Unconscious Cross-Modal Priming of Auditory Sound Localization by Visual Words

Ulrich Ansorge
University of Vienna

Shah Khalid
University of Vienna and University of Osnabrück

Bernhard Laback
Austrian Academy of Sciences, Vienna, Austria

Little is known about the cross-modal integration of unconscious and conscious information. In the current study, we therefore tested whether the spatial meaning of an unconscious visual word, such as up, influences the perceived location of a subsequently presented auditory target. Although cross-modal integration of unconscious information is generally rare, unconscious meaning stemming from only 1 particular modality could, in principle, be available for other modalities. Also, on the basis of known influences and dependencies of meaning on sensory information processing, such an unconscious meaning-based effect could impact sensory processing in a different modality. In 3 experiments, this prediction was confirmed. We found that an unconscious spatial word, such as up, facilitated position discrimination of a spatially congruent sound (here, a sound from above) as compared to a spatially incongruent sound (here, from below). This was found even though participants did not recognize the meaning of the primes. The results show that unconscious processing extends to semantic–sensory connections between different modalities.

Keywords: cross-modal priming, multisensory priming, unconscious processing, visual auditory integration, auditory localization

Human information processing draws on both conscious and unconscious processing modes, but how these two modes interact is not entirely understood (Dehaene & Changeux, 2011). One way of investigating this question is by studying the capabilities and limits of unconscious processing (Ansorge, Kunde, & Kiefer, 2014; Atkinson, Thomas, & Cleeremans, 2000; Dijksterhuis & Aarts, 2010). In perception, unconscious processing can be studied with masked priming. In vision, masking denotes a method of reducing prime visibility by the presentation of a masking stimulus (Breitmeyer & Ogmen, 2006). In masked priming, an unconscious prime is presented prior to a task-relevant target. Participants respond to the consciously perceived target, and the unconscious (masked) prime influences the responses (Dehaene et al., 1998). The prime’s imperceptibility, together with the priming effect, offers an excellent tool for studying unconscious processing.

Past studies demonstrated unconscious priming with primes and targets in the same modality—for example, visual primes and targets (Dehaene & Changeux, 2011)—but little is known about unconscious priming across modalities (Lamy, Mudrik, & Deouell, 2008). Cross-modal priming research is mostly restricted to conscious priming (cf. Baars, 1988; Dehaene & Naccache, 2001). However, unconscious word and number priming are exceptions. A handful of studies have shown that unconscious visual words (or numbers) prime auditory target words (or numbers; Faivre, Mudrik, Schwartz, & Koch, 2014; Grainger, Diependaele, Spinelli, Ferrand, & Farioli, 2003; Kiyonaga, Grainger, Midgley, & Holcomb, 2007; Kouider & Dehaene, 2009; Kouider & Dupoux, 2001; Nakamura et al., 2006).

Yet, unconscious cross-modal priming should not be restricted to interactions between visual and auditory words or numbers because semantic memory also connects (word) meaning with sensory information from different modalities (Barsalou, 1999; Chen & Spence, 2011; Glaser & Glaser, 1989). Take the example of perceived space (Regier & Carlson, 2001). All perceived stimuli are located in space and, accordingly, sensory representations include a spatial index specifying stimulus location. Yet, to use such a sensory spatial index for various purposes such as actions, attention shifts, or judgments, a semantic (or conceptual) spatial reference frame has to be imposed onto the spatial index (Logan, 1995). For instance, if you locate two objects in the sky—a plane and a missile approaching the plane from below—both these objects are above you in a conceptual spatial frame with an origin that is centered on you. However, the missile would be below the plane, so to direct your attention from the plane to the missile, you
would have to shift the origin of the conceptual spatial frame from yourself as a center onto the plane as its center. Obviously, the usage of sensory spatial indices for such purposes as locating a stimulus relative to you or to another object requires the projection of a particular conceptual spatial frame onto the object in question so as to specify, for example, the deictic relation between two objects (e.g., as above, below, right, or left of one another). Because of such close connections between the sensory and the semantic domain (Glaser & Glaser, 1989; Logan, 1995) and because unimodal unconscious semantic priming by visual words has been demonstrated (Kiefer, 2002; McCauley, Parmelee, Sperber, & Carr, 1980), unconscious cross-modal priming between a prime word’s meaning in one modality and the perception of sensory features of a nonword target in a different modality should be possible once there is a semantic connection between a word’s meaning and a sensory feature. Here, we used the spatial discrimination of auditory target locations to study this hypothesis.

**Experiment 1**

In Experiment 1, we studied unconscious cross-modal priming between seen words and heard locations. In the first stage of the experiment, participants localized an auditory stimulus as coming from above or below. We used a taxing localization task in which participants had to discriminate between relatively small elevation differences based on participant-specific head-related transfer functions (HRTFs). Prior to this auditory target, participants saw a conscious or unconscious visual word. This prime had a spatial meaning that was equally likely to be congruent or incongruent to the target’s position. For example, the word *up* was congruent if the following target was a sound from above, but it was incongruent if the target was from below. Participants were informed that primes were not informative to the sound localization task. However, word primes were presented in the participants’ line of sight. This could be sufficient for a semantic priming effect of unconscious task-irrelevant word primes (Naccache & Dehaene, 2001). We used a relatively difficult auditory task so that performance was not perfect. In this way, we hoped to provide sufficient room for a subliminal priming effect.

In the second stage of the experiment, we evaluated the visibility of the primes. Here, the task of the participants was to judge the prime–target relation as congruent versus incongruent. As compared to a pure prime discrimination, this task has the advantage of asking for the participants’ awareness of exactly that characteristic of the stimuli (i.e., the congruence vs. incongruence between prime and target) that was expected to lead to a behavioral priming effect (cf. Desender, Van Opstal, & Van den Bussche, 2014). The task also has a disadvantage. It is more demanding than a simple prime discrimination and thus may not be as sensitive as other conceivable visibility tests. Therefore, across experiments, the visibility test varied, and simpler tasks were used in Experiments 2 and 3.

**Method**

**Participants.** Twenty-two participants were tested. Here and in the following experiments, participants had normal or corrected-to-normal vision and were paid €7/hr. Also, informed consent was obtained from all participants, and the participants were treated in accordance with American Psychological Association standards and the rules of the Declaration of Helsinki. One participant had to be excluded because of too high a number of correct prime–target congruence judgments in masked conditions. He had a d’ of 1.1 in the masked condition that exceeded the group mean (d’ = 0.2) by more than 2 standard deviations (SD = 0.3). Six additional participants had to be excluded because of chance performance (around 50%) in the sound localization task, indicating that they were unable to localize the targets. The remaining 15 participants (12 female, three male, M<sub>age</sub> = 28.2 years, age range: 24–42 years) were analyzed.

**Apparatus, stimuli, and procedure.** Visual stimuli were presented on a 17-in. (43-cm) color flat screen display with a refresh rate of 59.1 Hz via an NVIDIA GeForce GT 220 (with 1,024 MB) graphics adapter. The participants sat at a distance of 57 cm from the screen in a quiet, dimly lit room, with their head resting in a chin rest to ensure a constant viewing distance and a straight-ahead gaze direction. Target responses were registered through the keys C and M labeled as *left* and *right* of a standard keyboard. Participants operated the keys with their left and right index fingers. The targets were two spatial sound stimuli generated using a virtual acoustics technique. A 300-ms broadband white Gaussian noise token, temporally shaped with a Tukey window with a 10-ms rise–fall time, was filtered with participant-specific directional transfer functions (DTFs). DTFs were derived from HRTFs, which were measured in a prior session, using the experimental setup and postprocessing techniques described in Appendix A. Targets’ spatial positions were above (with an elevation angle of +30° in the median plane) and below (with an elevation angle of −30° in the median plane). Targets were presented using Sennheiser 380 Pro headphones (with <0.1% total harmonic distortion). Visual prime stimuli were 20 German words denoting directions or positions on the vertical axis. All primes were high-frequency words, with more than 60 instances in 1 million words. The 10 spatial up primes were the words *oben* (on top), *darüber* (above), *hinauf*, *aufwärts*, *empor* (upward), *hoch*, *gehoben*, *erhöht* (elevated), *aufsteigend*, and *steigend* (rising), with a mean word length of Ø = 6.6 letters (range of four to 11 letters). The 10 spatial down primes were the words *unten* (down), *darunter* (below), *hinab*, *abwärts*, *herab* (downward), *niedrig* (low), *gesenkt* (lowered), *abfallend*, *sinkend* (descending), and *tief* (deep), with a mean word length of Ø = 6.3 letters (range of four to nine letters). Up and down words thus had a comparable length and frequency, and they were easily and equally discriminable by their spatial category membership (ensured by empirical pretesting; see Ansorge & Böhner, 2014). (For further details, refer to Ansorge, Khalid, & König, 2013.)

Each prime word was shown equally often (16 times per each of two stages of the experiment). For the creation of the prime–target pairs, each of the equally often presented sound locations was randomly combined with each of the 20 spatial primes. The resulting prime–target pairs were equally likely to be congruent and incongruent.

_Figure 1_ shows the sequence of events in a trial. The fixation cross, prime, and mask were all presented at the screen center. They were black (<1 cd/m²) against a gray background (24 cd/m²). A trial started with the fixation cross presented for 750 ms. In masked trials, next a forward mask was presented. It consisted of 10 randomly drawn uppercase letters and was shown for 200 ms. Then, the prime word occurred for 34 ms (in the short stimulus onset asynchrony [SOA] condition) or for 68 ms (in the long SOA...
condition). Two different SOAs were used because, due to a lack of fitting prior studies, we were not sure how quickly the priming effect would develop and dissipate. The prime was always in lowercase letters. A backward mask was presented after the prime. The backward mask also consisted of 10 uppercase letters, which were drawn independently of the letters of the forward mask. The backward mask was shown for 34 ms (leading to an SOA of 68 ms in the short SOA condition) or for 68 ms (leading to an SOA of 136 ms in the long SOA condition). Finally, the target was presented. The intervals between successive events were always 0 ms. In unmasked trials, everything was the same, with the exception that both forward and backward masks were omitted and replaced by blank screens.

The experiment consisted of two blocked stages, beginning with a sound localization task followed by a prime–target congruence discrimination task. During the localization task, participants judged the target stimulus as coming from above or below. In psychophysical terms, this corresponds to a sound source position discrimination task. Half of the participants were instructed to press the right key for positions above and the left key for positions below. The other half of the participants got the reversed stimulus–response (S–R) mapping. After each incorrect response or if a response exceeded 1,250 ms, participants received feedback about their error or their too-slow responses. Feedback lasted for 750 ms. Thus, keeping high accuracy and a fast response was mildly rewarded (i.e., saved 750 ms per trial that would otherwise be used for the feedback).

The second stage was a prime visibility task. For half of the participants, this task required that the right key be pressed in trials in which the prime was congruent to the target and that the left key be pressed if the prime was incongruent to the target. The other half of the participants got the reverse S–R assignment. Before the prime visibility test, the levels of the variable prime–target congruence were carefully explained to the participants, with relevant examples in the instructions. In the prime visibility stage, the temporal structure of events was identical to the localization discrimination stage so that the processing requirements with respect to the targets were the same in the two tasks. As a prime visibility test, the task of discriminating congruent from incongruent trials requires processing of the prime and target and thus tests the critical dimension responsible for the priming effect in the localization task. No feedback was given on the correctness of responses.

Each stage consisted of 320 trials. In total, this involved 40 trials of each combination of the two levels of the variables of Prime Visibility (masked, unmasked) × 2 Prime–Target SOAs (short, 68 ms; long, 136 ms) × 2 Prime–Target Congruence Relations (congruent, incongruent). The conditions were presented in a completely randomized order. Prior to both stages, participants were carefully instructed about their upcoming task. Before the localization task, participants were also informed about the fact that the primes were uninformative about the targets, and before the prime visibility task, participants were informed about the fact that the specific target locations were uninformative about the congruence versus incongruence of the primes. Also, the participants practiced the task for a minimum of 32 trials and, if they wanted, they could extend the practice by another 20 trials. We used the same stimuli for the test trials as for data collection. In addition to the written instructions on the display, if necessary (i.e., if there were questions), before and during practice, the task was explained verbatim in more detail. Together with two short breaks at regular intervals within stages and one break between the stages, the experiment took about 1 hr.

Results

Sound localization. Out of all correct responses, 3.2% were excluded because their reaction times (RTs) deviated by more than 2 standard deviations from a respective condition’s and individual participant’s mean RT. See Appendix B and Figure 2 for mean correct RTs, error rates (ERs), and prime visibility indices. It can be observed that participants’ performance in the congruent conditions was better than in the incongruent conditions at both short and long SOAs and in the masked as well as unmasked conditions. This was also confirmed by formal analyses.

An initial analysis of variance (ANOVA) with the between-participants variable of S–R mapping did not show a significant main effect of mapping nor any significant interaction with this variable. Data were therefore collapsed across mappings for a repeated-measures ANOVA of the means of the correct RTs with the within-participants variables of congruence (congruent, incongruent), prime–target SOA (short, long), and prime visibility (masked, unmasked). Bonferroni adjustments for multiple comparisons and an alpha level of .05 were applied here and throughout the study.

A significant main effect of congruence, $F(1, 14) = 6.52, p < .02$, Cohen’s $f = 0.69$, was found, reflecting faster RTs in congruent (607 ms, $SD = 38$) than incongruent (619 ms, $SD = 37$) conditions. No other significant main effect or interaction was found, all $F$s < 1.0. An additional pair sampled $t$ test confirmed significantly faster responses in the masked congruent condition ($M = 600$ ms, $SD = 36$) than in the masked incongruent condition ($M = 616$ ms, $SD = 42$), $t(14) = 2.65, p < .02$. A similar $t$ test for the unmasked congruent condition ($M = 614$ ms, $SD = 43$) and
incongruent condition ($M = 623 \text{ ms}, SD = 41$) was not significant, $t(14) = 0.92, p = .38$.

A repeated-measures ANOVA of the ERs with the same variables showed a lower mean accuracy in incongruent (ER = 15.0%, $SD = 7.2$) than congruent (ER = 12.2%, $SD = 5.5$) conditions. This was reflected in an almost significant main effect of congruence, $F(1, 14) = 4.43, p = .054, \text{Cohen's } f = 0.56$. There was also a significant main effect of prime visibility, $F(1, 14) = 5.61, p < .03$, Cohen’s $f = 0.64$, with higher ERs for unmasked (15.3%, $SD = 7.0$) than masked primes (11.9%, $SD = 6.0$). No other significant main effect or interaction was found, all $F$s < 1.0.

**Prime visibility.** To test whether participants failed to see the masked primes but identified the unmasked primes, we computed $d'$, an index of stimulus visibility (Reingold & Merikle, 1988). Individual $d'$ was computed separately for masked and unmasked primes and for primes at shorter and longer SOAs. For calculation of $d'$, congruent trials counted as signals and incongruent trials counted as noise. Accordingly, correct (i.e., congruent) judgments in congruent trials figured as hits, and incorrect (i.e., incongruent) judgments in incongruent trials figured as false alarms (FAs).

The participants were not able to discriminate the masked primes’ prime–target pairs with better-than-chance accuracy. For the masked primes at the shorter SOA, $d'$ was 0.08 ($SD = 0.4$), $t(14) = 0.77, p = .30$, and at the longer SOA it was 0.16 ($SD = 0.31$), $t(14) = 1.97, p = .07$. The unmasked primes’ prime–target pairs were successfully discriminated with the shorter SOA, $d' = 1.31$ ($SD = 2.0$), $t(14) = 2.51, p < .02$, and the longer SOA, $d' = 1.33$ ($SD = 2.3$), $t(14) = 2.21, p < .04$.

**Congruence and prime visibility correlations.** We also tested whether individual mean RT congruence effects (incongruent RT − congruent RT) were correlated with individual prime visibility indices in the masked condition, as would be expected if prime visibility accounted for the RT congruence effect of the masked primes. We found no significant correlation between masked primes’ RT congruence effects and $d'$, Pearson’s $r(13) = -.16, p = .60$.

**Discussion**

Results showed a cross-modal priming effect. Participants were better at localizing the sounds preceded by congruent than by incongruent visual primes. This was found in particular for unconscious primes, meaning that our participants probably actively suppressed processing of the nonpredictive conscious primes (Kinoshita, Mozer, & Forster, 2011). In addition, the primes were not visible in the masked conditions but were visible in the unmasked conditions. Thus, it seems as if the cross-modal priming effect did not depend on the participants’ awareness of the primes. This conclusion was also supported by a second observation: No significant correlation was found between the RT congruence effect and the prime visibility indices of the masked primes.

Yet, how could an unconscious visual word have influenced processing of the auditory targets? We think that semantic priming could have influenced processing of the location of the auditory targets via a joint representation of semantic and sensory information. According to this explanation, both a visual prime’s meaning and an auditory target’s location would have activated units in a joint semantic network connecting word meaning with associated sensory features in other modalities (cf. Chen & Spence, 2011).

For instance, congruent spatial word meanings could have practivated units that were subsequently used for processing of auditory locations. Because of the word prime’s preactivation of a unit, the critical threshold activity of such a unit that was also used for successful processing of auditory positions would have required less additional activation by the congruently primed than by the incongruently primed target sounds.

An alternative explanation of the unconscious priming effect in terms of response priming is less likely. First, S–R mapping had no effect. Thus, the primes did not lead to an orthogonal compatibility effect, with faster responses to spatially compatible (e.g., the prime above before a right response) than spatially incompatible primes (the prime above before a left response; see Proctor & Cho, 2006). Second, the visual primes were very different from the auditory targets. Response priming on the basis of prime–target fusion (Norris & Kinoshita, 2008) or prime–target confusion (cf. Damian, 2001; Kunde, Kiesel, & Hoffmann, 2003) was therefore also unlikely.

**Experiment 2**

In Experiment 2, we included neutral nonspatial primes in addition to the spatial primes. The neutral primes denoted shapes (e.g., *angular*) or surface properties (e.g., *brittle*). In the auditory target discrimination task, this allowed us to test (a) whether congruent spatial primes facilitated target discrimination as compared to neutral primes and/or (b) whether incongruent spatial primes delayed target discrimination relative to neutral primes.
Also, in Experiment 2’s prime visibility test, we asked directly for the prime’s meaning category—that is, whether the prime was a spatial word or a nonspatial word. This test might not be as straightforward as the prime–target congruence discrimination that we used in Experiment 1. However, the present prime visibility test asks for prime meaning only and is therefore maybe easier and thus more sensitive for the participants’ residual awareness of the primes than the visibility test of Experiment 1.

Method

Participants. Sixteen participants were tested. Four participants had to be excluded by the same criterion of chance performance in the sound localization task as in Experiment 1. The remaining 12 participants (10 female, two male, M_age = 28.4 years, age range: 23–44 years) were analyzed.

Apparatus, stimuli, and procedure. These were the same as in Experiment 1, except that in addition to the 20 spatial words, 20 spatially neutral German words about the surface and shape properties of objects were included as primes. These neutral words were hart (hard), fest (tight), glatt (smooth), eben (firm), eckig (angular), kantig (edgy), stabil (stable), sprüde (brittle), steif (stiff), brüchig (fragile), weich (soft), lose (loose), rau (rough), rissig (cracked), rund (round), warm (warm), kalt (cold), bieg sam (bendable), klebrig (sticky), and haftend (adhesive), with a mean word length of 5.2 letters (range of three to seven letters) and moderate frequencies between $1.0 \times 10^{-7}$ (haftend) and $1.8 \times 10^{-4}$ (fest). Further, because the variable SOA had no significant influence in Experiment 1, only the shorter SOA of 68 ms was used. Thus, the number of trials was kept the same, although a larger number of primes was used. Moreover, here and in Experiment 3, we increased the number of practice trials. At the beginning, participants practiced the target discrimination task until their performance was at least 75% correct. Finally, in the prime discrimination task, participants were asked to categorize the prime words as either spatial or neutral (nonspatial), and before this task it was carefully explained to the participants which words were used and which of them were of a spatial meaning and which were spatially neutral. The S–R mapping was balanced across participants.

Half of the participants pressed the left key for the spatial words and the right key for neutral words, and the other half of the participants got the reverse response assignment. The experiment took about 1 hr.

Results

Sound localization. Out of all correct responses, 3.5% were excluded because RTs deviated by more than 2 standard deviations from a respective condition’s and individual participant’s mean RT. For mean values, see Appendix C and Figure 3. As can be seen, the participants performed better in the congruent than in the incongruent conditions in both the masked and unmasked conditions. This was confirmed by our formal analyses (see below). The neutral conditions lay between the congruent and incongruent conditions.

As in Experiment 1, an initial ANOVA with the between-participants variable of S–R mapping did not show a significant main effect of mapping nor any significant interaction with this variable. Data were therefore again collapsed across mappings for a repeated-measures ANOVA of the means of the correct RTs, with the within-participant variables of congruence (congruent, neutral, incongruent) and prime visibility (masked, unmasked).

A significant main effect of prime visibility, $F(1, 11) = 5.09, p < .05$, Cohen’s $f = 0.69$, was found, reflecting faster RTs in masked (626 ms, SD = 62) than unmasked (644 ms, SD = 79) conditions. Importantly, there was also a significant main effect of congruence, $F(1, 11) = 5.59, p < .01$, Cohen’s $f = 0.72$. The participants performed faster in the prime–target congruent condition (617 ms, SD = 64) than in the neutral (635 ms, SD = 74), t(11) = 2.27, $p < .04$, and incongruent (653 ms, SD = 79) conditions, t(11) = 2.67, $p < .02$, but not in the neutral as compared to the incongruent condition, t(11) = 1.77, $p = .10$. No other significant effect or interaction was found, all Fs < 1.0. Additional t tests showed that when the spatial words were presented as primes prior to the targets, participants were not significantly faster in the masked congruent condition (612 ms, SD = 57) than in the masked neutral condition (629 ms, SD = 77), t(11) = 2.02, $p = .07$, but they were significantly faster in the masked congruent than the masked incongruent condition (636 ms, SD = 57), t(11) = 2.61, $p < .02$. The t tests also showed that in the unmasked congruent condition (622 ms, SD = 73), participants

![Figure 3. Results of Experiment 2: mean reaction times (RTs) in milliseconds (upper panel) and error rates (ERs) as a percentage (lower panel) on the ordinate plotted as a function of congruence (congruent, neutral, and incongruent) and prime visibility on the abscissa. Standard errors are indicated by bars. Details about the mean RTs and ERs in each of the conditions as well as results of the analysis of the prime’s visibility are summarized in Appendix C. The dark symbols represent the masked condition, and the light symbols represent the unmasked condition.](image-url)
were not faster than in the unmasked neutral condition (641 ms, \(SD = 74\)), \(t(11) = 1.82, p = .10\), but again they were significantly faster in the unmasked congruent than the unmasked incongruent condition (670 ms, \(SD = 104\)), \(t(11) = 2.50, p < .03\).

A repeated-measures ANOVA of the ERs with the same variables showed a significant main effect of congruence, \(F(1, 11) = 8.47, p < .01\), Cohen’s \(f = 0.89\). ERs were lower in the congruent (15.2\%, \(SD = 10.2\)) than in the neutral (18.1\%, \(SD = 10.9\)), \(t(11) = 2.41, p < .04\), and in the incongruent (20.5\%, \(SD = 13.2\)) condition, \(t(11) = 3.52, p < .01\), as well as in the neutral as compared to the incongruent condition, \(t(11) = 2.13, p < .05\). There was also a significant two-way interaction, \(F(1, 11) = 6.26, p < .01\), Cohen’s \(f = 0.75\). A follow-up ANOVA for the unmasked condition showed a significant main effect of congruence, \(F(1, 11) = 10.44, p < .01\), Cohen’s \(f = 0.98\). ERs were significantly lower in the congruent (12.7\%, \(SD = 10.0\)) than in the neutral (18.0\%, \(SD = 10.7\)), \(t(11) = 2.61, p < .02\), and in the incongruent (24.0\%, \(SD = 15.6\)) condition, \(t(11) = 3.53, p < .01\), as well as in the neutral than the incongruent condition, \(t(11) = 3.01, p < .01\). A similar ANOVA for the masked condition did not show a significant congruence effect, \(F < 1.0\).

**Prime visibility.** As in Experiment 1, individual \(d’\) was computed separately for masked and unmasked primes. For the calculation of \(d’\), spatial words counted as signals and neutral words counted as noise. Accordingly, correct judgments of spatial words figured as hits, and incorrect judgments of neutral words figured as FAs.

The participants were not able to discriminate the masked primes with better-than-chance accuracy: Mean \(d’\) was 0.01 (\(SD = 0.29\)), \(t(11) = 0.14, p = .89\). In the unmasked condition, the participants successfully discriminated the primes: Mean \(d’\) was 2.82 (\(SD = 1.03\)), \(t(11) = 9.48, p < .001\).

**Congruence and prime visibility correlations.** Once again, we found no significant correlation between RT congruence effects and prime visibility scores for the masked primes, Pearson’s \(r(10) = -.13, p = .68\).

**Discussion**

Results confirmed the congruence effect already found in Experiment 1. This time, we found the congruence effect in the masked and in the unmasked conditions. In fact, the congruence effect was numerically stronger in the unmasked than in the masked condition. The latter finding indirectly supports our suppression interpretation: Because in half of all trials of Experiment 2 visible nonspatial word primes were used and because these primes would never interfere with target processing, participants might have suppressed all of the priming words less. Also, by comparison to the neutral condition, we were able to show that facilitation in congruent conditions as compared to interference in incongruent conditions contributed to a stronger degree to the net RT congruence effect, but the evidence for this stronger facilitation effect was restricted to the unmasked priming conditions in which the difference between congruent and neutral conditions was significant. In contrast, in the masked priming conditions, the only significant difference was that between the congruent and the incongruent conditions.

Importantly, we also replicated the chance performance in our visibility test of the masked primes, although this test was based on a categorization of the primes only. Which test was used to assess prime visibility is therefore not crucial for the conclusion that the masked primes were not seen. Using the arguably more demanding judgment about the prime–target congruence (as in Experiment 1) and the maybe simpler judgment about prime meaning (as in the present experiment) as prime visibility tests led to the same conclusion: The primes could not be seen. Again, the lack of a significant correlation between the RT congruence effect and prime visibility measures in the masked condition supported the conclusion that this congruence effect was independent of the participants’ awareness of the primes.

Following Experiment 1, we speculated that the congruence effect might have reflected a facilitation of the representation of target sound locations by the congruent meaning of the masked prime words. Yet, in Experiments 1 and 2, prime word meanings might have also fitted to corresponding response triggers that were represented as part of our participants’ task sets (Kunde et al., 2003). For example, participants might have represented an S–R rule, with an up S (a sound from above) as an action trigger for a rightward response. The up prime might have fitted to such an action trigger, thereby activating a response. In congruent conditions, such a prime-activated response would have been the same as the finally required response, but in incongruent conditions, the prime-activated response would have been different from the required target response. Therefore, it is possible that the congruence effect reflected prime-activated responses rather than prime-activated sensory representations. Whether action triggering through the masked primes could have accounted for the congruence effect was tested in the final experiment.

**Experiment 3**

In Experiment 3, we changed the target discrimination task. Participants had to distinguish between two sound types by pressing one button for white noise sounds (from above or from below) and an alternative button for click train sounds (from above or below). As a consequence, up and down primes were neutral with respect to the response triggers. If the congruence effect reflected response triggering by the primes, the congruence effect should therefore be eliminated. However, if the congruence effect reflected a more or less tight fit between the prime word’s meaning and the sound location, a congruence effect should persist, with faster responses in congruent conditions (i.e., with up–up and down–down prime–target pairs) than in incongruent conditions (i.e., with up–down and down–up prime–target pairs).

We also took the opportunity for yet another prime visibility test. To obtain a fuller picture of the participants’ residual awareness of the primes, in Experiment 3 we conducted yet another complementary prime visibility test. We asked our participants to discriminate between up primes and down primes. If we find a prime–target congruence effect and if this congruence effect is truly based on subliminal input, participants should perform at chance level in the discrimination of the masked primes’ different spatial meanings also.

**Method**

**Participants.** Sixteen participants were tested. Two participants had to be excluded by the same criterion of chance-level
performance in the sound localization task as in Experiments 1 and 2. The remaining 14 participants (12 female, two male, $M_{\text{age}} = 27.9$ years, age range: 23–37 years) were analyzed.

**Apparatus, stimuli, and procedure.** These were the same as in Experiment 1, except for the following. First, four target sounds were used: two 300-ms broadband white noise sounds (one from above, one from below), as in Experiments 1 and 2, and two 300-ms click train sounds (again, one from above and one from below). The click train consisted of 200 clicks per second. The broadband noise and click train are characterized by their noisy and tonal sound percept, respectively. All sounds were created individually for each participant using the same procedure as in Experiment 1. Here, as in Experiment 2, task practice (until 75% correct responses were achieved) and only the short SOA of 68 ms were used. The participants were asked to categorize the targets as tonal or noisy. In the prime discrimination task, the participants were asked to categorize the primes as up or down words. The S–R mapping was balanced across the participants.

**Results**

**Sound categorization.** Out of all correct responses, 3.9% were excluded because RTs deviated by more than 2 standard deviations from a respective condition’s and individual participant’s mean RT. As can be seen in Appendix D and Figure 4, the overall ERs were quite low, showing that the listeners had no difficulty discriminating the two sound types. With respect to the RT, participants once again responded faster in the congruent conditions than in the incongruent conditions. This was again confirmed by formal analyses.

As in Experiments 1 and 2, an initial ANOVA with the between-participants variable of S–R mapping did not show a significant main effect of mapping nor any significant interaction with this variable. Data were therefore collapsed across mappings for a repeated-measures ANOVA of the means of the correct RTs, with the within-participant variables of congruence (congruent, incongruent), target type (noise sound, click train sound), and prime visibility (masked, unmasked).

We found a significant main effect of congruence, $F(1, 13) = 6.60, p < .02$, Cohen’s $f = 0.72$, reflecting faster RTs in congruent (604 ms, $SD = 61$) than incongruent (615 ms, $SD = 71$) conditions. No other significant main effect or interaction was found, all $Fs < 1.0$. An additional $t$ test confirmed that participants were significantly faster in the masked congruent condition ($M = 601$ ms, $SD = 65$) than in the masked incongruent condition ($M = 614$ ms, $SD = 76$), $t(13) = 2.35, p < .03$. A similar $t$ test of the unmasked congruent condition ($M = 607$ ms, $SD = 68$) and incongruent condition ($M = 615$ ms, $SD = 67$) was not significant, $t(13) = 1.07, p = .30$.

The same ANOVA with ERs showed only a significant two-way interaction of congruence and target type, $F(1, 13) = 8.27, p < .01$, Cohen’s $f = 0.80$. Subsequent $t$ tests showed that neither the noise sounds’ congruent condition (2.7%, $SD = 2.1$) was significantly different from its incongruent (1.6%, $SD = 1.3$) condition, $t = 1.55, p = .15$, nor that the click train sounds’ congruent condition (1.8%, $SD = 1.5$) was significantly different from its incongruent (3.2%, $SD = 2.8$) condition, $t = 1.71, p = .11$. No other significant main effect or interaction was found, all $Fs < 1.0$.

**Prime visibility.** The participants were not able to discriminate the masked primes with better-than-chance accuracy: Mean $d’$ was $-0.01$ ($SD = 0.20$), $t(13) = 0.11, p = .92$. However, they successfully discriminated the unmasked primes: Mean $d’$ was 2.45 ($SD = 1.04$), $t(13) = 8.79, p < .001$.

**Congruence and prime visibility correlations.** As in Experiments 1 and 2, we found no significant correlation between the RT congruence effect and prime visibility for the masked primes, Pearson’s $r(12) = .29, p = .31$.

**Discussion**

The results confirmed a significant RT congruence effect of the masked primes, although response triggering was ruled out and the primes were presented subliminally. When taken together, these results are in line with an influence of subliminal prime meaning on the sensory representation of the sound locations. As in Experiment 1, active suppression probably prevented the congruence effect in the unmasked condition (Kinoshita et al., 2011).

**Analysis Across the Three Experiments**

We also conducted a final analysis covering data from all three experiments. This was done to get a better estimate of the true effect size of the congruence effect. For this analysis, we considered the conditions that were the same across Experiments 1, 2, and 3: the spatial primes and noise targets with the short SOA.
Results

A repeated-measures ANOVA, with the within-participants variables of prime–target congruence (congruent, incongruent) and prime visibility (masked, unmasked) and the between-participants variable of experiment (Experiment 1, Experiment 2, Experiment 3), was conducted on RTs. It showed a main effect of prime–target congruence, $F(1, 40) = 12.80$, $p < .01$, Cohen’s $f = 0.58$. Mean RTs were shorter in congruent ($M = 613$ ms, $SD = 59$) than in incongruent conditions ($M = 632$ ms, $SD = 67$). We also found a significant main effect of prime visibility, $F(1, 40) = 6.48$, $p < .02$, Cohen’s $f = 0.41$. Mean RTs were shorter in masked ($M = 614$ ms, $SD = 60$) as compared to unmasked conditions ($M = 631$ ms, $SD = 68$). No other significant effect or interaction was found, all $Fs < 2.7$.

Discussion

In the current analysis across experiments, with its larger $n$, we achieved a more trustworthy and lower measure of the true effect size of the congruence effect than in the individual experiments. The analysis also showed that masked conditions were slightly easier than unmasked conditions (see also Experiment 2). Maybe the visible primes elicited some form of additional suppression. This would explain why RTs were longer in unmasked than masked conditions and why unmasked primes tended to elicit a quantitatively weaker congruence effect than masked primes, although this was not reflected in a significant interaction between prime visibility and congruence and although this was not true of Experiment 2. We will get back to this question at the end of the General Discussion.

General Discussion

Based on theories connecting semantic word representations with corresponding sensory representations within one joint network (Barsalou, 1999; Chen & Spence, 2011; Logan, 1995), we hypothesized that subliminal prime words in the visual modality should impact sensory discrimination between auditory targets. In the current experiments, this prediction was borne out with masked visual word primes and auditory targets. Our participants had to judge the vertical direction of an auditory target relative to their own position. In congruent conditions, the masked word prime’s spatial meaning fitted to the location of the auditory target. For example, the word *up* was presented before an auditory target from above. In the incongruent conditions, the masked word prime’s spatial meaning did not fit to the location of the auditory target. For example, the word *up* was presented before an auditory target from below. In line with the predictions, we found a congruence effect: faster responses in congruent than in incongruent conditions (in Experiments 1 to 3). The results support the existence of a cross-modal link between masked visual primes and auditory targets.

Importantly, all these cross-modal congruence effects were found where the participants could not recognize the masked primes. This was shown in our prime visibility (or prime awareness) tests. Certainly, each of these visibility tests had its drawbacks as well as its advantages. For example, our visibility test in Experiment 1 asked for prime-target congruence discrimination. This is a straightforward question when one wants to demonstrate a dissociation between congruence effects and a lacking awareness for the underlying congruence relation, but the task is more demanding than asking for the prime meaning alone (as we did in Experiments 2 and 3). Therefore, the prime visibility test of Experiment 1 might not have been optimally sensitive for the participants’ residual awareness of the primes, and easier tasks were used in Experiments 2 and 3. In Experiment 2, we asked the participants to only discriminate if the prime did or did not have a spatial meaning, and in Experiment 3, we directly asked the participants to only decide if the prime was an up or a down word. Both of these prime visibility tasks were probably easier than the congruence discrimination in Experiment 3, but arguably have the drawback of not asking an entirely correct question to demonstrate the dissociation. However, as we varied the visibility tests between experiments and found that the specifics of the visibility tests were immaterial for our conclusion that the masked primes were not seen, we are relatively sure that the cross-modal congruence effects were independent of the participants’ awareness of the primes. This conclusion was also supported by lacking correlations between masked primes’ visibility scores and RT congruence effects.

This leads us to the question as to how this awareness-independent influence of the masked primes was brought about. A possible answer can be given on the basis of Logan’s (1995) theory. First of all, our participants had to give a deictic judgment: They had to decide whether an auditory target was presented from above or from below. This required that the participants locate the targets relative to one another or to their own (i.e., the participant’s) position. According to Logan’s theory, this can only be done by the application of a conceptual spatial reference frame to the stimuli by which the spatial relation between the targets or between the participant’s own body (or here: his or her ears) and each target is determined. In this situation, the prime words could have worked in one of the following ways. First, each prime word could have facilitated the usage of one particular spatial frame, an up frame or a down frame, corresponding to the prime’s meaning. If that was the case, a congruent prime would have preactivated a conceptually fitting frame for the target, whereas an incongruent prime would have preactivated a conceptually nonfitting frame for the target, with the resulting necessity to adjust the spatial frame to the target in a time-consuming manner only in the incongruent conditions. Second, it is also possible that the participants had set up a joint conceptual spatial frame for all deictic judgments as a top-down template and in anticipation of the auditory targets. As a consequence of such a top-down template, participants could then have inadvertently processed the prime word’s meaning in accordance with the template, just as this is known to be the case for other forms of top-down contingent processing, such as motor activations or attention shifts (Folk, Remington, & Johnston, 1992; Kunde et al., 2003). Note that the spatial frame is a semantic concept, so it is easy to understand why a prime word’s meaning could have specified one particular spatial relation within the frame. Also, because the same concept of a spatial frame had to be used for the judgment of the deictic relation between targets between target and participant, an incongruent prime specifying a spatial relation different from that to the target could have delayed the correct judgment about target position. In general agreement with this possibility, application of a top-down template to process unconscious stimuli is possible once the top-down template has
been set up in advance of conscious targets (for a review, see Ansorge et al., 2014).

From a more general view, the awareness-independent cross-modal priming effect that we found is in line with views of word understanding that assume a tight link between word meaning and sensory processing, such as the embodied cognition view (e.g., Barsalou, 1999) or theories that reckoned that readers construct situation models out of the words they read (e.g., Kintsch & van Dijk, 1978; Zwaan & Radvansky, 1998). For instance, according to the embodied cognition view, when understanding an abstract word, a reader has to simulate (part of) the fitting past sensory and motor experiences that lay aground of understanding word meaning in the first place (Barsalou, 1999; Glenberg & Kaschak, 2002). Because embodied processing is assumed to be an obligatory consequence of reading a word and would link abstract word meaning and perceptual processing, it would be clear why even subliminal visual words could prime sensory processing in a different modality: Awareness independence is one hallmark of the automatic or obligatory processing (e.g., Posner & Snyder, 1975).

In contrast to these potential explanations in terms of semantic or conceptual priming, several alternative explanations of the cross-modal congruence effect were ruled out. According to one potential explanation, the masked primes might have activated responses via an orthogonal Simon effect (Proctor & Cho, 2006). If an orthogonal Simon effect would have created the congruence effect, we would have expected a masked congruence effect in the S–R compatible conditions (in which targets from above required a right-hand response and targets from below required a left-hand response) but not in the S–R incompatible conditions (in which targets from above required a left-hand response and targets from below required a right-hand response). However, the between-participants variable of S–R mapping did not significantly interact with the congruence effect in any of the experiments. Also, for Experiments 1 and 2, it could be argued that the masked prime could have directly triggered a response based on a fit of the primes’ meanings to the action triggers (Kunde et al., 2003). In these experiments, the participants had to give one response to targets from above and an alternative response to targets from below. However, this possibility was ruled out in Experiment 3 in which primes did not fit to the action triggers—that is, the primes’ meanings were not related to the responses.

In addition to an account of cross-modal connections between representations of word meaning and sensory stimulus characteristics in a single joint network, it may also be possible that the primes exerted their influence via an attention shift (Gibson & Kingstone, 2006; Logan, 1994). According to this explanation, the masked word would have triggered an attention shift with a direction corresponding to its spatial meaning, so that in congruent conditions attention would have been directed toward the target’s spatial position and in incongruent conditions it would have been directed away from the target’s position. In line with this possibility, previous research demonstrated (a) intermodal attentional effects of visual cues during the discrimination of auditory targets (Driver & Spence, 1998; Eimer & Schröger, 1998) as well as (b) attentional effects on the basis of subliminal visual cues (Ansorge, Kiss, & Eimer, 2009; McCormick, 1997). Although it is thus theoretically possible that masked primes led to an attention shift, this remains to be proven in future studies because, to our knowledge, so far there has not been a single study demonstrating attention shifts based on masked visual words.

Before we conclude, we need to quickly discuss the size of the congruence effect. The cross-modal congruence effects (incongruent RT – congruent RT) of 12 and 11 ms in the current Experiments 1 and 3 (see Figures 2 and 4) were quantitatively small as compared to the cross-modal congruence effect of 36 ms in Experiment 2 and to the intramodal congruence effect of 25 ms that was found when using the same spatial words as primes and targets in intramodal priming experiments (Ansorge et al., 2013). Two interconnected reasons for the small congruence effect might have been (a) the nonpredictive nature of the primes and (b) the intermixed presentation of the masked and unmasked primes. Concerning the first reason, because congruent and incongruent primes were equally likely and did thus not predict target identity, participants would have been well advised to ignore the primes. Also, it might be that the participants did not register the prime–target contingencies in the masked conditions (but see, e.g., Reuss, Desender, Kiesel, & Kunde, 2014), but at least in the unmasked conditions, participants should have noticed the nonpredictive nature of the primes. As a consequence, participants might have set up a mind-set to actively suppress the unmasked primes as much as possible given the constraint that they had to gaze at the experimental screen presenting the prime words. In line with this possibility, other studies sometimes observed zero congruence effects with unmasked nonpredictive primes (Kinoshita et al., 2011). Concerning the second reason, in the current study, the participants could have applied their suppressive mind-set for the unmasked primes inadvertently to the masked primes, too, because the masked primes were presented randomly intermixed with the unmasked primes. In other words, if the participants prepared to suppress the visible primes, masked primes could have also triggered their own suppression by their fit to a suppressive mind-set, at least in some of the trials, and in a way similar to their ability to trigger other forms of inhibitory processes (Experiment 5 of Ansorge, 2004). Therefore, in Experiments 1 and 3, suppression would have counteracted the congruence effect of the clearly seen unmasked primes, but the congruence effect of the masked prime could also have suffered from this suppression to some extent. This would explain both the fact that the congruence effect tended to be small and the fact that the congruence effect tended to be even smaller with unmasked than masked primes.

However, a notable exception from this pattern was Experiment 2. In Experiment 2, a quantitatively stronger overall RT congruence effect was found, and this was mostly due to the unmasked primes rather than the masked primes (see Appendix C). We think that this finding neatly dovetails with our explanation in terms of active prime suppression counteracting the congruence effect. In Experiment 2, half of the primes were spatially neutral and thus completely unrelated to target classification. As a consequence, on average, less suppression of the primes was necessary in Experiment 2. Yet, clearly, such a context effect would have had a better chance to express itself in the conditions with unmasked primes, too, so that the RT congruence effect of the unmasked primes was affected to a larger degree (in this case, it was boosted due to a
release from inhibition) than the congruence effect of the masked primes.

The size of the masked congruence effect of 24 ms in the present Experiment 2 (see Appendix C) is also interesting for yet another reason: This cross-modal congruence effect was almost as strong as the congruence effect of masked intermodal visual word–word priming of 27 ms in Experiment 2 of Ansorge et al. (2013). This similarity of the congruence effects suggests that almost all of the masked words’ congruence effects could be due to modality-unspecific (or amodal) semantic priming: The fact that (a) only in the study of Ansorge et al. (2013) but not in the present study could the visual words have primed the target modality, the letters, or the words of the target displays and that (b) these potential sources for differences did not affect the size of the congruence effect suggests that a principle that was shared between both studies, such as semantic priming, could have accounted for the congruence effect of the masked primes. This conclusion, however, is tentative. The results could be different where visual and auditory targets have to be discriminated from one another or where a different temporal interval between the prime and target is used. Therefore, answering the question if the priming effect in the current study was entirely amodal has to await a whole series of future experiments and is beyond the scope of the present study.

Conclusion

We have shown that the meaning of an unconscious visual word can cross-modally prime the perceived location of an auditory target. This unconscious cross-modal priming effect appears to reflect semantic processing.

References


Appendix A

Method for Obtaining Directional Transfer Functions

We describe only those aspects of the method that are relevant for the current application. Head-related transfer functions (HRTFs) were measured individually for each listener in a session prior to the experiment and on a different day. Twenty-two loudspeakers were mounted on a vertical circular arc with a spacing of 5°. The listener was seated in the center point of the arc, with a distance between the center point and each speaker of 1.2 m. Microphones were inserted into the listener’s ear canals, and their output signals were directly recorded. The test signal was a 1,729-ms exponential frequency sweep from 0.05 to 20 kHz. The acoustic influence of the equipment was removed by equalizing the HRTFs with the transfer functions of the equipment. For deriving the directional transfer functions (DTFs), first the magnitude of the common transfer function (CTF) was calculated by averaging the log-amplitude spectra of all HRTFs for each individual listener. The phase spectrum of the CTF was set to the minimum phase corresponding to the amplitude spectrum. The DTFs were finally obtained by filtering the HRTFs with the inverse complex CTF. Finally, the impulse responses of the DTFs were windowed with an asymmetric Tukey window (fade-in of 0.5 ms and fade-out of 1 ms) to a 5,33-ms duration. More details on this method can be found in Majdak, Walder, and Laback (2013).
## Appendix B

### Mean Reaction Times, Error Rates, and Prime Visibility Indices of Experiment 1

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Reaction times (ms)</th>
<th>Error rates (%)</th>
<th>Prime visibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean RTs</td>
<td>Net congruence</td>
<td>Mean ERs</td>
</tr>
<tr>
<td>Visible Masked</td>
<td>Short</td>
<td>600</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>600</td>
<td>14</td>
</tr>
<tr>
<td>Unvisible Masked</td>
<td>Short</td>
<td>623</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>604</td>
<td>15</td>
</tr>
</tbody>
</table>

*Note.* SOA = stimulus onset asynchrony; RT = reaction time; ER = error rate. Mean RTs and error rates are compared in the congruent and incongruent conditions, and the net congruence effect is calculated as the mean performance in the incongruent condition minus the mean performance in the congruent condition.

## Appendix C

### Mean Reaction Times, Error Rates, and Prime Visibility Indices of Experiment 2

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Reaction times (ms)</th>
<th>Error rates (%)</th>
<th>Prime visibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean RTs</td>
<td>Net congruence</td>
<td>Mean ERs</td>
</tr>
<tr>
<td>Visible Masked</td>
<td>Congruent</td>
<td>612</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>629</td>
<td>17.7</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>636</td>
<td>17.7</td>
</tr>
<tr>
<td>Unvisible Masked</td>
<td>Congruent</td>
<td>622</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>641</td>
<td>18.0</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>670</td>
<td>24.0</td>
</tr>
</tbody>
</table>

*Note.* RT = reaction time; ER = error rate. Mean RTs and error rates are compared in the congruent and incongruent conditions, and the net congruence effect is calculated as the mean performance in the incongruent condition minus the mean performance in the congruent condition.
## Appendix D

### Mean Reaction Times, Error Rates, and Prime Visibility Indices of Experiment 3

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Reaction times (ms)</th>
<th>Error rates (%)</th>
<th>Prime visibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean RTs</td>
<td>Net congruence</td>
<td>Mean ERs</td>
</tr>
<tr>
<td>Masked</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noise</td>
<td>Congruent</td>
<td>604</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>616</td>
<td></td>
</tr>
<tr>
<td>Click</td>
<td>Congruent</td>
<td>59.814</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>612</td>
<td></td>
</tr>
<tr>
<td>Unmasked</td>
<td>Noise</td>
<td>619</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Congruent</td>
<td>625</td>
<td>1.4</td>
</tr>
<tr>
<td>Click</td>
<td>Congruent</td>
<td>595</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Incongruent</td>
<td>605</td>
<td>1.6</td>
</tr>
</tbody>
</table>

**Note.** RT = reaction time; ER = error rate. Mean RTs and error rates are compared in the congruent and incongruent conditions, and the net congruence effect is calculated as the mean performance in the incongruent condition minus the mean performance in the congruent condition.