Research Article

Psychological Science
2016, Vol. 27(7) 945–956

DOI: 10.1177/0956797616642525
pss.sagepub.com

Scene Integration Without Awareness: No Conclusive Evidence for Processing Scene Congruency During Continuous Flash Suppression

Pieter Moors, David Boelens, Jaana van Overwalle, and Johan Wagemans
Department of Brain and Cognition, University of Leuven

Abstract
A recent study showed that scenes with an object-background relationship that is semantically incongruent break interocular suppression faster than scenes with a semantically congruent relationship. These results implied that semantic relations between the objects and the background of a scene could be extracted in the absence of visual awareness of the stimulus. In the current study, we assessed the replicability of this finding and tried to rule out an alternative explanation dependent on low-level differences between the stimuli. Furthermore, we used a Bayesian analysis to quantify the evidence in favor of the presence or absence of a scene-congruency effect. Across three experiments, we found no convincing evidence for a scene-congruency effect or a modulation of scene congruency by scene inversion. These findings question the generalizability of previous observations and cast doubt on whether genuine semantic processing of object-background relationships in scenes can manifest during interocular suppression.

Keywords
visual perception, consciousness, subliminal perception, open data

Received 12/18/15; Revision accepted 3/10/16

Understanding the scope and limits of unconscious visual processing has become a central research topic in cognitive neuroscience (Dehaene & Changeux, 2011). Mudrik, Breska, Lamy, & Deouell (2011) claimed to have obtained evidence that complex, high-level visual scene processing can happen unconsciously. These authors presented participants with scenes that were rendered invisible through continuous flash suppression (CFS; Tsuchiya & Koch, 2005). Scene congruency was manipulated, and participants had to indicate when an initially suppressed scene broke suppression (i.e., when a *perceptual breakthrough* occurred). Mudrik et al. (2011) observed that incongruent scenes broke suppression faster than congruent scenes. This led the authors to argue that conscious awareness of a scene is not required for high-level scene-processing mechanisms to unfold and hence to extract the congruency relation between the object and the background.

Because the results of Mudrik et al. (2011) have profound implications for theories on the extent of unconscious visual processing during CFS, it is of utmost importance that the congruency effect be attributed to genuine scene-processing mechanisms rather than to differences in low-level visual aspects of the scenes. Image analyses (Itti & Koch, 2000; Neumann & Gegenfurtner, 2006) of the stimulus set used in these experiments did not seem to reveal any consistent bias regarding low-level visual aspects for either the congruent or the incongruent category. However, including an experimental condition in which the scenes were inverted would be a stronger control for image-related characteristics; inversion would dramatically reduce the scenes' identifiability but fully preserve the low-level image properties.

Corresponding Author:
Pieter Moors, Laboratory of Experimental Psychology, Department of Brain and Cognition, KU Leuven, Tiensestraat 102, Box 3711, B-3000 Leuven, Belgium
E-mail: pieter.moors@kuleuven.be

For the current study, we had three goals. First, we wanted to assess the replicability of the original findings. Second, by including a scene-inversion condition, we wanted to rule out any potential low-level confounds related to the particular stimulus set. Third, we complemented the traditional repeated measures analysis of variance (ANOVA) with a Bayesian analysis based on linear mixed-effects modeling with crossed random effects for participants and stimuli (Clark, 1973; Judd, Westfall, & Kenny, 2012). Given that the experiment consisted of presenting various exemplars of congruent and incongruent scenes to participants, we included stimulus as a random effect to generalize to the population of congruent and incongruent scenes. Furthermore, Bayesian statistics allows us to quantify the evidence for the absence of an effect (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

Experiment 1

Experiment 1 consisted of a replication experiment of Mudrik et al. (2011), using the same stimuli and methods, yet also including a scene-inversion condition. If the results observed in Mudrik et al. (2011) are genuinely attributable to unconscious scene processing, we should observe a scene-congruency effect in the upright condition but not in the inverted condition. Conversely, if the effect does not pertain to processing the semantic aspects of the suppressed scenes, we predict a similar congruency effect in the inverted condition and, critically, no statistical interaction between scene inversion and scene congruency.

Method

Participants. Forty-five people participated in the study in return for money or course credit. All participants were naive with respect to the purposes of the study and had normal or corrected-to-normal vision. Because the original sample size of 18 participants used in Mudrik et al. (2011) yielded a post hoc power of 75%, we decided to substantially increase the sample such that, given the effect size reported in the original study, the power for this experiment would be 99%. This increase in sample size was further motivated by the fact that one needs sufficient measurements for each item to fit a linear mixed-effects model with crossed random effects.

Apparatus. Stimuli were shown on two 19.8-in. CRT monitors (Trinitron GDM F500-R; Sony, Tokyo, Japan) driven by a personal computer (Precision T3400; Dell, Round Rock, TX) with a 2.5-GHz processor (Core 2 Quad Q9300; Intel, Santa Clara, CA) running Windows XP. The monitors' resolution was 2,048 × 1,536 pixels, and their refresh rate was 60 Hz. Binocular presentation was achieved by a custom-made stereo setup. The two CRT monitors were located to either side of the participant's head and faced each other at a distance of 220 cm. Two mirrors were placed in front of the participant at 110 cm from the monitors. Each mirror was angled so that the participant could see the left monitor with his or her left eye and the right monitor with his or her right eye. A combination head and chin rest (15 cm from the mirrors) was used to stabilize fixation. The effective viewing distance was 125 cm. Stimulus presentation, timing, and keyboard responses were controlled with custom software programmed in Python using the PsychoPy library (Peirce, 2007, 2009).

Stimuli. The background of the display consisted of a random checkerboard pattern to achieve stable binocular fusion. The size of the individual elements of the checkerboard was equal to 0.34°. In both monitors, a gray frame (10° × 10°) was superimposed on the checkerboard pattern to present the stimuli. A white fixation cross (0.5° × 0.5°) was continuously present during the experiment (except during the eye-dominance measurement, which used a black cross). In the eye-dominance-measurement phase, the target consisted of an arrow (maximal width = 4°, maximal height = 2°), and the CFS mask consisted of 150 gray squares of random size (between 1° and 2°), each with a random luminance value.

The scene stimuli (2.86° × 2.03°) were the same as in Mudrik et al. (2011), and a detailed description of the stimulus set is available in that article. In short, the scenes depicted various human actions involving a certain object. Both congruent and incongruent versions of the scene were created by adding an object into the scene that was related or unrelated to the action (Mudrik et al., 2011). In the main experiment, the CFS mask consisted of 200 square elements with a randomly chosen color and size (between 0.75° and 1.5°) for each element. The elements were generated in a 5.26° × 5.26° square window centered at fixation. Because the maximal size of each element was 1.5°, the effective size of the CFS mask was thus at least 5.26° × 5.26° and no larger than 6.76° × 6.76° (compared with the constant 5.26° × 5.26° window in the original study). In both the eye-dominance phase as well as the main experiment, the refresh rate of the CFS mask was set at 10 Hz.

Procedure. In the first part of the experiment, participants' eye dominance was measured using the method of Yang, Blake, and McDonald (2010). On every trial, a fixation cross was presented for 1 s. Next, an arrow that gradually increased from 0% to 100% contrast over 2 s was presented to one eye, and the CFS mask was presented to the other eye. The arrow pointed either left or right. On breakthrough of the arrow, participants had to indicate the arrow's direction by pressing "1" for left or "3" for right on a numerical keyboard. The CFS mask was randomly presented to the left or right eye (40 trials per eye) for a total of 80 trials. The dominant eye was determined as the eye in which mean suppression time was shortest when the arrow was presented to that eye. In this phase, 44% of the observers were determined to be left-eye dominant, rather than right-eye dominant. In all subsequent phases of the experiment, the CFS mask should have been presented to each participant's dominant eye. Because of a programming error, however, the mask was always presented to the right eye. Thus, some participants in Experiment 1 received the mask in the nondominant eye.

The main experiment consisted of 160 CFS trials divided into four blocks of 40 trials. In each block, all four conditions (all combinations of scene congruency and scene inversion) were balanced. Only one version of a scene was presented in each block. Within each block, the ordering of conditions was completely randomized, whereas the original study was constrained such that items of the same type could not be presented on 4 or more consecutive trials. Before the start of the main experiment, participants completed 16 practice trials using four scenes that were not included in the main experiment.

On each trial, a CFS mask was presented to the right eye while the scene was presented to the left eye. The scene gradually increased from 0% to 100% contrast in steps of 10% every 100 ms. After the scene had reached full contrast, the CFS mask began to decrease in contrast to 0% over the course of 5.1 s. On breakthrough, participants had to indicate as quickly as possible whether the scene was presented to the left or right of fixation (for an overview of the trial sequence, see Fig. 1).

In the original study, there was a postexperimental rating session in which participants were asked to categorize all scene stimuli as being unusual or not, and incorrectly categorized stimuli were then removed before the start of the analysis. Because categorization performance could influence the results, we later invited these participants to an online rating session (see Experiment 3).

Results

All analyses were performed on correct trials (M = 0.99, SD = 0.01) for which the suppression time did not exceed the time at which the mask reached 0% contrast (M = 0.92, SD = 0.13), after removing outliers (defined as

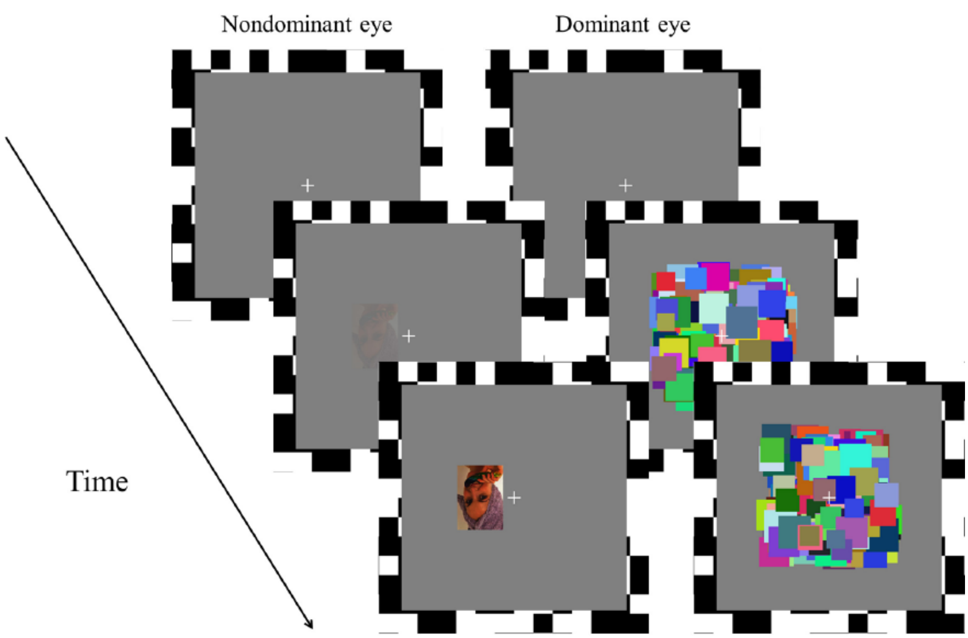

Fig. 1. Trial sequence for all experiments. A fixation cross was presented to both eyes for 1 s, after which the scene stimulus was presented to the nondominant eye, and the CFS mask was presented to the dominant eye. (Note that a programming error during Experiments 1 and 2 caused the mask to always be presented to participants' right eyes; thus, some participants received the mask in the nondominant eye.) The scene stimulus gradually increased in contrast for 1 s, after which the CFS mask decreased in contrast for 5.1 s. On breakthrough, participants had to indicate as quickly as possible whether the scene stimulus was presented to the left or the right of the fixation cross.

suppression times ≥ 3 *SD* higher than the mean; for each observer separately, $M = 0.005$, $SD = 0.007$). For the Bayesian analysis, we log-transformed the suppression times to account for the positive skew in the suppression-time distributions. Bayes factors (BFs) were computed to quantify the evidence for the presence or absence of a main effect or interaction. To compute the BFs, we used the BayesFactor package (version 0.9.11-1; Morey & Rouder, 2015) for the R software environment (R Development Core Team, 2015); we used default settings: the medium prior scale for fixed effects and the nuisance prior scale for random effects. All models were linear mixed-effects models with crossed random effects, including a random intercept for both participants and stimuli (Rouder, Morey, Speckman, & Province, 2012). To compute the BFs for the main effects and interaction, we compared a full model (including the two main effects and interaction) with a reduced model in which the effect of interest was not included (i.e., similar to the traditional repeated measures ANOVA).

According to the classification provided by Jeffreys (1961), a BF of 3 constitutes substantial evidence that the data are in line with one model compared with another, and a BF of 10 is considered strong evidence that the data are more likely under one model compared with the other. It is important to stress that BFs constitute a relative measure of evidence (i.e., it is a ratio of marginal likelihoods computed for two statistical models). In this article, we always report BFs with the values for the reduced model in the numerator and that for the full model in the denominator. Thus, BFs greater than 3 indicate evidence for the absence of a main effect or interaction under consideration, and BFs less than 0.3 indicate evidence for the presence of the effect under consideration. Because the BF is asymmetric around 1, we graphed the BFs after $\log_{10}$-transforming them such that a $\log_{10}$ BF of 1 or −1 indicates strong evidence for the absence or presence of an effect, respectively.

Before analyzing the data, we first checked the order in which stimuli broke suppression, from fastest to slowest, and whether this order was consistent across observers. That is, if the images were being processed during suppression, the low-level image characteristics should at least have an influence on suppression times, and this should be apparent from calculating Cronbach's α for the suppression times. Indeed, Cronbach's α was high for the suppression durations ($\alpha = .83$), indicating that there was consistency across observers for the order in which stimuli broke suppression, from fastest to slowest.

A two-way repeated measures ANOVA on the mean correct suppression times revealed no main effect of scene congruency (congruent scene: $M = 2.53$, $SD = 0.88$; incongruent scene: $M = 2.56$, $SD = 0.86$), $F(1, 44) = 1.08$, $p = .30$, $d = -0.11$, a main effect of scene inversion (upright scene: $M = 2.49$, $SD = 0.87$; inverted scene: $M = 2.61$, $SD = 0.86$), $F(1, 44) = 16.1$, $p = .0002$, $d = -0.44$, and no interaction between scene congruency and scene inversion, $F(1, 44) = 0.022$, $p = .88$, $d = -0.02$. As shown in Figure 2, inverted scenes broke suppression more slowly, on average, than did upright scenes. Furthermore, congruent scenes broke suppression numerically faster, on average, than did incongruent scenes, yet the *p* value associated with this difference did not reach the significance threshold.

The results of the repeated measures ANOVA were complemented by the results of the Bayesian analysis. Specifically, the BFs indicated strong evidence for (a) the absence of a scene-congruency effect (BF = 17), (b) the presence of a scene-inversion effect (BF < 0.01), and (c) the absence of an interaction effect (BF = 24).

Figure 3 depicts a sequential analysis of the data for both statistical techniques. The top panels depict how the *p* values of the repeated measures ANOVA evolved as more participants were tested. As shown, the main effect of congruency passed the significance threshold of .05 several times during data collection. In contrast, the interaction never reached significance. The evolution of the BFs tells a different story. Here, the main effect of congruency nearly always hovers around a $\log_{10}$ BF value of 1 (i.e., a BF of 10), indicating strong evidence for the absence of a congruency effect. The BF of the interaction effect also shows a gradual increase toward more evidence in favor of the absence of an interaction effect. For the main effect of inversion, both patterns show a gradual increase in evidence for an inversion effect.

We reanalyzed the data by excluding all participants for whom the CFS mask was presented to the nondominant eye. These reanalyses indicated that the results did not change when we included only participants for whom the CFS mask was presented to the dominant eye (see Supplemental Analyses in the Supplemental Material available online).

Discussion

The goal of Experiment 1 was to replicate the scene-congruency effect observed in Mudrik et al. (2011) and to assess whether it would be influenced by scene inversion. If genuine integration between the objects of a scene and its background can still proceed during CFS, one would predict that inverting the scenes would reduce or diminish the scene-congruency effect. The results of Experiment 1 showed no convincing evidence for a main effect of scene congruency or an interaction between scene congruency and scene inversion. In contrast, a reliable scene-inversion effect was found: Upright scenes broke suppression faster than inverted scenes.

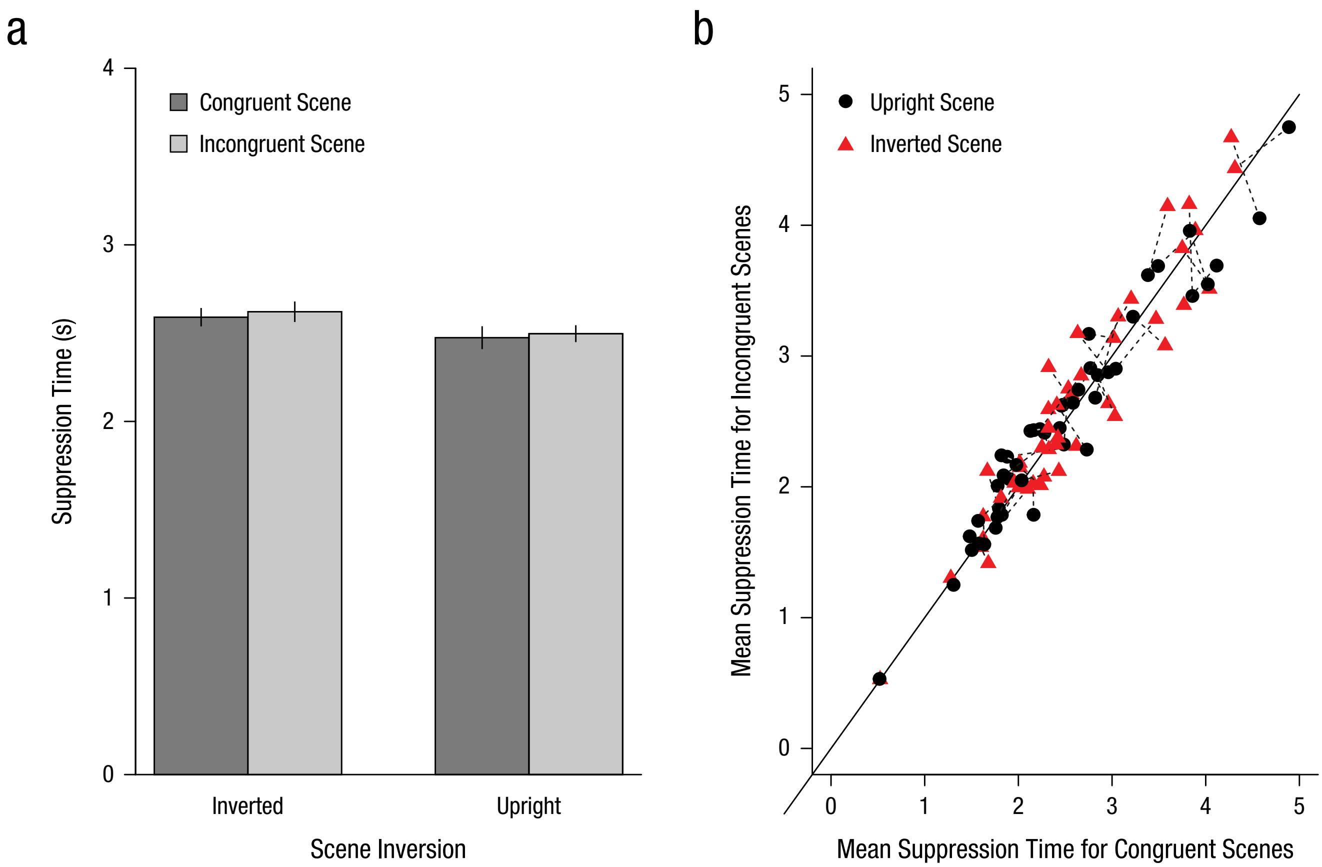

Fig. 2. Results of Experiment 1. Mean suppression time (a) is graphed as a function of scene inversion, separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). The relationship between mean suppression times for congruent and for incongruent scenes (b) is shown separately for the two scene types. Each dashed line connects the data points for upright and inverted scenes for a single participant. The diagonal line represents equal suppression durations for congruent and incongruent scenes.

Experiment 2

To increase our confidence in the absence of a scene-congruency effect, we ran a second experiment in which we increased the number of trials for each observer. Furthermore, we dropped the inversion condition. This allowed us to run a quasi-exact replication of the original study and to assess whether increasing the number of trials for each observer would give us more power to detect a scene-congruency effect. Furthermore, we decided to drop the mask-fade-out procedure because it had forced us to exclude a high number of trials among observers for whom suppression was very effective.

Method

Participants. Twenty-four new people participated in the study in return for money or course credit. All participants were naive with respect to the purposes of the study and had normal or corrected-to-normal vision.

Apparatus and stimuli. The setup and stimuli were exactly the same as in Experiment 1, except that inverted scenes were no longer included.

Procedure. The procedure was the same as in Experiment 1, except as noted. In this experiment, 50% of participants were identified as left-eye dominant. Thus, the CFS mask should have been presented to the nondominant eye for half of the participants. Because of a programming error, however, the mask was always presented to participants' right eyes. Thus, some participants in Experiment 2 received the mask in the nondominant eye. The experiment was set up such that each block constituted a fairly exact replication of the original study (Mudrik et al., 2011). That is, the experimental procedure was the same as in the original study, and this procedure was repeated four times. We eliminated the mask-fade-out procedure to ensure that the absence of a scene-congruency effect could not be attributed to limiting the observations for observers for whom suppression was

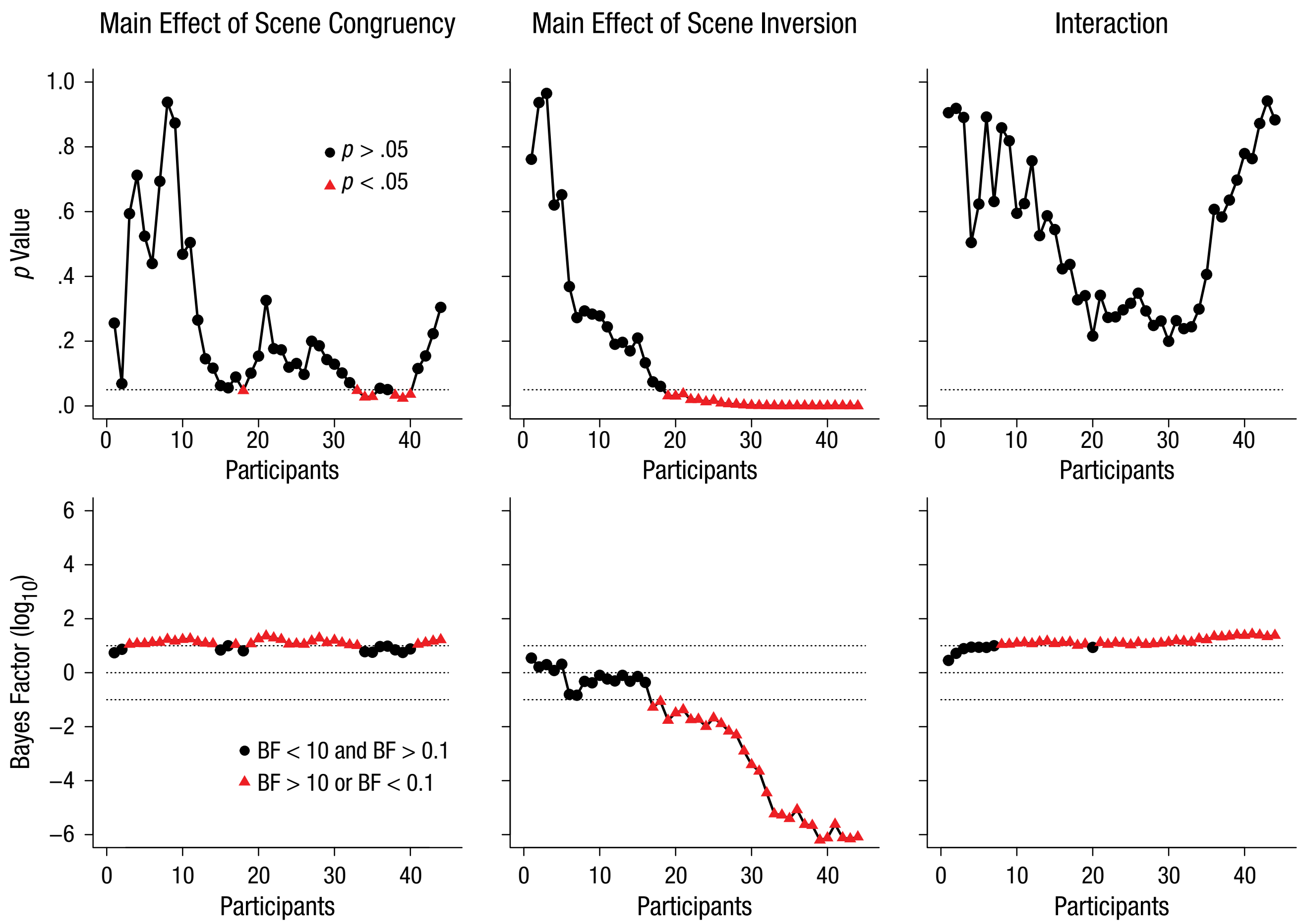

Fig. 3. Sequential analysis of the data from Experiment 1. The graphs show *p* values (top row) and $\log_{10}$ Bayes factor (BF) values (bottom row) as a function of the cumulative number of participants included in the analysis, separately for the main effect of scene congruency, the main effect of scene inversion, and their interaction. In the top row, the dotted lines indicate the significance threshold ($p = .05$). In the bottom row, the dotted lines indicate, from top to bottom, $\log_{10}$ BF = 1, $\log_{10}$ BF = 0, and $\log_{10}$ BF = −1; $\log_{10}$ BF values greater than 0 indicate evidence for the absence of an effect, whereas $\log_{10}$ BF values smaller than 0 indicate evidence for the presence of an effect.

very effective and thus excluding the upper tail of the suppression-time distributions for these participants. We also eliminated the postexperimental rating session in Experiment 2, but we contacted the participants later and asked them to complete the task online, as we did for Experiment 1 (see Experiment 3).

Design. Scene congruency was the only factor that was manipulated. Participants completed 8 practice trials, and the main experiment consisted of four repetitions of the full stimulus set, amounting to 320 trials in total.

Results

The results were analyzed in the same way as in Experiment 1. That is, all analyses were performed on the correct trials (first block: $M = 0.98$, $SD = 0.02$; all blocks: $M = 0.99$, $SD = 0.02$), after removing outliers (defined as suppression times ≥ 3 *SD* higher than the mean; for each observer separately, first block: $M = 0.007$, $SD = 0.01$; all blocks: $M = 0.02$, $SD = 0.008$). We report results separately for the first block alone (quasi-exact replication of the original study) and for all four blocks combined.

First, we again analyzed the consistency of suppression times for images across observers. This again indicated high reliability for the suppression times of the images across observers (first block: $\alpha = .67$; all blocks: $\alpha = .81$). The order in which stimuli broke suppression, from fastest to slowest, was highly correlated across Experiments 1 and 2 ($r = .89$, BF < 0.01).

No effect of scene congruency was observed in either the first block (congruent scenes: $M = 3.06$, $SD = 1.16$; incongruent scenes: $M = 3.08$, $SD = 1.13$), $t(23) = −0.21$, $p = .83$, $d = −0.04$, or among all blocks (congruent scenes: $M = 2.34$, $SD = 0.77$; incongruent scenes: $M = 2.36$, $SD = 0.77$), $t(23) = −0.69$, $p = .50$, $d = −0.14$. Likewise, a BF analysis of the data always indicated convincing evidence for the absence of a scene-congruency effect (first block:

BF = 17; all blocks: BF = 32). These results were very similar to those for Experiment 1.

As shown in Figure 4, increasing the number of trials per observer never led to a significant main effect of congruency. Again, and in contrast with the results of the traditional analysis, the BF analysis showed a gradual increase in evidence for the absence of a scene-congruency effect as the data accumulated.

Supplemental analyses (see the Supplemental Material) indicated that the results did not change when we included only participants for whom the CFS mask was presented to the dominant eye.

Discussion

Although we obtained no evidence for a scene-congruency effect in either Experiment 1 or Experiment 2, one might still raise objections to these attempts. First, compared with the original study (Mudrik et al., 2011), we did not implement a postexperimental rating procedure in which participants were asked to categorize the images as being unusual or not (i.e., incongruent or congruent) in the first two experiments. Second, the CFS mask was not presented in every participant's dominant eye in the first two experiments. Third, the addition of an inversion condition in Experiment 1 and omitting the mask-fade-out procedure, as well as repeating the images more than once in Experiment 2, might have obscured a subtle congruency effect.

Experiment 3

We addressed these concerns in a third experiment, in which we presented every scene upright, only once across two experimental blocks, and always in the participants' nondominant eye. We included the mask-fade-out procedure

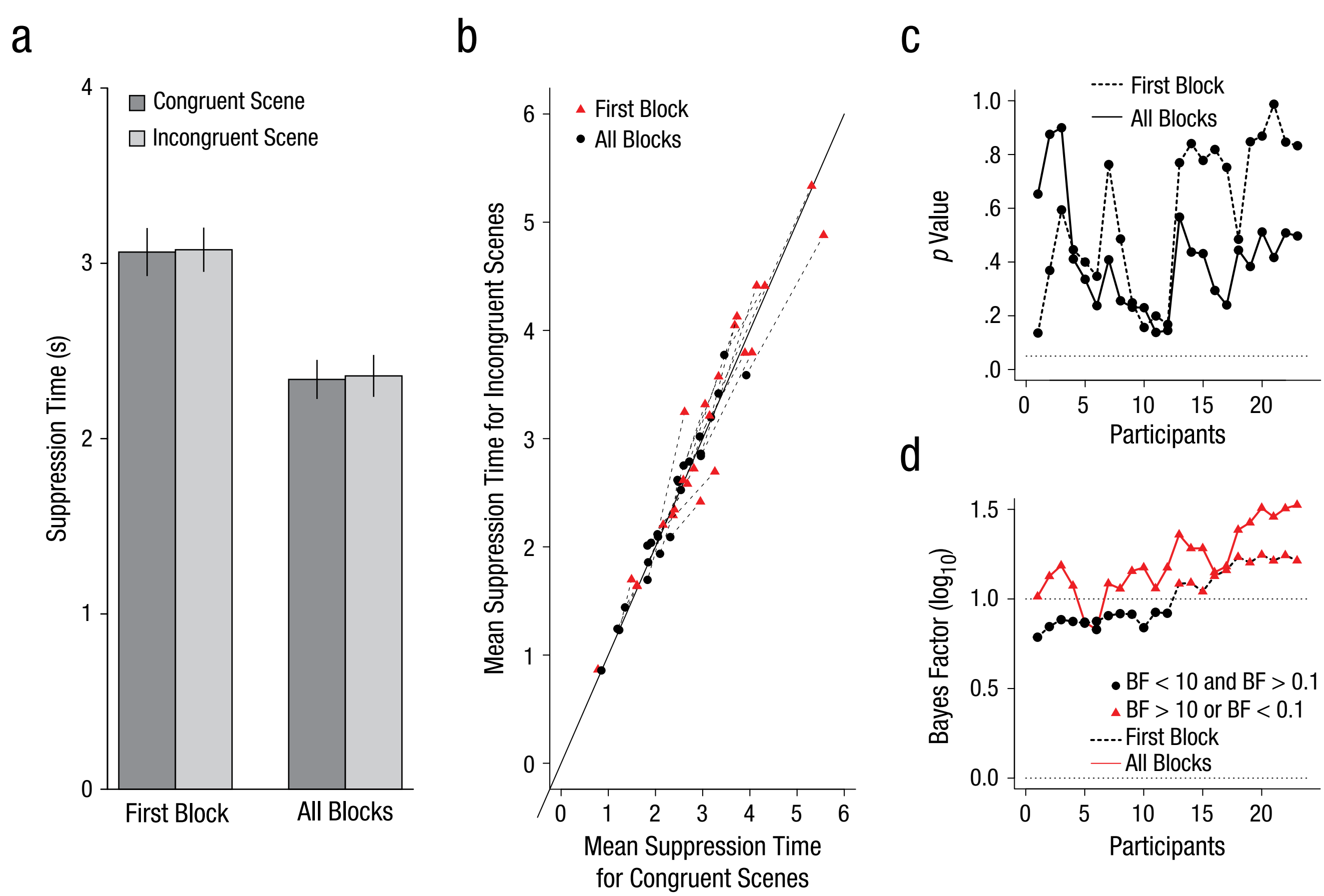

Fig. 4. Results of Experiment 2. Mean suppression time (a) is graphed for the first block and for all blocks, separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). In (b), the relationship between mean suppression times for congruent and for incongruent scenes is shown for each participant's performance in the first block and in all blocks. Each dashed line connects the data points for a single participant. The diagonal line represents equal suppression durations for congruent and incongruent scenes. The two other graphs show (c) p values and (d) $\log_{10}$ Bayes factor (BF) values as a function of the cumulative number of participants in the analysis, separately for the first block and for all blocks. In (c), the dotted line indicates the significance threshold (p = .05). In (d), the dotted lines indicate, from top to bottom, $\log_{10}$ BF = 1 and $\log_{10}$ BF = 0; $\log_{10}$ BF values greater than 0 indicate evidence for the absence of an effect, whereas $\log_{10}$ BF values smaller than 0 indicate evidence for the presence of an effect.

again, and this time, we implemented a postexperimental rating session.

Method

Participants. A new sample of 50 people participated in the experiment in return for money or course credit. All participants were naive with respect to the design of the study and had normal or corrected-to-normal vision. All participants provided written informed consent before the start of the experiment.

Apparatus and stimuli. The experimental setup and stimuli were exactly the same as in Experiment 2.

Procedure and design. Until the postexperimental rating session, the procedure was very similar to that in Experiment 2 (except for the use of the mask-fade-out procedure). Participants first completed the eye-dominance experiment, after which they completed two blocks of trials in which the scenes were presented to the nondominant eye and the CFS mask was presented to the dominant eye. After participants completed the CFS experiment, each scene was presented binocularly, and participants were asked to indicate whether they thought the presented scene was unusual or not (i.e., incongruent or congruent).

The experiment consisted of a within-subjects design with two conditions (i.e., congruent and incongruent). In the main experiment, participants completed 8 practice trials and 80 experimental trials. Within each block, the order in which trials were presented was completely randomized. The postexperimental rating session also consisted of 80 trials, presented in random order. At this point, we also invited participants from Experiments 1 and 2 to perform the rating task online, and we reanalyzed the data from Experiments 1 and 2 on the basis of their responses.

Results

The data were analyzed in the same way as in Experiments 1 and 2. That is, all analyses were performed on the correct trials (M = 0.98, SD = 0.03) for which the suppression time did not exceed the time at which the mask reached 0% contrast (M = 0.91, SD = 0.18), after removing outliers (again defined as suppression times ≥ 3 SD higher than the mean; for each observer separately, M = 0.005, SD = 0.01). Furthermore, only stimuli that were correctly categorized after the experiment (i.e., congruent as congruent and incongruent as incongruent) were included in the analysis (M = 0.71, SD = 0.06). A reanalysis including the incorrectly rated scenes (see Supplemental Analyses) did not change the results of the main analysis.

The speed with which given stimuli broke suppression was consistent across observers and was similar to what was observed in Experiments 1 and 2 (α = .85). Furthermore, the ordering of such stimuli (from fastest to slowest) again correlated very strongly with the order observed in Experiment 1 (r = .89, BF < 0.01) and Experiment 2 (r = .94, BF < 0.01), indicating that the images were processed similarly (irrespective of their congruency) across all three experiments.

As in Experiments 1 and 2, an analysis of mean correct suppression times revealed no effects of scene congruency (congruent scenes: M = 2.65, SD = 1.14; incongruent scenes: M = 2.65, SD = 1.09), $t(49) = -0.03$, p = .97, d = −0.005. Likewise, the BF analysis indicated strong evidence for the absence of a congruency effect (BF = 21). Figure 5 depicts the results of the sequential analysis. We observed a significant congruency effect in the traditional analysis early on during data collection, yet the BF never crossed the boundary for indicating evidence for a congruency effect and again gradually accumulated evidence for the absence of a congruency effect.

When we accounted for the categorizations provided in the online postexperimental rating session by participants from Experiments 1 and 2, the results were no different from those in the main analyses of Experiments 1 and 2 (see Supplemental Analyses).

In another supplemental exploratory analysis (see Supplemental Analyses), we examined the relationship between various statistical image properties and the mean suppression times for the images in all three experiments. That is, given the absence of a scene-congruency effect and the presence of high consistency regarding the order in which stimuli broke suppression, from fastest to slowest, we were interested in exploring whether some statistical properties of the images would predict suppression times. The results indicated that a measure of spatial coherence (an indicator of scene fragmentation) correlated positively with mean suppression duration. That is, when images were more cluttered, suppression times were higher on average.

Discussion and Conclusion

The goal of this study was to assess the replicability of the findings reported in Mudrik et al. (2011) and to gauge whether the scene-congruency effect was attributable to low-level differences between the stimuli by including a scene-inversion condition. In Experiment 1, using a larger sample than in the original study, we observed no effect of scene congruency and, critically, no interaction between scene inversion and scene congruency. Thus, scene inversion did not modulate the effect of scene

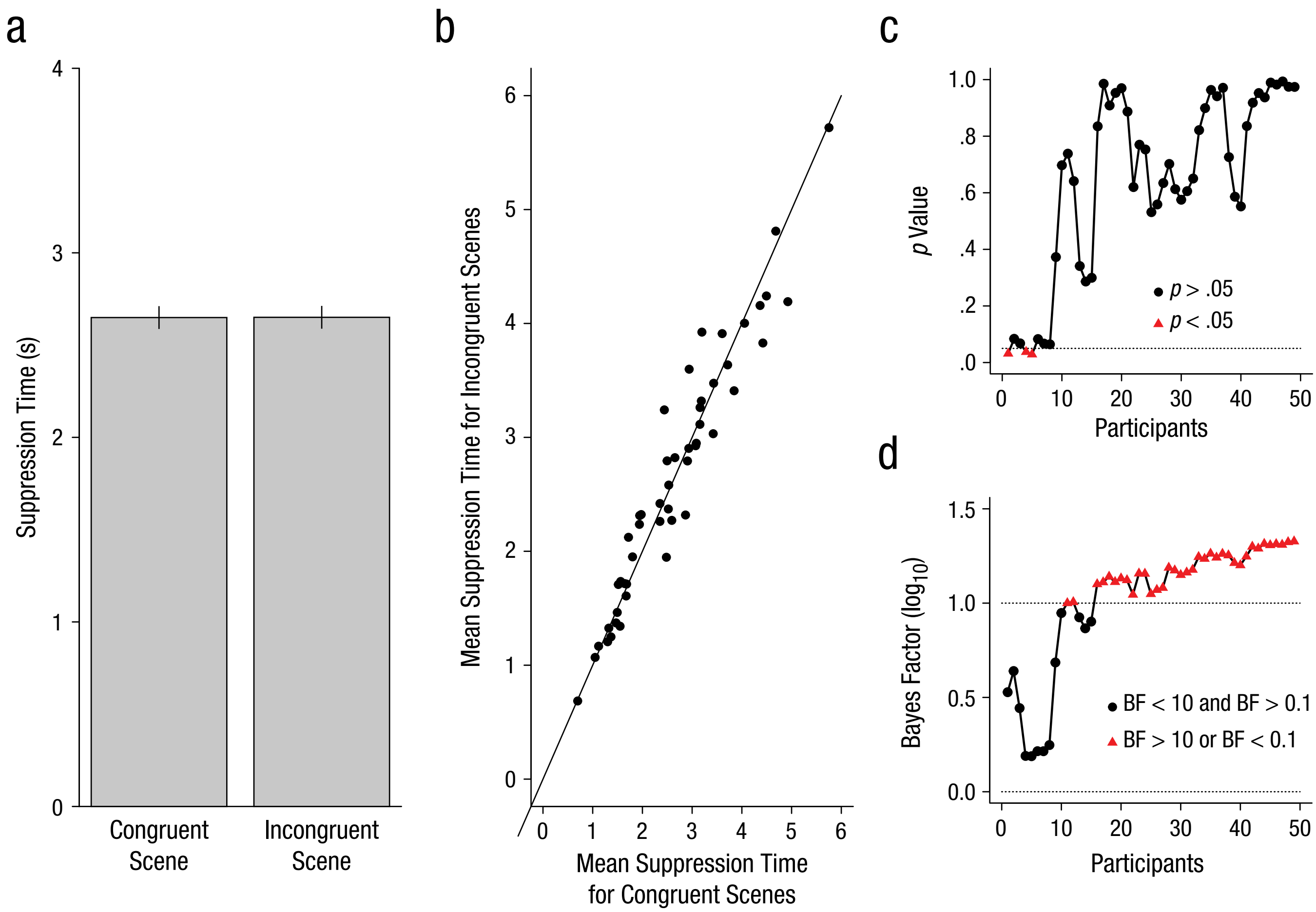

Fig. 5. Results of Experiment 3. Mean suppression time (a) is graphed separately for congruent and incongruent scenes. Error bars represent within-subjects 95% confidence intervals with the adjustment suggested by Morey (2008). The relationship between mean suppression times for congruent and for incongruent scenes (b) is shown. The diagonal line represents equal suppression durations for congruent and incongruent scenes. The two other graphs show (c) p values and (d) $\log_{10}$ Bayes factor (BF) values as a function of the cumulative number of participants in the analysis. In (c), the dotted line indicates the significance threshold ($p = .05$). In (d), the dotted lines indicate, from top to bottom, $\log_{10}$ BF = 1 and log 10 BF = 0; $\log_{10}$ BF values greater than 0 indicate evidence for the absence of a congruency effect, whereas $\log_{10}$ BF values smaller than 0 indicate evidence for the presence of a congruency effect.

congruency, which would have been expected if the scene-congruency effect were a semantic effect. Furthermore, a BF analysis relying on linear mixed-effects models with crossed random effects showed convincing evidence for the absence of a scene-congruency effect. A sequential analysis of our data highlighted a strong discrepancy between the inferences from a traditional repeated measures ANOVA and the BF analysis. Indeed, at several steps throughout our data collection, the traditional analysis yielded a significant main effect of scene congruency, whereas this was never the case for the BF analysis, indicating that failing to consider the random-item variation yields a statistical procedure that is too liberal (Clark, 1973).

In the case of the scene-inversion effect, both the traditional and the Bayesian analysis converged toward the same conclusion. In Experiment 2, we increased the number of trials fourfold for each observer to assess whether increasing the precision of the effect-size estimate for each observer would reveal a more consistent scene-congruency effect. Again, we observed no scene-congruency effect. Experiment 3 consisted of a third high-powered replication attempt in which we also included a postexperimental rating task to exclude congruent stimuli that were perceived as incongruent (and incongruent stimuli that were perceived as congruent) by our participants. The results of this last experiment also indicated strong evidence for the absence of a scene-congruency effect.

What do the results of these experiments tell us? Could it be that the images were not processed at all? We contend that several aspects of our results argue against such an interpretation. In all experiments, we observed high internal consistency across observers regarding the order in which stimuli broke suppression, from fastest to slowest. Moreover, this pattern of slow and fast breakthrough correlated strongly across all experiments in independent sets of observers. Third, a measure of spatial coherence correlated with suppression times (with a similar magnitude) in all experiments. These aspects indicate that the scene stimuli did not break suppression in a completely random fashion.

In Experiment 1, we observed a consistent scene-inversion effect. Could this be an indication that the stimuli were somehow interpreted during suppression? Although inversion reduces the identifiability of the scenes and preserves low-level image features, it may also influence higher-order image features to which the visual system is sensitive (Okazawa, Tajima, & Komatsu, 2015). Therefore, although an inversion effect does not by definition indicate that a stimulus is processed up to a semantic level, it could also reflect the sensitivity of the visual system to natural input.

The original scene-congruency effect was interpreted as evidence for unconscious integration; nevertheless, our results cannot be interpreted as providing evidence against unconscious integration per se. That is, there remain some differences between our study and the original. Besides the obvious differences in hardware, experimental environment, and pools of observers, there were also slight differences in trial randomization procedure and the size of the CFS mask display. Our findings therefore show that the results of Mudrik et al. (2011) do not generalize across these particular testing differences, indicating that any scene-congruency effect is particularly fragile.

However, this lack of generalizability is hardly compatible with the conclusions derived from the original result. If unconscious integration can manifest under CFS, such an effect should not be dependent on factors such as particular testing conditions or a different participant pool. Indeed, our results show that there is no evidence for scene integration without awareness under CFS. Although CFS initially proved to be a promising technique to assess the limits of unconscious visual processing (Bahrami et al., 2010; Jiang, Costello, Fang, Huang, & He, 2006; Jiang, Costello, & He, 2007; Sklar et al., 2012), our findings fit well in a series of more recent findings highlighting rather limited visual processing during CFS (Hedger, Adams, & Garner, 2015; Hesselmann & Knops, 2014; Heyman & Moors, 2014; Moors, Huygelier, Wagemans, de-Wit, & van Ee, 2015; Moors, Wagemans, van Ee, & de-Wit, 2016). In hindsight, this also is not too surprising given that binocular rivalry is known to disrupt processing of the suppressed stimulus beyond early visual areas (Fogelson, Kohler, Miller, Granger, & Tse, 2014; Hesselmann & Malach, 2011; Yuval-Greenberg & Heeger, 2013). This does not imply, however, that unconscious integration per se cannot take place. Indeed, some forms of unconscious integration have been shown to exist (for a review, see Mudrik, Faivre, & Koch, 2014) yet often rely on different suppression paradigms.

Although this study highlights the importance of replication and the inclusion of appropriate control conditions, it also reveals a much broader point: Different statistical methods can produce strongly divergent results throughout the data-collection process. Indeed, although both types of analysis yielded the same conclusion at the end of data collection, the traditional repeated measures ANOVA more than once indicated a significant scene-congruency effect as the data accumulated. The BF analysis, however, yielded a more consistent picture in that it always provided evidence for the absence of a scene-congruency effect while the evidence also gradually accumulated when more data were collected. Furthermore, in Experiment 1, the BF indicated convincing evidence for the presence of an inversion effect, and the results of the repeated measures ANOVA converged with those of the BF analysis. This highlights the idea that for experimental designs in which the dependent measure can vary across participants and items, using the traditional repeated measures ANOVA might be too liberal an approach (Clark, 1973), and an approach based on crossed random effects is recommended (Baayen, Davidson, & Bates, 2008).

In sum, our study questions the replicability and generalizability of the findings reported in Mudrik et al. (2011) by obtaining strong evidence for the absence of a scene-congruency effect across three experiments and, moreover, showing that scene congruency was not modulated by scene inversion. Therefore, it is unlikely that complex high-level scene processing can ensue during CFS.

Action Editor

Alice J. O'Toole served as action editor for this article.

Author Contributions

P. Moors developed the study concept. Testing and data collection were performed by D. Boelens and J. van Overwalle. P. Moors performed the data analysis and interpretation. P. Moors and J. Wagemans drafted the manuscript, and D. Boelens and J. van Overwalle provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgments

We are grateful to Liad Mudrik for sharing the original stimulus set.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by a Fund for Scientific Research Flanders (FWO) doctoral fellowship awarded to P. Moors and by the Methusalem program of the Flemish government (Grants METH/08/02 and METH/14/02) awarded to J. Wagemans.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

Open Practices

All data have been made publicly available via Figshare and can be accessed at https://figshare.com/s/cccac39f891998509c85. The complete Open Practices Disclosure for this article can be found at http://pss.sagepub.com/content/by/supplemental-data. This article has received the badge for Open Data. More information about the Open Practices badges can be found at https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/ and http://pss.sagepub.com/content/25/1/3.full.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005

Bahrami, B., Vetter, P., Spolaore, E., Pagano, S., Butterworth, B., & Rees, G. (2010). Unconscious numerical priming despite interocular suppression. *Psychological Science*, *21*, 224–233. doi:10.1177/0956797609360664

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359. doi:10.1016/S0022-5371(73)80014-3

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*, 200–227. doi:10.1016/j.neuron.2011.03.018

Fogelson, S. V., Kohler, P. J., Miller, K. J., Granger, R., & Tse, P. U. (2014). Unconscious neural processing differs with method used to render stimuli invisible. *Frontiers in Psychology*, *5*, Article 601. doi:10.3389/fpsyg.2014.00601

Hedger, N., Adams, W. J., & Garner, M. (2015). Autonomic arousal and attentional orienting to visual threat are predicted by awareness. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 798–806. doi:10.1037/xhp0000051

Hesselmann, G., & Knops, A. (2014). No conclusive evidence for numerical priming under interocular suppression. *Psychological Science*, *25*, 2116–2119. doi:10.1177/0956797614548876

Hesselmann, G., & Malach, R. (2011). The link between fMRI-BOLD activation and perceptual awareness is "stream-invariant" in the human visual system. *Cerebral Cortex*, *21*, 2829–2837. doi:10.1093/cercor/bhr085

Heyman, T., & Moors, P. (2014). Frequent words do not break continuous flash suppression differently from infrequent or nonexistent words: Implications for semantic processing of words in the absence of awareness. *PLoS ONE*, *9*(8), Article e104719. doi:10.1371/journal.pone.0104719

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Jeffreys, H. (1961). *Theory of probability*. Oxford, England: Oxford University Press.

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences, USA*, *103*, 17048–17052. doi:10.1073/pnas.0605678103

Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: Advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychological Science*, *18*, 349–355. doi:10.1111/j.1467-9280.2007.01902.x

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54–69.

Moors, P., Huygelier, H., Wagemans, J., de-Wit, L., & van Ee, R. (2015). Suppressed visual looming stimuli are not integrated with auditory looming signals: Evidence from continuous flash suppression. *i-Perception*, *6*, 48–62. doi:10.1068/i0678

Moors, P., Wagemans, J., van Ee, R., & de-Wit, L. (2016). No evidence for surface organization in Kanizsa configurations during continuous flash suppression. *Attention, Perception, & Psychophysics*, *78*, 902–914. doi:10.3758/s13414-015-1043-x

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (Version 0.9.11-1) [Software]. Retrieved from https://github.com/richarddmorey/BayesFactor/tree/0.9.11-1

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science*, *22*, 764–770. doi:10.1177/0956797611408736

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, *18*, 488–496. doi:10.1016/j.tics.2014.04.009

Neumann, D., & Gegenfurtner, K. R. (2006). Image retrieval and perceptual similarity. *ACM Transactions on Applied Perception*, *3*, 31–47. doi:10.1145/1119766.1119769

Okazawa, G., Tajima, S., & Komatsu, H. (2015). Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proceedings of the National Academy of Sciences, USA*, *112*, E351–E360. doi:10.1073/pnas.1415146112

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13. doi:10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*, Article 10. doi:10.3389/neuro.11.010.2008

R Development Core Team. (2015). R: A language and environment for statistical computing (Version 3.2.2) [Computer software]. Retrieved from https://www.r-project.org/index.html

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374. doi:10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237. doi:10.3758/PBR.16.2.225

Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences, USA*, *109*, 19614–19619. doi:10.1073/pnas.1211645109

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

Yang, E., Blake, R., & McDonald, J. E. (2010). A new interocular suppression technique for measuring sensory eye dominance. *Investigative Ophthalmology & Visual Science*, *51*, 588–593. doi:10.1167/iovs.08-3076

Yuval-Greenberg, S., & Heeger, D. J. (2013). Continuous flash suppression modulates cortical activity in early visual cortex. *The Journal of Neuroscience*, *33*, 9635–9643. doi:10.1523/JNEUROSCI.4612-12.2013