AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

© 2023 American Psychological Association
ISSN: 0096-3445

2024, Vol. 153, No. 1, 102–121
https://doi.org/10.1037/xge0001483

Interaction Between Top-Down Decision-Driven Congruency Effect and Bottom-Up Input-Driven Congruency Effect Is Correlated With Conscious Awareness

Ze-Fan Zheng[1, 2], Shu-Yue Huang[1], Shena Lu[1], and Yong-Chun Cai[1]
[1] Department of Psychology and Behavioural Sciences, Zhejiang University
[2] Research Group Neural Circuits, Consciousness, and Cognition, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany

In a conventional (Stroop) priming paradigm, it was well documented that objective prime-target incongruency delays response time (RT) to target compared to prime-target congruent condition. Recent evidence suggests that incongruency between the target and subjectively reported prime identity also delays RT over and above the classic congruency effect. When the prime is rendered invisible, the former effect is fundamentally a bottom-up (BU) stimulus-driven congruency effect and the latter a top-down (TD) guess-driven congruency effect. An influential theory of consciousness, global neuronal workspace theory, postulates that the long-lasting simultaneous and reciprocal interaction between TD decision network and BU input network is preserved during conscious processing and disabled during unconscious processing. Current study is focused on testing this theoretical postulation using two behavioral experiments. Our results showed that indeed TD-congruency and BU-congruency produced additive RT effects on prime-invisible trials, which implies that TD and BU prime representations are activated in independent neuronal populations. Meanwhile, an underadditive interaction effect was observed as prime visibility rose, which is a signature that TD and BU prime representations recruited overlapping neuronal populations during conscious perception. In addition, we suggest that current behavioral paradigm might be a financially friendly alternative to detect the presence of representational overlap in the brain between a wide range of mental representations, such as expectation, prediction, conscious/unconscious perception, and conscious/unconscious working memory.

Public Significance Statement

When the perceptual environment is noisy, what we think we hear or see (subjective identity) often trumps what we in fact hear or see (objective identity). Previous literature has shown that both subjective and objective identities of previous experience could significantly influence our current behavior. In this study, we demonstrated that these two identities of previous experience impact our current behavior completely independently when we have no conscious experience of seeing the sensory input. However, they shape our current behavior in an interactive fashion as the visibility of the sensory input increases. This observation is highly congruent with the global neuronal workspace theory of consciousness, which predicts representations from both BU stimulus-driven processing (objective identity) and TD decision-driven processing (subjective identity) will be activated in shared overlapping neural networks during conscious perception and in largely distinct networks during unconscious perception. We believe our approach could be a very financially friendly way to detect representational overlap between expectation, conscious/unconscious working memory and conscious/unconscious perception in the brain.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This article was published Online First October 5, 2023.

Ze-Fan Zheng and Shu-Yue Huang contributed equally to this study.

This work was supported by the Natural Science Foundations of China (61876222), Project of Humanities and Social Sciences, Ministry of Education (18YJA190001, 18YJA190011), and Zhejiang Provincial Natural Science Foundation of China (LY18C090001). The preliminary version of this study was presented at the 24th meeting of the Association for the Scientific Study of Consciousness (ASSC). All authors declare that they have no conflicts of interest. We would like to extend our sincere gratitude toward Li Ang for her critical early experimental works on this project. We also owe a big thank you to Umberto Noe and his colleagues for their precious feedback on the statistical modeling employed in the current article. We thank Darinka Truebutschek and Alex Lepauvre for their critical feedback on the rationale of this work. Lastly, we also highly appreciate the feedback we got from ASSC in Tel Aviv. All data and analysis code of all experiments in this article have been made publicly available on Open Science Framework (OSF, https://osf.io/tfwg6/; Zheng et al., 2022). Experiment 2 is a reanalysis of Sand and Nilsson (2017a), and we originally retrieved the raw data of Experiment 2 from https://osf.io/4cg2s/.

Ze-Fan Zheng served as lead for data curation, formal analysis, and writing–original draft. Shu-Yue Huang contributed equally to data curation, formal analysis, and investigation. Shena Lu contributed equally to resources and supervision. Yong-Chun Cai served as lead for funding acquisition, investigation, methodology, resources, and supervision. Ze-Fan Zheng, Shu-Yue Huang, Shena Lu, and Yong-Chun Cai contributed equally to conceptualization and writing–review and editing. Ze-Fan Zheng, Shu-Yue Huang, and Shena Lu contributed equally to methodology. Shu-Yue Huang, Shena Lu, and Yong-Chun Cai contributed equally to writing–original draft.

Correspondence concerning this article should be addressed to Yong-Chun Cai or Shena Lu, Department of Psychology and Behavioral Sciences, Zhejiang University, No. 388 Yuhangtagn Road, Hangzhou 310058, People's Republic of China. Email: yccai@zju.edu.cn or snlu@zju.edu.cn

Keywords: access consciousness, top-down, bottom-up, Stroop priming

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Since Crick and Koch (1990) advocated that it was finally time to tackle the "hard problem" of human mind—the neural correlates of consciousness (NCC), many philosophers, psychologists, and neuroscientists have contributed to theory building on this issue in the past three decades (Aru et al., 2020; Baars, 2005; Chalmers, 1996; Dehaene & Naccache, 2001; Lamme, 2006; Lau & Rosenthal, 2011; Tononi, 2004). Various theories have been proposed to account for the neural mechanism of conscious experiences from very different disciplinary stances, such as neurocognitive perspectives (Baars, 2005; Dehaene & Naccache, 2001), neurobiological perspectives (Aru et al., 2020; Lamme, 2006; Tononi, 2004), and philosophical perspectives (Lau & Rosenthal, 2011). Despite the wide variety of theories, many of them seem to have a strong consensus on the fact that conscious experiences emerge when top-down (TD) and bottom-up (BU) information streams somehow meet up in the brain (Aru et al., 2020; Cauller, 1995; Dehaene & Naccache, 2001; Lamme, 2006; Lamme & Roelfsema, 2000; Tononi, 2004). Specifically, a highly simplified interpretation of global neuronal workspace theory suggests that the global feedback from frontal–parietal network is critical to conscious perception in the brain (Dehaene & Naccache, 2001; Mashour et al., 2020), whereas recurrent processing theory has similar postulates, but insists that the local recurrent activities within the primary sensory cortices matter more (Lamme, 2006; Lamme & Roelfsema, 2000). A more recent proposal based on evidence in neurobiology points out that the connection between TD and BU compartment on the cellular level is the true NCC (Aru et al., 2020). Finally, integrated information theory, which claims maximal irreducible integrated information as a measure of consciousness, also postulates that a pure feedforward network has very low integrated information (therefore, low consciousness, see Tononi, 2004; Tononi et al., 2016). Notably, BU refers to the feedforward activities caused by external sensory input. TD refers to feedback activities carrying internal thoughts, prior knowledge, goals, or task demands in this work.

Observing the effect of spontaneous interaction between two information streams using neuroimaging methods has been proven particularly difficult and costly (de Lange et al., 2018; McMains & Kastner, 2011; Ogawa & Komatsu, 2006). Many methods have been proposed to indirectly infer the presence of information exchange across the cortical hierarchy, such as temporal characteristics of activity (e.g., pre- or post- 200 ms after stimulus onset; Gwilliams & King, 2020), causal modeling methods like dynamic causal modeling (Dijkstra, Bosch, & van Gerven, 2017; Stephan et al., 2010) or Granger causality (e.g., Gaillard et al., 2009; Seth et al., 2015). However, all these neural imaging methods are based on various contestable assumptions (de Lange et al., 2018). A recently proposed more promising method termed laminar profiling is based on much more solid neurological foundations (de Lange et al., 2018; Stephan et al., 2019). However, the application of a laminar functional magnetic resonance imaging (fMRI) is too costly for most psychology labs in the world. In the current study, instead of diving deep into this difficult issue using neuroimaging method, we aimed to propose a novel behavioral approach to isolate the effects of TD, BU, and their interaction during conscious and unconscious processing based on global neuronal workspace theory (Dehaene & Naccache, 2001). Crucially, this global workspace framework points out that it is the global broadcast of the feedforward sensory information in the frontal–parietal decision-based networks that enable conscious percept of the sensory input.

According to the global neuronal workspace theory, any external BU stimulus input must reach the global workspace to be made conscious (Dehaene, 2014; Dehaene & Naccache, 2001; Mashour et al., 2020). This global workspace, supported by the long-axon neurons in the prefrontal cortices, is able to broadcast the information to the TD networks in the brain, which host memory, attention, evaluation, value, and action systems. These systems also serve to generate and report the high-level perceptual decisions (Dehaene et al., 2006). However, in case external input information fails to reach the global workspace, they would remain unconsciously processed in the local BU networks without sustained and reciprocal connection with TD networks. In this case, the BU network has access to the external stimulus input (e.g., a word) so that a BU representation of the stimulus, if possible, could be generated in this isolated local network. Meanwhile, despite that the TD network is hosting information about goal, context, and task, it has no access to the external stimulus input so that only a TD inference, that is, a random guess of the identity of external stimulus could be generated. According to this theory, only the representations, which are propagated from the TD or BU modular networks to the global workspace, are consciously experienced and reportable. As intensity of the external stimulus rises, more BU external information would be made available for the generation of the TD representation. Ultimately, the BU and TD representations would merge to the same representation when the stimulus is fully visible and correctly identified. In other words, it is expected that BU representation (i.e., identity of external input) and TD representation (i.e., identity of discrimination decision) would be established and activated in independent networks when the stimulus is rendered invisible and shared networks when it is visible (Dehaene et al., 2006).

However, it is to note that global neuronal workspace theory does not deny all interactions between TD and BU networks during unconscious processing. For example, global neuronal workspace theory does stress that TD attention or task relevance could boost the depth of unconscious processing (Berkovitch et al., 2018; Dehaene et al., 2006; Koivisto et al., 2009; Kouider & Dehaene, 2007; Naccache et al., 2002). Global neuronal workspace theory also accepts the remarkable effect of unconscious influences on TD control (Dehaene, 2014; Pessiglione et al., 2007). However, such interaction is typically unidirectional and initiated from one end of the cortical hierarchy to the other without a sustaining reciprocal reverberation loop between TD and BU networks. Such unidirectional interaction has been found to fail to induce neural overlap between TD mental representations and unconscious BU representation, whereas there is significant overlap between TD imagined representation and conscious BU representation (Dijkstra et al., 2021).

The TD and BU pathways of object recognition have been well supported by the empirical literature (Carreiras et al., 2014; Dehaene et al., 2001; Gwilliams & King, 2020; Heilbron et al., 2020; Magnuson et al., 2018; Melloni et al., 2007; Mostert et al., 2016; Wyatte et al., 2012) and theoretical frameworks (Bar, 2003, 2007; Bar et al., 2006; Fenske et al., 2006; Panichello et al., 2013). For example, during noisy or ambiguous perceptual experiences, it has been found that our brain must recruit the recurrent

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

activities in frontal–parietal areas for the computation of subjective stimulus identity (Gwilliams & King, 2020; Mostert et al., 2016); meanwhile, objective processing of external sensory input is established in occipital and temporal cortices through the conventional ventral pathway in a feedforward pattern (Bar et al., 2006; Dehaene et al., 2001; Gwilliams & King, 2020). However, no previous study examined how TD networks in the front and BU networks in the back jointly contribute to the object recognition as a function of conscious awareness.

As discussed above, the BU representation (i.e., original stimulus representation) is established and activated in distinct networks from those that support the generation of TD representation (i.e., reported stimulus representation) during unconscious processing. An interesting and testable prediction could be made that these two representations would influence the processing of subsequent information in an independent and additive fashion. Reversely, as during conscious perception, these representations would be expected to be activated in shared neural networks (i.e., the global workspace), they should influence the processing of subsequent information in an interactive fashion. The current study is motivated to use behavioral experiments to test these hypotheses.

The Current Study

We employed Stroop priming paradigm in the two experiments of the current study. A prime color word was presented shortly prior to a target color frame. Prime word and target color are either semantically congruent or incongruent with each other. Response time (RT) to target and subjective visibility and identification of prime were recorded. In Experiment 1, a prime word in Chinese, meaning either *yellow* or *blue*, was presented under continuous flash suppression (CFS; Tsuchiya & Koch, 2005). In Experiment 2, a secondary analysis on the data of Sand and Nilsson (2017b) was conducted, in which backward masks were presented shortly prior to a color target. In the priming literature, it has been well documented that it takes subjects longer time to respond to the target when it follows an incongruent prime compared to a congruent prime (e.g., Dehaene et al., 1998). The input prime representation is named as *BU-representation* in the current work. We call the congruency between prime and target as *BU-congruency*. Recently, Sand and Nilsson (2017a) found, in addition to that, the incongruency between target and reported prime representation in the two-alternative-force-choice (2AFC) discrimination task could also influence RT to the target. In the current work, we name the reported prime representation *TD-representation* and its congruency with target *TD-congruency*. As such, all the trials could be grouped into four conditions in accordance with TD- and BU-congruency as outlined in Figure 1. They are both-TD-and-BU-congruent condition, both-TD-and-BU-incongruent condition, only-TD-incongruent condition, and only-BU-incongruent condition, respectively. By contrasting the RT of the first condition with the latter three, both-TD-and-BU-incongruent effect, only-TD-incongruent effect, and only-BU-incongruent effect could be computed, respectively.

The rationale and prediction of the current study are illustrated in Figure 1. On the unconscious trials (Figure 1B and D), as TD and BU representations of the prime are predicted to be established and activated in distinct networks, it is hypothesized that these representations would influence the target processing in an independent fashion so that additive effect of TD- and BU-congruency should be observed. In other words, the summation of TD-incongruent effect ($X_{TD} - X_0$), and BU-incongruent effect ($X_{BU} - X_0$), should be equal to both-TD-and-BU-incongruent effect ($X_{BU–TD} - X_0$). On the conscious trials, TD and BU representations are expected to be activated in shared networks. In both-TD-and-BU-incongruent condition, TD representation is always the same as BU representation. It is assumed that TD and BU representations would converge to a unified representation in the overlapped networks when they are the same, which could be realized by coupling between TD (apical) and BU (basal) compartments during conscious states (Aru et al., 2020; Suzuki & Larkum, 2020). Putatively, this representational cancel-out should also lead to cancel-out in congruency effect. Therefore, both-TD-and-BU-incongruent effect ($X_{BU–TD} - X_0$) is predicted to be smaller than the summation of BU-incongruent effect ($X_{BU} - X_0$) and TD-incongruent effect ($X_{TD} - X_0$) on the conscious trials. Based on an assumption that it is the interference effect that mainly contribute to the Stroop priming effect rather than the facilitation effect, as evidenced by previous studies (Glaser & Glaser, 1982; MacLeod, 1998; MacLeod & MacDonald, 2000), we reason that TD and BU congruency would produce underadditive interaction effects in the conscious trials.

Notably, the TD representation employed in the current work is report-based. It is possible that our approach might only have the sensitivity to detect the global interaction between TD decision networks and BU sensory networks as proposed by global neuronal workspace theory, as this theory postulates a strong bond between consciousness and reportability. In other words, it may not be sensitive enough to detect the local recurrent activities within the primary sensory cortices as proposed by other recurrent theories, as they are more targeted at explaining phenomenal consciousness that may not be available for report at all.

We tested these speculations in two experiments, using an objective measure of consciousness (i.e., prime identification accuracy) in CFS (Experiment 1) and a subjective measure (i.e., perceptual awareness scale; Sandberg et al., 2010) in backward masking (Experiment 2). Moreover, physical contrast of the prime varied across different visibility conditions in Experiment 1 but was held the same in Experiment 2 so that we can rule out the confounding factor of prime physical contrast if the same interaction effect pattern is observed in both experiments. Our results showed that, indeed, TD- and BU-congruency produced underadditive interaction effects when the prime was consciously perceived. This underadditive interaction effect was a strong correlate of both subjective and objective consciousness measures. Meanwhile, TD- and BU-congruency showed additive effects in the low objective performance trials of Experiment 1 or when the prime is subjectively invisible in Experiment 2.

Experiment 1

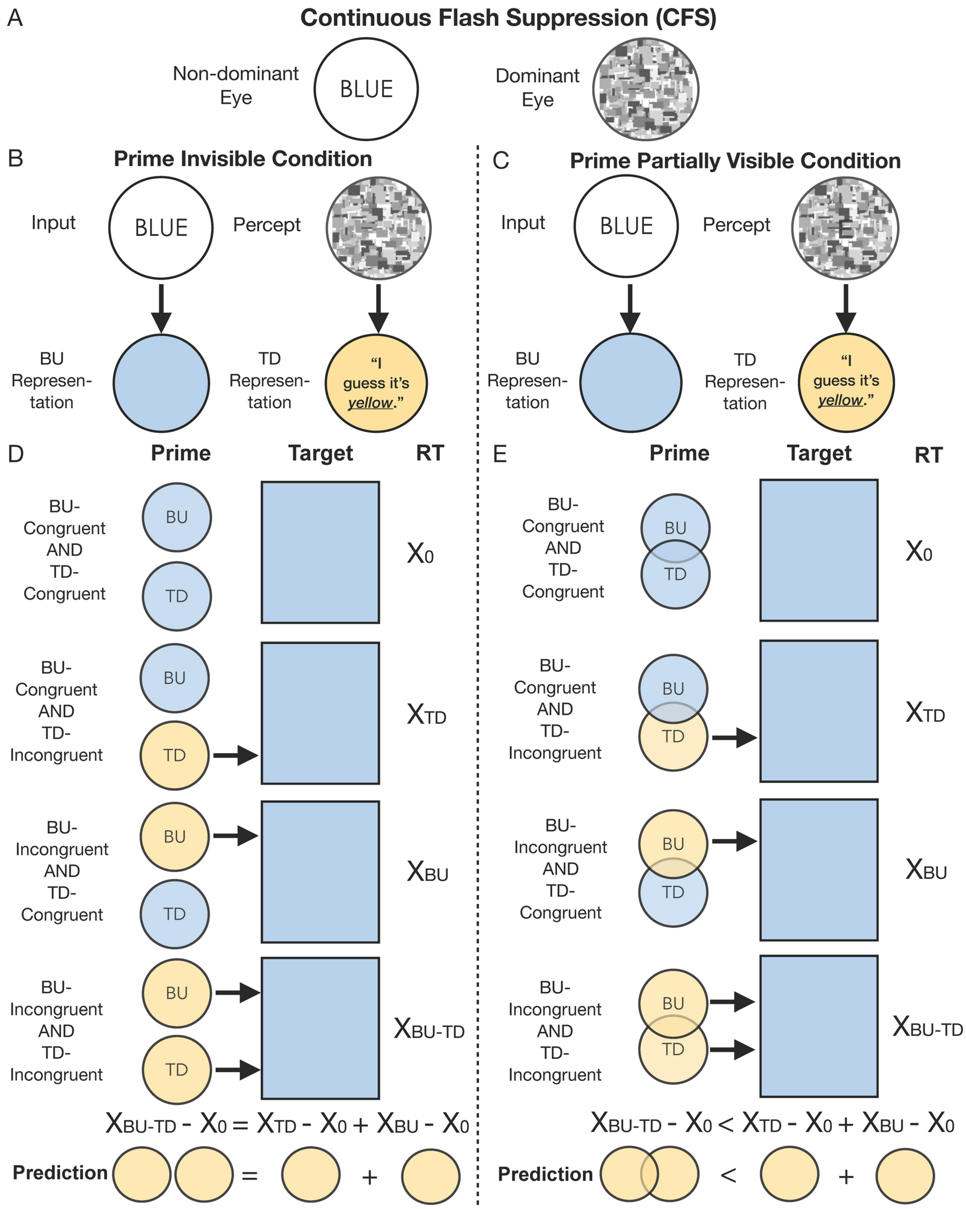
In Experiment 1, we employed CFS paradigm to mask the prime stimulus, whose visibility (i.e., consciousness) was manipulated by varying the physical contrast of the prime stimulus. Conscious awareness of the prime was indexed by the accuracy of the prime discrimination. We predicted that the more accurately the prime was identified, the stronger the interaction effect between TD- and BU-congruency would be.

Method

Participants

Sixty healthy participants with normal or corrected-to-normal vision were recruited for this experiment. All participants were

Figure 1
Schematics of Rationale and Prediction



Note. CFS was employed to manipulate prime visibility, in which prime is presented to the nondominant eye while continuously flashing mask is presented to the dominant eye (A). Subjects could generate the TD representation by randomly guessing prime identity when the prime is rendered invisible (B), or draw the inference based on limited available information, for example, a visible letter "E," when prime is partially visible (C). BU representation refers to original prime representation, while TD representation refers to reported prime representation (i.e., forced-choice discrimination decision). BU-congruency refers to the congruency between BU stimulus-driven prime representation and target. TD-congruency refers to the congruency between TD decision-driven prime representation and target. All trials in each visibility condition could be grouped into four different congruency conditions, respectively. BU and TD representations are hypothesized to be activated in independent networks when prime is invisible (D), and in overlapping networks when prime is visible (E). Circles represent TD and BU neural networks, and the filled colors (i.e., blue/yellow in the online version and gray/light in the printed version) refer to the activated prime representation. Squares represent the target and the filled color (i.e., blue/yellow in the online version and gray/light in the printed version) refers to target representation. Arrows denote the putative presence of incongruency effect, which could be estimated by contrasting the RT of the corresponding congruency condition and X_0. It is predicted that TD- and BU-congruency produce additive effect when prime is rendered invisible, and underadditive effect when prime is visible. CFS = continuous flash suppression; TD = top-down; BU = bottom-up; RT = response time. See the online article for the color version of this figure.

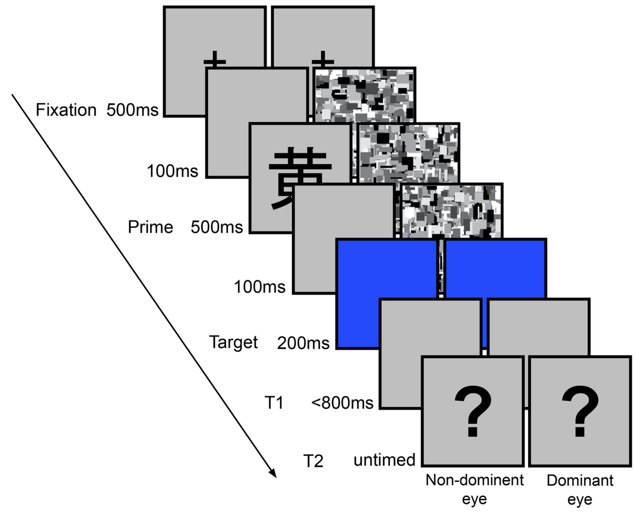
This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

undergraduate or postgraduate students at Zhejiang University and were either paid or rewarded course credits for the participation. Among them, 29 participants were excluded from further data analysis due to the three criteria which were orthogonal to effect of interests: the accuracy of the lowest prime contrast level should be lower than 60%, the accuracy of the highest prime contrast level should be higher than 75%, and the residuals R in fitting accuracy as a sigmoid function of prime contrast level should be lower than 0.3. The first two criteria were served to make sure we successfully manipulated the subjective visibility within participants. The third criterion was dedicated to making sure participants were giving optimal performance across conditions of prime contrast, which might be violated by loss of attention or strategic responding. All participants gave their written informed consent for the procedures. The experimental designs of all the experiments included in the current study have been approved by the Research Ethics Board of Zhejiang University. The data of 31 subjects (18 female and 13 male) at the average age of 20.1 (standard deviation, $SD = 2.2$) years were forwarded to final data analysis. The sample size was determined by a priori power analysis based on an experiment with similar design (Sand & Nilsson, 2017a). A minimum of 21 participants was sufficed to detect a similar effect (Cohen's $d = 0.66$) with adequate power of 80% and α of .05.

Apparatus, Stimuli, and Procedure

All stimuli were generated using Matlab (MathWorks, Natick, Massachusetts, United States) with Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and were displayed on a gamma-corrected cathode-ray tube monitor (21″ Dell UltraScan |P1130; 1,024 × 768 resolution; 100-Hz refresh rate). Throughout the experiment, participants were comfortably seated. A pair of dichoptic displays was viewed through a mirror stereoscope. The effective viewing distance was 82 cm. A black square frame (4.5° × 4.5° visual angle) was continuously presented on the gray screen with a luminance of 30 cd/m^2 in each display to promote stable binocular alignment throughout a trial. The trial events are shown in Figure 2. Each trial began with a black central fixation cross (0.6° × 0.6°) which was presented in each display. After 500 ms, a square patch of luminance Mondrian noise was presented in the dominant eye for 700 ms. The Mondrian mask was produced by positioning 1,000 white, black, and medium gray squares of variable sizes randomly sampled between 0.28° and 1.27° (Hesselmann et al., 2011). In the meantime, a dark gray prime word (2.0° × 2.0°), either Chinese character 黄 (meaning yellow) or 蓝 (meaning blue), was presented to the nondominant eye for 500 ms, and was preceded and followed by a 100-ms blank interval to prevent the possibility that the abrupt onset and offset of the word might make the word suddenly visible. The luminance contrast of the prime word against the background varied from trial to trial. Here, the luminance contrast was Weber contrast, which was defined as $(L_{\text{background}} - L_{\text{word}})/L_{\text{background}}$. Subsequently, either a yellow or a blue (Commission Internationale de l'éclairage; CIE x, y chromaticity coordinates of 0.419/0.150 or 0.505/0.060, respectively) colored target patch (4.3° × 4.3°) was equally likely to be presented in both eyes for 200 ms, and was followed by a blank interval which remained until participants performed Task 1 (for brevity, T1) or for 800 ms if no response was made. After then, a question mark "?" (1.0° × 1.5°) would be presented until subjects make response to Task 2 (for brevity, T2). In each trial, subjects had to perform two tasks. In T1,

Figure 2
Trial Event of Experiment 1



Note. The prime word, either 黄 *yellow* or 蓝 *blue* in simplified Chinese, was presented to the nondominant eye, while Mondrian mask was presented to the dominant eye. Color target, either in yellow or in blue (i.e., light or dark in the printed version), was presented to both eyes since mask offset. Immediate response to target color (i.e., T1) was required. Short warning tone occurred immediately when wrong responses or late responses (longer than 1 s) were made. Question mark showed up after T1 response, upon which response to prime word identity without time pressure (i.e., T2) was required. Short warning tone would also occur upon incorrect prime identification. T1 = Task 1; T2 = Task 2. See the online article for the color version of this figure.

subjects were asked to make speeded and accurate color discrimination response for the target (i.e., either yellow or blue) by pressing specified buttons on a four-button response box using two thumbs of their hands. RT to the target was recorded since the onset of the color target. Incorrect responses or no responses within 1,000 ms (200 ms of target display and 800 ms of blank screen) would be immediately followed by a 1,000-Hz-beep feedback for 200 ms. Afterward, subjects were instructed to make nonspeeded 2AFC response to identify the prime word (i.e., either "黄" meaning yellow or "蓝" meaning blue) presented previously using two index fingers of their hands. Similarly, a 500-Hz-beep feedback for 200 ms would occur in case incorrect response to T2 was made. No audial feedback is given following correct responses in both T1 and T2.

This experiment contains training session, threshold testing session, and main test session. The procedure of each trial was identical across all sessions. In the training session, 20–50 trials were used for the subjects to familiarize with the stimulus and tasks. Afterward, a threshold testing session was undertaken to test the objective threshold (i.e., 70% of prime discrimination accuracy) of each participant prior to the main test session. The threshold testing session included 100 trials. Among the 100 trials, the contrast value of the prime words varied between 11.71% and 74.21% in four equal steps, with 20 trials on each contrast level. We determined this prime contrast range for two underlying motivations. Firstly, we observed that changes in contrasts higher than 74.21% are hardly subjectively distinguishable and any contrast lower than 11.71% is too weak even to detect with bare eyes without CFS suppression. Secondly, we also

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

aim to probe prime visibility on a range of prime contrast that is as wide as possible. A sigmoid nonlinear function with prime contrast as independent variable and prime identification accuracy as dependent variable was fitted after subjects finished the session. The contrast which corresponds to 70% accuracy of prime identification was noted as the threshold of the specific subject.

In the main test session, contrast values were determined for each participant based on the threshold of 70% accuracy obtained in the threshold testing session. For each participant, contrast values varied within a range where the highest value was by 16% Weber contrast higher than the threshold and its lowest value was by 30% Weber contrast lower than the threshold. The 16%-higher upper limit makes sure that on the highest trials participants perform well above chance level but cannot generate a clear percept, whereas the 30%-lower limit ensures more invisible trials are included. Therefore, in order to ensure that the lowest and highest contrast values determined for each subject in the main test session fall within an appropriate boundary, subjects whose 70%-accuracy thresholds obtained in threshold testing session are lower than 31.25% or higher than 58.59% were refused to take the main test session. There were 30 contrast levels for each participant with a step of 1.56% Weber contrast between adjacent levels. Each contrast level was repeated 18 times, resulting in 540 trials in total. Contrast values were randomly assigned across the whole session. A 5-min rest was arranged after every 105 trials until the whole session was finished.

Data Analysis

As mentioned above, there were two types of congruencies—BU-congruency and TD-congruency. BU-congruency referred to the consistency between meaning of the prime word and the target color, and TD-congruency referred to the consistency between the discrimination judgment for the prime in T2 and the target color. A reliable estimation of the TD- and BU-congruency main effects and their interaction effects requires enough trials for all four congruency conditions in each visibility condition. A sliding window analysis of the three effects was conducted in the current study, as shown in Figure 3. Each window consists of ($18 \times 15 =$) 270 trials on 15 contrast levels. Window 1 included the lowest 15 contrast levels (from Contrast Level 1 to Contrast Level 15), whereas the Window 16 included the highest 15 contrast levels (from Contrast Level 16 to Contrast Level 30). Every higher window is generated by moving the lower window one contrast level upward. As such, 16 windows were generated in total for the analysis.

All trials with wrong or slow answers (RT $>$ 1,000 ms) to Task 1, which was the timed response to the color of the target, were excluded from further data analysis. All data analysis was conducted in R (Version 3.0-6; R Core Team, 2019). TD- and BU-congruency of each trial were contrast-coded (i.e., -0.5 means congruent, while 0.5 means incongruent) based on the assumption that interference of the incongruent condition mainly contribute to the Stroop priming effect (Glaser & Glaser, 1982; MacLeod, 1998; MacLeod & MacDonald, 2000). Plots were created by *ggplot2* package (Version 3.3.5; Wickham, 2016). The number of TD-congruent trials (8,347) is not significantly different from the number of TD-incongruent trials (8,393), $\chi^2(1) = 0.13$, $p = .722$. This shows that prime identification response was not biased by the identity of the target but genuinely reflecting the subjective identification of the prime in this experiment.

In the current study, we applied generalized linear mixed modeling (GLMM) for statistical inference of TD- and BU-congruency main effects and their interaction effects in each window using *glmer* function in *lme4* package (Version 1.1-23; Bates, Mächler, et al., 2015). This is motivated by a recent article which suggests that transformed RT data would distort the original RT effects, especially when interaction effects are the main effect of interest (Lo & Andrews, 2015). In GLMM, only the specification of the most approximate distribution of the dependent variable is needed so that this distribution will be used to fit the data of dependent variable. As such, data transformation is no more necessary. A distribution test of dependent variable in the current data set would be a mandatory prerequisite for GLMM. We employed such an analysis using *fitdistrplus* package to seek the best distribution fit to the present RT data among normal distribution, Gamma distribution, and Weibull distribution (Version 1.1-1; Delignette-Muller & Dutang, 2015). Results demonstrated that Gamma distribution outperformed the others in all model fits indices, which is in line with the results of Lo and Andrews (2015) (see Appendix A). Thus, Gamma distribution with identity link function was specified as the distribution of the dependent variable in all the GLMM models reported below. Estimation method applied in the analysis is maximum likelihood method, which is the default estimation method of *glmer* function. We did not attempt to remove any outliers to fully respect the original distribution of the RT data (Rabagliati et al., 2018).

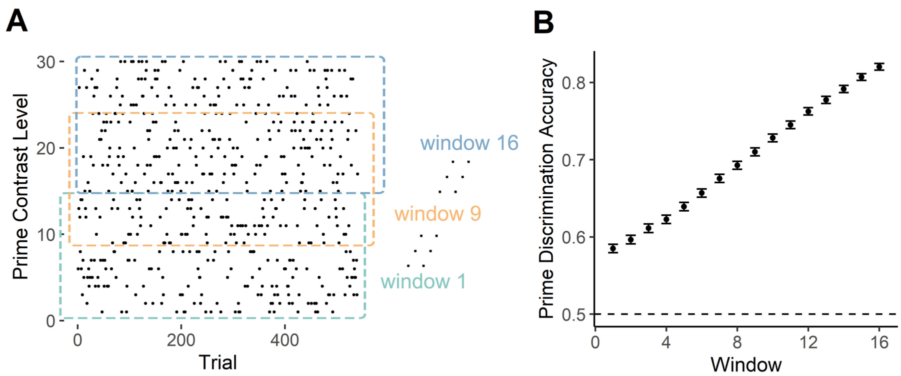
A GLMM model which contains maximal random effect structure was estimated, since maximal random effect structure was found to be able to suppress the likelihood of false positive results most effectively (Barr et al., 2013). In this model, TD-congruency, BU-congruency, and their interaction were included as the fixed effects. As all fixed effect variables vary within participants, all of them were also included as random slope effects. Only by-subject analysis was conducted since there were only two items (i.e., yellow and blue color block) in the current design. However, this maximal model might fail to converge, which typically occurs when the current sample size is insufficient for the estimation of a model with too complicated random effect structure. Convergence issues imply that the model output might not be robust, and random effect structure must be simplified. Hence, in case the maximal model in the analysis of a window fail to converge, we would follow the parsimonious suggestions from Bates, Kliegl, et al., 2015 to remove the correlation effects between random intercepts and slopes effects from the maximal model. All models successfully converged after this minor adjustment. Type III Wald chi-square was calculated using the *Anova* function in *car* package for statistical inference report (Version 3.0.-6; Fox & Weisberg, 2018). The formula of the maximal model used to compute three effects of interests in each window in the following results section is as followed: RT $\sim$ BU-congruency + TD-congruency + BU-congruency $\times$ TD-congruency + (BU-congruency + TD-congruency + BU-congruency $\times$ TD-congruency|subject).

Transparency and Openness

We report design and analysis following JARS (Kazak, 2018), including how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data and analysis codes and research materials used in the study have been made publicly available and can be accessed at https://osf.io/tfwg6/(Zheng et al., 2022). Experiment 2 was a reanalysis of data set of Sand and

Figure 3

The Specification of (A) Prime Contrast and (B) Prime Discrimination Accuracy of Each Sliding Window in the Sliding Window Analysis



Note. (A) This figure demonstrates an example prime contrast matrix. Each black dot represents a specific prime contrast assigned to a specific trial. Sliding window analysis was based on this contrast matrix. Each window consisted of trials on 15 contrast levels. Window 1 started with the trials on the 15 lowest contrast levels, and Window 16 included the trials on the 15 highest contrast levels. A higher window is by one contrast level higher than the adjacent lower window. (B) Prime discrimination accuracy rises as a function of window number. Dash line represents chance-level performance (i.e., 50%). Error bars represent standard error. See the online article for the color version of this figure.

Nilsson (2017a), which can be found at https://osf.io/tfwg6/. We reported the software used to analyze data above. This study's design and its analysis were not preregistered.

Results and Discussion

The BU main effect refers to RT difference between BU-incongruent conditions (including BU-incongruent–TD-congruent condition and BU-incongruent–TD-incongruent condition) and BU-congruent conditions (including BU-congruent–TD-congruent condition and BU-congruent–TD-incongruent condition) computed by the GLMM syntax mentioned above. Similarly, the TD main effect was the RT difference between TD-incongruent and TD-congruent conditions. Trial counts of each congruency in each window are attached to Appendix D.

Figure 3B showed that the higher the window sequence number was, the higher prime identification accuracy was. Our following analysis assumes that prime visibility rises as window sequence number rises. As shown in Figure 4A, TD- and BU-congruency produced additive effects (i.e., significant TD- and BU-congruency main effects without significant interaction effect) in the low accuracy windows (e.g., Windows 1–4). As window sequence number rose (e.g., Windows 8–16), TD- and BU-congruency started to exhibit underadditive interaction (i.e., significant TD and BU main effects with negative and significant interaction effect). More importantly, the magnitude of interaction effect increased with the sliding window sequence. In general, these descriptive patterns perfectly aligned with our theoretical predictions that the interaction effect between TD- and BU-congruency is a correlate of prime visibility.

These descriptive patterns were indeed supported by results of statistical inferences drawn from GLMM. The TD and BU interaction effect keeps insignificant from Window 1 to Window 4; Figure 4D, Window 1, $\chi^2(1) = 2.13$, $p = .145$; Window 2, $\chi^2(1) = 0.23$, $p = .631$; Window 3, $\chi^2(1) = 3.16$, $p = .076$; Window 4, $\chi^2(1) = 2.16$, $p = .141$. Meanwhile, we observed significant TD main effects in all these windows; Figure 4C, Window 1, $\chi^2(1) = 42.70$, $p < .001$; Window 2, $\chi^2(1) = 52.42$, $p < .001$; Window 3, $\chi^2(1) = 52.91$, $p < .001$; Window 4, $\chi^2(1) = 105.94$, $p < .001$, significant BU main effects from Window 2 to Window 4; Figure 4B, Window 2, $\chi^2(1) = 5.07$, $p = .024$; Window 3, $\chi^2(1) = 9.22$, $p = .003$; Window 4, $\chi^2(1) = 5.86$, $p = .015$ and marginally significant in Window 1, $\chi^2(1) = 3.40$, $p = .065$. To sum, TD- and BU-congruency indeed exhibited additive effects from Window 1 to Window 4. Moreover, all three effects reached significance since Window 5 onwards (ps $< .05$), which suggest that TD- and BU-congruency produced underadditive effects. Pearson correlations between window sequence number and all three effects were found significant (BU main effect, $r = .97$, $p < .001$; TD main effect, $r = .58$, $p = .021$; Interaction effect, $r = -.98$, $p < .001$), indicating that all these effects were statistically correlated with prime visibility.

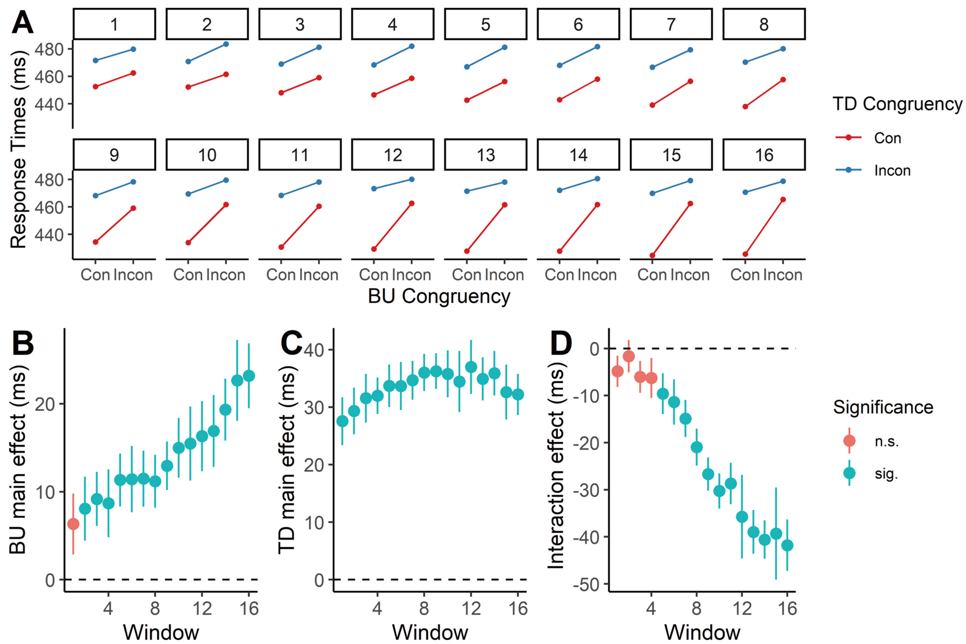
Experiment 2

In Experiment 1, we successfully found empirical support for one of the predictions of our hypothesis that TD and BU interaction effect occurs during conscious perception and increases as consciousness rises. However, neither did we measure subjective visibility nor did the prime identification accuracy in low contrast condition reached chance-level. We could not address the other prediction of our hypothesis, that is TD and BU interaction effect would be absent when the stimulus is suppressed completely outside of awareness, although no significant interaction between TD and BU was observed for low accuracy windows (Windows 1–4). Furthermore, as we modulate discrimination accuracy and visibility by manipulating the physical contrast of the prime, arguments might well be made that the increasing TD and BU interaction effect observed in Experiment 1 might be driven by rising prime physical

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 4

Results of Sliding Window Analysis in Experiment 1



Note. BU-congruency refers to the congruency between BU stimulus-driven prime representation and target. TD-congruency refers to the congruency between TD decision-driven prime representation and target. Con refers to the condition when BU- and/or TD-congruency is congruent while Incon refers to the condition when BU- and/or TD-congruency is incongruent. (A) Mean RTs of four congruency conditions in the 16 sliding windows. The estimates of BU main effect (B), TD main effect (C), and interaction effect (D) as a function of window sequence number based on GLMM analysis. α set at .05 for significance inference. Error bar depicts standard error. TD = top-down; BU = bottom-up; Con = Congruent; Incon = Incongruent; n.s. = nonsignificant; sig. = significant; RT = response time. See the online article for the color version of this figure.

contrast rather than rising prime visibility. Therefore, experimental designs that could vary subjective experiences while holding physical stimulus input unchanged across all trials are highly appreciated. Moreover, consciousness is fundamentally a subjective phenomenological experience. All scientific endeavors undertaken to quest for deeper understanding of human consciousness should undoubtedly trust the subjective report from subjects and take it as one of the most important approaches to quantitatively measure consciousness (Dehaene, 2014). Thus, the follow-up experiment or data should be able to demonstrate the robustness of our results when subjective report is taken as the criteria of consciousness. Lastly, besides CFS used in our experiment, there have been many experimental paradigms developed to render a visual stimulus invisible. One of the most classic and popular one is backward masking. To further demonstrate the robustness and generalizability of our findings, we should also aim to test our hypothesis on the data set of such paradigms, in which similar experimental procedure was conducted.

The experimental design of Sand and Nilsson (2017a) met all the requirements mentioned above. Therefore, we aim to conduct a secondary analysis on this existing publicly available data set. Meanwhile, we are convinced that testing our hypothesis in a data set that was collected independently outside of our laboratory would provide stronger evidence for the replicability and robustness of our findings. Sand and Nilsson (2017) used backward masking to induce invisibility to the word prime in a Stroop priming task, in which the prime was either English word *RED* or *BLUE*, and the target is a color block either in red or in blue. RT to target color was recorded, and 2AFC prime discrimination and subjective prime visibility were jointly reported subsequently without time pressure. Sand and Nilsson (2017a) held the strength of mask and prime stimulus unchanged across all critical trials and subjects. The idea is that when the stimulus was presented on the perceptual threshold, subjects would sometimes have visible percept of the prime while at other times less or no visible percept at all. Thus, if our hypothesis still holds in backward masking paradigm, subjective visibility reports, and when the prime physical strength remains unchanged across visible and invisible conditions, we should be able to observe no TD and BU interaction effect in invisible trials and significant underadditive interaction effect in visible trials.

Method

Participants, Stimuli, and Procedure

Sixty-seven participants (52 female, 15 male) with an average age of 23 years ($SD = 3$) from various background participated in the study, and only one participant was excluded from data analysis in the original study of Sand and Nilsson (2017a) due to not following task instructions. The data of 29 subjects were not included for final data analysis in the current study. The only exclusion criterion is that

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the number of invisible trials and visible trials of each participant should exceed 40, as current analysis requires enough trials to make robust estimation of the RT mean of four congruency conditions in each visibility condition, respectively.

In each trial, a 0.6° fixation cross was presented at the center of the screen for 500–900 ms, following which a 1° prime English word, either *RED* or *BLUE*, in capital letters appeared on the screen for 6 ms only. This prime word was masked by a subsequently presented mask which consisted of six-pound signs (#) on the screen. Notably, there was a 25 ms interstimulus interval (ISI) between prime offset and mask onset. The target, a 2° rectangle colored frame either in red or in blue, was presented since mask offset for 140 ms. Speeded responses to target color were required by pressing buttons "F" or "J" on a keyboard when a 1.5° question mark appeared on the screen. A 2 × 3 prime assessment grid was presented following response to target. In this grid, two columns represented two options of prime word identity, which is either RED or BLUE, while three rows corresponded to three forced choice of subjective visibility rating, which are "no percept," "unclear percept," and "clear percept," respectively. Participants had to report the prime identity and prime visibility jointly by choosing one of the six cells in the grid displayed on the screen using mouse click. Such 25-ms-ISI trials were presented 240 times in total for each subject intermixed with sixty 100-ms-ISI trials and thirty 25-ms-ISI catch trials with no prime stimulus presented. Thus, each subject finished 330 trials in total. Only the data of the 25-ms-ISI trials were included in the data analysis of current study. In total, 4,258 "no percept" trials (50.4%), 3,436 "unclear percept" trials (40.6%), and 762 "clear percept" trials (9.0%) were included in the final analysis (see Table E1 in Appendix E for this statistics of each individual). More details of experimental procedure and apparatus could be seen in Sand and Nilsson (2017a).

Data Analysis

Similar to the analysis protocol in Experiment 1, we contrast coded the BU-congruency and TD-congruency in each trial (congruent: −0.5; incongruent: 0.5). Considering that the number of clear percept trials is very low for all participants and we need more than 40 trials for each visibility condition, we grouped all the "unclear percept" trials and "clear percept" trials into the visible condition and all the "no percept" trials into the invisible condition. As such, we got invisible and visible conditions in Experiment 2. Prime discrimination performance in the invisible trials was not significantly higher than chance-level ($M_{\text{invisible}} = 50.1\%$, standard error, $SE = 0.03$, $p = .244$), whereas there was far above chance-level performance in visible trials ($M_{\text{visible}} = 70.8\%$, $SE = 0.11$, $p < .001$), as shown in Figure 5. Moreover, we observed significant response bias in prime identification toward the target identity, TD-congruent: 4,879 trials; TD-incongruent: 3,577 trials; $\chi^2(1) = 200.47$, $p < .001$.

We employ the same approach as has been applied in Experiment 1. Gamma distribution is employed as default distribution for dependent variable (see Appendix B for distribution check). Two GLMM models with the same syntax of the model reported in Experiment 1 were computed to estimate the TD- and BU-congruency main effects and their interaction effect in each visibility condition. A third one was constructed and estimated, in which we contrast-coded visibility condition (invisible: −0.5; visible: 0.5) and included the visibility variable into the fixed, random slope and random intercept effects of the GLMM. Similarly, convergence issue is resolved after removing correlation terms between random slopes from the maximal model. This zero-correlation model is reported in the following session. The syntax of third model reported in the following results section is as followed: RT ~ BU-congruency + TD-congruency + Visibility + BU-congruency × TD-congruency + BU-congruency × Visibility + TD-congruency × Visibility + BU-congruency × TD-congruency × Visibility + (BU-congruency + TD-congruency + Visibility + BU-congruency × TD-congruency + BU-congruency × Visibility + TD-congruency × Visibility + BU-congruency × TD-congruency × Visibility || subject).

Results and Discussion

As shown in Figure 6, TD and BU interaction effect was completely absent on invisible trials, $\chi^2(1) = 0.01$, $p = .908$, but strongly present on visible trials, $\chi^2(1) = 8.25$, $p = .004$. The three-way interaction fixed effect in the model reported above suggested that TD and BU interaction effect significantly varied across two visibility conditions with the visible condition showing stronger underadditive effect, $\chi^2(1) = 8.66$, $p = .003$, see Appendix C for estimate parameters of the main GLMM model reported above.

Furthermore, BU-congruency main effect failed to approach significance on the invisible trials, $\chi^2(1) = 0.74$, $p = .389$. It is to note that this result is in line with the results reported in Sand and Nilsson (2017a), in which a by-subject *t* test between median RT of incongruent trials and congruent trials was computed. This implies that we failed to find evidence for unconscious processing effect in this backward masking experiment. In contrast, BU-congruency main effect was very strong on the visible trials, $\chi^2(1) = 22.14$, $p < .001$, even though the physical strength of the prime is completely the same as that on invisible trials. Moreover, TD-congruency main effect was robust in both visibility conditions as in Experiment 1, invisible trials, $\chi^2(1) = 16.52$, $p < .001$; visible trials, $\chi^2(1) = 104.37$, $p < .001$. Stronger TD-congruency main effect was found on visible trials compared to invisible trials as shown in Appendix C, $\chi^2(1) = 90.52$, $p < .001$.

To conclude, this secondary analysis of Sand and Nilsson's (2017a) data revealed that TD and BU interaction effect was not a correlate of prime physical strength. Instead, this interaction effect would reliably occur when the prime is visible and would be absent, that is showing additive pattern, when the prime stimulus is suppressed invisible. Most importantly, the effect size of this underadditive interaction effect is reliably a strong correlate of subjective visibility.

General Discussion

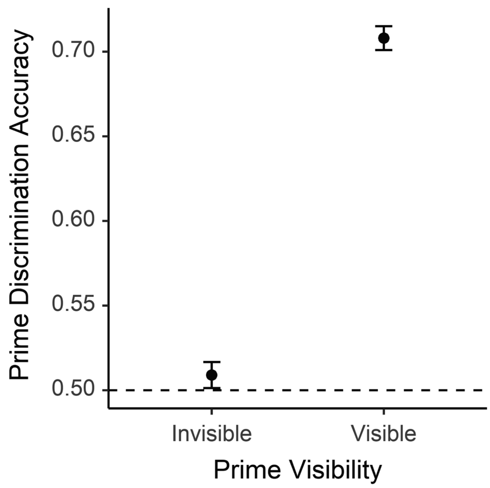
As predicted, we indeed observed that TD- and BU-congruencies produce additive effects in the unconscious conditions and underadditive interaction effects in the conscious conditions across subjective and objective indexes of consciousness and across CFS and backward masking paradigms. These results are highly consistent with our hypotheses derived from global neuronal workspace theory.

The Interaction Between TD and BU Is Underadditive or Overadditive?

In Figure 1, the deduction of results prediction assumed that RT to target is delayed by the interference of incongruent prime representation

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 5
Prime Discrimination Performance of Two Visibility Conditions in Experiment 2



Note. Prime discrimination performance of two visibility conditions in Experiment 2. The chance level for discrimination performance is 0.5. Error bar depicts standard error.

(i.e., interference account). Therefore, we took the congruent condition as the baseline condition by coding the congruent condition as −0.5 and the incongruent conditions as 0.5 during the GLMM modeling and observed underadditive interaction of RT delaying between TD- and BU-congruency in the conscious condition (Figures 4 and 6). However, it is also well possible that Stroop priming is driven by the facilitation of congruent prime representation (i.e., facilitation account). If facilitation effects were the main source of the Stroop priming effect, we should set the incongruent condition as the baseline and reverse the GLMM coding (i.e., the congruent condition was coded as 0.5 and the incongruent condition as −0.5). This would lead to overadditive interaction of RT speeding: the RT reduction in both-TD-and-BU-congruent condition relative to both-TD-and-BU-incongruent condition would be larger than the summation of RT reduction in only-TD-congruent condition and reduction in only-BU-congruent condition.

Our decision is based on the empirical literature on the interference or facilitation nature of Stroop priming over the past four decades. All studies have been dedicated to resolving the issue by adding a neutral condition in addition to the congruent/incongruent conditions. Early studies employing meaningless symbols as neutral condition (e.g., ####) reported that the Stroop priming effect was mainly attributed to the interference of incongruent condition, rather than the facilitation of congruent condition (Glaser & Glaser, 1982; MacLeod, 1998; MacLeod & MacDonald, 2000). More recent studies have challenged the previous designs and recommended employing neutral meaning words as neutral condition (e.g., table). But they did also find that interference effect is larger than facilitation effect (Brown, 2011; Parris et al., 2022; Quétard et al., 2023). Therefore, we assumed that Stroop priming is dominated by interference effect in the current study. Underadditive is more likely to be the true way of interaction between TD- and BU-congruency in conscious condition.

In Figure 1, we have only deduced the prediction of interaction results based on interference-dominant account of Stroop priming. In addition, we have also tried to deduce the prediction of this experiment based on facilitation-dominant account of Stroop priming (see Appendix F). Our empirical result is only congruent with the prediction of interference-dominant account. Hence, the underadditive interpretation of the interaction effect perfectly aligns the previous Stroop priming literature with Global Neuronal Workspace hypothesis.

Potential Visibility Confounds Could Not Explain the Underadditive Results

Worries might be voiced that the underadditive effects produced by TD- and BU-congruencies in the two experiments might involve some visibility confounds. For example, we categorized all the trials according to TD- and BU-congruencies into four conditions, which are both-congruent condition, both-incongruent condition, only-TD-incongruent condition, and only-BU-incongruent condition. The former two conditions are those in which correct responses to prime identification were made, while the latter two conditions were those in which the prime was incorrectly identified. If incorrect responses reflect lower subjective visibility compared to the correct response trials, underadditive effects might well be attributable to the possibility that lower prime subjective visibility induces longer reaction time to the target.

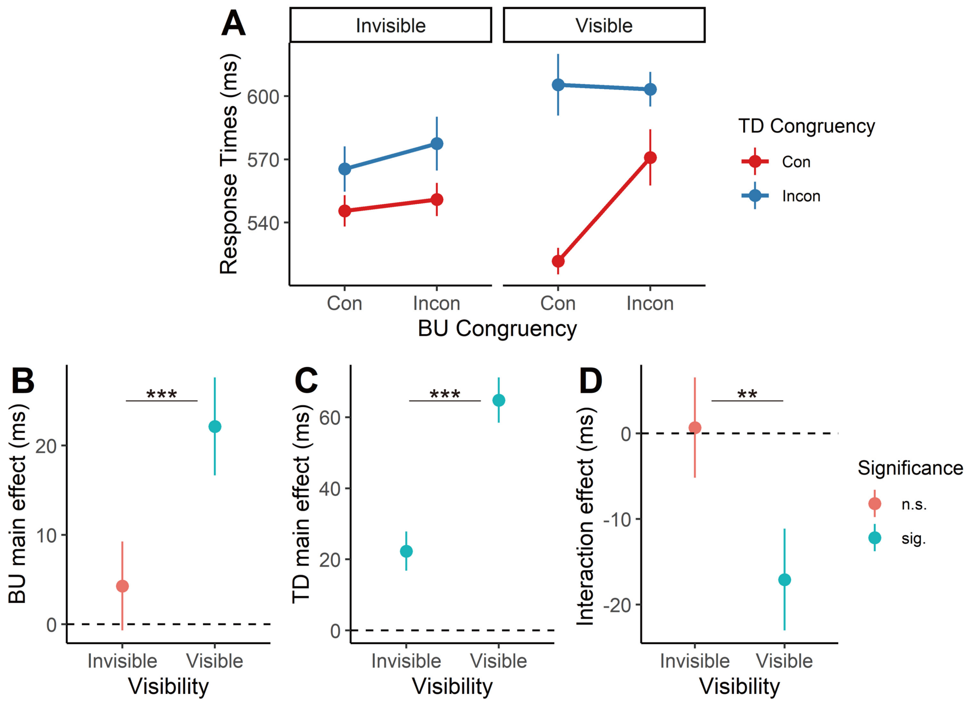
However, this possibility is not supported by our current data. Firstly, strong negative correlation between subjective visibility and RT was well established only in the experiments in which subjective visibility and RT were measured on the same stimulus (Yap et al., 2008). In other words, lower visibility to a target could indeed result in longer recognition and RT to the target itself, whereas it remains unclear whether lower visibility to a *prime*, on the whole, could delay the recognition time to a *target* (a *negative* correlation) in general regardless of prime-target congruency. Secondly, this hypothesis received no support in the current data as significant underadditive interaction effect occurred on the visible trials of Experiment 2 when there was a significant *positive* correlation between prime visibility and target RT, which putatively should lead to overadditive interaction effects instead of underadditive effects.

Additive Effect or the Lonely TD-Congruency Effect

In the current study, we postulated that TD- and BU-congruency should produce additive effects when the prime is rendered invisible. Regarding this hypothesis, we interpreted the null TD and BU interaction effect as the evidence for the disabled interaction between TD and BU processing network in two experiments. However, it is to note that we did not find significant BU main congruency effect on the invisible trials of Experiment 2 that could be rigorously categorized as an unconscious condition (i.e., no subjective visibility and chance level performance). Therefore, another plausible interpretation of the null TD and BU interaction effect is that BU information processing was completely inhibited so that TD processes have nothing to interact with. In other words, how could we say TD- and BU-congruency produce *additive* effects when BU-congruency main effect is not there so that TD-congruency effect has no other effects to add with? Thus, it may well be argued that TD and BU representations of the prime

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 6
Results of Experiment 2



Note. BU-congruency refers to the congruency between BU stimulus-driven prime representation and target. TD-congruency refers to the congruency between TD decision-driven prime representation and target. Con refers to the condition when BU- and/or TD-congruency is congruent while Incon refers to the condition when BU- and/or TD-congruency is incongruent. (A) Mean RT (ms) of four congruency conditions on invisible and visible trial types, respectively. Estimates of BU-congruency main effect (B), TD-congruency main effect (C), and interaction effect (D) in invisible and visible conditions, respectively. Error bar depicts standard error. Sig. = significant; *ns* = nonsignificant; TD = top-down; BU = bottom-up; Con = Congruent; Incon = Incongruent; RT = response time. See the online article for the color version of this figure.
*** $p < .001$. ** $p < .01$.

may always be activated in shared neural network regardless of conscious awareness.

Currently, we do not have concrete data available to argue against this lonely-TD-effect hypothesis in a rigorous subliminal condition, unless we are accessible to another data set in which both TD- and BU-congruency main effects are significant in rigorous unconscious conditions but not their interaction effect. However, it is notable that we did observe the occurrence of insignificant interaction effect but significant two main effects in the sliding window analysis of Experiment 1. In the analysis, we showed that Windows 1, 2, 3, and 4 conformed rigorously to such additive pattern. We argue that true additive scenario could well occur. As prime identification accuracy in these four windows were significantly above chance level (ranging from 58.5% to 62.3%, all $ps < .05$) and subjective report of consciousness was not recorded, it is hard for us to be sure whether such above chance-level performance is the consequence of blindsight phenomenon in healthy subjects (see Hesselmann et al., 2011; Lau & Passingham, 2006) or consequence of residual conscious awareness. But we are sure that these results demonstrate strong dissociation between discrimination performance and interaction effect. Also, if TD and BU representations are always established in overlapping networks regardless of conscious awareness, we should never observe such additive effect in any window conditions in Experiment 1. Thus, TD and BU interaction effect is indeed a true correlate of conscious access. Whether we observe additive effect (i.e., significant two main effects and insignificant interaction effect) or lonely TD effect in the unconscious condition only depends on whether unconscious information processing could be preserved in the experimental paradigm employed. It is not reasonable to argue that the interaction effect could re-occur in the chance-level performance condition when it has already disappeared in the marginally above-chance-level conditions.

Moreover, if the trials in Windows 1, 2, 3, and 4 are categorized as unconscious trials, one might draw the conclusion that the significant BU main effects might be the evidence of unconscious semantic processing during CFS. However, Abrams and Greenwald (2000) found that unconscious priming effect might result from low-level stimulus-response mapping established on the conscious trials. Since the same stimulus set was used on conscious and unconscious trials in both experiments, we believe that the weak BU main effect might also be subject to low-level stimulus-response mapping rather than genuine unconscious semantic activation. In contrast, the much

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

stronger BU effects on the higher windows are more likely to be induced by semantic processing.

TD Main Effect, BU Main Effect, Interaction Effect, and Consciousness

In both experiments, we have observed that all three effects are significantly correlated with the level of visibility. If a true behavioral correlate of consciousness needs to meet two standards: exhibiting a null effect outside of conscious awareness and increasing as a function of conscious awareness, we believe only the interaction effect is a true correlate of conscious awareness, whereas the other two are influenced by other factors.

TD main effect remained relatively stable in Experiment 1 across different windows, but still increased significantly as a function of window number. In Experiment 2, a sharp increase of TD main effect was observed from invisible trials to visible trials. A major difference between Experiments 1 and 2 is that prime visibility was passively manipulated in the former experiment (i.e., objectively varying prime contrast) and subjective reported in the latter (i.e., subjective reporting on the liminal threshold). As TD main effect was never eliminated in unconscious conditions, it fails to be a valid behavioral marker of conscious awareness. Instead, a premature speculation is that it tracks the level of confidence in prime identification decision that was acquired both on the current trial (confidence in correct prime identification on the current trial) and in the global context (confidence in correct prime identification in the statistics of the context). This hypothesis awaits serious investigation of future research.

BU main effect was only observed significant in conscious conditions in both experiments (higher windows in Experiment 1 and visible report in Experiment 2) but insignificant in unconscious conditions (lower windows in Experiment and invisible report in Experiment 2). Given ample evidence of preserved feed-forward processing outside of conscious awareness, BU main effect also fails to be a valid index of conscious awareness (Abrams & Greenwald, 2000; de Gardelle et al., 2011; Dehaene, 2014; Dehaene et al., 1998, 2001; Hesselmann et al., 2011; Naccache et al., 2002).

In contrast, we observed null interaction effect in unconscious conditions and increasing interaction effect as prime visibility increased. These observations indicate that this interaction effect may be an index of conscious awareness. Moreover, these results are highly in line with the previous works that showed additive, linear, and gradual nature of unconscious processing effects, whereas conscious processing effects always see a nonlinear all-or-none boost (de Gardelle et al., 2011; Gwilliams & King, 2020; King & Dehaene, 2014). In the existing literature, the relationship between TD and BU interaction and consciousness has only been investigated using neural data in the previous studies (Dehaene et al., 2001; Dijkstra et al., 2021; Melloni et al., 2007; Suzuki & Larkum, 2020). Especially, Dijkstra and colleagues recently found that TD-based conscious mental imagery activates overlapping neural network with conscious perception (Dijkstra, Bosch, & van Gerven, 2017). However, conscious mental imagery failed to show any such neural overlapping with unconscious representations (Dijkstra et al., 2021). This evidence is highly congruent with current findings that TD feedback activities are absent during unconscious visual processing. As during unconscious conditions subjects could only make random guess of what the prime was in the current experiments, it is to note that this guessed representation, also termed as TD representation in this work, is very similar but not necessarily the same as the mental imagery representations. They are similar in the sense that both rely solely on high-level feedback-based computations in the absence of relevant visual input (Dijkstra, Zeidman, et al., 2017; Mostert et al., 2016), while they are dissimilar in the sense that our randomly guessed representation might not necessarily involve imagery-based visual realization, which could result in strong involvement of low-level visual areas (Bergmann et al., 2016; Chen et al., 1998; Dijkstra, Bosch, & van Gerven, 2017). In general, in line with previous literature and global neuronal workspace theory, the interaction effect between TD- and BU-congruency is a valid behavioral correlate of consciousness.

To sum, the main contribution of the current study is to showcase how to study the neural activations of TD and BU representations and their overlap using pure behavioral approach, as such approaches are much more financial-friendly relative to the expensive neural imaging approaches. Specifically, two representations should produce additive effect on subsequent information processing when they are activated in independent neural networks. Underadditive effect should occur when there is neural overlap in activating the two representations. Using this approach, one might study the representational overlap between expectation, prediction, conscious/unconscious working memory or conscious/unconscious perceptual representations across a wide range of experimental conditions.

Future Directions

To sum, we point out three directions for future studies.

Firstly, further studies might ultimately test the additive (noninteractive) prediction in a rigorous unconscious condition (no subjective visibility and chance-level performance), in which both significant TD and BU main effects could occur, and mental imagery might be employed as TD representation. Specifically, we point out that the most critical and tricky factors is to make sure BU main effect (also known as unconscious priming effect in the literature) is robustly present in the experimental paradigm. We suggest future work validate our additive findings on invisible trials using lower-level features which are expected to induce stronger BU-congruency effect outside of awareness.

Secondly, the current approach also provides clear predictions regarding the anatomical features of the activation of TD representation and BU representation across different visibilities. Future behavioral studies might systematically investigate the respective factors that influence TD main effect, BU main effect, and interaction effect. Furthermore, neuroimaging methods, such as magnetoencephalography or fMRI could be employed to record the neural mechanisms underlying TD and BU representations and their overlaps. A clear testable prediction would be that TD and BU representations of the prime should be decodable from distinct neural populations in the unconscious condition and from shared ones in the conscious condition.

Lastly, our work is also highly relevant to the predictive coding account of conscious perception, which explains how TD prediction and BU prediction error jointly shape our conscious perception (Hohwy & Seth, 2020; Seth & Bayne, 2022). This account has been proven particularly useful in explaining some individual differences in abnormal conscious perception, for example, psychosis (see Sterzer et al., 2018, for a review). However, it has been a hot debate

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

regarding what is the mechanistic cause of psychotic experiences. Some views suggest that psychosis results from sharpened precision of TD inference, others argue for the opposite view that psychotic symptoms are attributed to weakened precision of BU input (Sterzer et al., 2018). Conventional accuracy-based approaches could barely disentangle these accounts (i.e., probing the accuracy of ambiguous object recognition given relevant prior knowledge across psychosis patients and healthy controls, see Teufel et al., 2015, for an example). But we argue that using the current design might help reveal how different populations differentially weight high-level TD representation, low-level BU sensory representation, and their interaction during ambiguous perceptual experiences. For example, some convincing empirical research has found that unconscious BU processing remains intact on schizophrenia patients, but TD attention amplification is impaired compared to healthy controls (Berkovitch et al., 2017, 2018). If this is the case, we should observe equally strong BU-congruency main effect but stronger TD-congruency main effect on healthy controls than schizophrenic patients using the current behavioral paradigm.

Constraints on Generality Statement

The results of Experiment 1 are consistent with the reanalysis result of Sand and Nilsson (2017a). Experiment 1 employed CFS paradigm with Chinese characters as prime stimuli, whereas Sand and Nilsson (2017a) employed backward masking paradigm with English words as the prime stimuli. Therefore, we expect our results to generalize more broadly to other masking procedures and written languages that follows the same reading direction. Participants in Experiment 1 were students at a Chinese university while participants in Sand and Nilsson (2017a) were with different backgrounds. Therefore, we expect our results to generalize to healthy participants with varying background. However, we expect a slightly different pattern of BU and TD main effect to occur to schizophrenic patients given the predictions of predictive coding framework, but the absence of interaction effect in unconscious conditions and its presence in conscious conditions should still hold.

Conclusion

In this study, we found that both congruency between target and TD prime representation and congruency between target and BU prime representation influence reaction time to the target. Importantly, two congruencies delayed RT in an additive fashion when prime is suppressed outside of conscious awareness, whereas they produced underadditive interaction effect when prime is consciously perceived. This underadditive interaction effect was found to be a strong correlate with prime visibility. We suggest that this phenomenon results from the fact that TD and BU processes of object recognition are established in distinct neural networks during unconscious processing, and the interaction between two networks is recovered during conscious processing. Our finding is highly in line with the critical postulates of Global Neuronal Workspace theory (see Dehaene, 2014).

Context of the Research

We started the current project with the aim to find unconscious semantic processing effect under CFS. Meanwhile, we found evidence of how TD prime representation influences RT to the target regardless of prime visibility. By then, our analysis was based on the assumption that TD- and BU-congruency produce additive effects on RT. However, post hoc analysis on all existing datasets shockingly revealed the presence of strong interaction between two congruency effects on the visible trials and absence of interaction on the invisible ones. Given that some major theories of consciousness, especially global neuronal workspace theory would lead to the same prediction, as they predict TD-decision driven representations recruit independent neuronal populations from BU-input driven representations during unconscious processing and shared populations during conscious processing. Therefore, we came to the idea of formally testing this hypothesis using a sliding window approach (Experiment 1). Confirmation of this effect in external existing data set, which employed backward masking paradigm instead of CFS, further validated the robustness and generalizability of our effect (Experiment 2). More importantly, we propose that our approach might be extended to test representational overlap in the brain in a much broader context, such as expectation, prediction, conscious/unconscious working memory, or conscious/unconscious perceptual representations.

References

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science*, *11*(2), 118–124. https://doi.org/10.1111/1467-9280.00226

Aru, J., Suzuki, M., & Larkum, M. E. (2020). Cellular mechanisms of conscious processing. *Trends in Cognitive Sciences*, *24*(10), 814–825. https://doi.org/10.1016/j.tics.2020.07.006

Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Progress in Brain Research*, *150*, 45–53. https://doi.org/10.1016/S0079-6123(05)50004-9

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. https://doi.org/10.1162/089892903321662976

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Schmidt, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv preprint. http://arxiv.org/abs/1506.04967

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

Bergmann, J., Genç, E., Kohler, A., Singer, W., & Pearson, J. (2016). Smaller primary visual cortex is associated with stronger, but less precise mental imagery. *Cerebral Cortex*, *26*(9), 3838–3850. https://doi.org/10.1093/cercor/bhv186

Berkovitch, L., Dehaene, S., & Gaillard, R. (2017). Disruption of conscious access in schizophrenia. *Trends in Cognitive Sciences*, *21*(11), 878–892. https://doi.org/10.1016/j.tics.2017.08.006

Berkovitch, L., Del Cul, A., Maheu, M., & Dehaene, S. (2018). Impaired conscious access and abnormal attentional amplification in schizophrenia. *NeuroImage: Clinical*, *18*, 835–848. https://doi.org/10.1016/j.nicl.2018.03.010

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brown, T. L. (2011). The relationship between Stroop interference and facilitation effects: Statistical artifacts, baselines, and a reassessment. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 85–99. https://doi.org/10.1037/a0019252

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, *18*(2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005

Cauller, L. (1995). Layer I of primary sensory neocortex: Where top-down converges upon bottom-up. *Behavioural Brain Research*, *71*(1–2), 163–170. https://doi.org/10.1016/0166-4328(95)00032-1

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Chen, W., Kato, T., Zhu, X. H., Ogawa, S., Tank, D. W., & Ugurbil, K. (1998). Human primary visual cortex and lateral geniculate nucleus activation during visual imagery. *NeuroReport*, *9*(16), 3669–3674. https://doi.org/10.1097/00001756-199811160-00019

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, *2*, 263–275. https://doi.org/10.1016/B978-0-12-185254-2.50021-8

de Gardelle, V., Charles, L., & Kouider, S. (2011). Perceptual awareness and categorical representation of faces: Evidence from masked priming. *Consciousness and Cognition*, *20*(4), 1272–1281. https://doi.org/10.1016/j.concog.2011.02.001

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211. https://doi.org/10.1016/j.tics.2006.03.007

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2

Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J. F., Poline, J. B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4*(7), 752–758. https://doi.org/10.1038/89551

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600. https://doi.org/10.1038/26967

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Delignette-Muller, M. L., & Dutang, C. (2015). Fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1–34. https://doi.org/10.18637/jss.v064.i04

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *Journal of Neuroscience*, *37*(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dijkstra, N., van Gaal, S., Geerligs, L., Bosch, S. E., & van Gerven, M. A. J. (2021). No evidence for neural overlap between unconsciously processed and imagined stimuli. *eNeuro*, *8*(5), Article ENEURO.0228-21.2021. https://doi.org/10.1523/ENEURO.0228-21.2021

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A., & Friston, K. (2017). Distinct top-down and bottom-up brain connectivity during visual perception and imagery. *Scientific Reports*, *7*(1), Article 5677. https://doi.org/10.1038/s41598-017-05888-8

Fenske, M. J., Aminoff, E., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: Object-based and context-based contributions. *Progress in Brain Research*, *155*, 3–21. https://doi.org/10.1016/S0079-6123(06)55001-0

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression* (3rd ed.). Sage Publications.

Gaillard, R., Dehaene, S., Adam, C., Clémenceau, S., Hasboun, D., Baulac, M., Cohen, L., & Naccache, L. (2009). Converging intracranial markers of conscious access. *PLoS Biology*, *7*(3), Article e1000061. https://doi.org/10.1371/journal.pbio.1000061

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(6), 875–894. https://doi.org/10.1037/0096-1523.8.6.875

Gwilliams, L., & King, J. R. (2020). Recurrent processes support a cascade of hierarchical decisions. *eLife*, *9*, Article e56603. https://doi.org/10.7554/eLife.56603

Heilbron, M., Richter, D., Ekman, M., Hagoort, P., & de Lange, F. P. (2020). Word contexts enhance the neural representation of individual letters in early visual cortex. *Nature Communications*, *11*(1), Article 321. https://doi.org/10.1038/s41467-019-13996-4

Hesselmann, G., Hebart, M., & Malach, R. (2011). Differential BOLD activity associated with subjective and objective reports during "blindsight" in normal observers. *The Journal of Neuroscience*, *31*(36), 12936–12944. https://doi.org/10.1523/JNEUROSCI.1556-11.2011

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*(II), 1–34. https://doi.org/10.33735/phimisci.2020.II.64

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

King, J.-R., & Dehaene, S. (2014). A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641), Article 20130204. https://doi.org/10.1098/rstb.2013.0204

Koivisto, M., Kainulainen, P., & Revonsuo, A. (2009). The relationship between awareness and attention: Evidence from ERP responses. *Neuropsychologia*, *47*(13), 2891–2899. https://doi.org/10.1016/j.neuropsychologia.2009.06.016

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 857–875. https://doi.org/10.1098/rstb.2007.2093

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501. https://doi.org/10.1016/j.tics.2006.09.001

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences*, *103*(49), 18763–18768. https://doi.org/10.1073/pnas.0607716103

Lau, H. C., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, Article 1171. https://doi.org/10.3389/fpsyg.2015.01171

Macleod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory and Cognition*, *26*(2), 201–211. https://doi.org/10.3758/BF03201133

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*(10), 383–391. https://doi.org/10.1016/S1364-6613(00)01530-8

Magnuson, J. S., Mirman, D., Luthra, S., Strauss, T., & Harris, H. D. (2018). Interaction in spoken word recognition models: Feedback helps. *Frontiers in Psychology*, *9*, Article 369. https://doi.org/10.3389/fpsyg.2018.00369

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*(5), 776–798. https://doi.org/10.1016/j.neuron.2020.01.026

McMains, S., & Kastner, S. (2011). Interactions of top-down and bottom-up mechanisms in human visual cortex. *Journal of Neuroscience*, *31*(2), 587–597. https://doi.org/10.1523/JNEUROSCI.3766-10.2011

Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W., & Rodriguez, E. (2007). Synchronization of neural activity across cortical areas correlates with conscious perception. *Journal of Neuroscience*, *27*(11), 2858–2865. https://doi.org/10.1523/JNEUROSCI.4623-06.2007

Mostert, P., Kok, P., & de Lange, F. P. (2016). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, *5*(1), Article 18253. https://doi.org/10.1038/srep18253

Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological Science*, *13*(5), 416–424. https://doi.org/10.1111/1467-9280.00474

Ogawa, T., & Komatsu, H. (2006). Neuronal dynamics of bottom-up and top-down processes in area V4 of macaque monkeys performing a visual search. *Experimental Brain Research*, *173*(1), 1–13. https://doi.org/10.1007/s00221-006-0362-5

Panichello, M. F., Cheung, O. S., & Bar, M. (2013). Predictive feedback and conscious visual experience. *Frontiers in Psychology*, *3*, Article 620. https://doi.org/10.3389/fpsyg.2012.00620

Parris, B. A., Hasshim, N., Wadsley, M., Augustinova, M., & Ferrand, L. (2022). The loci of Stroop effects: A critical review of methods and evidence for levels of processing contributing to color-word Stroop effects and the implications for the loci of attentional selection. *Psychological Research*, *86*(4), 1029–1053. https://doi.org/10.1007/s00426-021-01554-x

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., & Frith, C. D. (2007). How the brain translates money into force: A neuroimaging study of subliminal motivation. *Science*, *316*(5826), 904–906. https://doi.org/10.1126/science.1140459

Quétard, B., Spatola, N., Parris, B. A., Ferrand, L., & Augustinova, M. (2023). A mouse-tracking study of the composite nature of the Stroop effect at the level of response execution. *PLoS ONE*, *18*(1), Article e0279036. https://doi.org/10.1371/journal.pone.0279036

Rabagliati, H., Robertson, A., & Carmel, D. (2018). The importance of awareness for understanding language. *Journal of Experimental Psychology: General*, *147*(2), 190–208. https://doi.org/10.1037/xge0000348

R Core Team. (2019*). R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Sand, A., & Nilsson, M. E. (2017a). When perception trumps reality: Perceived, not objective, meaning of primes drives Stroop priming. *Psychological Science*, *28*(3), 346–355. https://doi.org/10.1177/0956797616684681

Sand, A., & Nilsson, M. E. (2017b). *When perception trumps reality: Perceived, not objective, meaning of primes drives Stroop priming* [Data set]. Retrieved February 2021 from https://osf.io/4cg2s/ and now available at https://osf.io/tfwg6/

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition*, *19*(4), 1069–1078. https://doi.org/10.1016/j.concog.2009.12.013

Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, *35*(8), 3293–3297. https://doi.org/10.1523/JNEUROSCI.4399-14.2015

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*, *23*(7), 439–452. https://doi.org/10.1038/s41583-022-00587-4

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage*, *49*(4), 3099–3109. https://doi.org/10.1016/j.neuroimage.2009.11.015

Stephan, K. E., Petzschner, F. H., Kasper, L., Bayer, J., Wellstein, K. V., Stefanics, G., Pruessmann, K. P., & Heinzle, J. (2019). Laminar fMRI and computational theories of brain function. *NeuroImage*, *197*, 699–706. https://doi.org/10.1016/j.neuroimage.2017.11.001

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, *84*(9), 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015

Suzuki, M., & Larkum, M. E. (2020). General anesthesia decouples cortical pyramidal neurons. *Cell*, *180*(4), 666–676.e13. https://doi.org/10.1016/j.cell.2020.01.024

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., Goodyer, I. M., & Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences*, *112*(43), 13401–13406. https://doi.org/10.1073/pnas.1503916112

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5*(1), Article 42. https://doi.org/10.1186/1471-2202-5-42

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–1101. https://doi.org/10.1038/nn1500

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer.

Wyatte, D., Curran, T., & O'Reilly, R. (2012). The limits of feedforward vision: Recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, *24*(11), 2248–2261. https://doi.org/10.1162/jocn_a_00282

Yap, M. J., Balota, D. A., Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 495–513. https://doi.org/10.1037/0278-7393.34.3.495

Zheng, Z., Huang, S., Lu, S., & Cai, Y. (2022). *Top-down and bottom-up interaction as a correlate of conscious access* [Data set]. https://doi.org/10.17605/OSF.IO/TFWG6

(*Appendices follow*)

Appendix A

Table A1

Goodness-of-Fit Indices for Distribution Analysis of RT Data in Experiment 1

Statistics	Distribution: Normal	Weibull	Gamma
Kolmogorov–Smirnov	0.08	0.09	0.04
Cramer–von Mises	44.57	54.40	10.24
Anderson–Darling	291.67	368.43	73.47
Akaike's information criterion	214,911.5	215,320.7	211,978.9
Bayesian information criterion	214,926.9	215,336.1	211,994.4

Note. We fit the RT data in Experiment 1 among normal, Gamma, and Weibull distributions, respectively, and obtain the statistics displayed in the column Statistics. The smaller the statistic or criterion values are, the better the distribution fits the RT data. RT = response time.

Appendix B

Table B1

Goodness-of-Fit Indices for Distribution Analysis of RT Data in Experiment 2

Statistics	Distribution: Normal	Weibull	Gamma
Kolmogorov–Smirnov	0.20	0.19	0.14
Cramer–von Mises	135.73	113.25	54.50
Akaike's information criterion	120,554.7	117,923.5	114,550.1
Bayesian information criterion	120,568.8	117,937.6	114,564.2

Note. We fit the RT data in Experiment 2 among normal, Gamma and Weibull distributions, respectively, and obtain the statistics displayed in the column Statistics. The smaller the statistic or criterion values are, the better the distribution fits the RT data. Anderson–Darling statistic was not reported due to failure of valid estimation. RT = response time.

Appendix C

Table C1

Estimate Parameters of the GLMM in Experiment 2

Fixed effect	Estimate	*SE*	*t*	*p*
Intercept	602.24	3.62	166.27	<.001
BU-congruency	13.06	3.10	4.21	<.001
TD-congruency	42.18	6.66	6.33	<.001
Visibility	24.12	4.15	5.81	<.001
BU-Congruency × TD-Congruency	−5.93	6.60	−0.90	.369
BU-Congruency × Visibility	15.23	4.34	3.51	<.001
TD-Congruency × Visibility	42.44	4.46	9.51	<.001
BU-Congruency × TD-Congruency × Visibility	−20.49	6.96	−2.94	.003

Note. BU-congruency referred to the BU stimulus-driven congruency effect, and TD-congruency referred to the TD decision-driven congruency effect. TD-congruency, BU-congruency, and visibility were contrast-coded. GLMM = generalized linear mixed model; TD = top-down; BU = bottom-up.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix D

Table D1
Trial Count of Each Congruency Condition in Each Window in Experiment 1

Window	BU congruency	TD congruency	Trial count
1	Con	Con	2,343
1	Con	Incon	1,646
1	Incon	Con	1,667
1	Incon	Incon	2,328
2	Con	Con	2,358
2	Con	Incon	1,628
2	Incon	Con	1,588
2	Incon	Incon	2,397
3	Con	Con	2,444
3	Con	Incon	1,549
3	Incon	Con	1,550
3	Incon	Incon	2,429
4	Con	Con	2,463
4	Con	Incon	1,533
4	Incon	Con	1,468
4	Incon	Incon	2,494
5	Con	Con	2,553
5	Con	Incon	1,450
5	Incon	Con	1,421
5	Incon	Incon	2,537
6	Con	Con	2,603
6	Con	Incon	1,397
6	Incon	Con	1,335
6	Incon	Incon	2,627
7	Con	Con	2,708
7	Con	Incon	1,298
7	Incon	Con	1,287
7	Incon	Incon	2,680
8	Con	Con	2,751
8	Con	Incon	1,259
8	Incon	Con	1,189
8	Incon	Incon	2,767
9	Con	Con	2,850
9	Con	Incon	1,165
9	Incon	Con	1,145
9	Incon	Incon	2,810
10	Con	Con	2,904
10	Con	Incon	1,113
10	Incon	Con	1,051
10	Incon	Incon	2,896
11	Con	Con	2,983
11	Con	Incon	1,033
11	Incon	Con	998
11	Incon	Incon	2,957
12	Con	Con	3,029
12	Con	Incon	987
12	Incon	Con	905
12	Incon	Incon	3,049
13	Con	Con	3,117
13	Con	Incon	911
13	Incon	Con	866
13	Incon	Incon	3,088
14	Con	Con	3,159
14	Con	Incon	861
14	Incon	Con	800
14	Incon	Incon	3,149
15	Con	Con	3,233
15	Con	Incon	796
15	Incon	Con	745
15	Incon	Incon	3,211
16	Con	Con	3,267
16	Con	Incon	756
16	Incon	Con	676
16	Incon	Incon	3,269

Note. TD = top-down; BU = bottom-up; Con = congruent; Incon = incongruent.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix E

Table E1
Number of Trials With Each Subjective Visibility Rating in Experiment 2

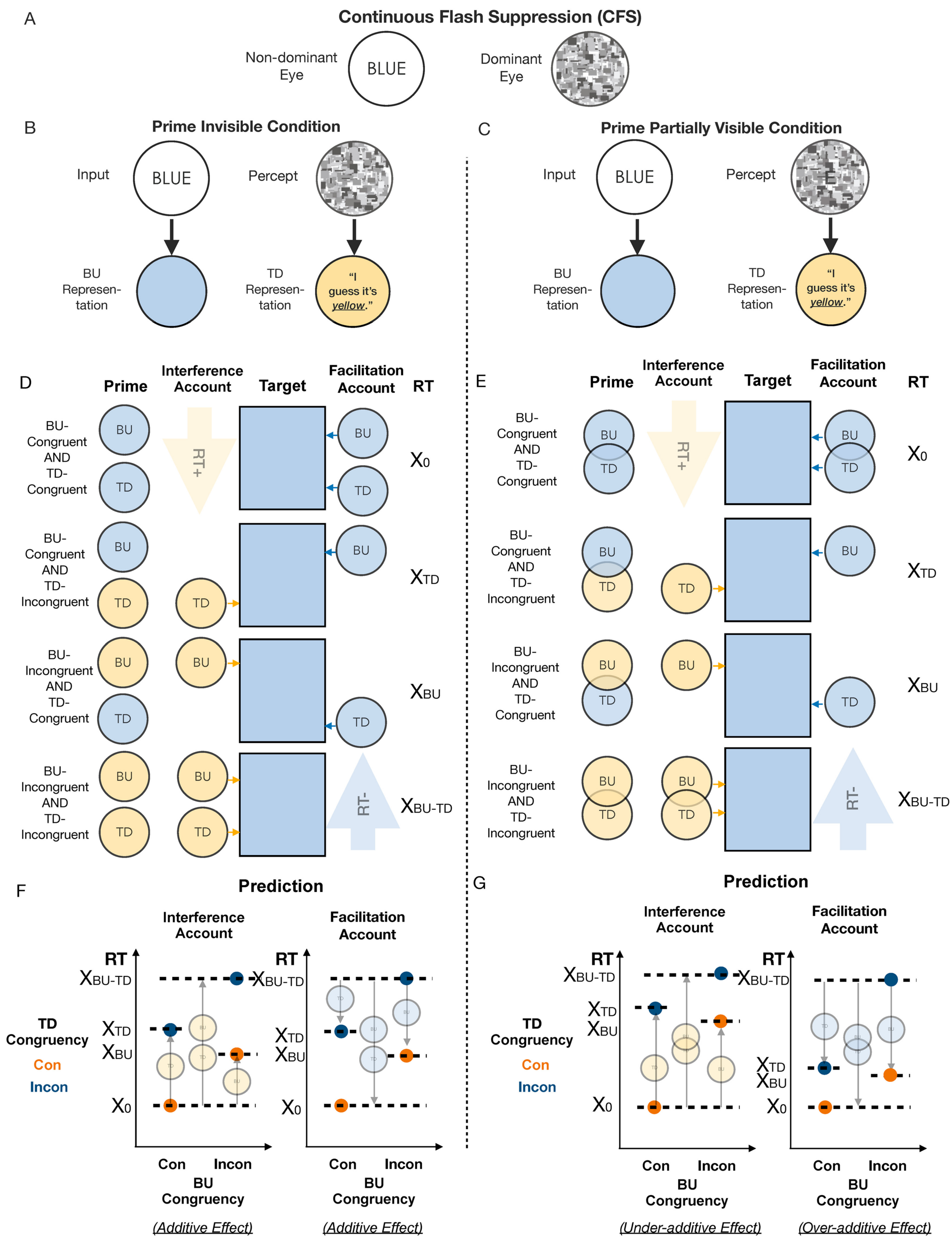
Subject ID	No percept	Unclear percept	Clear percept
103	165	60	10
105	58	118	2
106	107	54	7
107	51	77	39
109	171	44	1
110	98	105	11
111	133	104	0
112	144	77	4
113	115	100	19
115	132	78	12
116	173	46	3
117	67	157	2
119	94	124	5
120	154	59	4
122	135	69	25
124	62	156	7
126	147	83	4
127	55	94	83
128	191	42	3
129	114	76	29
134	127	88	8
135	111	101	2
137	85	135	18
139	44	80	108
140	73	141	24
144	132	70	22
145	110	108	8
148	134	80	13
153	109	94	29
155	133	65	15
156	113	97	21
157	193	41	3
158	76	116	46
161	79	87	73
164	67	94	54
165	149	57	6
168	87	116	24
169	70	143	18

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix F

Figure F1

Prediction Deduction of Interference and Facilitation Account



(*Continued on next page*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure F1 (*Continued*)

Note. (A) CFS was employed to manipulate prime visibility, in which prime is presented to the nondominant eye while continuously flashing mask is presented to the dominant eye. Subjects could generate the TD representation by randomly guessing prime identity when the prime is rendered invisible (B), or draw the inference based on limited available information, for example, a visible letter "E," when prime is partially visible (C). BU representation refers to original prime representation, while TD representation refers to reported prime representation (i.e., forced-choice discrimination decision). BU-congruency refers to congruency between BU representation and target. TD-congruency refers to congruency between TD representation and target. All trials in each visibility condition could be grouped into four different congruency conditions, respectively. BU and TD representations are hypothesized to be activated in independent networks when prime is invisible (D), and in overlapping networks when prime is visible (E). Interference hypothesis postulates that response priming is fundamentally a RT-delaying effect (RT+) driven by the incongruent representation, therefore, the congruent representation is equivalent to baseline. Facilitation hypothesis, reversely, predict that response priming is a RT-facilitating effect (RT−) driven by the congruent representation, therefore, the incongruent representation is equivalent to baseline. Global workspace model of consciousness predicted TD- and BU-congruency produce additive effect when prime is rendered invisible (F), and interaction effect when prime is visible (G). Specifically, interference hypothesis predicts underadditive effect, whereas facilitation hypothesis postulates overadditive effect in the prime (partially) visible conditions. Circles represent TD and BU neural networks, and the filled colors (i.e., blue/yellow in the online version and gray/light in the printed version) refer to the activated prime representation. Squares represent the target and the filled color (i.e., blue/yellow in the online version and gray/light in the printed version) refers to target representation. Arrows denote the putative presence of incongruency/congruency effect. CFS = continuous flash suppression; TD = top-down; BU = bottom-up; RT = response time; Con = Congruent; Incon = Incongruent. See the online article for the color version of this figure.

Received April 27, 2022
Revision received July 25, 2023
Accepted July 28, 2023 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.