Sound-symbolism effects in the absence of awareness. A replication study.

Tom Heyman¹, Anne-Sofie Maerten¹, Hendrik Vankrunkelsven¹, Wouter Voorspoels¹, &
Pieter Moors¹

¹ University of Leuven

Author Note

The manuscript was written in R (R Core Team, 2016) using the packages `papaja` (Aust & Barth, 2017) and `rmarkdown` (Allaire et al., 2016). A full list of the packages with version information is provided as Supplemental Material. The experiment, including the analysis code, was pre-registered on Open Science Framework (<https://osf.io/3f7ta>). On the project's page (<https://osf.io/kqwcg>), one can find the `.Rmd` file, the material, and the data.

Correspondence concerning this article should be addressed to Tom Heyman, Department of Experimental Psychology, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium. E-mail: tom.heyman@kuleuven.be

Abstract

People have been shown to link particular sounds with particular shapes. For instance, the round-sounding non-word *bouba* tends to be associated with curved shapes, whereas the sharp-sounding non-word *kiki* is deemed to be related to angular shapes. This tendency of people to associate sounds and shapes has been observed across different languages. In the present study, we re-examined the claim of Hung, Styles, and Hsieh (2017) that such sound-shape mappings can occur before becoming aware of the visual stimuli. More precisely, we replicated their first experiment in which congruent and incongruent stimuli (e.g., *bouba* presented in a round or an angular shape, respectively) were rendered invisible through continuous flash suppression. The results showed that congruent combinations, on average, broke suppression faster than incongruent stimuli, thus providing converging evidence for Hung and colleagues' assertions. Collectively, these findings now provide a solid basis from which to explore the boundary conditions of the effect.

Keywords: sound-shape mapping; bouba-kiki effect; automaticity; replication; continuous flash suppression

Word count: 1582

Sound-symbolism effects in the absence of awareness. A replication study.

Sound-symbolism, defined as the non-arbitrary link between the sound and meaning of a word, has been the topic of much research in cognitive and developmental (neuro)psychology (see Lockwood & Dingemanse, 2015, for a review). A broad array of sound-meaning links has been established empirically, ranging from obvious and direct onomatopoeia like *hiss* or *splash*, over particular sounds marking grammatical categories, to individual letters being associated with very specific meanings (e.g., *i* tends to be linked with smaller objects as compared to the large-sounding vowel *a*; Sapir, 1929). Particularly well-studied in sound symbolism is the *maluma-takete* or *bouba-kiki* effect, which entails that round-sounding non-words (e.g., *maluma* or *bouba*) tend to be associated with round shapes, whereas sharp-sounding non-words (e.g., *takete* or *kiki*) are linked with angular shapes (Köhler, 1947; Ramachandran & Hubbard, 2001). This effect has been shown to emerge very early in children’s development (Ozturk, Krehm, & Vouloumanos, 2013), and replicates consistently across different languages and cultures (Bremner et al., 2013).

Recently, the fundamental nature of the *bouba-kiki* effect has been further explored by Hung, Styles, and Hsieh (2017), who claimed that sound-shape mappings affect cognitive processing even before people become aware of the stimuli. Relying on a technique called continuous flash suppression (henceforth CFS; Tsuchiya & Koch, 2005), Hung and colleagues found that congruent combinations of shapes and words (i.e., *kiki* presented in an angular shape or *bubu* presented in a round shape) gained preferential access to awareness over incongruent combinations (i.e., *kiki* in a round shape or *bubu* in an angular shape). In the CFS paradigm, a critical stimulus presented to one eye is masked by a dynamic noise pattern presented to the other eye. Though initially rendered invisible, the critical stimulus will eventually break through the mask. Hung et al. observed that congruent sound-shape combinations, such as *kiki* in an angular shape, were detected faster on average than incongruent combinations.

In light of the debate in the literature on what kind of information can be extracted from words and non-words under CFS, Hung et al.'s (2017) finding is intriguing. For instance, Yang and Yeh (2011) found that neutral words broke suppression faster than negative words, but Sklar et al. (2012) reported the opposite for two-word combinations (e.g., *black eye* versus *sand box*). Furthermore, using the same breaking CFS paradigm, Heyman and Moors (2014) were unable to obtain standard word recognition phenomena like the effect of word frequency and lexicality. A recent paper by Rabagliati, Robertson, and Carmel (2018) featuring 10 studies found no evidence that the meaning of a phrase or a word could be processed under CFS. Taken together, there is reason to believe that CFS fundamentally disrupts the processing of word-like stimuli, yet the effect reported by Hung et al. (2017) arguably depends on word recognition and interpretation processes to occur under CFS. Indeed, Hung and colleagues proposed that phonological information gets extracted from (non-)words under CFS to the extent that the sound-shape mapping causes congruent stimuli to break suppression faster.

To generalize their findings, Hung and colleagues (2017) also conducted one experiment that did not involve CFS. Instead, the non-words *bubu* or *kiki* were auditorily presented, and the threshold at which briefly presented masked shapes became visible, was measured. The visibility threshold turned out to be significantly lower when a congruent shape was presented 150 ms after onset of the auditory stimulus. The particular methodological characteristics of this experiment, however, do not allow for unequivocal conclusions regarding the unconscious nature of sound-symbolism processing. First, incongruent and congruent trials were presented in a blocked fashion, opening the door for strategy and practice effects, though participants reportedly had no conscious expectations about the sound-shape mappings. Second, the unusually short SOA (compared with Kovic, Plunkett, & Westermann, 2010, for instance), implies that the shape appeared *before* the utterance of the auditory stimulus was finished, thus raising doubts about the temporal dynamics of the effect.

As such, Hung and colleagues' (2017) general claim that sound-shape mapping occurs

prior to people becoming aware of the visual stimuli, rests on their two CFS studies. The first one involved the basic *bouba/bubu-kiki* effect, whereas the second experiment used an unfamiliar script to rule out that visual similarity between words and shapes was responsible for the effect. As Experiment 2 was an elaboration of, and crucially hinged on the effect observed in Experiment 1, we focus on the latter.

Obtaining evidence for the effect in a large, independent replication study would strongly support the idea that sound-shape mapping does not require awareness. It would also suggest that suppressed word(-like) stimuli do get processed to a higher level than previously assumed (Heyman & Moors, 2014), though some researchers have posited that even semantic information can get extracted from single words under CFS (e.g., Yang & Yeh, 2011, but see Rabagliati et al., 2018, for evidence to the contrary). Failing to find a sound-symbolism congruency effect, on the other hand, would further support the idea that CFS precludes the processing of word(-like) stimuli. Moreover, the combination of a (potential) null effect with the methodological issues regarding Hung et al.’s non-CFS experiment, would undermine the empirical basis of their claims. Indeed, assertions about the unconscious nature of sound-symbolism effects would become unfounded at this point. Given these theoretical implications and the relatively high prevalence of non-replications in the CFS literature (Moors, Hesselmann, Wagemans, & van Ee, 2017), we consider it imperative to conduct an independent replication of Hung et al.’s study.

Method

For the main experiment, we mimicked Hung and colleagues’ (2017) Experiment 1 as closely as possible, and any deviations are explicitly mentioned. Similar to Hung et al., we followed the usual CFS protocol: A dynamic noise pattern was presented to the participants’ dominant eye, while a target stimulus was presented to the other eye, either to the left or the right of a fixation cross. The target stimulus was initially invisible, but at some point broke

suppression. Participants were asked to press a key as soon as any part of the target stimulus became visible, and to report its location (i.e., left or right of fixation).

To validate our experimental settings, we conducted a control experiment, which was not part of Hung et al.'s (2017) protocol. It immediately followed the main experiment and was completely identical, except that the target stimuli were upright or inverted faces. It is well-established in the CFS literature that upright faces break suppression faster than inverted faces (e.g., Jiang, Costello, & He, 2007; Stein, Sterzer, & Peelen, 2012), and replicating this effect would thus validate the presently employed paradigm. While the primary focus of our replication naturally lies on the main experiment, replicating the face-inversion effect would strengthen the conclusions regarding the nature of the *bouba-kiki* effect. Note, though, that the sample size calculation, and participant exclusion criteria described below, were entirely determined by, and geared towards, (the data from) the main experiment. More specifically, sample size was determined using a sequential testing procedure. Rather than fixing sample size a priori, data collection terminated depending on the (provisional) outcome of the study. For the sake of transparency, we describe the full decision tree associated with this procedure, rather than just the particular branch we ended up taking.

Participants

A total of 193 participants took part in the experiment (149 women, 44 men, mean age = 22) in return for eight euro. All participants declared to have normal or corrected-to-normal visual acuity, as well as typical color and face perception. In order to determine sample size, we performed a Bayes factor design analysis (Schönbrodt & Wagenmakers, 2018). More specifically, we opted for a sequential design with a maximal sample size, which entails that data were gathered until enough evidence had been accumulated, or the pre-defined maximum number of participants had been tested. That is,

we continuously calculated the Bayes factor (BF) comparing the fit between the observed data and the predictions of two models, one assuming that congruent stimuli on average break suppression faster than incongruent stimuli, the other positing that congruent stimuli on average break suppression equally fast as incongruent stimuli (see Data analysis for more details). Data collection terminated as soon as the BF reached a certain threshold or when sample size maxed out.

Such a sequential design requires one to specify the BF boundaries, the maximal sample size (N_{max}), the sample size at which the BF is computed for the first time (N_{min}), and the number of participants that will be added each time until N_{max} or one of the BF boundaries is reached ($N_{increment}$). For the BF boundaries, we choose 6 and $\frac{1}{6}$, values also used in Schönbrodt and Wagenmakers (2018). $N_{increment}$ was set to 10 for practical purposes (i.e., scheduling participants), whereas N_{max} and N_{min} were determined through simulations with the BFDA package (Schönbrodt, 2018).

In a first step, we obtained a conservative estimate of the standardized effect size Cohen’s d by conducting a kind of meta-analysis on the outcome of Hung et al.’s (2017) Experiment 1 and 2 (i.e., the ones involving CFS). Using the `meta.ttestBF` function of the BayesFactor package (Morey & Rouder, 2015), we performed a one-sided Bayesian meta-analytic t -test with the function’s default priors. Through sampling 10,000 times from the posterior distribution of the model assuming an effect in the expected direction (i.e., congruent stimuli breaking suppression faster than incongruent stimuli), we obtained the 95% highest posterior density interval for the standardized effect size. We then used the lower bound, 0.24, as a conservative estimate of effect size to derive N_{max} and N_{min} .

Initially, N_{max} was set to an arbitrarily high value, 1,000, and N_{min} to 40, as in the example of Schönbrodt and Wagenmakers (2018), though they compared two group means from independent samples with N_{min} per group being 20. Next, we simulated 10,000 hypothetical datasets assuming sequential testing with the parameters defined above under two scenarios: one with an assumed standardized effect size of 0.24, and the other supposing

that there is no effect (i.e., the null hypothesis). We subsequently estimated the required maximal sample size such that one of the Bayes factor boundaries would be reached in 80% of the cases, and this for both scenarios. N_{max} was tentatively set to the maximum of both these values. Finally, we examined the rate of false negative and false positive evidence under this combination of N_{min} and N_{max} . If both rates were equal to or smaller than .05, we would settle on those values for N_{min} and N_{max} . If not, N_{min} would be increased by 10, which would then be used to simulate new hypothetical datasets as described above, and this until the false evidence criterion was met (i.e., both rates $\leq .05$). Ultimately, the entire procedure resulted in an N_{min} of 90 and an N_{max} of 180¹.

Thus far, we assumed all participants would perform the task adequately. That is, participants had to indicate the location of the target stimulus, either to the left or to the right of a central fixation cross. Following Hung et al. (2017), we removed and replaced participants whose accuracy on this task was 3 *SDs* or more below the group mean. Furthermore, participants who did not finish the entire experiment were removed and replaced. Finally, during debriefing, participants' knowledge of the *bouba-kiki* effect was assessed. Anyone familiar with the phenomenon was excluded and replaced as well. Thus, prior to the analyses, we made sure that all participants met these criteria. If not, additional participants were recruited until the pre-determined sample size was reached (i.e., 90 participants or 100, 110, ..., 180 participants in case $\frac{1}{6} < BF < 6$).

¹Note that we used Bayes factor boundaries of 6 and $\frac{1}{6}$ rather than Jeffreys' substantial evidence criteria of 3 and $\frac{1}{3}$ (Jeffreys, 1961), in order to limit the rate of false positive and false negative evidence. Furthermore, Bayes factors express how to change one's prior beliefs in light of the data. Should we obtain a Bayes factor of approximately $\frac{1}{3}$, and stop collecting data, one might be largely agnostic as to whether unconscious sound-shape mappings occur, given that one's prior odds were probably somewhat greater than 1 based on Hung and colleagues' (2017) findings. Therefore, we opted for the stricter boundaries of 6 and $\frac{1}{6}$, as in Schönbrodt and Wagenmakers (2018).

Material

There were four target stimuli, created by placing a non-word (i.e., either *bubu* or *kiki*) inside an angular or rounded shape, presented against a black background (see Figure 1). The non-words were displayed in Arial typeface with a maximum height of 0.40° , and a total width of 0.53° , for *kiki*, or 0.80° , for *bubu*. The shapes, including the background, subtended 1.8° by 1.8° . On each trial, one of the four target stimuli was presented to the non-dominant eye, either to the left or the right of a red, central fixation cross, also in Arial typeface with a maximum height of 0.36° . At the same time, a dynamic noise pattern consisting of Mondrian stimuli (i.e., a number of blue, red, and yellow-colored rectangles) flashing at 10 Hz, was projected to the dominant eye such that both the left and the right side of the fixation point was covered (see Figure 2). These dynamic Mondrian stimuli comprised 500 rectangles, each of a randomly picked color (blue, red, or yellow), and a random height and width between 0.06° and 0.6° . The rectangles were displayed in an aperture of 1.8° by 1.8° , thus matching the dimensions of the critical stimuli.

Stimuli were shown on two 19.8-in. Sony Trinitron GDM F500-R (each 2,048 x 1,536 pixels at 60 Hz) monitors driven by a DELL Precision T3400 computer with an Intel Core Quad CPU Q9300 2.5 GHz processor running on Windows XP. Binocular presentation was achieved by a custom made stereo set-up. Two CRT monitors, standing opposite to each other with a distance of 220 cm, projected to the left and right eye respectively via two mirrors placed at a distance of 110 cm from the screen. In order to stabilize fixation, a head- and chin rest was placed 15 cm from the mirrors. Thus, the effective viewing distance was 125 cm. Stimulus presentation, timing, and response recording were controlled with custom software programmed in Python using the PsychoPy library (Peirce, 2007, 2009).

In order to achieve proper binocular fusion, we used a random checkerboard pattern (28 by 28 squares) as the background throughout the experiment. Each square of the checkerboard subtended 0.2° by 0.2° . Two black frames (4° by 4°), superimposed on the

Figure 1. Example of the critical stimuli. Congruent and upright stimuli are shown on the left-hand side, incongruent and inverted stimuli are on the right.

checkerboard, served as the immediate background for the target and Mondrian stimuli, leaving the outer four rows of the checkerboard visible. Note that Hung et al. (2017) presented stimuli against a black background with a narrow, white frame to ensure stable fusion. Finally, for the control experiment, two neutral faces (one male and one female) were created using the FaceGen software (Singular Inversions Inc., 2017). The faces, shown in Figure 1, also subtended 1.8° by 1.8° (including the black background frame). All stimuli were presented 1.075° to the left or right of the fixation cross, counting from the center of the stimuli to the center of the fixation cross.

Testing took place in a sound-attenuated, dimly lit room. Based on a number of pilot runs, the entire study lasted approximately one hour. The Social and Societal Ethics Committee of the University of Leuven reviewed and approved this study (approval number G-2017 10 967).

After signing an informed consent form and providing demographic information (age and gender), participants' eye dominance was assessed using the Dolman method. That is,

participants were asked to look at an object through a card with a small aperture, and to then gradually bring the card to the head, thus revealing the dominant eye. When eye dominance was determined, participants proceeded to the main experiment. On each trial, they saw a red fixation cross for a variable time between 0 and 1 second, with the exact duration being randomly drawn from a uniform distribution. Next, the target stimulus was presented to the non-dominant eye while the dynamic mask was presented to the dominant eye. The contrast of the target stimulus increased linearly during a 10-second interval from 0% to 75%, whereas the Mondrian stimuli were always displayed at full contrast.

Participants were asked to press the “5” button of a numeric keypad when any part of the target stimulus became visible. After detection, the target disappeared, but the Mondrian stimuli were still shown for another 500 ms to prevent, or at least reduce, the occurrence of afterimages. Participants were subsequently prompted to indicate whether the target stimulus had appeared to the left or the right of fixation by pressing, respectively, the “4” or “6” button of the numeric keypad. If participants did not respond within 10s after stimulus onset, the target stimulus, but not the dynamic Mondrians, remained on the screen for another 500 ms. Participants were still required to perform the location judgment task, which should become fairly easy given that the target was no longer masked. There was no time limit on the location task, and the next trial started 500 ms after participants had responded. During the intertrial interval, participants saw a black screen without a fixation cross. This is different from Hung and colleagues’ procedure, in which participants had to press a key to start the next trial.

The main experiment comprised 160 trials, split up in four blocks of 40 trials. Between blocks, participants could take a self-paced break. Each of the four target stimuli (i.e., *bubu*-round shape, *kiki*-round shape, *bubu*-angular shape, and *kiki*-angular shape) was presented 40 times in total, either to the left of fixation, on 20 trials, or to the right of fixation, on the other 20 trials. Trial order was randomized within every block for each participant separately. Before the main experiment, participants completed 16 practice trials,

two of each non-word x shape x location combination. After finishing the main experiment, participants performed a control task, which was completely analogous in terms of its design, procedure, and number of trials, except that target stimuli were upright or inverted, male or female faces.

During debriefing, participants' knowledge of the *bouba-kiki* effect was probed (see the Supplemental Material for the complete questionnaire). Response sheets were independently coded, by the second author (ASM) and another researcher with relevant expertise but no ties to the project, prior to any other form of data inspection. Participants whose answers suggested that they had knowledge of the effect or the research hypothesis, were excluded from subsequent analyses.

Data analysis

In this section, we describe the exclusion criteria, the sequential testing procedure, and the statistical analyses. We start by describing the main experiment's protocol, and then turn to the control experiment.

First, participants whose accuracy on the location judgment task was 3 *SDs* or more below the group mean were removed and replaced before conducting the crucial analyses². Following Hung et al. (2017), accuracy was calculated based on all location judgments, also those given to targets that did not break suppression within 10s. Because these trials do not provide a reliable estimate of suppression time, they were excluded from all further analyses. The same holds for trials on which the location judgment was incorrect, and trials with response times below 0.5s or above 9.5s³. Next, the average suppression time was calculated

²In the pre-registered analysis code, we recalculated the performance cutoff after removing the outlying participants. This resulted in a cycle of removing participants, shifting the criterion, removing additional participants, shifting the criterion again, . . . As a result, a lot of participants would get thrown out, which, in hindsight, would not be sensible. Therefore, we decided to calculate a fixed cutoff based on the data of the initial 90 participants who finished the experiment and indicated not having any prior knowledge.

³These criteria are admittedly somewhat arbitrary. Furthermore, no such response time trimming procedure

for each condition (i.e., congruent and incongruent) x participant combination. As in Hung et al., suppression times more than 3 *SDs* above or below the corresponding by-participant conditional mean were also removed from the analyses.

Upon securing 90 participants who met the inclusion criteria and after completing the data-cleaning procedure, the average suppression time was calculated for each condition x participant combination. Subsequently, a difference score was obtained for each participant by subtracting the average suppression time in the congruent condition from the average suppression time in the incongruent condition. Notice that this was the other way around in the pre-registered protocol (i.e., congruent - incongruent), but we switched the variables in order to facilitate the interpretability and comparability of the effect sizes⁴.

The resulting difference scores were then subjected to a Bayesian one-sample *t*-test, assuming the Jeffreys-Zellner-Siow (JZS) default prior (see Rouder, Speckman, Sun, Morey, & Iverson, 2009) with $M_0 : \mu_{incongruent} - \mu_{congruent} = 0$ and $M_1 : \mu_{incongruent} - \mu_{congruent} > 0$. For M_1 , we assumed a positive half-Cauchy prior with a default scale parameter $r = \sqrt{2}/2$ on the standardized effect size. This analysis, carried out with the BayesFactor package (Morey & Rouder, 2015), yielded a Bayes factor BF_{10} that addressed the question of whether the congruency effect was present or absent in the replication study. It also served as our sequential testing criterion. That is, after the initial batch of 90 participants, we checked whether $\frac{1}{6} < BF_{10} < 6$. If that was the case, 10 participants who met the inclusion criteria would be added and BF_{10} would be recalculated taking the data from all participants into account. This procedure would repeat itself until BF_{10} reaches/crosses either threshold or until 180 participants meeting the inclusion criterion would have taken part in the

was mentioned in Hung et al. (2017). However, personal communication (S. Hung, October 9, 2017) revealed that this was the procedure actually employed by Hung et al. in their analysis of Experiment 1. Therefore, we decided to use the same approach in the present replication study.

⁴From classical psychometrics, we know that the reliability of difference scores can be quite low. However, the current study does not seek to predict individual observations, but instead focuses on the aggregate level, so this is not a concern (e.g., Thomas & Zumbo, 2012).

experiment. To be clear, no additional data would be gathered if BF_{10} was already greater than 6 or smaller than $\frac{1}{6}$ for the initial 90 participants.

Once the final sample size was established, we computed the conditional mean and standard deviation from the participant-specific average suppression times. Besides the JZS Bayes factor test, we also performed an analogous, frequentist one-sample directional t -test, and a replication Bayes factor test (Verhagen & Wagenmakers, 2014). The latter yielded a Bayes factor, BF_{r0} , that indicated whether the congruency effect was either similar to what Hung et al. (2017) found, or absent. To derive BF_{r0} , we used the ReplicationBF package (Harms, 2017) with the reported t -value and sample size from Hung et al.’s Experiment 1 as input (respectively, 2.12 and 20). Finally, we also computed the conditional mean and standard deviation from the participant-specific average suppression times of those participants who reported having knowledge about the *bouba-kiki* effect or the research hypothesis.

The analysis of the control experiment was completely analogous, with the exact same participants contributing data to both experiments. This means that there were no additional participant exclusion criteria and that no sequential testing procedure was employed for the control experiment. In other words, the sample size entirely depended on the outcome of the JZS Bayes factor test for the main experiment. The statistical tests themselves were equivalent across both experiments, only the condition changed: congruent was replaced by upright and incongruent by inverted. The sole difference was that we used the t -value and sample size reported by Jiang et al. (2007) to determine BF_{r0} (respectively, 5.53 and 10).

Results

The results are summarized in Figure 3. Congruent stimuli broke suppression faster than incongruent stimuli ($M_{congruent} = 3.56$ with $SD_{congruent} = 1.00$, versus $M_{incongruent} =$

Figure 3. Summary of the main experiment’s results. Every dot represents a participant. Those above the diagonal identity line (colored light blue) show a congruency effect in the expected direction based on Hung et al.’s (2017) findings.

3.61 with $SD_{incongruent} = 0.98$). The JZS Bayes factor test on the difference scores obtained by subtracting the by-participant conditional means (i.e., $M_{i,incongruent} - M_{i,congruent}$, with i being the participant identifier) yielded a BF_{10} of 6.38. An analogous, frequentist

one-sample directional t -test gave the following result: $t(178) = 2.75$, $p = .003$, 95% CI $[0.02, \infty]$, Cohen's $d = 0.21$. Finally, the Bayesian replication test resulted in a BF_{r0} of 6.79. As for the participants who reported having knowledge about the *bouba-kiki* effect or the research hypothesis ($N = 10$), average suppression times in the congruent and incongruent condition were, respectively, 3.80 ($SD = 1.42$) and 3.75 ($SD = 1.18$).

Notice that the sequential Bayes factor procedure did not terminate prior to reaching the pre-determined maximal sample size (i.e., 180)⁵, though the Bayes factor ultimately did cross the evidential threshold of 6. In addition, the replication Bayes Factor provided evidence that the congruency effect was similar to what Hung et al. (2017) found, rather than being absent. However, the observed effect size did fall outside of the meta-analytic 95% credibility interval based on Hung et al.'s Experiment 1 and 2, which had a lower bound of 0.24. As an aside, in a correction of their paper, Hung et al. changed the measure of effect size:

$$\frac{M_{incongruent} - M_{congruent}}{\frac{SD_{incongruent} + SD_{congruent}}{2}} \text{ instead of } \frac{t}{\sqrt{N}}$$

The pre-registered analysis code used the latter, but for ease of comparison we also calculated the former, non-registered metric. Overall, this yielded considerably smaller effect sizes, and, most importantly, the effect in the current experiment turned out to be about half the size of that observed by Hung and colleagues: 0.05 and 0.10, respectively.

Turning to the control experiment (see figure included as Supplemental Online Material), we found that upright faces broke suppression faster than inverted faces ($M_{upright} = 4.54$ with $SD_{upright} = 1.19$, versus $M_{inverted} = 4.62$ with $SD_{inverted} = 1.22$). The JZS Bayes factor test on the difference scores obtained by subtracting the by-participant conditional means (i.e., $M_{i,inverted} - M_{i,upright}$) yielded a BF_{10} of 52.15. A directional one-sample t -test on the same data gave the following result: $t(179) = 3.47$, $p < .001$, 95% CI $[0.04, \infty]$, Cohen's $d = 0.26$. Finally, the Bayesian replication test resulted in a BF_{r0} of

⁵One participant, who met all pre-registered inclusion criteria, never broke suppression in this experiment. Hence, the effective sample size was 179.

0.43, indicating that the data were neither consistent with the original effect, as observed by Jiang et al. (2007), nor with a null effect. In light of the large BF_{10} , one could therefore conclude that there is compelling evidence for a face inversion effect, but of a more modest magnitude compared with earlier studies (see also Stein et al., 2012).

Discussion

The current study showed that congruent sound-shape stimuli, on average, broke suppression faster than incongruent combinations. Even though the observed effect size was appreciably smaller than in Hung et al. (2017), the face inversion effect also shrunk relative to earlier findings. Thus, it is possible that the present study underestimates the true size of both phenomena. However, many replication attempts have yielded smaller effect sizes, presumably because biased reporting inflated the estimates in the original studies (Open Science Collaboration, 2015). So, an alternative explanation is that previous studies *overestimated* the magnitude of both the congruency and the inversion effect.

Taken together, we obtained converging evidence for the notion that congruent sound-shape stimuli gain preferential access to awareness. We believe this provides a solid basis to now explore the boundary conditions of the effect, and the underlying processes. For example, it is possible that the similarity between the shapes and the visual properties of the letters is (partly) responsible for the congruency effect (i.e., *k* looks more angular than *b*). Even though Hung et al.’s follow-up experiment with newly learned stimuli in an unknown script seemingly debunked this explanation, one could also envision an experiment in which the non-words are presented upside down. Stimulus inversion hinders letter identification, hence the corresponding sounds will not be able to elicit any shape associations. Consequently, the congruency effect should disappear, unless visual similarity between the shapes and non-words is the driving factor.

Another aspect worthy of further examination is the role of top-down attention

(Lupyan & Ward, 2013). Indeed, Hung et al. (2017), as well as the present study, relied exclusively on the typical *bouba-kiki* example. However, due to repeated stimulus presentation, participants might develop certain expectations, which could influence the congruency effect. Systematically varying the stimuli would address this concern, thereby allowing us to establish whether unconscious sound-shape mappings are indeed an ubiquitous phenomenon.

Author contributions

TH, PM, and HV contributed to the study design. TH, PM, and WV developed the data analytical approach. ASM gathered the data. TH drafted the manuscript, and all authors provided critical revisions and approved the final version for submission.

Acknowledgments

TH and PM are postdoctoral fellows of the Research Foundation-Flanders (FWO-Vlaanderen). WV is a postdoctoral fellow at the University of Leuven. HV and ASM are graduate students. We would like to thank Marijke Casteleyn for assisting in the data collection, and Anne White for coding the response sheets.

References

- Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., . . . Hyndman, R. (2016). *rmarkdown: Dynamic Documents for R*. Retrieved from <https://CRAN.R-project.org/package=rmarkdown>
- Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Bremner, A. J., Caparos, S., Davidoff, J., Fockert, J. de, Linnell, K. J., & Spence, C. (2013). “Bouba” and “Kiki” in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, 126(2), 165–172. doi:10.1016/j.cognition.2012.09.007
- Harms, C. (2017). *ReplicationBF: Calculating replication Bayes factors for different scenarios*. Retrieved from <https://github.com/neurotroph/ReplicationBF>
- 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), e104719. doi:10.1371/journal.pone.0104719
- Hung, S.-M., Styles, S. J., & Hsieh, P.-J. (2017). Can a word sound like a shape before you have seen it? Sound-shape mapping prior to conscious awareness. *Psychological Science*, 28(3), 263–275. doi:10.1177/0956797616677313
- Jeffreys, H. (1961). *Theory of probability*. Oxford, UK: Oxford University Press.
- 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(4), 349–355. doi:10.1111/j.1467-9280.2007.01902.x
- Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, 114(1), 19–28. doi:10.1016/j.cognition.2009.08.016
- Köhler, W. (1947). *Gestalt psychology* (2nd ed.). New York, NY: Liveright.
- Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, 6. doi:10.3389/fpsyg.2015.01246
- Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, 110(35), 14196–14201. doi:10.1073/pnas.1303312110
- Moors, P., Hesselmann, G., Wagemans, J., & van Ee, R. (2017). Continuous flash suppression: Stimulus fractionation rather than integration. *Trends in Cognitive Sciences*. doi:10.1016/j.tics.2017.06.005
- Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs*. Retrieved from <https://CRAN.R-project.org/package=BayesFactor>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. doi:10.1126/science.aac4716
- Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, 114(2), 173–186. doi:10.1016/j.jecp.2012.05.004
- Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of*

- Neuroscience Methods*, 162(1), 8–13. doi:10.1016/j.jneumeth.2006.11.017
- Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2, 10. doi:10.3389/neuro.11.010.2008
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rabagliati, H., Robertson, A., & Carmel, D. (2018). The importance of awareness for understanding language. *Journal of Experimental Psychology: General*, 147(2), 190–208. doi:10.1037/xge0000348
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia - a window into perception, thought and language. *Journal of Consciousness Studies*, 8(12), 3–34.
- 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(2), 225–237. doi:10.3758/pbr.16.2.225
- Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, 12(3), 225–239. doi:10.1037/h0070931
- Schönbrodt, F. D. (2018). *BFDA: An R package for Bayes factor design analysis*. Retrieved from <https://github.com/nicebread/BFDA>
- Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128–142. doi:10.3758/s13423-017-1230-y
- Singular Inversions Inc. (2017). *FaceGen*. Toronto, ON., Canada. Retrieved from

<https://www.facegen.com>

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*, *109*(48), 19614–19619. doi:10.1073/pnas.1211645109

Stein, T., Sterzer, P., & Peelen, M. V. (2012). Privileged detection of conspecifics: Evidence from inversion effects during continuous flash suppression. *Cognition*, *125*(1), 64–79. doi:10.1016/j.cognition.2012.06.005

Thomas, D. R., & Zumbo, B. D. (2012). Difference scores from the point of view of reliability and repeated-measures anova: In defense of difference scores for data analysis. *Educational and Psychological Measurement*, *72*(1), 37–43. doi:10.1177/0013164411409929

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

Verhagen, J., & Wagenmakers, E.-J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General*, *143*(4), 1457–1475. doi:10.1037/a0036731

Yang, Y.-H., & Yeh, S.-L. (2011). Accessing the meaning of invisible words. *Consciousness and Cognition*, *20*(2), 223–233. doi:10.1016/j.concog.2010.07.005