Expected but omitted stimuli affect crossmodal interaction

Marcello Costantini[1,2], Daniele Migliorati[2], Brunella Donno[2], Miroslav Sirota[1] & Francesca Ferri[1]

[1] Centre for Brain Science, Department of Psychology, University of Essex, Colchester, UK

[2] Laboratory of Neuropsychology and Cognitive Neuroscience, Department of Neuroscience and Imaging, University G. d'Annunzio, Chieti, Italy & Institute for Advanced Biomedical Technologies - ITAB, Foundation University G. d'Annunzio, Chieti, Italy.

Running title: On expectation and multisensory processing

Corresponding authors:

Francesca Ferri, Centre for Brain Science, Department of Psychology, University of Essex; Wivenhoe Park, Colchester CO4 3SQ, United Kingdom; Voice: ++01206 873710; e-mail: fferri@essex.ac.uk

Or

Marcello Costantini, Centre for Brain Science, Department of Psychology, University of Essex; Wivenhoe Park, Colchester CO4 3SQ, United Kingdom; Voice: ++01206 873493; e-mail: mcostaa@essex.ac.uk

Abstract

One of the most important ability of our brain is to integrate input from different sensory modalities to create a coherent representation of the environment. Does expectation affect such multisensory integration? In this paper, we tackled this issue by taking advantage from the crossmodal congruency effect (CCE). Participants made elevation judgments to visual target while ignoring tactile distractors. We manipulated the expectation of the tactile distractor by pairing the tactile stimulus to the index finger with a high-frequency tone and the tactile stimulus to the thumb with a low-frequency tone in 80% of the trials. In the remaining trials we delivered the tone and the visual target, but the tactile distractor was omitted (Study 1). Results fully replicated the basic crossmodal congruency effect. Strikingly, the CCE was observed, though at a lesser degree, also when the tactile distractor was not presented but merely expected. The contingencies between tones and tactile distractors were reversed in a follow-up study (Study 2), and the effect was further tested in two conceptual replications using different combinations of stimuli (Studies 5 and 6). Two control studies ruled out alternative explanations of the observed effect that would not involve a role for tactile distractors (Studies 3, 4). Two additional control studies unequivocally proved the dependency of the CCE on the spatial and temporal expectation of the distractors (Study 7, 8). An internal small-scale meta-analysis showed that the crossmodal congruency effect with predicted distractors is a robust medium size effect. Our findings reveal that multisensory integration, one of the most basic and ubiquitous mechanisms to encode external events, benefits from expectation of sensory input.

1. Introduction

Two key mechanisms help us to cope with an overwhelming amount of sensory inputs coming from the environment: sensory expectation and crossmodal interaction. The former refers to the idea that we do not solely react to external stimuli; rather we constantly create predictions about forthcoming sensory events (Engel, Fries, & Singer, 2001). The latter refers to the idea that we do not use sensory systems one at a time, rather we simultaneously process information coming from different sensory modalities. These two mechanisms can be observed already in non-human primates (Amemori & Sawaguchi, 2006; Siemann et al., 2014), suggesting that they are unlikely related to the privileged cognitive status of humans. Conversely, they might represent a fundamental prerequisite for an efficient interaction with the environment.

Models of predictive brain have been used to explain how expectation of upcoming stimuli is generated (for reviews: (Clark, 2013; Friston, 2010). According to these models, expectation at the neural level takes the form of increased baseline neural activity (i.e., biased by the probability of stimulus occurrence) and increased evoked response (i.e, similar for expected and actual stimuli; for a review see: (Summerfield & de Lange, 2014). For instance, previous research has shown that cues predicting a forthcoming visual stimulus lead to increases in BOLD signal in category-specific visual regions. For example, when the word 'house' predicts the subsequent occurrence of a house, it triggers higher BOLD signals in the parahippocampal place area (Puri, Wojciulik, & Ranganath, 2009). Behavioural evidence demonstrates that expectation is beneficial for processing and responding to external stimuli. For instance, expectation of low-level features (e.g., colour, direction of motion) leads to facilitated processing of stimuli containing those features (Ball & Sekuler, 1981; Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Saenz, Buracas, & Boynton, 2002).

Only recent research has begun to investigate the relationship between expectation and crossmodal interaction. Examining this issue is critical for the understanding of how we perceive and react to environmental stimuli. Indeed, in daily life we usually do not perceive external events through only

one sensory modality. Instead, information from the environment reaches us via multiple sensory systems. Integrating information across the senses improves a wide range of behavioral outcomes, including detection (Lovelace, Stein, & Wallace, 2003; Stein & Wallace, 1996), localization (Nelson et al., 1998; Wilkinson, Meredith, & Stein, 1996), and speed of response (Diederich & Colonius, 2004; Hershenson, 1962).

To date, however, research on the relationship between expectation and crossmodal interaction has focused predominantly on the extent to which top-down expectations impact on actual multimodal events (Gau & Noppeney, 2016; Nahorna, Berthommier, & Schwartz, 2012) as in the case of the McGurk effect. In the McGurk effect, participants are presented with the auditory phoneme /ga/ synchronous with an incongruent lip movement /ba/. This leads to the illusory perception of a different syllable /da/ (McGurk & MacDonald, 1976). Nahorna and colleagues (Nahorna et al., 2012) manipulated participants' top-down congruency expectations by presenting McGurk stimuli embedded in blocks of congruent or incongruent syllables. They showed that the multimodal McGurk effect was largely reduced when the constituent unisensory stimuli were preceded by an incoherent audiovisual context. Other studies (Stekelenburg & Vroomen, 2012) have shown similar top-down effects on crossmodal interaction employing different pair of stimuli, e.g. audio-tactile or visuo-tactile. For instance, by using a spatial cuing task, Spence and Driver (Spence & Driver, 1996) showed that participants were faster at judging the elevation of visual or auditory targets when the location of the upcoming stimulus was cued by a stimulus in either the same or different sensory modality. This finding is intriguing and clearly suggests that expectation of sensory events might occur across modalities.

So far, little research has investigated whether crossmodal interaction between an actual stimulus and an expected, but omitted stimulus, could occur. In support of this working hypothesis, there is evidence showing that expectations affect the sensory response in the absence of sensory input (den Ouden, Friston, Daw, McIntosh, & Stephan, 2009; Kok, Rahnev, Jehee, Lau, & de Lange, 2012;

SanMiguel, Widmann, Bendixen, Trujillo-Barreto, & Schroger, 2013; Todorovic, van Ede, Maris, & de Lange, 2011; Wacongne et al., 2011). For instance, Kok and colleagues (Kok, Failing, & de Lange, 2014) showed that expectation of a specific visual stimulus evokes a pattern of activity in the visual cortex with similar features as those evoked by real stimuli.

Starting from this evidence, we seek to investigate whether expectation of a tactile event is a sufficient condition to elicit crossmodal interaction in a modified version of the crossmodal distractor congruency task (Driver & Spence, 1998; Spence, 2010, 2011; Spence, Pavani, & Driver, 2000, 2004; Spence & Walton, 2005). In a typical study, participants hold two foam blocks, one in either hand, provided with vibrotactile stimulators and light emitting diodes (LEDs) in the upper and lower surfaces. On each trial, a vibrotactile and a visual stimulus are presented randomly from any one of the four possible stimulus locations. Participants are required to make speeded elevation (up/down) discriminations for each vibrotactile target stimulus, presented to either the index finger or the thumb, while simultaneously ignoring any visual distractor. The common finding is that participants are significantly faster at discriminating the elevation of tactile targets when visual distractors are presented at congruent elevation. That is, a spatially non-predictive visual cue enhances judgments for tactile targets presented near to it, relative to those presented elsewhere (Driver & Spence, 1998). The effect is consistently found also when participants respond to the visual stimulus trying to ignore the tactile stimulus (Spence & Walton, 2005).

In this study we conceived a modified version of the classic crossmodal distractor congruency task. Participants made elevation judgments to visual target while ignoring tactile distractors. We manipulated the expectation of the tactile distractor by pairing the tactile stimulus to the index finger with a high-frequency tone and the tactile stimulus to the thumb with a low-frequency tone in 80% of the trials. In the remaining trials we delivered the tone and the visual target, but the tactile distractor was omitted.

Based on evidence suggesting that expected stimuli evoke a pattern of activity with similar features as that evoked by the real stimulus (Kok et al., 2014), we predict a crossmodal congruency effect not only when the *actual tactile distractor* is spatially incongruent with the visual target, but also when the *expected tactile distractor* is spatially incongruent with the visual target, even if omitted (study 1).

The contingencies between tones and tactile distractors were reversed in a follow-up study (study 2). Two additional control studies (studies 3 & 4) aimed at testing the dependency of our effect on the expectation of the tactile distractor. These studies ruled out the possibility that the mere association between the auditory cue and visual target could account for by our results (study 3) by investigating the time course of the audio-visual crossmodal congruency effect (study 4). Two conceptual replications of study 1 (studies 5 & 6) aimed at assessing the generalizability of our effect to other stimulus combinations. In particular, in study 5 the cue, the distractor and the target were auditory, visual and tactile, respectively; while in study 6 the cue, the distractor and the target were visual, visual and tactile, respectively. Two further control studies (studies 7 & 8) aimed at testing the dependency of our effect on the spatial and temporal predictability of the distractor. In study 7 the cue, the distractor and the target were auditory, visual and tactile, respectively, as in study 5, but in this case the auditory cue predicted the spatial location of the forthcoming distractor with a 50% of accuracy, thus making it "spatially" unpredictable. Finally, in study 8 the cue, the distractor and the target were auditory, visual and tactile, respectively. In this study, the expectation of the distractor dissipated over time in specific trials, due to the delayed presentation of the target. Thus, the crossmodal congruency effect should not be observed in these trials.

2. Materials and Methods

2.1. Methodological disclosure and description of the analysis

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in each study. We report all the studies conducted for this project (Simmons, Nelson, &

Simonsohn, 2011). We have used two inferential frameworks to assess the evidence for the critical effects: null hypothesis significance testing and a Bayesian inference framework – Bayes factor analysis. The latter framework enabled us to quantified relative evidence to support the null effect model against models assuming an effect (or vice versa). The crucial notion here is a Bayes factor (BF), which is the ratio of the probability of the data given model A (e.g., the null model) to the probability of the data given model B (e.g., a model assuming a certain distribution of effects). Bayes factors allows us to quantify how many more times are the data likely to occur under the assumption of the model A compared to the assumptions of model B (or vice versa). For instance, $BF_{01} = 20$ means that the data are 20 times more likely to occur under the model A (i.e., the null model here) relative to the model B. A Bayes Factor, BF_{01} with a value lower than 1 indicates that the model assuming the effect (model B) is more likely relative to the null model (model A) and with value greater than 1 indicates that the model assuming no effect is more likely relative to the model assuming the effect. Furthermore, the Bayes Factor values may also be interpreted as evidence categories, for example, BF_{01} values between 1 to 3 indicate *anecdotal evidence* to support the null model relative to the competing model, whereas values greater than 100 indicate *extreme evidence* to support the null model (Jeffreys, 1961; Lee & Wagenmakers, 2014). Finally, if the prior probability odds are defined, then Bayes factors can be combined into the posterior odds and can thus quantify support for tested hypotheses. For instance, if we assume prior odds of the two competing models to be 1 (i.e., equally likely) before running a study, then $BF_{01} = 100$ can be combined into the posterior odds (1*100 = 100), which will mean that the null model is 100 more likely relative to the compared model (i.e., assuming a specific distribution of the effects). Here, we calculated a default Bayes factors using JASP and R package Bayes Factor (Love et al., 2015; Morey & Rouder, 2015) for the critical effects of the crossmodal congruency effects (Studies 1, 2, 5, 6) – they are synthesized in the subsection "Evidence synthesis" – and for possible alternative explanations of the effects – they are reported in the respective results sections of studies 3, 4, 7 and

8. All the studies were approved by the local ethical committee and carried out in accordance with ethical guidelines laid down in the Declaration of Helsinki.

2.2. Study 1

2.2.1. Participants

Thirty right-handed healthy volunteers (mean age = 21 years old, 18 female) were recruited by advertisement to take part in this study from the student pool. We determined our sample size a-priori based on our experience with similar studies. Sensitivity analysis showed that such a sample size would be sufficient to detect a medium size effect of dz = 0.53, while assuming α = 0.05, 1 – β = 0.80 and a two-tailed matched paired t-test. The participants in this and all subsequent studies in this paper reported normal hearing, normal touch sensitivity and normal or corrected-to-normal vision, and were naive as to the purpose of the study.

2.2.2. Stimuli

The study was performed in a dark room. The participant rested her forearms on a table and held a foam block (width 6 cm; depth 6 cm; height 6.5 cm) between the index finger and the thumb of the right hand. The foam block was devised to deliver visual and vibro-tactile stimuli. For this purpose, two tapper solenoids (diameter 9 mm; length 11.5 mm; weight: ~12.5 g; coil resistance: 36Ω +/-1Ω) and two round red light-emitting diodes (LEDs; diameter 5 mm; luminance 14 mcd) were embedded on the foam block in order to provide vibro-tactile and visual stimuli, respectively. LEDs were embedded on the top (Upper Led) and the bottom end (Lower Led) of an imaginary vertical axis bisecting one face of the foam block. The tapper solenoids were hidden to reveal the rim containing a magnetic vibrator cone close to the LEDs. Specifically, each solenoid was placed just above the topside, and below the bottom side, of the face that exhibits the LEDs, aligned with them. Furthermore, participants wore a pair of stereo headphones (AKG K-514; rated impedance 32 ohms, frequency range 18-22.000 Hz, sensitivity 112dB/V) to hear one of two pure-tone sound frequencies consisting in a high frequency sound (1000 Hz) or a low frequency sound (375 Hz).

The audio output was balanced at a common level of comfortable audibility and it remained unchanged throughout the study and across participants. Reaction times were recorded, from the onset of the visual target, using a response pad (Cedrus RB-834, Cedrus Corporation, San Pedro, USA) located 70 cm to the left from the participant's midline. Auditory, vibro-tactile and visual stimuli, as well as participants' responses, were controlled by a PC running Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997) implemented in MATLAB (The MathWorks Inc., Natick, MA, USA).

2.2.3. Procedure

Each experimental block was composed of 192 (80%) *conditioning trials* in which the auditory cue, the vibro-tactile distractor and the visual target were presented, and 48 (20%) *expectation trials* in which the auditory cue and the visual target were presented, while the vibro-tactile distractor was expected but actually omitted. Each block also included 24 *catch trials* in which only the auditory cue and the vibro-tactile distractor were presented. Trials were randomized between blocks. In total each block was made up of 264 trials. Moreover, each block was binned into four blocks (hereafter time bin) of 66 trials, in which all the experimental conditions were equally presented. The experimental session lasted approximately 45 minutes and consisted of two experimental blocks for a total of 528 trials.

Each *conditioning trial* started with an auditory cue (200ms), consisting of either a low-frequency (375 Hz) or a high-frequency (1000 Hz) tone, which predicted the stimulation of the thumb or the index finger respectively with 100% validity (see Figure 1, panel A). The auditory cue was followed after 500ms by a vibro-tactile stimulus. The vibro-tactile stimulus was followed, after 30ms, by the visual target. Hence, the vibro-tactile stimulus acted as a distractor of the visual event. Half of the trials were congruent (96, upper LED/Index Finger or lower LED/Thumb); the other half was incongruent (96, Upper LED/thumb, or lower LED/index finger). Intertrial interval (ITI) ranged between 1800 and 2200ms in 100ms steps.

The structure of *expectation trials* was identical to the structure of *conditioning trials* with the exception of the vibro-tactile distractor, which was omitted (see Figure 1, panel A). Half of the expectation trials were congruent (24, upper LED/expected index finger or lower LED/expected thumb); the other half was incongruent (24, upper LED/expected thumb or lower LED/expected index finger).

Participants were invited to hold the foam block with the right hand, and to respond with the left index or left middle finger to indicate whether the visual target was presented on the upper location or the lower location. Half of the participants were instructed to use the index finger to indicate upper locations and the middle finger to indicate lower locations. The opposite was true for the other half. To avoid spatial compatibility effects with the vertical arrangement of the visual targets, the response buttons were arranged horizontally.

The experimental design was a 2x2 factorial. The two within-subjects factors were the congruency of the vibro-tactile distractor with respect to the elevation of the visual target (congruent vs. incongruent) and the type of trial (conditioning vs. expectation). The crossing of these factors yielded 4 possible conditions, i) congruent – conditioning, ii) congruent – expectation, iii) incongruent – conditioning, iv) incongruent – expectation, that were included in a 2x2 analysis of variance (ANOVA).

2.2.4. Results

Shapiro-Wilk's test and visual inspection of the data showed that data violated the assumption of normality, thus data were log-transformed. In this study as in all the other studies, the procedure was effective at reducing the skewness of the distributions ($p > .18$). Missed responses (0.20%), anticipatory responses (RTs faster than 120ms, 0.09%), errors (1.9%), and outliers, defined as RTs below or above 2 standard deviations from the individual's mean (4.47%), were not included in the analysis. A 2x2 ANOVA with congruency of the vibro-tactile distractor (congruent vs. incongruent) and type of trial (conditioning vs. expectation) was run on log-transformed RTs. Simple effect

analyses were run using two-tailed paired sample t-tests, when necessary, using Bonferroni correction method. As we were interested in testing the difference between congruent and incongruent trials in conditioning and expectation trials, alpha value was divided by 2 (p< .025). The analysis revealed a main effect of congruency [$F_{(1,29)}$= 62.3; p< .001; ηp^2= .68], with faster RTs to congruent (5.77log(ms) ± .16log(ms)) than incongruent trials (5.82log(ms) ± .14log(ms)). The main effect was further qualified by the two-way interaction [$F_{(1,29)}$= 36.8; p< .001; ηp^2= .55; see Figure 1, panel B]. Simple effect analyses revealed faster RTs to congruent than incongruent trials in both the conditioning ($t_{(29)}$= 8.5; p< .001; CI [0.06 0.09]) and the expectation trials ($t_{(29)}$= 3.4; p= .002; CI [0.008 0.03], See Table 1 and Figure 1, panel B). The main effect of type of trial was not significant [$F_{(1,29)}$= 0.05 ; p= .86; ηp^2= .06].

Figure 1

2.3. Study 2

Previous research has shown that pitch height biases spatial attention (Chiou & Rich, 2012; Spence & Deroy, 2013), in that high-frequency sounds bias spatial attention towards the upper visual field, while low frequency sounds bias spatial attention towards the lower visual field. Consequently, in Study 1 the auditory cue itself might have lead to a CCE, by virtue of an association between pitch height and spatial location (elevation) of the visual target. To rule out this alternative explanation, in study 2 the contingency between pitch and vibro-tactile distractors was reversed.

2.3.1. Participants, Stimuli and Procedure

The sample size was defined as per study 1. Thirty right-handed healthy volunteers (mean age = 21.2 years old, 20 female) were recruited by advertisement to take part in this study from the student pool. Stimuli and procedures were the same as in study 1. However, in this study the contingency between tones and vibro-tactile distractors was reversed, in that, high-frequency sounds predicted the forthcoming thumb stimulation, while low-frequency sounds predicted the forthcoming index stimulation (See figure 2, panel A).

2.3.2. Results

Shapiro-Wilk's test and visual inspection of the data showed that data violated the assumption of normality, thus data were log-transformed. Missed responses (0.10%), anticipatory responses (RTs faster than 120ms, 0.07%), errors (2.2%), and outliers, defined as above, were not included in the analysis (4.6%). A 2x2 ANOVA with congruency of the vibro-tactile distractor (congruent vs. incongruent) and type of trial (conditioning vs. expectation) was run on log-transformed RTs. Simple effect analyses were run as for study 1 (Bonferroni corrected alpha value; $p < .025$). The ANOVA revealed a main effect of congruency [$F(1,29) = 141.9$ $p < .001$; $\eta p^2 = .83$], with faster RTs to congruent ($5.85\log(ms) \pm .17\log(ms)$) than incongruent trials ($5.90\log(ms) \pm .16\log(ms)$). The main effect was further qualified by the two-way interaction [$F(1,29) = 38.7$; $p < .001$; $\eta p^2 = .57$; see Figure 2, panel B]. Simple effect analyses revealed faster RTs to congruent than incongruent trials in both the conditioning ($t(29) = 10.3$; $p < .001$; CI [0.07 0.11]) and the expectations trials ($t(29) = 2.4$; $p = .02$; CI [0.002 0.02]; see Table 1). The main effect of type of trial was not significant [$F(1,29) = 0.09$; $p = .76$; $\eta p^2 = .06$].

Overall, the results confirmed the findings of the first study, in that they showed the crossmodal congruency effect in both the conditioning and expectation trials, regardless of stimulus contingencies.

Figure 2

2.4. Study 3

Results from Study 2, in which we reverted the contingencies between the cue and the distractor, suggest that it is not the auditory cue itself to lead to a CCE, at least with the timing we used. Study 3 was conducted to further corroborate this finding.

Participants

The sample size was determined as above. Thirty right-handed healthy volunteers (mean age = 22.5 years old, 22 female) were recruited by advertisement to take part in this study from the student pool.

2.4.1. Stimuli and Procedure

The apparatus and stimuli were as in studies 1 and 2. The design was identical to studies 1 and 2, with the following exceptions: the vibro-tactile distractor was never presented.

Each trial started with an auditory stimulus (200ms), consisting of either a low-frequency (375 Hz) or high-frequency (1000 Hz) tone. The auditory stimulus was followed, after 530ms, by a visual stimulus. Hence, the auditory stimulus acted as a distractor of the visual event (Figure 3, panel A). ITI ranged between 1800 and 2200ms in 100ms steps.

The participants held the foam block and responded, as in the previous study, to the elevation of the visual target, trying to ignore the auditory stimulus. The sound did not predict the target location. The study was made up of 216 trials. Ninety-six trials were congruent (high-frequency sound/upper LED or low-frequency sound/lower LED), while 96 trials were incongruent (high-frequency sound/lower LED or low-frequency sound/upper LED). Twenty-four were catch trials in which the visual stimulus was not presented. Trials were randomized across participants.

2.4.2. Results

Shapiro-Wilk's test and visual inspection of the data showed that data violated the assumption of normality, thus data were log-transformed. Missed responses (0.01%), anticipatory responses (RTs faster than 120ms, 0.07%), errors (4.4%) and outliers, defined as above, were not included in the analysis (1.5%).

Reaction times to congruent and incongruent trials were analysed using a paired-sample t-test. The comparison was not significant (t(29)= 1.61; p= .11; CI [-0.002 0.02] see Figure 3, panel B). This result suggests that with the timing we used, that is, a delay of 530ms between the two stimuli,

auditory events do not impact on correct localization of visual target (Congruent: 5.97 ± .16log(ms); Incongruent: 5.98 ± .17log(ms)). To further investigate this null effect, a default Bayes factor paired-sample *t*-test was conducted. The data were 1.4 most likely to occur under the null effect model relative to the alternative effect model, $BF_{01}= 1.4$, and were more in favour of the null model with increasing Cauchy prior width (which puts more emphasis on bigger effects): with wide prior, $BF_{01}= 1.8$ and ultrawide prior $BF_{01}= 2.4$. Hence, we found only anecdotal evidence to support the null effect model; evidence that is clearly inconclusive.

Figure 3

2.5. Study 4

Study 3 might seems at odds with previous research showing that crossmodal congruency effect for audio-visual pairs does exist (Bernstein & Edelstein, 1971). Already in the early seventies Bernstein and Edelstein (Bernstein & Edelstein, 1971) demonstrated how irrelevant high frequency binaural tones facilitate detection of upper visual targets, while low frequency binaural tones facilitate detection of lower visual targets with regard to a fixation. In their study, however, the temporal offset between the auditory and the visual target was much shorter (range 0 - 45ms) than in our study (530ms). This suggests that the temporal offset between the auditory and the visual target might play a crucial role in the audio-visual crossmodal congruency effect. To test this hypothesis, in study 4, we investigated the time course of the audio-visual crossmodal congruency effect using delays ranging from 0 to 750ms.

Participants were to judge the elevation of a visual target, while trying to ignore a task irrelevant high frequency or low frequency sound. The temporal offset between the two stimuli ranged between 0 and 750ms in 250ms steps.

2.5.1. Participants

Twenty-five right-handed healthy volunteers (mean age = 21 years old, 21 female) were recruited by advertisement to take part in this study from the student pool.

2.5.2. Stimuli and Procedure

The apparatus and stimuli were as in study 3. The experimental design was a 2x4 factorial. The two within-subjects factors were the congruency of the auditory distractor with respect to the elevation of the visual target (congruent vs. incongruent) and the delay (0, 250, 500 and 750ms, see Figure 4, panel A). ITI ranged between 1800 and 2200ms in 100ms steps. The participants held the foam block and responded, as in the previous studies to the elevation of the visual target, trying to ignore the auditory stimulus. Each experimental condition was repeated 32 times for a total of 256 experimental trials. Fifty catch trials were also included in which the visual stimulus was not presented. Trials were randomized across participants.

2.5.3. Results

Shapiro-Wilk's test and visual inspection of the data showed that data violated the assumption of normality, thus data were log-transformed. One participant was excluded due to few valid trials (less than 60%). There were not missed responses. Anticipatory responses (RTs faster than 120ms, 0.02%), errors (0.9%) and outliers (1.9%) computed as in the previous studies were not included in the analysis. A 2x4 ANOVA with congruency (congruent vs incongruent) and delay (0, 250, 500, 750ms) was performed on log transformed RTs. The analysis revealed a main effect of delay [$F_{(3,69)}= 6.3$; $p< .001$, $\eta p^2= 0.21$]. This main effect was further qualified by the interaction between congruency and delay [$F_{(3,69)}= 9.4$; $p< .001$, $\eta p^2= 0.29$, see Figure 4, panel B]. Simple effects analyses (Bonferroni correction method; $p< .012$), showed that the crossmodal congruency effect was significant only in the 250ms condition (Congruent: $5.83 \pm .18$log(ms); Incongruent: $5.88 \pm .20$log(ms) $t_{(23)}= 4.2$; $p< .001$; CI [0.02 0.06]; see Figure 4, panel B). The same effect was absent, or did not survive correction for multiple comparisons, in the 0ms delay condition (Congruent: $5.90 \pm .17$log(ms); Incongruent: $5.89 \pm .18$log(ms); $t_{(23)}= 1.1$; $p= .28$; CI [-0.03 0.01]) and 750ms delay conditions (Congruent: $5.89 \pm .17$log(ms); Incongruent: $5.91 \pm .17$log(ms); $t_{(23)}= 2.4$; $p= .024$; CI [0.003 0.04]). An opposite trend was observed at 500ms delay condition with faster reaction times

to incongruent than congruent trials (Congruent: 5.88 ± .19log(ms); Incongruent: 5.85 ± .17log(ms); t(23)= 2.6; p= .016; CI [-0.05 -0.005]; see Figure 4, panel B). However, this effect did not survive correction for multiple comparisons. This was further supported by the Bayes factor analysis. We found a moderate support for the effect in the opposite direction relative to the null effect model, BF_{10}= 3.0. When we restricted the prior probability mass of this opposite direction, a Bayes factor supported strongly the null effect, since BF_{0+}= 14.2 and increased with wide and ultrawide priors, BF_{0+}= 20.2 and BF_{0+}= 28.7, respectively. Taken together with the results of the study 3, there is relative evidence supporting the null effect using delays longer than 250ms.

Figure 4

2.6. Study 5

Study 1 (and 2) suggested that the crossmodal congruency effect can be observed even when the distractor is not actually presented but merely expected. To test for the consistency of the effect and to rule out the possibility that the observed effect is bound to the used sensory modalities, we ran a first conceptual replication of study 1, with a different stimulus combination. We inverted the sensory modalities of the distractor and the target – i.e., a visual stimulus was used as a distractor and participants were to respond to a vibro-tactile target. The temporal structure of the study was left unchanged.

2.6.1. Participants and Stimuli

Thirty right-handed volunteers (mean age = 22.7 years old, 24 female) were recruited by advertisement to take part in this study from the student pool. The sample size was determined based on a priori power analysis of the critical crossmodal congruency effect in expectancy trials averaged across the two studies (Studies 1 and 2), dz = -0.53, while assuming α = 0.05, 1 – β = 0.80, and a two-tailed significance test for a matched-sample t-test.

2.6.2. Procedure

Procedure was the same as for study 1 except for the following: participants were required to judge the elevation of the vibro-tactile stimulus instead of the visual stimulus. Hence, in this case, the visual stimulus acted as a distractor. Consequently, the contingency was created between the auditory stimulus and the visual stimulus. Each experimental block was composed of 192 (80%) *conditioning trials* in which the auditory cue, the visual distractor and the vibro-tactile target were presented, and 48 (20%) *expectation trials* in which the auditory cue and the vibro-tactile target were presented, while the visual distractor was expected but actually omitted. Each block also included 24 *catch trials* in which only the auditory cue and the visual distractor were presented. Trials were randomized between blocks. In total each block was made up of 264 trials. The experimental session lasted approximately 45 minutes and consisted of two experimental blocks for a total of 528 trials (See figure 5, panel A).

Each *conditioning trial* started with an auditory cue (200ms), consisting of either a low-frequency (375 Hz) or a high-frequency (1000 Hz) tone, which predicted the presentation of the lower or the upper visual stimulus with 100% validity. The visual stimulus was followed, after 30ms, by the vibro-tactile target. Hence, the visual stimulus acted as a distractor of the vibro-tactile event. Half of the trials were congruent (96, upper LED/Index Finger or lower LED/Thumb); the other half was incongruent (96, Upper LED/thumb, or lower LED/index finger). ITI ranged between 1800 and 2200ms in 100ms steps.

The structure of *expectation trials* was identical to the structure of *conditioning trials* with the exception of the visual distractor, which was omitted. Half of the expectation trials were congruent (24, expected upper LED/ index finger or expected lower LED/ thumb); the other half was incongruent (24, expected upper LED/ thumb or expected lower LED/ index finger, See figure 5, panel A).

Responses were provided as per study 1. The experimental design was a 2x2 factorial. The two within-subjects factors were the congruency of the visual distractor with respect to the elevation of

the vibro-tactile target (congruent vs. incongruent) and the type of trial (conditioning vs. expectation). The crossing of these factors yielded 4 possible conditions, i) congruent – conditioning, ii) congruent – expectation, iii) incongruent – conditioning, iv) incongruent – expectation, that were included in a 2x2 ANOVA.

2.6.1. Results

Shapiro-Wilk's test and visual inspection of the data showed that data violated the assumption of normality, thus data were log-transformed. Missed responses (1.3%), anticipatory responses (RTs faster than 120ms, 0.72%), errors (4.9%), and outliers (4.09%), defined as above, were not included in the analysis. A within-subjects 2x2 ANOVA with congruency of the visual distractor (congruent vs. incongruent) and type of trial (conditioning vs. expectation) was run on log-transformed RTs. The analysis revealed a main effect of congruency [$F_{(1,29)}$= 43.8; p< .001; ηp^2= .60], with faster RTs to congruent (5.68log(ms) ± .27log(ms)) than incongruent trials (5.78log(ms) ± .26log(ms)). The main effect was further qualified by the two-way interaction [$F_{(1,29)}$= 22.4; p< .001; ηp^2= .43; see Figure 5, panel B]. Simple effect analyses revealed faster RTs to congruent than incongruent trials in both the conditioning ($t_{(29)}$= 6.0; p< .001; CI [0.11 0.22]) and the expectation trials ($t_{(29)}$= 2.9; p= .007; CI [0.008 0.04], See Table 1 and Figure 5, panel B). The main effect of type of trial was not significant [$F_{(1,29)}$= 0.03 ; p= .86; ηp^2= 0.06]. Results from study 5 suggest that the effect of expectation on crossmodal interaction observed in study 1 can be generalized to other sensory modalities of the distractor and the target.

Figure 5

2.7. Study 6

Study 6 was run as a second conceptual replication of study 1 with a different stimulus combination. We used visual, rather than auditory cues. Hence, we had visual, visual and tactile stimuli as cue, distractor and target, respectively. The temporal structure of the study was left unchanged.

2.7.1. Participants

Sample size was determined as for study 5. Thirty right-handed healthy volunteers (mean age = 24.2 years old, 20 female) were recruited by advertisement to take part in this study from the student pool.

2.7.2. Stimuli and procedure

The procedure was the same as in study 5. The only difference was the modality of the cue stimulus. The auditory cue was replaced by a visual stimulus (green LED, see figure 6, panel A). The visual stimuli used as cue (green LED) and target (red LED) were actually the same bi-colour LED.

2.7.3. Results

Shapiro-Wilk's test and visual inspection of the data showed that data violated the assumption of normality, thus data were log-transformed. Missed responses (2.11%), anticipatory responses (RTs faster than 120ms, 0.35%), errors (4.4%), and outliers (4.22%), defined as above, were not included in the analysis. A within-subjects 2x2 ANOVA with congruency of the visual distractor (congruent vs. incongruent) and type of trial (conditioning vs. expectation) was run on log transformed RTs. The analysis revealed a main effect of congruency [$F(1,29)= 48.5$; $p< .001$; $\eta p^2= .62$], with faster RTs to congruent (5.73log(ms) ± .23log(ms)) than incongruent trials (5.83log(ms) ± .24log(ms)). The main effect was further qualified by the two-way interaction [$F(1,29)= 15.75$; $p< .001$; $\eta p^2= 0.35$; see Figure 6, panel B]. Simple effect analyses with Bonferroni correction method revealed faster RTs to congruent than incongruent trials in both the conditioning ($t(29)= 6.41$; $p< .001$; CI [0.10 0.20]) and the expectation trials ($t(29)= 3.09$; $p= .004$; CI [0.015 0.07], See Table 1 and Figure 6, panel B). The main effect of type of trial was not significant [$F(1,29)= 0.06$; $p= .81$; $\eta p^2= 0.06$].

Figure 6

2.8. Study 7

This study tested whether the *spatial* predictability of the distractor is necessary to induce the crossmodal congruency effect, when the distractor is not actually presented. To this aim, we made the distractor "spatially" unpredictable by using auditory cues with a validity of 50%. The cue, the distractor and the target were auditory, visual and tactile, respectively, as in study 5.

2.8.1. Participants

Sample size was defined as for studies 5 and 6. Thirty right-handed volunteers (mean age = 21.8 years old, 28 female) were recruited by advertisement to take part in this study from the student pool.

2.8.2. Stimuli and procedure

Procedure was the same as in study 5. The only difference was that the auditory stimulus predicted the location of the distractor only 50% of the times. Each experimental block was composed of 192 (80%) *conditioning trials* in which auditory cue, visual distractor and the vibro-tactile target were presented. In half of these trials (96 trials) the auditory cue predicted the location of the vibro-tactile distractor, in the other half (96 trial) the auditory cue did not predict the location of the vibro-tactile distractor. The *expectation trials* (20%) and the 24 *catch trials* were the same as before. In brief, the auditory cue was not effective in inducing expectation of the subsequent visual distractor (See figure 7, panel A). It should be noted here that we labelled the trials as *conditioning* and *expectation* even if no expectation is induced, because the validity of the cue was at chance level. Similarly, as the non-existent expectation can neither be congruent nor incongruent with the target, the factor congruency referred to the spatial correspondence between the cue and the target.

Results

Shapiro-Wilk's test and visual inspection of the data showed that data violated the assumption of normality, thus data were log-transformed. Missed responses (0.56%), anticipatory responses (RTs faster than 120ms, 0.56%), errors (3.5%), and outliers (4.14%), defined as above, were not included in the analysis. A within-subjects 2x2 ANOVA with congruency of the visual distractor (congruent vs. incongruent) and type of trial (conditioning vs. expectation) was run on log transformed RTs. The analysis revealed a main effect of congruency [$F(1,29)= 72.8$; $p< .001$; $\eta p^2= .71$], with faster RTs to congruent (5.65log(ms) ± .24log(ms)) than incongruent trials (5.76log(ms) ± .27log(ms)). The main effect was further qualified by the two-way interaction [$F(1,29)= 58.4$; $p< .001$; $\eta p^2= .66$; see Figure 7, panel B], owing to faster RTs to congruent than incongruent trials in the conditioning trials ($t(29)= 8.54$; $p< .001$; CI [0.15 0.25]), but not in the expectations trials ($t(29)= 1.10$; $p= .27$; CI [-0.007 0.02]; see Table 1 and Figure 7, panel B). The main effect of type of trial was not significant [$F(1,29)= 2.7$; $p= .11$; $\eta p^2= 0.35$].

We further quantified the evidence for support of the interaction effect model relative to null effect model. We ran a Bayesian repeated measures ANOVA with the two main factors and their interaction. The null model comprised an intercept, congruency, and presence of the signal as nuisance terms (we use the term "nuisance" in a more generic way as used in the Bayesian statistics literature, e.g., Wagenmakers et al, 2017). The null model was compared with the model including in addition the interaction term between the two main factors. Thus, we compared two models, featuring the interaction or not, while all the other factors were included. We found extreme evidence to support the interaction model against the null model assuming no interaction effect, $BF_{01}= 3.8*10^9$. To unpack the interaction, we conducted two default Bayesian paired samples *t*-tests. We found anecdotal evidence supporting the model assuming no congruency effect relative to the model assuming the congruency effect in expectation trials, $BF_{01} = 2.9$ (for Cauchy scale 0.707, and moderate evidence, $BF_{01} = 3.9$ and $BF_{01} = 5.4$, for wide and ultrawide priors, respectively), and extreme evidence supporting the model assuming congruency effect relative to the null effect model

in conditioning trials, $BF_{10} = 5.8*10^6$ for (Cauchy scale 0.707, and extreme evidence, $BF_{10} = 6.9*10^6$ and $BF_{10} = 7.4*10^6$, for wide and ultrawide priors, respectively).

Figure 7

2.9. **Study 8**

This study was designed to support the finding that the actual expectation of the distractor, which is supposed to occur close in time to the target stimulus (30ms), is causing the CCE. If the CCE would occur due to the cue stimulus alone, as the cue itself may become a distractor stimulus by association over the course of the experiment, then the CCE should persist over the cue-to-target interval (>530ms). If the CCE, on the other hand, depends upon expectation of the distractor, it must be only short-lived, as expectation quickly dissipates over time. To this aim, in Study 8, in addition to the "expectation trials" (expected distractor-to-target interval = 30ms), we also included "expectation trials with delayed target" (expected distractor-to-target interval = 530ms). The cue, the distractor and the target were auditory, visual and tactile, respectively, as in study 5.

2.9.1. Participants

Sample size was defined as for the previous studies. Thirty right-handed volunteers (mean age = 26.3 years old, 18 female) were recruited by advertisement to take part in this study from the student pool.

2.9.2. Stimuli and procedure

Each experimental block was composed of 192 *conditioning trials* in which an auditory cue, a visual distractor and a vibro-tactile target were presented, and 48 *expectation trials* in which the auditory cue and the vibro-tactile target were presented, while the visual distractor was expected but actually omitted. Conditioning and expectation trials were the same as in the previous studies. Each block also included 48 *expectation trials with delayed target,* in which, the auditory cue and the vibro-tactile target were presented, while the visual distractor was expected but actually omitted.

Most importantly, in these trials the delay between the cue and the target stimulus was 1030 milliseconds instead of 530ms. As for the previous studies, trials could be congruent or incongruent. 24 *catch trials* were also included, in which only the auditory cue and the visual distractor were presented. Trials were randomized between blocks, and responses were provided as in the previous studies.

2.10.   Results

Shapiro-Wilk's test and visual inspection of the data showed that data violated the assumption of normality, thus data were log-transformed. Missed responses (0.19%), anticipatory responses (RTs faster than 120ms, 0.83%), errors (2.1%), and outliers (6.3%), defined as above, were not included in the analysis. A within-subjects 2x3 ANOVA with congruency of the visual distractor (congruent vs. incongruent) and type of trial (conditioning vs. expectation vs. delayed) was run on log transformed RTs. The analysis revealed a main effect of congruency [$F(1,29)= 138.6$; $p< .001$; $\eta p^2= 0.82$], and type of trials [$F(1,29)= 30.01$; $p< .001$; $\eta p^2= 0.50$]. These main effects were further qualified by the two-way interaction between type of trials and congruency [$F(1,58)= 47.7$; $p< .001$; $\eta p^2= 0.62$; see Figure 8, panel B]. Simple effect analyses revealed faster RTs to congruent (5.67log(ms) ± .28log(ms)) than incongruent trials in conditioning trials (5.85log(ms) ± .27log(ms); $t(29)= 9.15$; $p< .001$; CI [0.14 0.22]). The CCE was observed also in the expectation trials ($t(29)= 7.16$; $p< .001$; CI [0.026 0.047]; Congruent: 5.72log(ms) ± .24log(ms); Incongruent: 5.76log(ms) ± .25log(ms)), but not in the expectation trials with delayed target ($t(29)= 0.069$; $p= .945$; CI [-0.01 0.017]; Congruent: 5.84log(ms) ± .26log(ms); Incongruent: 5.84log(ms) ± .26log(ms)).

These findings were further supported by the Bayes factor analysis. First, we conducted Bayesian repeated measures ANOVA, which yielded extreme evidence to support the model including the interaction between congruency and type of trial relative to the null model including intercept and the two main effects, $BF_{10} = 9.3*10^{10}$. Second, to unpack this interaction a set of default Bayesian paired-samples t-tests (i.e., Cauchy scale of 0.707) was conducted. We found extreme evidence to

support the model assuming congruency effect relative to the null effect model in the conditioning

trials, BF_{10} = 212003 as well as in the expectation type of trials, BF_{10} = $2.3*10^7$. In contrast, but

aligned with our predictions, we found moderate evidence to support the null effect model relative

to the model assuming congruency effect in the expectation trials with delayed target, BF_{01} = 5.1.

Figure 8

3. Evidence synthesis and robustness of the crossmodal congruency effect in expectation trials

We synthesized evidence concerning the critical crossmodal congruency effect for expectation trials

across studies, which share the same design structure, namely 1, 2, 5 and 6 using an internal small

scale meta-analysis and Bayes factor meta-analysis. The purpose of the former analysis was to

provide more precise estimate of the effect and demonstrate the effect with increased power

(Cumming, 2014); the rationale behind conducting the latter analysis was to offer quantified

evidence to relative support for the null and alternative hypothesis (Rouder & Morey, 2011).

3.1. Evidence synthesis

To synthesize the evidence we used two methods. First, we have run a random-effects meta-

analytical model using restricted maximum likelihood estimation method to estimate standardized

change scores for the crossmodal congruency effect in expectation trials across the studies which

share the same design structure (Experiment 1, 2, 5, 6; R package metaphor was used, Viechtbauer,

2010). The studies were very similar with almost no heterogeneity, τ^2= 0, *SE*= 0.03, Cochran's

Q(3)= 0.35, *p*= .950. Critically, the overall meta-analytical effect was -0.53, 95% CI[-0.72, -0.34];

the effect was statically significant, *z*= -5.41, *p*< .001 (Figure 9).

Second, we quantified the relative evidence of the null model (assuming no effect) and alternative

models (assuming existence of the effect with different distribution of its values) using a default

Bayes factor meta-analysis with three different priors (R package BayesFactor was used, Morey &

Rouder, 2015). Bayes factor re-analyses of the individual studies provided evidence supporting

existence of the effect (using a default prior, r scale = 0.707) relative to the null model but of different strength: study 1: BF_{10} = 18.5; study 2: BF_{10} = 2.8; study 5: BF_{10} = 9.5; study 6: BF_{10} = 5.8. The meta-analysis of these data provided *extreme evidence* to support models assumed by the alternative hypothesis supporting existence of the crossmodal congruency effect relative to the null model, BF_{10}= 428*10^3 (i.e., using a wide prior of r scale= 0.707). This means that the data under the assumption of the alternative hypothesis model are around 428 thousand more likely than under the assumption of the model derived from the null hypothesis. Sensitivity analysis with different priors reached the same conclusion: BF_{10}= 446*10^3 (medium prior, r scale= .5), and BF_{10}= 370*10^3 (ultra-wide prior, r scale= 1). To conclude, both approaches yielded evidence supporting existence of the effect; we estimated the effect to be of a medium size.

Figure 9

3.2. Exclusion sensitivity

We also tested robustness of our conclusion for the critical crossmodal congruency effect in expectation trials in terms of their sensitivity to exclusion criteria (recall, a-priori set-up 2 SD rule was applied here). Specifically, we tested whether our conclusions about the effect were sensitive to the trimming that we have applied consistently across the studies 1, 2, 5 and 6, which share the same design structure, when aggregating the trials for each participant within each condition. To do so, we first excluded the incorrect trials and anticipatory responses, and we transformed logarithmically the trial-level data; then we set and applied a-priori trimming levels: no trimming (0%), 2.5% trimming, 5% trimming, 7.5% trimming, 10% trimming values. For instance, 5% trimming remove 5% most extreme values from lower and upper end of the sample of trial data (within condition and participant) and then the mean was calculated. Then we aggregated the data for each participant in each study using mean reaction times and re-tested the crossmodal congruency effect for expectation trials using an internal meta-analysis with the same setting as applied above (Table 2). In other words, to aggregate the data we have used a trimmed mean of

different levels – a standard tool of robust statistics – that can boost precision, power and control for Type I error (Erceg-Hurn & Mirosevich, 2008). We can see that the point estimates of the meta-analytical effects for the six various methods of trimming are medium sized and virtually identical (ranging from -0.44 to -0.53); they vary only slightly between each other. With the increasing level of trimming, the 95% confidence intervals are narrowing slightly but the gain in precision is not dramatic. To conclude, the trimming method did not affect our conclusion about the existence of the effect (i.e., 95% confidence intervals of meta-analytical effects do not contain zero) neither about the effect size of the effect. The crossmodal congruency effect in expectation trials is robust.

Table 2

4. Discussion

Endorsing the view that the brain is essentially a predictive machine (Friston, 2010), we hypothesized that the congruency effect between visual and tactile events might occur even when a tactile event is expected, but actually omitted. In Study 1, participants held a foam block with the right hand that housed one LED on the top and one LED on the bottom. Participants responded to the elevation (up or down) of the visual target while tactile distractors were presented at the same or opposite elevation as the visual stimulus. Despite the instruction to ignore tactile distractors, and in agreement with previous findings (for a review: (Spence & Deroy, 2013) participants' responses were faster when tactile distractors occurred at the same elevation as visual target (congruent trials) than when distractors occur at opposite elevations (incongruent trials). Difference in performance between congruent and incongruent trials is known as the congruency effect (CCE). Importantly, we found a CCE when the tactile distractor was cued by an auditory stimulus, but actually omitted. Hence, our data suggest that expectation is capable of leading to CCE. It should be noted that we used a modified version of the classic CCE. In particular, we used only one foam cube instead of two (Spence & Driver, 1996). Such simplification could explain why the magnitude of the CCE we observed is less than half compared to what found before (Spence & Deroy, 2013).

Possibly, one might argue that our results could be explained by the well-known, naturally occurring spatial mapping between pitch elevation and elevation of visual targets. In that, high and low tones induce attention shifts to upper or lower locations, depending on the pitch height (Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Chiou & Rich, 2012; Evans & Treisman, 2010). Indeed, research has shown that responding to an upper (versus lower) visual stimulus is faster when the same stimulus is preceded or accompanied by a high (versus low) tone, compared to when it is preceded or accompanied by a low tone (Ben-Artzi & Marks, 1995; Bernstein & Edelstein, 1971; Chiou & Rich, 2012; Evans & Treisman, 2010). Therefore, the spatial mapping could have produced the congruency effect we observed in study 1, when the tactile stimulus was omitted. To rule out this alternative hypothesis we ran a second study in which the contingency between the auditory cue and the vibro-tactile distractor was reversed. Also in this second study we found an effect of tactile expectation on the CCE. In a third study, we investigated whether the CCE we observed in studies 1 and 2 could be explained by the congruency between the pitch of auditory cue and the elevation of the visual target (Bernstein & Edelstein, 1971; Chiou & Rich, 2012; Spence, 2011), regardless of the expectation of the distractor. We failed to find any significant effect. Results from study 3 might still seem surprising, as the spatial mapping has been replicated in a plethora of studies (for review see (Chiou & Rich, 2012; Spence & Deroy, 2013). However, all these studies have used fixed asynchronies in the range of few hundred milliseconds, often not larger than 400 ms. It is indeed believed that such effect dissipates quickly (Spence, 2010). So far, only few studies have successfully replicated the audio-visual spatial mapping using SOA larger than 400 ms (Chiou & Rich, 2012; Fernández-Prieto, Vera-Constán, García-Morera, & Navarra, 2012). However, in these studies the authors used a simple detection task rather than a discrimination task, and a relatively small sample size. Hence, such methodological differences might account for by the different results. In a fourth study we investigated whether the lack of significant effects in study 3 could be explained by the temporal relationship between auditory and visual stimuli. To investigate this issue, we manipulated the delay between the auditory and the

visual target from 0 to 750 ms in 250 ms steps. Results revealed a congruency effect only when the delay between the two stimuli was 250 ms. The other intervals did not produce any CCE, or did not survive correction for multiple comparisons. The null effects in studies 3 and 4, which suggest that the expectation of the distractor is crucial for the CCE observed in study 1 and 2, have been supported by Bayesian analyses. Furthermore, studies 5 and 6, in which different stimulus combinations were used, conceptually replicated the findings of studies 1 and 2, thus, suggesting that our results can be generalized to other sensory modalities of the cue, distractor and target. Finally, studies 7 and 8 demonstrated the relevance of spatial and temporal regularities in the induction of expectation. Specifically, Study 8 revealed that the CCE ceases if the time between expected distractor and target is too long (530 ms vs 30 ms). This interpretation of our findings is further supported by prior results from Shore and colleagues (2006). They investigated the temporal dynamics of the general (visuo-tactile) CCE and showed that the cost for incongruent visual distractors that preceded tactile targets did not significantly differ from the baseline at SOAs longer than 100 ms. Similarly, Poole and colleagues (Poole, Couth, Gowen, Warren, & Poliakoff, 2015; Poole, Gowen, Warren, & Poliakoff, 2015) reported that tactile distractors presented nearly simultaneously (i.e. <100ms) to visual target produced a significantly larger congruency effect than larger stimulus onset asynchronies. This evidence suggests that the null effect we observed in study 8 occurred because the time between cue and target, rather than the time between expected distractor and target, was too long for a CCE to occur.

To further support our results we have conducted an internal meta-analysis as recommended by the current practice (e.g., Cumming, 2014). The main conclusion of this analysis is that the crossmodal effect is statistical significant medium size effect -0.53, 95% CI [-0.72, -0.34]. Overall, our results demonstrate that CCE can be induced not only by real stimuli but also expected sensory events.

Before discussing the possible implications of our findings, we should mention an alternative interpretation of our results. One might argue that participants shifted spatial attention according to the location of the auditory cue (exogenous attention). Thus, even in trials without a tactile stimulus,

allocating spatial attention to the presumed location might induce the observed cuing effect. We believe this alternative interpretation is unlikely given the results of studies 3 and 4. If exogenous attentional mechanisms played a main role in our effect, we should have observed the same CCE also in studies 3 and 4.

But, how is it possible to integrate information when one stimulus is expected but actually omitted? A possible mechanism pertains the predictive processing in perception (Summerfield et al., 2006). As suggest by the predictive framework, processing of information does not occur through mere reaction to stimuli, rather we continuously create predictions about forthcoming sensory events (Friston, 2010). This holds true for both unimodal and crossmodal events. Theorizations of predictive processing are mainly grounded on Bayesian statistical inference (for reviews see (Brown & Brüne, 2012; Bubic, Von Cramon, & Schubotz, 2010). This method of inference can be used to determine the probability of a certain outcome, given a predetermined assumption, which can be subsequently updated according to the actual outcome.

Our findings complement and extend previous findings on the relationship between sensory encoding and expectation. For instance, Berger and Ehrsson (Berger & Ehrsson, 2013) found that imagination of a sound is enough to induce the cross-bounce illusion. The cross-bounce illusion is a crossmodal illusion, where sound affects vision. Two visual targets moving across each other can be perceived either to bounce off or to stream through each other. In 1997 Sekuler et al. (Sekuler, Sekuler, & Lau, 1997) demonstrated that a brief sound at the moment the targets coincide, biases perception toward bouncing. Berger and Ehrsson (Berger & Ehrsson, 2013) elegantly demonstrated that imagined sounds, at the moment the targets coincide, biases perception towards bouncing rather than streaming, as if the auditory stimulus was actually presented. Related to this, Spence and colleagues (Spence, Nicholls, & Driver, 2001) showed that sensory processing can be less efficient in an unexpected modality than in the same modality when it is expected. Hence, performance depends not only on what actually happens, but also on what is anticipated, expected or imagined.

Our results seem to support this view by showing crossmodal congruency effect between a visual target and an expected, but omitted distractor.

Concerning the specific neural mechanisms enabling the effect of an expected stimulus on a real stimulus, research in the neuroscience of expectation shows that the brain anticipates forthcoming events. For instance, Carlsson et al. (Carlsson, Petrovic, Skare, Petersson, & Ingvar, 2000) found that brain activation in response to the expectation of a tactile stimulus, as a tickle provided with a light touch of painter's brush on the foot sole, is similar to the engaged during actual somatosensory stimulation. More recently, Kok and colleagues (Kok et al., 2014), using functional magnetic resonance imaging, demonstrated that expectation of a visual stimulus induces an activation of the visual cortex which resembles the ones induced by the real stimulus. The mechanisms leading to the generation of such an accurate template seems to be related to low-frequency oscillatory activity of the brain.

Cravo and colleagues (Cravo, Rohenkohl, Wyart, & Nobre, 2013) asked participants to judge the orientation of Gabor patterns tilted clockwise or counterclockwise, which could be embedded within temporally regular or irregular streams of noise-patches. At the behavioral level, results revealed that expectation enhanced contrast sensitivity of visual targets. Indeed, participants were more accurate at discriminating the gabor patch embedded in predictable display rather than unpredictable display. At the neural level, results showed that the phase of delta oscillations overlying visual cortex was predictive of the quality of target processing only in regular streams of events. These results suggest that phase entrainment of low-frequency oscillations to external sensory cues can serve as an important and flexible mechanism for generating expectation. We speculate that a similar phase entrainment of low-frequency oscillations might account for our effect. This idea is supported by evidence showing that entrainment of low-frequency oscillations plays a pivotal role in sensory selection and multisensory integration (Lakatos, Chen, O'Connell,

Mills, & Schroeder, 2007; Schroeder & Lakatos, 2009; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Stefanics et al., 2010).

Overall, our findings enrich current knowledge on sensory expectation and crossmodal interaction by suggesting that our brain uses both actual and predicted stimuli to cope with overwhelming stimuli from the environment.

References

Amemori, K., & Sawaguchi, T. (2006). Contrasting effects of reward expectation on sensory and motor memories in primate prefrontal neurons. *Cereb Cortex, 16*(7), 1002-1015.

Ball, K., & Sekuler, R. (1981). Cues reduce direction uncertainty and enhance motion detection. *Percept Psychophys, 30*(2), 119-128.

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: role of stimulus difference. *Percept Psychophys, 57*(8), 1151-1162.

Berger, C. C., & Ehrsson, H. H. (2013). Mental imagery changes multisensory perception. *Curr Biol, 23*(14), 1367-1372.

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *J Exp Psychol, 87*(2), 241-247.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience, 6.*

Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience, 4.*

Carlsson, K., Petrovic, P., Skare, S., Petersson, K. M., & Ingvar, M. (2000). Tickling expectations: Neural processing in anticipation of a sensory stimulus. *Journal of Cognitive Neuroscience, 12*(4), 691-703.

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception, 41*(3), 339-353.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci, 36*(3), 181-204.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science, 248*(4962), 1556-1559.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *J Neurosci, 33*(9), 4002-4010.

Cumming, G. (2014). The New Statistics. *Psychological Science, 25*(1), 7-29.

den Ouden, H. E., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cereb Cortex, 19*(5), 1175-1185.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception and Psychophysics, 66*(8), 1388-1404.

Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philos Trans R Soc Lond B Biol Sci, 353*(1373), 1319-1331.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience, 2*(10), 704-716.

Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: an easy way to maximize the accuracy and power of your research. *American Psychologist, 63*(7), 591.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *J Vis, 10*(1), 6 1-12.

Fernández-Prieto, I., Vera-Constán, F., García-Morera, J., & Navarra, J. (2012). Spatial recoding of sound: Pitch-varying auditory cues modulate up/down visual spatial attention. *Seeing and Perceiving, 25*(0), 150-151.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci, 11*(2), 127-138.

Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *Neuroimage, 124*(Pt A), 876-886.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *J Exp Psychol Gen, 63*(3), 289-293.

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

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *J Cogn Neurosci, 26*(7), 1546-1554.

Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cereb Cortex, 22*(9), 2197-2206.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279-292.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*: Cambridge University Press.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Ly, A., Gronau, Q. F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2015). JASP (Version 0.7.5)[Computer software].

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research, 17*(2), 447-453.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: An R package for Bayesian data analysis. (Version 0.9.10-2).

Nahorna, O., Berthommier, F., & Schwartz, J. L. (2012). Binding and unbinding the auditory and visual streams in the McGurk effect. *J Acoust Soc Am, 132*(2), 1061-1077.

Nelson, W. T., Hettinger, L. J., Cunningham, J. A., Brickman, B. J., Haas, M. W., & McKinley, R. L. (1998). Effects of localized auditory information on visual target detection performance using a helmet-mounted display. *Human Factors, 40*(3), 452-460.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*(4), 437-442.

Poole, D., Couth, S., Gowen, E., Warren, P. A., & Poliakoff, E. (2015). Adapting the Crossmodal Congruency Task for Measuring the Limits of Visual-Tactile Interactions Within and Between Groups. *Multisens Res, 28*(3-4), 227-244.

Poole, D., Gowen, E., Warren, P. A., & Poliakoff, E. (2015). Investigating visual-tactile interactions over time and space in adults with autism. *J Autism Dev Disord, 45*(10), 3316-3326.

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Res, 1301*, 89-99.

Rouder, J., & Morey, R. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review, 18*(4), 682-689.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat Neurosci, 5*(7), 631-632.

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schroger, E. (2013). Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *J Neurosci, 33*(20), 8633-8639.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci, 32*(1), 9-18.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci, 12*(3), 106-113.

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*(6614), 308.

Siemann, J. K., Muller, C. L., Bamberger, G., Allison, J. D., Veenstra-VanderWeele, J., & Wallace, M. T. (2014). A novel behavioral paradigm to assess multisensory processing in mice. *Front Behav Neurosci, 8*, 456.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science, 22*(11), 1359-1366.

Spence, C. (2010). Crossmodal spatial attention. *Ann N Y Acad Sci, 1191*, 182-200.

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Atten Percept Psychophys, 73*(4), 971-995.

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Conscious Cogn, 22*(1), 245-260.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *J Exp Psychol Hum Percept Perform, 22*(4), 1005-1030.

Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Percept Psychophys, 63*(2), 330-336.

Spence, C., Pavani, F., & Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. *J Exp Psychol Hum Percept Perform, 26*(4), 1298-1319.

Spence, C., Pavani, F., & Driver, J. (2004). Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cogn Affect Behav Neurosci, 4*(2), 148-169.

Spence, C., & Walton, M. (2005). On the inability to ignore touch when responding to vision in the crossmodal congruency task. *Acta Psychol (Amst), 118*(1-2), 47-70.

Stefanics, G., Hangya, B., Hernadi, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J Neurosci, 30*(41), 13578-13585.

Stein, B. E., & Wallace, M. T. (1996). Comparisons of cross-modality integration in midbrain and cortex. *Progress in Brain Research, 112*, 289-299.

Stekelenburg, J. J., & Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Front Integr Neurosci, 6*, 26.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nat Rev Neurosci, 15*(11), 745-756.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive Codes for Forthcoming Perception in the Frontal Cortex. *Science, 314*(5803), 1311-1314.

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J Neurosci, 31*(25), 9118-9123.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1-48.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci U S A, 108*(51), 20754-20759.

Wilkinson, L. K., Meredith, M. A., & Stein, B. E. (1996). The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Experimental Brain Research, 112*(1), 1-10.

Figure captions

Figure 1: Study 1. Panel A) Temporal structure of conditioning trials (upper raw) and expectation trials (lower raw). The note represents the auditory cue (validity 100%), the red lightening bolt represents the actual vibro-tactile stimulus, the grey lightening bolt represents the expected vibro-tactile stimulus, the red circle represents the visual target. ISI: Inter Stimulus Interval. Panel B): Mean reaction times for congruent and incongruent trials in expectation and conditioning trials. Error bars represent standard errors. * indicates p < .05.

Figure 2: Study 2. Panel A) Temporal structure of conditioning trials (upper raw) and expectation trials (lower raw). The note represents the auditory cue (validity 100%), the red lightening bolt represents the actual vibro-tactile distractor, the grey lightening bolt represents the expected vibro-tactile distractor, the red circle represents the visual target. ISI: Inter Stimulus Interval. Panel B): Mean reaction times for congruent and incongruent trials in expectation and conditioning trials. Error bars represent standard errors. * indicates p < .05.

Figure 3: Study 3. Panel A) Temporal structure of the trials. The note represents the auditory distractor, the red circle represents the visual target. ISI: Inter Stimulus Interval. Panel B): Mean reaction times for congruent and incongruent trials. Error bars represent standard errors. ns: non significant.

Figure 4: Study 4. Panel A) Temporal structure of the trials. The note represents the auditory distractor, the red circle represents the visual target. ISI: Inter Stimulus Interval. Panel B): Mean reaction times for congruent and incongruent trials. Error bars represent standard errors. * indicates p < .05. § indicates comparisons not surviving correction for multiple comparisons. ns: non significant.

Figure 5: Study 5. Panel A) Temporal structure of conditioning trials (upper raw) and expectation trials (lower raw). The note represents the auditory cue (validity 100%), the red circle represents the actual visual distractor, the grey circle represents the expected visual distractor, the red lightning bolt represents the tactile target. ISI: Inter Stimulus Interval. Panel B): Mean reaction times for congruent and incongruent trials in expectation and conditioning trials. Error bars represent standard errors. * indicates p < .05.

Figure 6: Study 6. Panel A) Temporal structure of conditioning trials (upper raw) and expectation trials (lower raw). The green circle represents the visual cue (validity 100%), the red circle represents the actual visual distractor, the grey circle represents the expected visual distractor, the red lightning bolt represents the tactile target. ISI: Inter Stimulus Interval. Panel B): Mean reaction times for congruent and incongruent trials in expectation and conditioning trials. Error bars represent standard errors. * indicates p < .05.

Figure 7: Study 7. Panel A) Temporal structure of conditioning trials (upper raw) and expectation trials (lower raw). The note represents the auditory cue (validity 50%), the red circle represents the actual visual distractor, the grey circle represents the expected visual distractor, the red lightning bolt represents the tactile target. ISI: Inter Stimulus Interval. Panel B): Mean reaction times for congruent and incongruent trials in expectation and conditioning trials. Error bars represent standard errors. * indicates p < .05. ns: non significant.

Figure 8: Study 7. Panel A) Temporal structure of conditioning trials (upper raw), expectation trials (middle raw), and expectation trials with delayed target (lower raw). The note represents the auditory cue (validity 100%), the red circle represents the actual visual distractor, the grey circle represents the expected visual distractor, the red lightning bolt represents the tactile target. ISI: Inter Stimulus Interval. Panel B): Mean reaction times for congruent and incongruent trials in expectation, conditioning, and expectation trials with delayed target. Error bars represent standard errors. * indicates p < .05. ns: non significant.

Figure 9. Meta-analysis. A small-scale random-effect meta-analysis of the crossmodalcongruency effect for expectation trials (i.e., simple effects). Note. Weigth – weight (inverse variance) of each study in the overall effect, ES – effect size, specifically here standardised mean change, 95% CI – 95% confidence intervals for standardised mean change

Table 1. Mean reaction times for congruent and incongruent trials in conditioning trials, expectation trials, and expectation trials with delayed target. t: t values. ** indicate $p < .001$; * indicate $p < .05$. Italics indicate non-significant values.

Table 2. The robustness of the crossmodalcongruency effect across different trimming procedures. Note. ES – standardized effect size (standardized mean change), 95% CI – 95% confidence intervals for standardized mean change