyses, as well as trials with RTs faster than 200 ms and slower than 5000 ms¹⁴.

Trials that contained words that the participants did not know the meaning of (those participants checked ‘2’ or ‘1’ on a 4-point scale) were excluded from the analysis because this would have interfered with automatic word categorization. For the trials where the participants reported knowing the meaning of the particular target words but the meaning was the opposite of what they thought, conditions for the analysis were modified to reflect one’s original implicit association for those words.

Analysis of reaction times

Since the raw RTs obtained in simple decision tasks are skewed, we used an inverse Gaussian generalized linear mixed-effects model with the identity link function²⁵. The model included main effects and interactions for face (i.e., scramble, inverted, and neutral for Experiment 1 and scramble, neutral, happy, and angry for Experiment 2) and valence (positive and negative target words) as fixed factors with random intercepts for subjects. The random slopes were not included since the model could not converge. We used the ‘lme4’ package in

the R program for statistical computing^{26,27}. Fixed effects were tested for significance using Type III Wald chi-square tests²⁸. The R code is expressed as:

```
'glmer (Reaction time ~ face + valence + face: valence + (1 | person), data = df, family
= inverse.Gaussian (link = "identity"))'
```

Hierarchical diffusion modeling

The RT and accuracy data were fitted using a hierarchical diffusion model (HDM)²⁹ that incorporated fixed effects for faces-by-valence conditions (6 conditions for Experiment 1 and 8 conditions for Experiment 2) for estimation of the mean of the drift rate (v) and the mean of the nondecision time (θ). Hierarchical models are ideally suited to handle data sets with few trials per participant²⁹. We allowed the nondecision time τ and drift rate δ to differ between persons p , conditions i , and trials j . In other words, the drift rate parameter δ (p_{ij}) was cast as a random variable with a condition-by-person specific mean v , and the nondecision time τ (p_{ij}) followed condition-by-person-specific mean θ . These two parameters were chosen from our hypothesis based on the previous literature presented in the introduction. In addition to fixed effects, the hierarchical model allows participant-level random effects for boundary separation (α), and we set the bias parameter (β , the relative starting point of the diffusion process between the two boundaries) to 0.5 assuming an unbiased diffusion process following Vandekerckhove *et al.*^{15,29}. We fitted the diffusion model with within-trial variability of the drift rate at $s = 1$, as implemented in JAGS¹⁵. Figure 1D shows a graphical model representation of the HDM, and the full JAGS model code is available to download from the Open Science Framework: <https://osf.io/zqnbtf/>.

Statistical inference for the hierarchical diffusion model

To estimate best-fitting parameters, all models were fit using Markov chain Monte Carlo (MCMC) as implemented in JAGS³⁰ with the ‘matjags’ interface (<https://github.com/msteyvers/matjags>) for MATLAB 2017b (The MathWorks, Inc., Natick, MA). For each model, we ran three chains with a burn-in period of 2000 samples, and 2000 further samples were then retained for analysis. Chain convergence was assessed via the \hat{R} statistic, where we considered $\hat{R} < 1.1$ as an acceptable value³¹. To examine differences in the parameters of interest, we examined the 95% highest density interval (HDI, the smallest region of the posterior that contains the 95% proportion of its mass). If the HDI does not contain 0, it means that there is a 95% probability that the parameter is not 0²⁹.

Results

Experiment 1

Preanalysis data exclusion

One individual was excluded since he responded as having perceived something other than CFS in many trials, resulting in less than 30 trials per condition. After rejection of the data based on our criteria, 6.4% of trials were excluded – 5.1% due to outlier RTs and 1.3% due to prime perception.

Reaction time analyses

The generalized linear mixed-effects analysis of RTs yielded a significant main effect of target word valence, $\chi^2(1) = 10.6$, $p < 0.01$, but an insignificant main effect of faces, $\chi^2(2) =$

0.341, $p = 0.843$, as well as the interaction of both factors, $\chi^2(2) = 0.271$, $p = 0.873$.

Planned comparisons, directly encoded in the model, showed that when primed with scrambled faces, RTs were higher when judging the negative words than the positive words (Figure 3A; [$t = 3.261$, $p = 0.001$]): that is, when primed by the scrambled face, responses tended to be faster for the positive words than the negative words. This ‘positive bias’ was also observed in the subgroup analyses of neutral and inverted faces [$t = 3.909$, $p < 0.001$; $t(1,503) = 3.999$, $p < 0.001$, respectively]. On the other hand, there was no significant facial priming effect by neutral and inverted faces paired with positive (Figure 3A; [$t = -0.121$, $p = 0.903$; $t = 0.417$, $p = 0.676$, respectively]) or negative words (Figure 3A; [$t = 0.47$, $p = 0.638$; $t = 0.429$, $p = 0.667$, respectively]).

Hierarchical diffusion model analyses

Assessment of convergence and model fit

The \hat{R} statistic was below 1.02 for all variables, indicating convergence of the MCMC chains to stationary posterior distributions. The correlation between empirical data and the model’s predicted RT quantiles ranged from $r = 0.95$ to 0.98 for the 6 (3-by-2) conditions (Figure S1A).

Model parameter analysis of posterior estimates

Summary statistics of the drift rate (μ_v) and the mean of the nondecision time (μ_0) per condition are given in Table 1. We did not observe any difference between parameters based on the primes (Figure S2; HDI included the null value of 0). To investigate the valence effect

of the target words, we computed the difference in the mean of the drift rate and the mean of the nondecision time by valence for each face (Figure 4A). Relative to the positive target word condition, the drift rate tended to decrease, and the nondecision time increased. When the effects were added across faces (i.e., (Scramble, positive + Neutral, positive + Inverted, positive) – (Scramble, negative + Neutral, negative + Inverted, negative)), the HDI of the word valence effect did not contain 0 for either drift rate or nondecision time (Figure 4A, last row). This result indicated that the ‘positive bias’ effect observed in the RT analysis was due to both an increase in the drift rate and a decrease in the nondecision time.

Experiment 2

Preanalysis data exclusion

One participant was excluded from the analysis due to a programming error. Two participants with word categorization accuracy less than 90% were excluded from the analysis. They were 68.4% and 69.6% correct, while others were $93.8 \pm 1.6\%$ correct, and we assumed that those two outliers did not work on the task seriously. Eight individuals were rejected because they responded as having perceived something other than CFS in many trials, resulting in fewer than 30 trials per condition. Data from the remaining 20 participants were analyzed, and overall, 7.4% of these trials were excluded – 4.9% due to outlier RTs and 2.5% due to prime perception.

Reaction time analyses

The generalized linear mixed-effects analysis of RTs yielded a significant main effect of target word valence, $\chi^2(1) = 5.25$, $p = 0.021$, marginal significance of the interaction of both

factors, $\chi^2(3) = 7.72$, $p = 0.051$, but an insignificant main effect of faces, $\chi^2(3) = 4.91$, $p = 0.178$.

Planned comparisons, directly encoded in the model, showed that RTs were higher when positive target words were primed by the angry face (Figure 3B; [$t = 2.181$, $p = 0.029$]) than the scrambled face. Additionally, there was a significant interaction effect between the angry face prime and the negative target word (Figure 3B; [$t = -2.561$, $p = 0.008$]); the RT difference between positive and negative target words was significantly decreased when primed by the angry face versus the scrambled face. This interaction was due to the ‘positive bias’ of scrambled face priming. When primed with invisible scrambled faces, the response times were significantly faster with the positive target word than with the negative target word (Figure 3B; [$t = 2.304$, $p = 0.021$]); however, the effect was the opposite when primed with the angry face.

In summary, analysis of the RTs suggested that subliminal angry face priming delayed the response to positive words, which is an emotionally incongruent condition, and enhanced judgment when paired with negative words (emotionally congruent conditions). No other faces, including the happy and neutral faces, showed statistical significance. In addition, ‘positive bias’ was observed in the scrambled face prime conditions.

Hierarchical diffusion model analyses

Assessment of convergence and model fit

The \hat{R} statistic was below 1.01 for all variables, indicating convergence of the MCMC chains to stationary posterior distributions. The correlation between empirical data and the model’s predicted RT quantiles ranged from $r = 0.88$ to 0.98 for the 8 (4-by-2) conditions (Figure S1B).

Model parameter analysis of posterior estimates

Summary statistics of the drift rate (μ_v) and the mean of the nondecision time (μ_0) per condition are given in Table 1. The scrambled face prime showed a significantly higher drift rate when categorizing the positive target word than the negative target word (Figure 4B; first row, left column). However, a decrease in the drift rate was not observed with any of the three emotional face primes (Figure 4B; 2, 3, and 4th row, left column). In other words, a significant decrease in the drift rate was observed only in scrambled face priming (i.e., the HDI did not contain 0). In contrast, the means of the nondecision time (μ_0) were significantly increased when primed with happy or neutral faces (Figure 4B; 2, and 3rd row, right column), which was not observed with the angry and scrambled face primes. We further investigated the nondecision time difference between conditions, and only the ‘Happy, negative’ condition was greater than the ‘Angry, negative’ condition (HDI = [0.003 0.040]), indicating that a happy face interfered with judging negative words via an increase in the nondecision time.

Discussion

We conducted two experiments to reveal how unconscious facial priming affects subsequent decisions. In Experiment 1, we attempted to identify the facial identity effect using inverted and upright neutral faces. What was repeatedly found with the primes in Experiment 1 was the ‘positivity bias’ effect¹⁸, i.e., a processing advantage of positive over negative stimuli. Using diffusion modeling, we found that both the drift rate and nondecision time changed based on the target word valence (Figure 4A, last row). This result is similar to previous reports using a diffusion model to investigate emotional word processing¹⁹. Mueller *et al.* reported that the

drift rate was associated with a processing advantage of happy words over fear words¹⁹. However, they did not allow the nondecision time to vary between conditions. Our finding adds that not only drift rate but also nondecision time accounts for the ‘positive bias’ effect.

The weak effect of subliminal facial priming observed in Experiment 1 was different than what we hypothesized. In Experiment 2, we observed the ‘positive bias’ effect only in the scrambled face priming condition. In contrast, Experiment 1 showed a ‘positive bias’ effect in all facial priming conditions, implying that subliminal facial priming was not effective. We carefully suggest that awareness of the facial primes may be the underlying reason for this discrepancy between the two experiments. The age and sex of participants in both experiments were not significantly different ($t = 1.32$, $p = 0.19$; $\chi^2 = 2.31$, $p = 0.12$). However, we used different CFS masking methods in the two experiments (Figure 2). We changed to an open-source mask considering the possibility that the Mondrian mask we generated may hide or confound the effect of the experiment. When we used the CFS masking image shown in Figure 2B, the participants reported seeing something other than the flash in 20.9% of the total trials. In contrast, when we used Mondrian images (Figure 2A) as CFS masking, the participants reported awareness of subliminal faces in only 1.5% of the trials. Although we included only the trials that participants reported seeing only the flash, we might suggest that different levels of subconscious processing may have existed between the two experiments. In Experiment 1, most of the participants were not aware of the subliminal faces. We suggest that CFS masking was very effective, resulting in much degraded and unconscious, therefore weak, priming effects.

Experiment 2 was conducted to reveal the priming effects of facial expressions using subliminal face primes with positive and negative target words. Our results based on the RTs in Experiment 2 revealed an emotional congruency effect by the angry face prime but not the

happy and neutral face primes. The RTs were significantly higher when angry faces were paired with positive target words than when scrambled faces were paired with positive target words, and the interaction with target word valence was also significant. This result is in accordance with previous studies reporting the affective priming effect using CFS. Experiments using CFS have consistently shown that negative emotional expressions (i.e., fearful or angry faces) gain privileged access to visual awareness over faces with neutral or happy expressions^{10,32}.

We subsequently used the HDM to see how the underlying cognitive processes mapped onto different parameters. In the scrambled face priming condition, the mean drift rate was significantly higher with positive target words. However, this change in drift rate commonly disappeared with all three emotional faces (Figure 4B, left column), resulting in the disappearance of a ‘positive bias’. Since all three kinds of emotional faces have facial identity, this result suggests that the priming effect with a higher drift rate in the negative condition may be related to facial identity. Another possible explanation is the familiarity toward the prime: the scrambled images look similar to each other, while emotional faces are not.

With respect to the nondecision time, when judging negative words, the happy face prime showed the longest nondecision time, followed by the neutral and angry faces (Table 1, Experiment 2). This is in the order of emotional valence (positive – neutral – negative). Additionally, only the happy and neutral face primes showed significant increases in nondecision times (Figure 4B, right column), thereby showing an emotionally incongruent effect. This result implied that the emotional information in the facial expression was reflected in the nondecision time. The nondecision time represents the encoding at the stimulus stage; thus, the emotional content in the face provides a head start to the decision process.

In summary, we revealed that not only facial identity but also emotional information from emotional face priming is being processed under CFS. Our result from the simple RT

analysis replicated previous studies – only angry faces showed a significant priming effect. By using the diffusion model, we found further evidence that the two types of information in the emotional face were preserved under CFS. This is a main advantage of diffusion model analysis in that we can distinguish underlying processes even between conditions with similar RTs¹⁴. Additionally, comparing the two experiments, we suggest that the depth of unconsciousness may affect the CFS priming effect.

There are several limitations in this study. First, the sample size was relatively small to draw generalizations with statistical significance. However, our main goal was to detect differences between priming effects, which are within-subject measures, rather than individual differences. As discussed earlier, we used HDM, ideally suited to handle data sets with few trials per participant, to estimate parameters to obtain a reliable measure of parameters per condition. Second, the present study did not elucidate the neural mechanism regarding affective priming. In future studies, it is necessary to investigate the neural mechanisms underlying the generation of affective response bias using electroencephalography or functional MRI. Third, we did not collect or match the arousal ratings of target words. Although we matched the valence, the arousal of the words may confound since the arousal of the target can possibly influence the valence judgment. Fourth, we did not counterbalance the response mapping of the “negative” and “positive” responses on the direction keys. Fifth, in this task, we excluded perceived trials based on participants’ subjective reports. Some existing objective tasks to ensure this report could improve the reliability of the result^{5,33}. Sixth, although we used grayscale, elliptical, and low-contrast primes, the difference in race between participants and primes may limit the generalizability of this study. Seventh, we only analyzed RT and did not conduct an analysis of the error rates, following previous studies on similar topics^{5,9}. Since we make our data available, anyone who is interested can investigate.

Acknowledgments and Disclosures

This research was supported by the Brain Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science & ICT (NRF-2016 M3C7A1914448 and NRF-2017 M3C7A1031331). The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Open Practices Statement (Data and code availability)

Data and code used in the reported analyses are available to download from the Open Science Framework: <https://osf.io/zqnb/>.

Table 1. Posterior estimates (mean and highest density interval) for the mean of the nondecision time and drift rate in the Experiment 1 and 2.

Experiment 1		
Drift rate		
Scrambled	Positive	3.15 [2.89 3.53]
Scrambled	Negative	3.04 [2.71 3.38]
Inverted	Positive	3.20 [2.89 3.59]
Inverted	Negative	2.94 [2.67 3.23]
Neutral	Positive	3.21 [2.91 3.49]
Neutral	Negative	2.91 [2.63 3.20]
Nondecision time		
Scrambled	Positive	0.49 [0.48 0.50]
Scrambled	Negative	0.51 [0.50 0.52]
Inverted	Positive	0.49 [0.48 0.50]
Inverted	Negative	0.51 [0.50 0.52]
Neutral	Positive	0.50 [0.48 0.51]
Neutral	Negative	0.51 [0.50 0.51]
Experiment 2		
Drift rate		
Scrambled	Positive	2.69 [2.48 2.90]
Scrambled	Negative	2.40 [2.19 2.60]
Neutral	Positive	2.52 [2.32 2.72]
Neutral	Negative	2.70 [2.44 2.95]
Happy	Positive	2.72 [2.49 2.94]

Happy	Negative	2.80 [2.56 3.07]
Angry	Positive	2.52 [2.33 2.71]
Angry	Negative	2.64 [2.42 2.86]
Nondecision time		
Scrambled	Positive	0.46 [0.44 0.48]
Scrambled	Negative	0.46 [0.44 0.48]
Neutral	Positive	0.45 [0.43 0.47]
Neutral	Negative	0.48 [0.46 0.49]
Happy	Positive	0.46 [0.45 0.48]
Happy	Negative	0.49 [0.47 0.50]
Angry	Positive	0.46 [0.45 0.48]
Angry	Negative	0.47 [0.45 0.48]

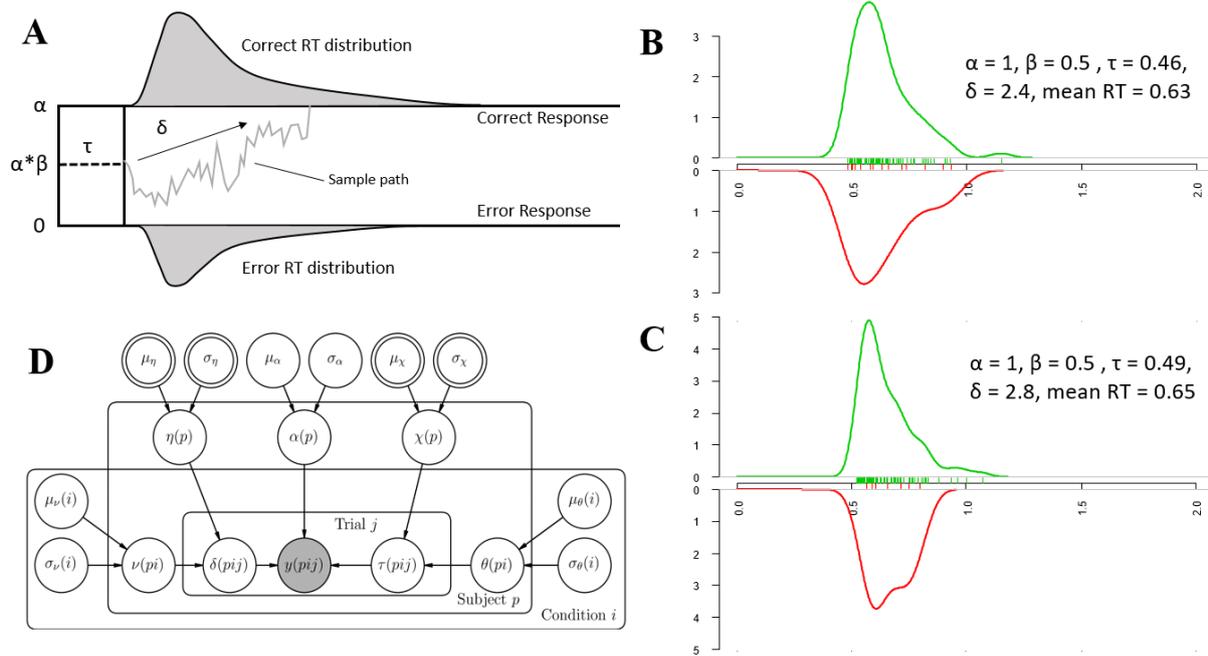


Figure 1. Graphical illustration representing the Wiener diffusion model and the hierarchical diffusion model (HDM). **A.** A graphical illustration of the Wiener diffusion model. α = boundary separation indicating the evidence required to make a response; β = initial bias indicating the a priori status of the evidence counter as a proportion of α ; δ = average rate of information uptake; τ = time used for everything except making a decision. The picture is drawn based on Vandekerckhove *et al.*²⁹. **B. and C.** Plots show generated reaction times (RT) via the ‘RWiener’ package³⁴ using the parameters shown in each figure. Although different parameters were used to generate the data, the mean reaction times were similar (two-sample t-test; $t = 1.02, p = 0.31$). **D.** Graphical notation of the model used in parameter estimation. We followed the notation method suggested in the book ‘Bayesian Cognitive Modeling’³⁵. The node distinctions between observed versus unobserved variables use shaded and unshaded nodes, and those between stochastic versus deterministic variables use single- and double-bordered nodes. Subscripts and plates indicate repetitions of the parameter across participants p , conditions i , and trials j .

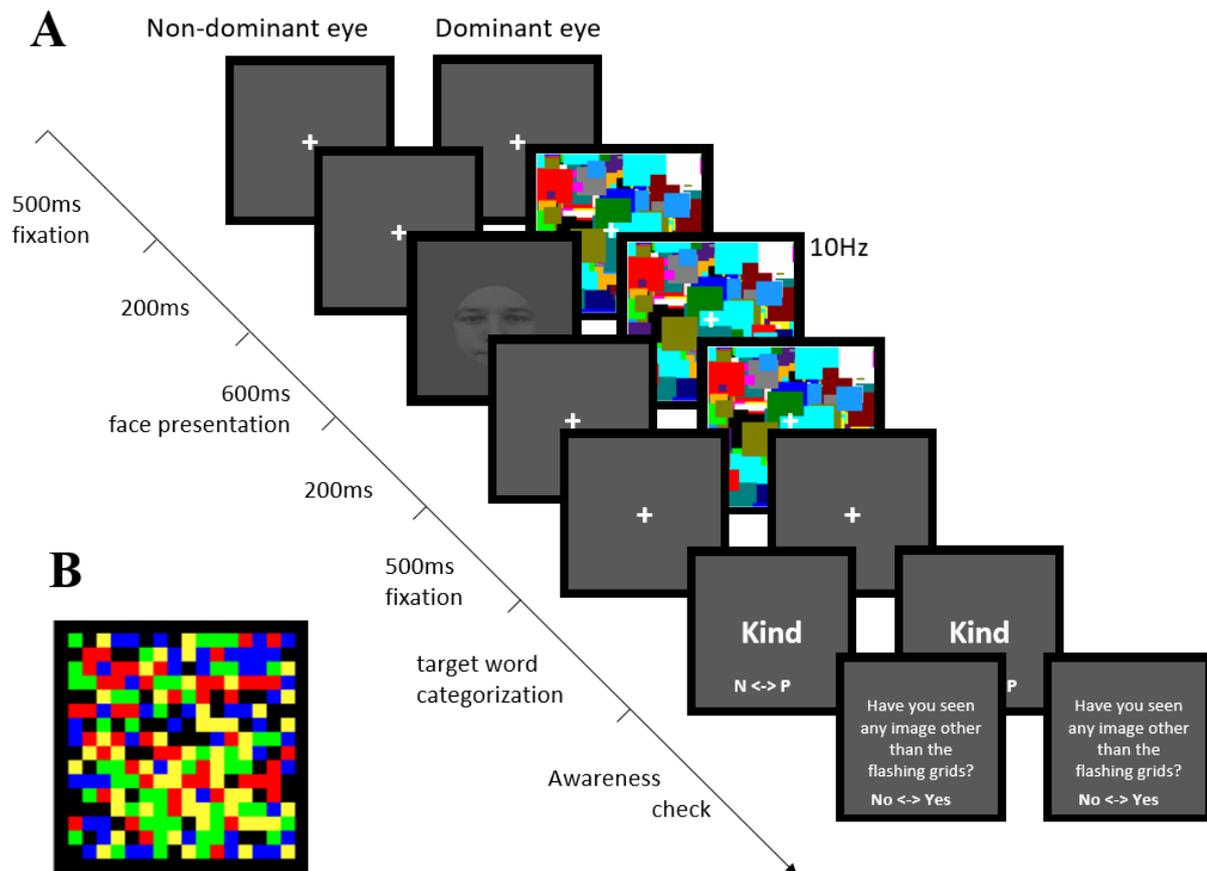


Figure 2. **A.** Schematic representation of the experimental trial structure. Following a fixation cross to both eyes, colorful flash grids were presented concurrently with face stimuli, inducing continuous flash suppression (CFS). Then, word categorization and awareness checks were performed in sequence. Face images were retrieved from the Karolinska Directed Emotional Faces database (KDEF, <http://www.emotionlab.se/resources/kdef>). **B.** CFS mask image used in Experiment 2.

Reaction time - estimates (Fixed effects) of generalized linear models

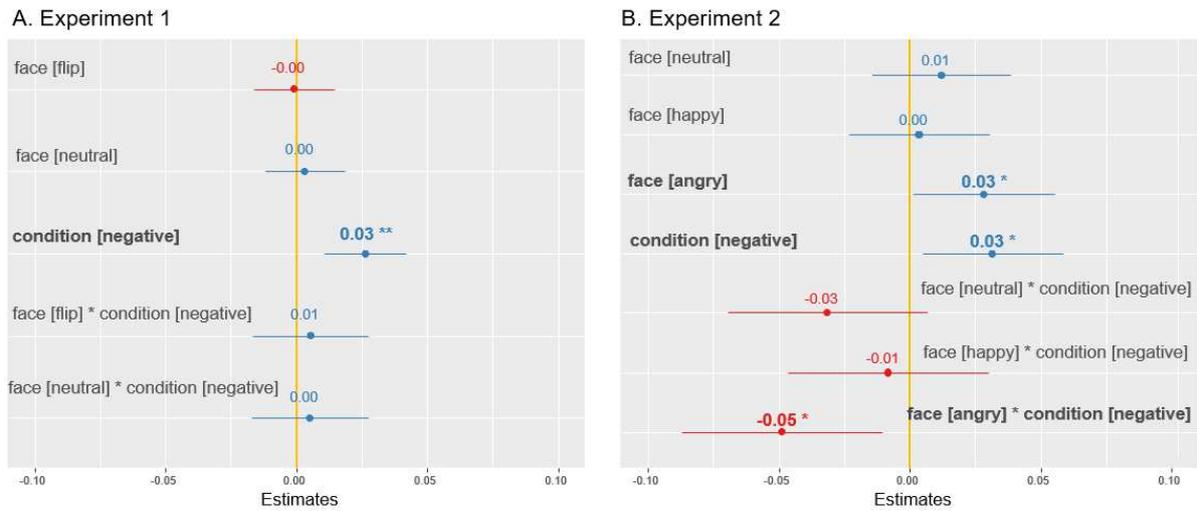


Figure 3. Estimates of the fixed effect on the reaction time in Experiment 1 (A) and Experiment 2 (B). The intercept is the mean of the reaction time at the base level (scrambled face prime with positive target word). The estimate indicates how much the reaction time increases with the fixed effect. The “neutral” line, the vertical intercept that indicates no effect, is drawn in yellow. Significant effects are in bold.

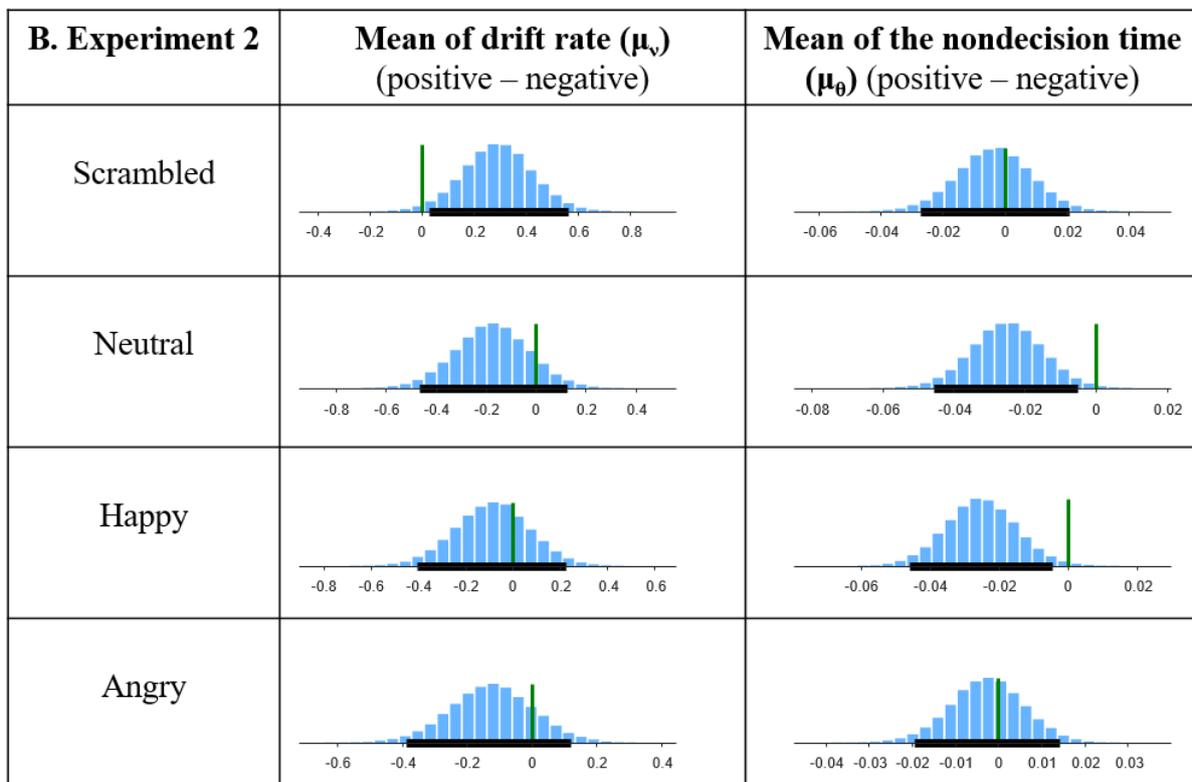
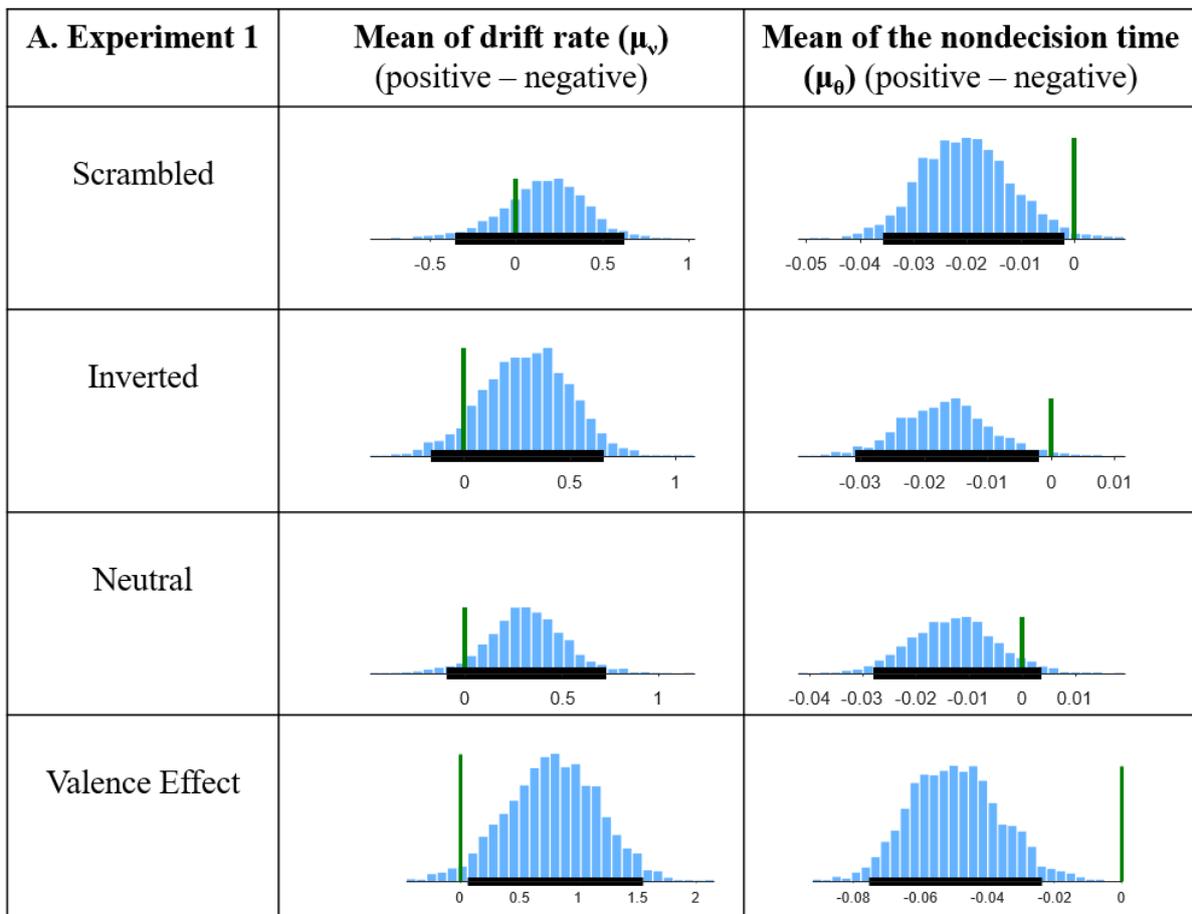


Figure 4. The posterior estimate differences in the mean of the drift rate and the mean of the nondecision time per condition The ‘valence effect’ in Figure 4A refers to ‘(Scramble, positive + Neutral, positive + Inverted, positive) – (Scramble, negative + Neutral, negative + Inverted, negative)’.

References

- 1 Lapate, R. C., Rokers, B., Li, T. & Davidson, R. J. Nonconscious emotional activation colors first impressions: A regulatory role for conscious awareness. *Psychological science* **25**, 349-357 (2014).
- 2 Schwarz, N. & Clore, G. L. Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of personality and social psychology* **45**, 513 (1983).
- 3 Fazio, R. H. On the automatic activation of associated evaluations: An overview. *Cognition & Emotion* **15**, 115-141 (2001).
- 4 Anderson, E., Siegel, E., White, D. & Barrett, L. F. Out of sight but not out of mind: Unseen affective faces influence evaluations and social impressions. *Emotion* **12**, 1210 (2012).
- 5 Yang, Y.-H. & Yeh, S.-L. Unconscious processing of facial expression as revealed by affective priming under continuous flash suppression. *Psychonomic bulletin & review* **25**, 2215-2223 (2018).
- 6 Yuan, J., Hu, X., Lu, Y., Bodenhausen, G. V. & Fu, S. Invisible own-and other-race faces presented under continuous flash suppression produce affective response biases. *Consciousness and cognition* **48**, 273-282 (2017).
- 7 Tsuchiya, N. & Koch, C. Continuous flash suppression reduces negative afterimages. *Nature neuroscience* **8**, 1096-1101 (2005).
- 8 Tsuchiya, N., Koch, C., Gilroy, L. A. & Blake, R. Depth of interocular suppression associated with continuous flash suppression, flash suppression, and binocular rivalry. *Journal of vision* **6**, 6-6 (2006).
- 9 Almeida, J., Pajtas, P. E., Mahon, B. Z., Nakayama, K. & Caramazza, A. Affect of the unconscious: visually suppressed angry faces modulate our decisions. *Cognitive, Affective, & Behavioral Neuroscience* **13**, 94-101 (2013).
- 10 Yang, E., Zald, D. H. & Blake, R. Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion* **7**, 882 (2007).
- 11 Vetter, P., Badde, S., Phelps, E. A. & Carrasco, M. Emotional faces guide the eyes in the absence of awareness. *ELife* **8**, e43467 (2019).
- 12 Posamentier, M. T. & Abdi, H. Processing faces and facial expressions. *Neuropsychology review* **13**, 113-143 (2003).
- 13 Ratcliff, R. A theory of memory retrieval. *Psychological review* **85**, 59 (1978).
- 14 Voss, A., Nagler, M. & Lerche, V. Diffusion models in experimental psychology: A practical introduction. *Experimental psychology* **60**, 385 (2013).
- 15 Wabersich, D. & Vandekerckhove, J. Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavior research methods* **46**, 15-28 (2014).
- 16 Houwer, J. D., Hermans, D., Rothermund, K. & Wentura, D. Affective priming of semantic categorisation responses. *Cognition & Emotion* **16**, 643-666 (2002).
- 17 Voss, A., Rothermund, K., Gast, A. & Wentura, D. Cognitive processes in associative and categorical priming: A diffusion model analysis. *Journal of Experimental Psychology: General* **142**, 536 (2013).
- 18 Kauschke, C., Bahn, D., Vesker, M. & Schwarzer, G. The role of emotional valence for the processing of facial and verbal stimuli—positivity or negativity bias? *Frontiers in psychology* **10**, 1654 (2019).
- 19 Mueller, C. J. & Kuchinke, L. Individual differences in emotion word processing: A diffusion model analysis. *Cognitive, Affective, & Behavioral Neuroscience* **16**, 489-501

- (2016).
- 20 Yap, M. J., Balota, D. A. & Tan, S. E. Additive and interactive effects in semantic priming: Isolating lexical and decision processes in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **39**, 140 (2013).
- 21 Hesselmann, G., Darcy, N., Ludwig, K. & Sterzer, P. Priming in a shape task but not in a category task under continuous flash suppression. *Journal of Vision* **16**, 17-17 (2016).
- 22 Almeida, J. *et al.* Grasping with the eyes: The role of elongation in visual recognition of manipulable objects. *Cognitive, Affective, & Behavioral Neuroscience* **14**, 319-335 (2014).
- 23 Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* **39**, 175-191 (2007).
- 24 Lundqvist, D., Flykt, A. & Öhman, A. The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet* **91**, 2.2 (1998).
- 25 Lo, S. & Andrews, S. To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in psychology* **6**, 1171 (2015).
- 26 Bates, D., Maechler, M., Bolker, B. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-4. [http. CRAN. R-project.org/package=lme4](http://CRAN.R-project.org/package=lme4) (2013).
- 27 Team, R. C. R: A language and environment for statistical computing. (2013).
- 28 Fox, J. & Weisberg, S. *An R companion to applied regression*. (Sage publications, 2018).
- 29 Vandekerckhove, J., Tuerlinckx, F. & Lee, M. D. Hierarchical diffusion models for two-choice response times. *Psychological methods* **16**, 44 (2011).
- 30 Plummer, M. in *Proceedings of the 3rd international workshop on distributed statistical computing*. 1-10 (Vienna, Austria).
- 31 Gelman, A. & Hill, J. *Data analysis using regression and multilevel/hierarchical models*. (Cambridge university press, 2006).
- 32 Tsuchiya, N., Moradi, F., Felsen, C., Yamazaki, M. & Adolphs, R. Intact rapid detection of fearful faces in the absence of the amygdala. *Nature neuroscience* **12**, 1224-1225 (2009).
- 33 Lin, Z. & He, S. Seeing the invisible: The scope and limits of unconscious processing in binocular rivalry. *Nature Precedings*, 1-1 (2008).
- 34 Wabersich, D. & Vandekerckhove, J. The RWiener Package: an R Package Providing Distribution Functions for the Wiener Diffusion Model. *R Journal* **6** (2014).
- 35 Lee, M. D. & Wagenmakers, E.-J. *Bayesian cognitive modeling: A practical course*. (Cambridge university press, 2014).

Author contributions

All authors designed the experiment. M.K. and J.K. collected the data. M.K. and J.K. analysed the data. M.K., J.K. and B.J. wrote the manuscript.

Figures

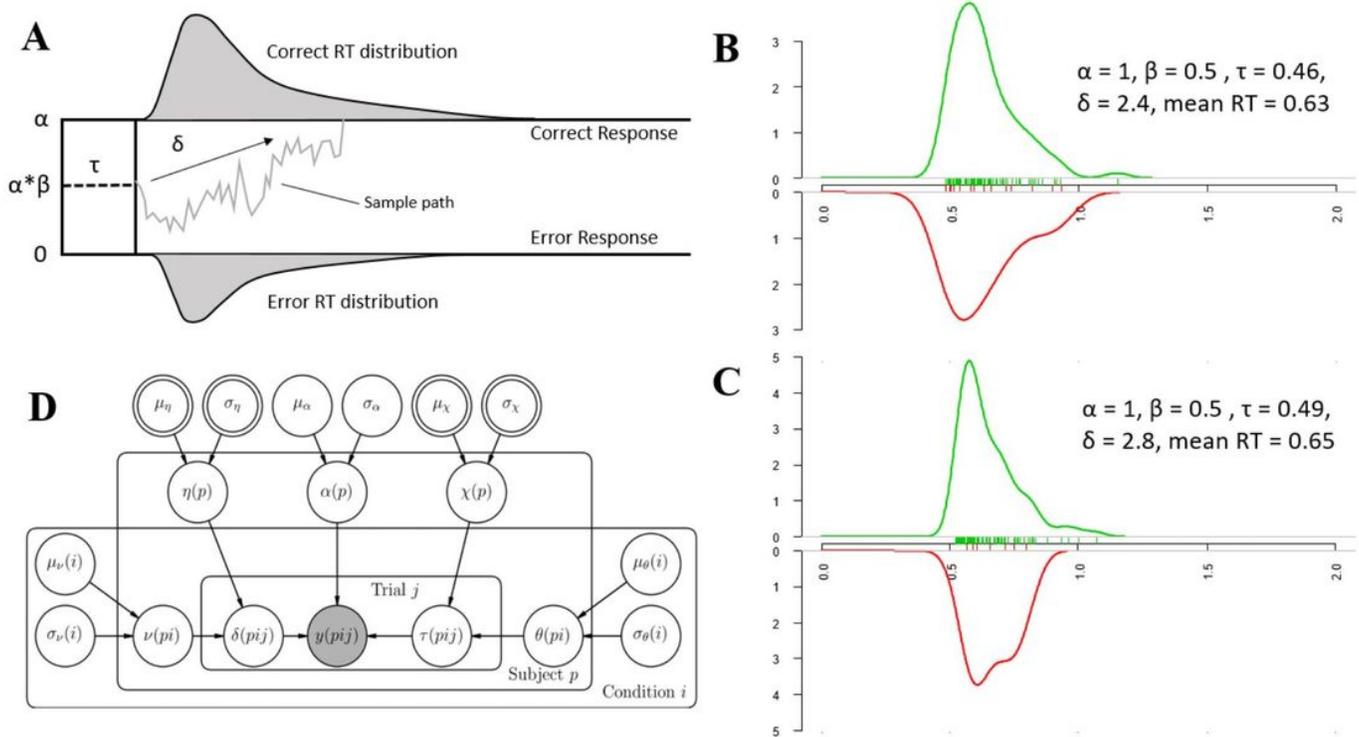


Figure 1

Graphical illustration representing the Wiener diffusion model and the hierarchical diffusion model (HDM). A. A graphical illustration of the Wiener diffusion model. α = boundary separation indicating the evidence required to make a response; β = initial bias indicating the a priori status of the evidence counter as a proportion of α ; δ = average rate of information uptake; τ = time used for everything except making a decision. The picture is drawn based on Vandekerckhove et al. 29. B. and C. Plots show generated reaction times (RT) via the 'RWiener' package 34 using the parameters shown in each figure. Although different parameters were used to generate the data, the mean reaction times were similar (two-sample t-test; $t = 1.02, p = 0.31$). D. Graphical notation of the model used in parameter estimation. We followed the notation method suggested in the book 'Bayesian Cognitive Modeling' 35. The node distinctions between observed versus unobserved variables use shaded and unshaded nodes, and those between stochastic versus deterministic variables use single- and double-bordered nodes. Subscripts and plates indicate repetitions of the parameter across participants p , conditions i , and trials j .

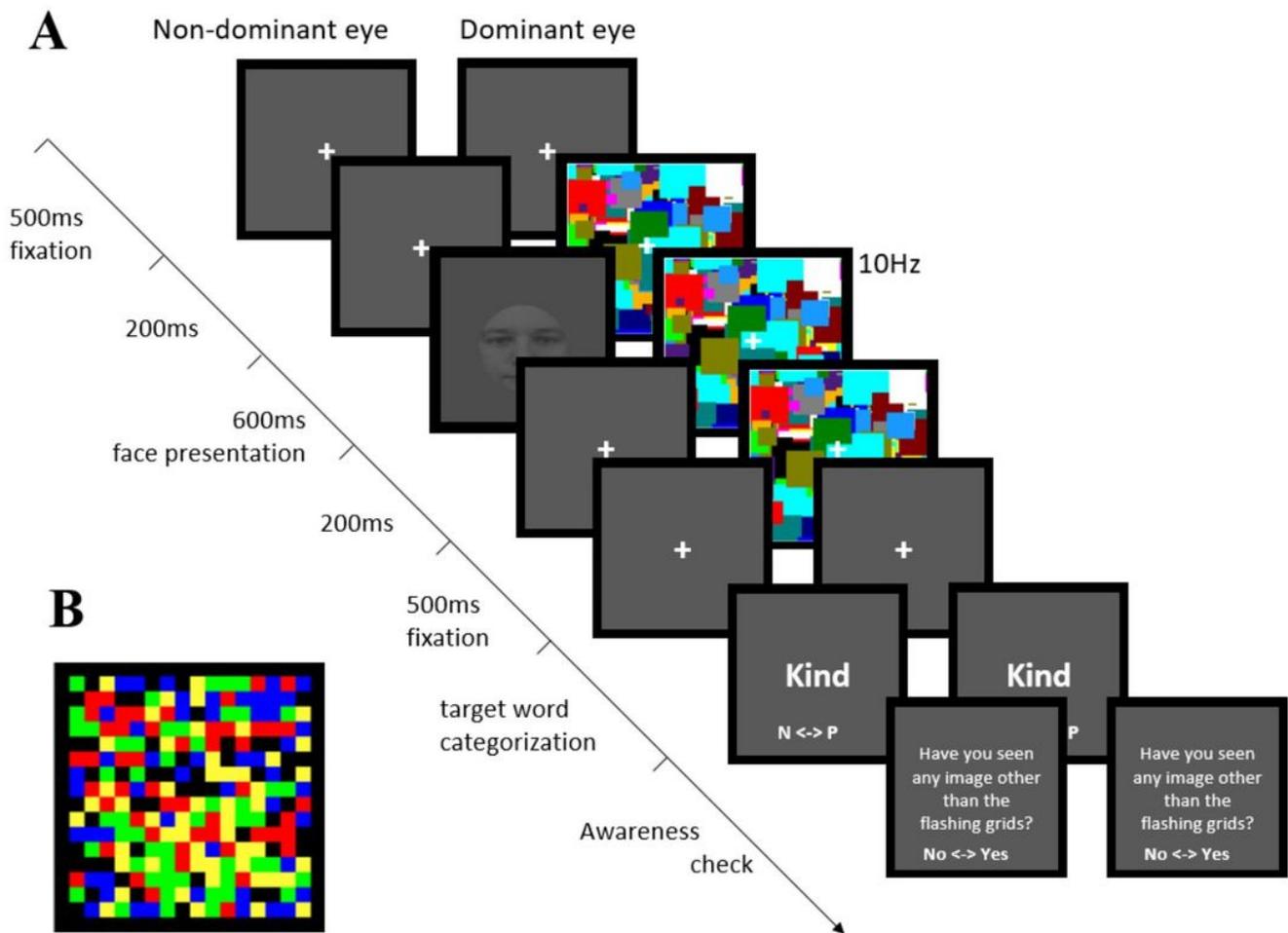
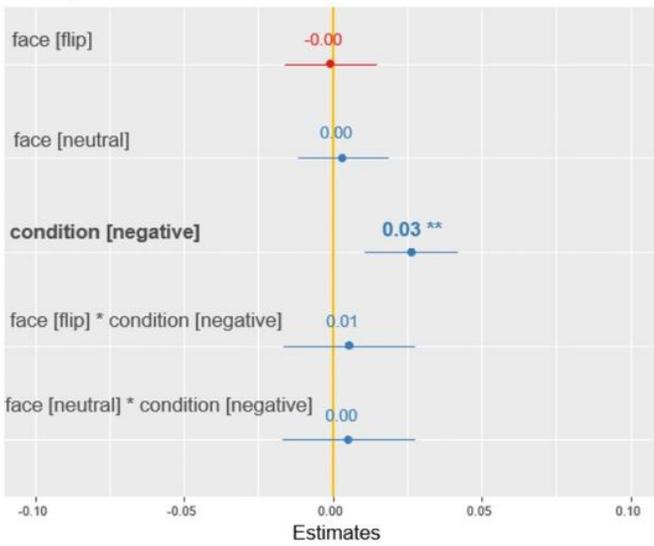


Figure 2

A. Schematic representation of the experimental trial structure. Following a fixation cross to both eyes, colorful flash grids were presented concurrently with face stimuli, inducing continuous flash suppression (CFS). Then, word categorization and awareness checks were performed in sequence. Face images were retrieved from the Karolinska Directed Emotional Faces database (KDEF, <http://www.emotionlab.se/resources/kdef>). B. CFS mask image used in Experiment 2.

Reaction time - estimates (Fixed effects) of generalized linear models

A. Experiment 1



B. Experiment 2

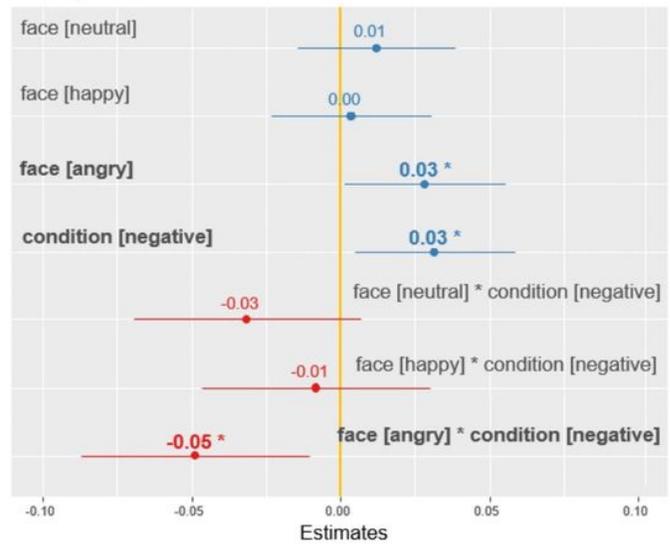


Figure 3

Estimates of the fixed effect on the reaction time in Experiment 1 (A) and Experiment 2 (B). The intercept is the mean of the reaction time at the base level (scrambled face prime with positive target word). The estimate indicates how much the reaction time increases with the fixed effect. The “neutral” line, the vertical intercept that indicates no effect, is drawn in yellow. Significant effects are in bold.

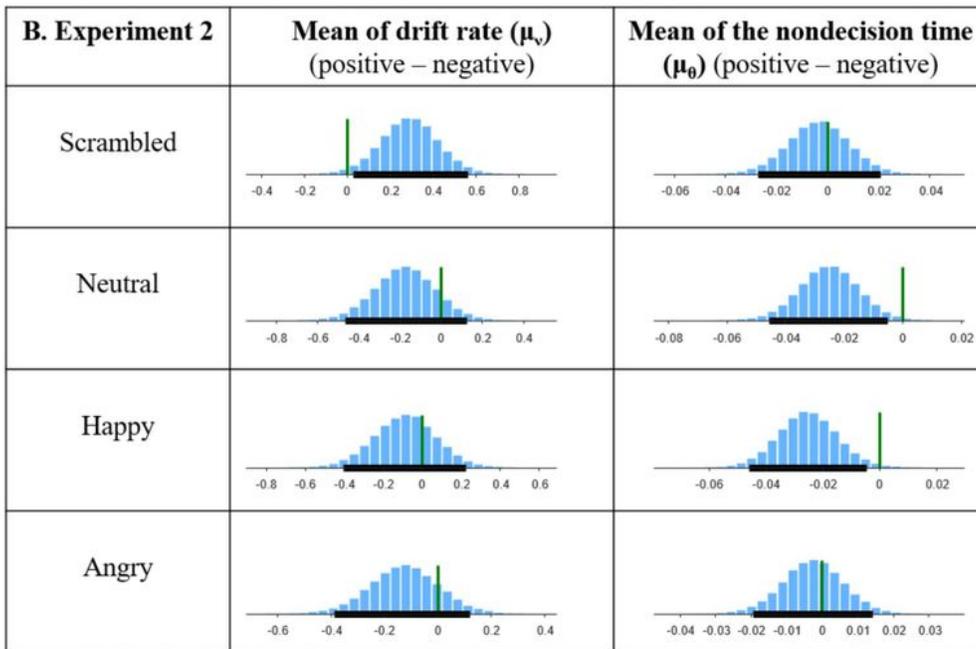
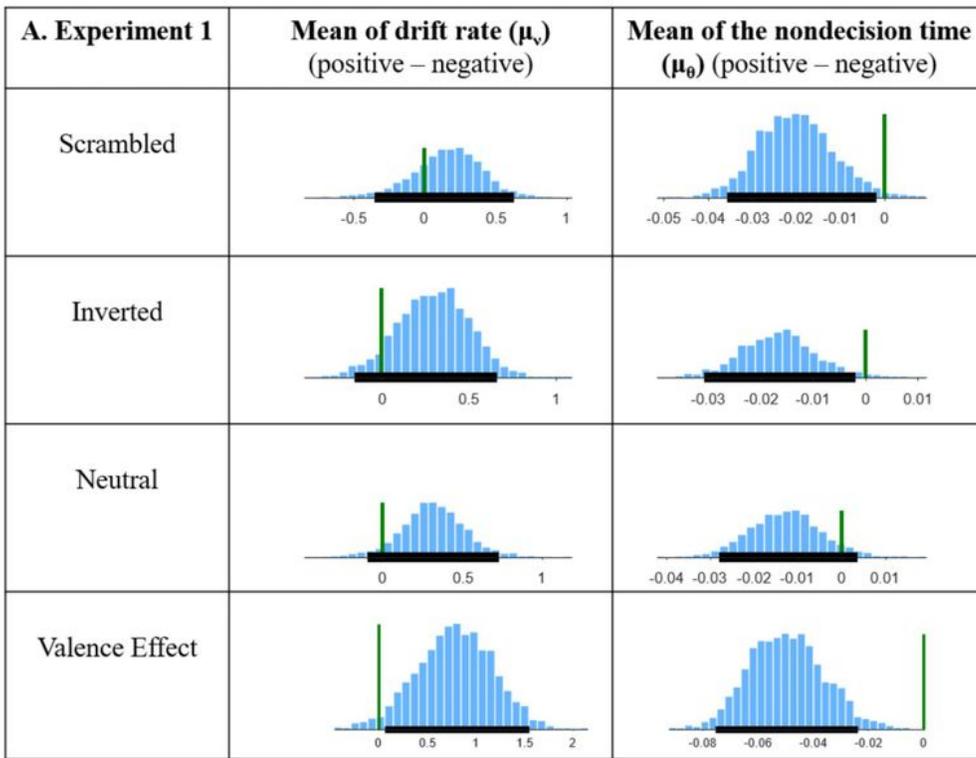


Figure 4

The posterior estimate differences in the mean of the drift rate and the mean of the nondecision time per condition. The 'valence effect' in Figure 4A refers to '(Scramble, positive + Neutral, positive + Inverted, positive) – (Scramble, negative + Neutral, negative + Inverted, negative)'.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplmentarymaterial.docx](#)