Title

Crossmodal Spatial Mappings as a Function of Online Relational Analyses?

Permalink

https://escholarship.org/uc/item/3bh2h6fv

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Zafirova, Yordanka
Petrova, Yolina
Petkov, Georgi

Publication Date

2019

Crossmodal Spatial Mappings as a Function of Online Relational Analyses?

Yordanka Zafirova (jovasileva@abv.bg)
Department of Cognitive Science and Psychology, New Bulgarian University,
21 Montevideo Street, 1618 Sofia, Bulgaria

Yolina Petrova (yolina.petrovaa@gmail.com)
Central and East European Center of Cognitive Science,
Department of Cognitive Science and Psychology, New Bulgarian University,
21 Montevideo Street, Sofia 1618, Bulgaria

Georgi Petkov (gpetkov@nbu.cogs.bg)
Central and East European Center of Cognitive Science,
Department of Cognitive Science and Psychology, New Bulgarian University,
21 Montevideo Street, Sofia 1618, Bulgaria

Abstract

Crossmodal correspondences are innate, language-based and statistically derived. They occur across all sensory systems and in different cultures. Despite their multiformity, they are exhibited analogously, mainly through robust congruency effects. One plausible explanation is that they rely on a common underlying mechanism, reflecting the fundamental ability to transfer relational patterns across different domains. We investigated the pitch-height correspondence in a bimodal sound-discrimination task, where the context of one relative sound pitch was changed online. The intermediate sound frequency was presented in successive blocks with lower *or* higher equidistant sounds and two squares at fixed up and down vertical positions. Congruency effects were transferred across sound contexts with ease. The results supported the assumption about the relational basis of the crossmodal associations. In addition, vertical congruency depended critically on the horizontal spatial representations of sound.

Keywords: crossmodal associations; relational mapping; pitch-height correspondence; SMARC effect

Introduction

Multidimensional information is integrated not only within the neural frame of one sensory system (e.g. Garner, 1974), but also across different modalities. Thus, certain features extracted from one perceptual realm interact with other, modality-specific attributes, and create coherent multimodal percepts, or intersensory Gestalts (for a review, see Spence, 2015). During the process of integration, particular aspects of the polysensory flow may modulate one another (like in McGurk & MacDonald, 1976, where visual stimuli modified auditory content, creating perceptual illusion), or bind together in bistable crossmodal entities with corresponding features. Examples of such corresponding features for pitch are shape (Melara & O'Brien, 1987; Walker et al., 2010); brightness (Marks, 1974; Martino &Marks, 1999; Melara, 1989); hue (Simpson, Quinn, Ausubel, 1956); smell (Belkin, Martin, Kemp, & Gilbert, 1997); size (Evans & Treisman, 2009; Mondloch & Maurer, 2004; Parise & Spence, 2012); height (Ben-Artzi & Marks, 1995; Chiou & Rich, 2012; Mudd, 1963; Patching & Quinlan, 2002; Rusconi, Kwan, Giordano, Umilta, & Butterworth, 2006) etc.

Apparently, invariable crossmodal associations might occur across all sensory systems. Examples of such associations are observed in different cultures (e.g. Bremner et al., 2013; Levitan et al., 2014; Parkinson, Kohler, Sievers, Wheatley, 2012; Wan et al., 2014) and are more or less implicit (Chen, Tanaka, Namatame, & Watanabe, 2016; Evans & Treisman, 2009; Parise & Spence, 2012). Given the broad scale and the diversity of the correspondence effect, at least three assumptions come to mind. First, these mappings should be nonrandom and in some way meaningful for perception. Second, they might comprise different processes and originate in different perceptual and cognitive networks. Third, they might be related to a common underlying mechanism, reflecting more general, inherent adaptive framework.

It was demonstrated that 4-month-olds are already sensitive to the associations between pitch and height, and pitch and sharpness (Walker et al., 2010). Lewkowicz and Turkewitz (1980) found evidence for mappings between brightness and loudness in infants 21 to 31 days of age. Along with that, crossmodal couplings are reported in nonhuman animals (see Ratcliffe, Taylor, & Reby, 2016, for a recent review). These findings are critical for the validity of the notion that multisensory associations are semantically mediated. It seems that they emerge on a lower, perceptual level and congruency effects are dependent on available attentional resources. More specifically, “attention is likely to play an important role in cross-modal perceptual organization” (Spence, 2015, p. 12). At the same time, it might be argued that even newborns have enough experience with environment, given the fact that they are extremely sensitive to certain statistical frequencies. In fact, 6-month-olds were better than adults in extracting implicit crossmodal information from the context (Rohlf, Habets, von Frieling, & Röder, 2017). Additionally, it was suggested that human perceptual system might foster the development of language by auxiliary crossmodal mappings of speech sounds to other percepts, for example shapes (Ozturk, Krehm, & Vouloumanos, 2013). The opposite is also true – linguistic experience systematically promotes progressive coupling of nonlinguistic dimensions (Martino & Marks, 1999). In short, the importance of language in crossmodal perception is undeniable. However, not all data can be explained with semantics (consider the animal studies). By all appearances, perceptual constraints, environment and language contribute synchronously to the establishment of stable crossmodal links. According to Spence (2011) certain structural correspondences might be innate, while statistical and semantic ones are evidently learned, and all are a function of environment.

A basic tool for exploring the nature and the strength of crossmodal associations is the speeded discrimination task, developed by Garner (1974), where participants respond to one task-relevant dimension while ignoring another, task-irrelevant dimension. Corresponding dimensions are integrated and trigger congruency effects that influence selective attention and performance. Apart from that, particular dimensions might interact on different levels. There is evidence that certain crossmodal couplings might actually prompt perceptual change (e.g. Evans & Treisman, 2009; McGurk & MacDonald, 1976). Others communicate on decisional level (e.g. Melara, 1989; Rusconi et al., 2006) or result from semantic inconsistencies (e.g. Martino & Marks, 1999). Notably, they can be correlated directly – for instance, psychophysical dimensions, like pitch and height usually covary in magnitude. On the other hand, it is generally accepted that these mappings are relative – one level of the first dimension can be mapped on different levels of the second dimension, depending on the context (for a review, see Spence, 2011). However, the mechanisms behind these relative mappings remain unclear.

In short, crossmodal correspondences engage all sensory systems, they are universal, innate, automatic, can be learned and assist learning, and interact on lower, bottom-up and higher, top-down levels. Considering the relative nature of the mappings, it is plausible to assume that they recruit the mechanisms of another, major cognitive process, reflecting the ability to build and compare relations. What is more, "the ability to pick out patterns, to identify recurrences of these patterns despite variation in the elements that compose them, to form concepts that abstract and reify these patterns, and to express these concepts in language" is considered a fundamental core of cognition (Holyoak, Gentner, & Kokinov, 2001, p. 2). Besides, relational analyses can be performed online and automatically, with or without utilizing attentional resources. Evidence is piling up that relations can be retrieved unconsciously and transferred across corresponding sets of data (Hristova, 2017; Li, Li, Zhang, Shi, & He, 2018). That being said, the capacity to construct and compare associations across modalities is one possible explanation for the pervasiveness of crossmodal correspondences. Thus, perceptual dimensions are represented not in their absolute values but as correlated dyads. To check this hypothesis, we investigated the congruency effects between one sound frequency and two vertical spatial positions – higher and lower. In other words, we measured the interaction between pitch and height in a speeded sound-discrimination task. Crucially, the context of the sound was changed during the task – it was presented with either higher or lower pitch, so that it was perceived as relatively lower or higher than the other sound. Previous studies demonstrated that pitch and height were positively correlated – higher frequencies were consistently associated with higher vertical positions (e.g. Ben-Artzi & Marks, 1995; Evans & Treisman, 2009; Rusconi et al., 2006 etc.). If crossmodal correspondences are indeed represented as relations, are we should expect comparable congruency effects in both contexts – i.e., one and the same sound should be mapped to different vertical positions in accordance with its relative frequency. Importantly, this shift should be effortless and almost instantaneous.

Experiment: Sound-Discrimination Task

Method

Participants 24 students (7 males) with mean age 23.8 years (standard deviation SD=6.3) from the Psychology Department were recruited for the task, after approval from the Cognitive Science and Psychology Ethics Committee. All signed the informed consent form and reported no problems performing the task.

Stimuli and design The stimuli were three sinusoidal sound waves with different frequencies, and a black square presented at two vertical positions. The sounds were generated on *Audacity* at 600 Hz, 900Hz and 1200Hz; 16 bits, mono, on an amplitude level of -2.5 dBFS (decibels relative to full scale) and duration 1000 ms. The square was solid black, 100x100 pixels (px) JPEG image. Participants had to perform a sound-discrimination task with bimodal presentation of the stimuli – i.e. the values of the task-relevant (sound pitch) and task-irrelevant (square height) dimensions were coupled randomly for each trial and presented simultaneously in both visual and auditory modalities.

Procedure The experiment was conducted in the presence of the experimenter in one of the booths of the Experimental Psychology Laboratory. Presentation and timing were controlled by the E-prime software (Schneider, Eschman, & Zuccolotto, 2012) and the multifunctional USB-based stimuli and response device *Chronos* which recorded accuracy and response times (RTs) with 1 ms resolution

(*Chronos operator manual*, 2015). The sounds in the experiment were presented via *Chronos* accessory headset.

The experiment started with onscreen instruction about the task requirements. Participants were asked to keep their eyes on the screen (as cued by the fixation cross) and to respond by pressing one button for the *thick* sound and another button for the *squeaky* sound, with the index finger of their dominant hand. The buttons of the response box (the first and the third, with the finger resting on the middle one between trials) were counterbalanced across participants. The words *low* and *high* were avoided in the instruction because of possible semantic priming. The sounds were presented in pairs – 600/900 Hz or 900/1200 Hz, in two continuous blocks with a pair change in between. Crucially, participants were informed that at one point the sounds would be changed but they should continue performing the same task. The experimenter urged the participants to look at the screen and monitored the execution of the task.

Each trial started with 500 ms fixation cross (*Consolas*, 22 pt, black, on a white background). Then participants heard one of two sounds – lower or higher, presented pseudorandomly, no more than four of the same pitch successively (to avoid motor fluency). The sounds were randomly coupled with a black square, presented on a white background below or above fixation – at 20% or 80% along the vertical midline of a 24-inch monitor with 60 Hz refresh rate and resolution 1920x1080 px (at 7.5 cm below or above the center of the display). Each pitch was accompanied by an equal number of high and low squares. As the viewing distance was approximately 60 cm, the side of the square corresponded to 2.5°-3° horizontal visual angle. The sound and the square were presented simultaneously (bimodally) for 1000 ms or until response, and were followed by 1500 ms intertrial interval (white screen). The sequence of one trial is represented in Figure 1.

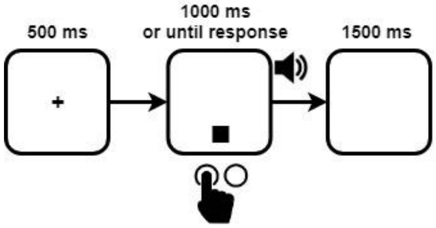

Figure 1: Schematic representation of one trial, not to scale.

For all participants the experiment started with one sound pair in the first block and ended with another sound pair in the second block, in counterbalanced order. There were 216 experimental trials. During the first 104 trials some participants responded to 600/900 Hz, and other – to 900/1200 Hz, and vice versa for the rest of the trials. The first 8 trials after the sound change were treated as practice and were analyzed separately. That way, the two blocks consisted of 104 trials each. Additionally, there were 8 practice trials before the experimental part, always with the sound pair of the following block. These trials were excluded from the analyses. Thus, there were 224 trials overall – 8 practice trials separated with a break from the experimental part, and 216 experimental trials (104 before the sound change, 8 practice trials after the sound change and 104 trials with the second sound pair) with no break. The experiment lasted about ten minutes.

Results

First, the overall accuracy was assessed (.94, range .83-1). No participant was excluded on that account. As 900 Hz was the pitch of interest, *only* responses to that pitch were considered in the subsequent analyses. Then, accuracy and RT were aggregated by trial in order to examine the nature of transition between the two contexts. The progress of the values over time is visualized on Figure 2.

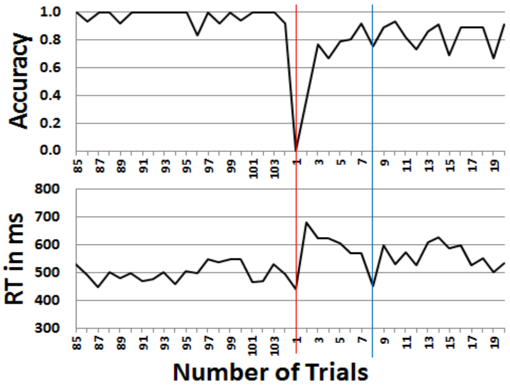

Figure 2: Accuracy and RT around context change, responses to relative pitch only (900 Hz), by trial. The red line indicates the point of change. The blue line marks the end of the online practice trials.

As participants had eight trials of practice before the experimental part, the first eight trials after the context change were also considered as practice and disregarded. Accuracy was analyzed as a function of experimental block and pitch-height congruency. There was no main effect of congruency – no difference in accuracy for congruent and incongruent responses (F(1,23)=2.32, *p*=.142; η^2_p=.092). However, there was main effect of experimental block (F(1,23)=8.12, *p*=.009; η^2_p=.261), i.e. more accurate responses in the first block; and an interaction between experimental block and congruency (F(1,23)=5.24, *p*=.032; η^2_p=.185). Newman-Keuls post-hoc revealed difference between the incongruent trials of the second block and the congruent and incongruent trials of the first block (*p*<.001), between the congruent trials of the first block and the second block (*p*=.045) and between the congruent and incongruent trials of the second block (*p*=.008). That is, participants made more mistakes in the second part of the experiment, especially in the incongruent trials.

There were a total of 5.6% errors in the experimental trials (.08% with no RT) which were also excluded. The data were then trimmed with 2.5 SDs from the subject means per condition (another 2.8%). Only trimmed data from the experimental blocks, concerning the correct responses to the relative pitch (900 Hz) were examined for differences in response times. The analyses were again performed with experimental block and pitch-height congruency as within factors. Means and standard deviations per condition are presented in Table 1.

Table 1: Descriptive statistics: mean RT (SD) per condition

First block		Second block	
Congruent	Incongruent	Congruent	Incongruent
512 (110)	529 (119)	497 (96)	522 (83)

There was no main effect of experimental block (F(1,23)=.761, p=.392; η^2_p=.032). At the same time, in accordance with the presumed crossmodal interaction between pitch and height, there was a substantial congruency effect (F(1,23)=5.65, p=.026; η^2_p=.197). Crucially for our hypothesis, there was *no* interaction between congruency and experimental block (F(1,23)=.533, p=.473; η^2_p=.023) – in other words, congruent responses were faster regardless of the relative sound context. Mind that only the context of the 900 Hz sound was changed between the blocks. And yet, in the presence of the higher pitch (1200 Hz) it was mapped to the lower vertical position, and subsequently to the higher vertical position in the presence of the lower pitch (600 Hz). Figure 3 illustrates the remapping of the sound across the two experimental blocks.

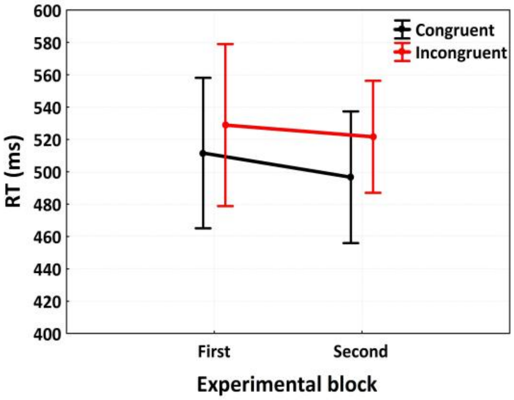

Figure 3: Transfer of vertical congruency effects across different sound contexts in the experimental blocks. Vertical bars denote .95 confidence intervals.

Due to the nature of the experiment, additional analyses were performed to account for alternative explanations. In the bimodal sound-discrimination task, sound pitch is coupled with visual stimuli presented along the vertical dimension. However, crossmodal correspondences were already demonstrated between pitch and the horizontal space (e.g. Rusconi et al., 2006). More specifically, lower pitch was mapped more readily to the left, and higher pitch was mapped more readily to the right. Our participants responded by pressing a left or a right button for the lower or higher pitch, in counterbalanced order. That is, for one half of the participants the sounds were horizontally congruent (they always pressed the left button for the lower pitch), while the other responded in horizontally incongruent manner.

To estimate the possible interaction between horizontal and vertical congruency, the mapping of the response was added as a categorical predictor in the above analyses. There was an interaction between horizontal and vertical congruency for accuracy (F(1,22)=5.75, p=.025; η^2_p=.207) and for RT (F(1,22)=25.13, p<.001; η^2_p=.533) (Figure 4).

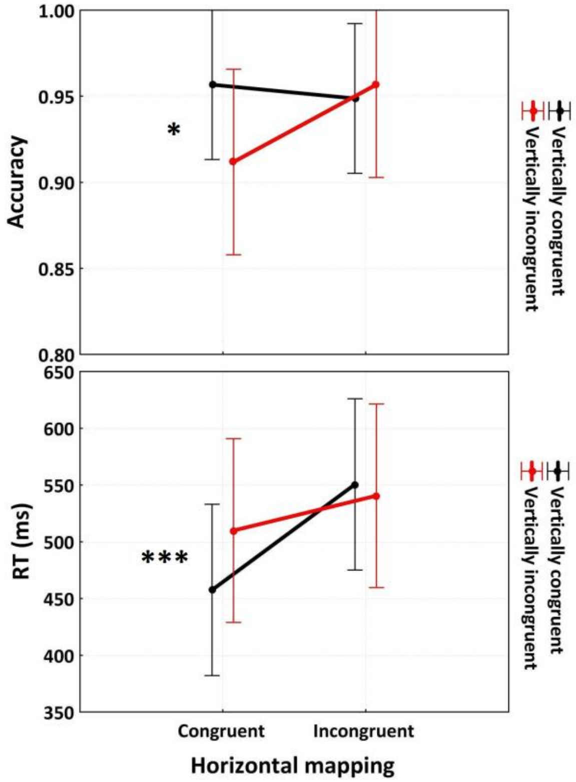

Figure 4: Interaction between horizontal and vertical congruency for accuracy (up) and RT (down). Vertical bars denote .95 confidence intervals.
* p<.05; *** p<.001

Newman-Keuls post-hoc revealed difference between the vertically congruent and incongruent trials (p=.041 for accuracy; and p<.001 for RT), but *only* when the responses were horizontally congruent. Remarkably, there was no difference when responses were mapped incongruently.

Discussion

On the whole, we did find the typical pitch-height congruency effects (Ben-Artzi & Marks, 1995; Chiou & Rich, 2012; Evans & Treisman, 2009; Patching & Quinlan, 2002; Rusconi et al., 2006 etc.). That way, we supported experimentally one of the major claims in the field – that the mappings between two dimensions are relative. For example, Gallace and Spence (2006) reported similar effects in a bimodal size-discrimination task with task-irrelevant sound frequencies, relatively mapped to the values of the visual stimuli. In another study, Smith, Grabowecky and Suzuki (2007) presented participants with the same visual stimuli (androgynous faces) accompanied by a sound in the male or female frequency speaking range, and found differences in perception of gender. Here, we demonstrated online remapping of one intermediate pitch to lower and higher vertical spatial positions in a speeded sound-discrimination task. Note that the intermediate pitch was presented throughout the whole experiment, but in a different context – with equidistant lower and subsequently higher pitch for one half of the participants, and the other way around for the other half of the participants. As follows, participants had to change their response as well – those who responded to the lower pitch with the *thick* button had to remap the same sound to the *squeaky* button, and vice versa. Moreover, in accordance with the expected congruency effects, responses to the same pitch were faster and more accurate when it was coupled with a square in the lower or higher visual field and perceived as *thick* or *squeaky*, respectively.

Crucially, the crossmodal correspondence between pitch and vertical space depended on the correspondence between pitch and horizontal space. Rusconi and colleagues (2006) provided conclusive evidence for explicit and implicit vertical and horizontal spatial mappings of pitch (but see Pitteri, Marchetti, Priftis, & Grassi, 2017). In their experiments responses were gathered vertically and horizontally, and participants performed the tasks with crossed and uncrossed hands. Higher pitch was mapped to upper and right buttons, and lower pitch was mapped to lower and left buttons (the so-called SMARC effect), even when responses did not require explicit processing of pitch, as in a wind vs. percussion sounds discrimination task. This implies that pitch-height correspondences are not solely semantically modulated, as horizontal mapping of sound is not linguistically promoted. Crossmodal correspondences depend mostly on failures in selective attention, especially within the speeded discrimination task (Spence 2011, 2015). When the task-irrelevant dimension is visuospatial, we might expect interaction between generated spatial and response codes (see Lu & Proctor, 1995, for a review of Simon and spatial Stroop effects). Interaction was reported also for mental representations and responses in horizontal space (Dehaene, Bossini, & Giraux, 1993). In our task, responses were gathered horizontally, while visual stimuli were presented along the vertical axis. And yet, we found substantial interaction between horizontal mental representation of sound and response side (unlike Pitteri et al., 2017, who reported the same effect for pitch and brightness, but only for musicians). It can be speculated that sound is represented both horizontally and vertically. That way, mentally generated horizontal and vertical spatial codes interact with stimuli-generated vertical spatial codes and modulate the crossmodal congruency effect. As a result, horizontal congruency emerged as an essential prerequisite for vertical congruency effects.

In addition, it seems that crossmodal mappings happen automatically and effortlessly. Our results are in line with previous findings, demonstrating that relations can be retrieved unconsciously and transferred across domains. (Hristova, 2017; Li, Li, Zhang, Shi, & He, 2018) Thus, as Holyoak, Gentner and Kokinov (2001) pointed out, the ability to manipulate relations might be basic for cognition.

That being said, the experiment is a beginning of a larger experimental work within the field of crossmodal correspondences. The hypothesis about the online relational analyses should be explored further. Additional experimental settings should investigate whether similar associations exist among isolated features, or are integrated in larger cognitive frameworks. Another major challenge would be to outline the dissimilarities between given crossmodal mappings and relating them to other forms of associations.

Acknowledgements

This research was supported financially by the European Office for Aerospace Research and Development under grant FA9550-15-1-0510 (Anticipating Future by Analogy-Making).

References

Belkin, K., Martin, R., Kemp, S. E., & Gilbert, A. N. (1997). Auditory pitch as a perceptual analogue to odor quality. *Psychological Science*, *8*(4), 340-342.

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

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition*, *126*(2), 165-172.

Chen, N., Tanaka, K., Namatame, M., & Watanabe, K. (2016). Color-Shape Associations in Deaf and Hearing People. *Frontiers in psychology*, *7*. Retrieved from https://doi.org/10.3389/fpsyg.2016.00355

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

Chronos operator manual. (2015). Pittsburgh, PA: Psychology Software Tools, Inc.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371-396.

Evans, K. K., & Treisman, A. (2009). Natural cross-modal mappings between visual and auditory features. *Journal of Vision*, *10*(1), 6:1-12.

Fastl, H. (2004). Audio-visual interactions in loudness evaluation. *Proceedings of the 18th International Congress on Acoustics*, *Kyoto, Japan* (pp. 1161–1166).

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191-1203.

Garner, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.

Holyoak, K. J., Gentner, D., & Kokinov, B. N. (2001). Introduction: The place of analogy in cognition. In D. Gentner, K. J. Holyoak & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science*, (1-19). Cambridge, MA: MIT press

Hristova, P. (2017). Can we think unconsciously via analogy?. In Z. Radman (Ed.), *Before consciousness: In search of the fundamentals of mind*, (270-295). Exeter, UK: Imprint Academic.

Levitan, C. A., Ren, J., Woods, A. T., Boesveldt, S., Chan, J. S., McKenzie, K. J., ... & van den Bosch, J. J. (2014). Cross-cultural color-odor associations. *PloS One*, *9*(7), e101651.

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. *Developmental Psychology*, *16*(6), 597-607.

Li, J., Li, X., Zhang, X., Shi, K., & He, Y. (2018). Can unconscious thought detect relational similarities?. *International Journal of Psychology*, 1-7. doi:10.1002/ijop.12550

Lu, C. H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, *2*(2), 174-207.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology*, 173-188.

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, *28*(7), 903-923.

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

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(1), 69-79.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, *116*(4), 323-336.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133-136.

Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tone. *Journal of Experimental Psychology*, *66*(4), 347-352.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173-186.

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research*, *220*(3-4), 319-333.

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception*, *41*(7), 854-861.

Patching, G. R., & Quinlan, P. T. (2002). Garner and congruence effects in the speeded classification of bimodal signals. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 755-775.

Pitteri, M., Marchetti, M., Priftis, K., & Grassi, M. (2017). Naturally together: pitch-height and brightness as coupled factors for eliciting the SMARC effect in non-musicians. *Psychological Research*, *81*(1), 243-254.

Ratcliffe, V. F., Taylor, A. M., & Reby, D. (2016). Cross-modal correspondences in non-human mammal communication. *Multisensory Research*, *29*(1-3), 49-91.

Rohlf, S., Habets, B., von Frieling, M., & Röder, B. (2017). Infants are superior in implicit crossmodal learning and use other learning mechanisms than adults. *eLife*, *6*, e28166.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*(2), 113-129.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime: User's guide.* Pittsburgh, PA: Psychology Software Tools, Inc.

Smith, E. L., Grabowecky, M., & Suzuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology*, *17*(19), 1680-1685.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971-995.

Spence, C. (2015). Cross-modal perceptual organization. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (649-664). Oxford, UK: Oxford University Press.

Simpson, R. H., Quinn, M., & Ausubel, D. P. (1956). Synesthesia in children: Association of colors with pure tone frequencies. *The Journal of Genetic Psychology*, *89*(1), 95-103.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, *21*(1), 21-25.