Journal of Experimental Child Psychology 162 (2017) 72–88

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

The impact of semantically congruent and incongruent visual information on auditory object recognition across development

Rhiannon L. Thomas [a,b], Marko Nardini [b,c], Denis Mareschal [b,*]

[a] *Sensorimotor Development Research Unit, Department of Psychology, Goldsmiths College, University of London, London SE14 6NW, UK*
[b] *Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck College, University of London, London WC1E 7HX, UK*
[c] *Department of Psychology, University of Durham, Durham DH1 3LE, UK*

ARTICLE INFO

Article history:
Received 15 July 2016
Revised 21 April 2017

Keywords:
Audition
Vision
Multisensory
Congruency
Stroop
Semantic

ABSTRACT

The ability to use different sensory signals in conjunction confers numerous advantages on perception. Multisensory perception in adults is influenced by factors beyond low-level stimulus properties such as semantic congruency. Sensitivity to semantic relations has been shown to emerge early in development; however, less is known about whether implementation of these associations changes with development or whether development in the representations themselves might modulate their influence. Here, we used a Stroop-like paradigm that requires participants to identify an auditory stimulus while ignoring a visual stimulus. Prior research shows that in adults visual distractors have more impact on processing of auditory objects than vice versa; however, this pattern appears to be inverted early in development. We found that children from 8 years of age (and adults) gain a speed advantage from semantically congruent visual information and are disadvantaged by semantically incongruent visual information. At 6 years of age, children gain a speed advantage for semantically congruent visual information but are not disadvantaged by semantically incongruent visual information (as compared with semantically unrelated visual information). Both children and adults were influenced by associations between auditory and visual stimuli, which they had been exposed to on only 12 occasions during the learning phase of the study. Adults showed a significant speed

---

* Corresponding author.
E-mail address: d.mareschal@bbk.ac.uk (D. Mareschal).

http://dx.doi.org/10.1016/j.jecp.2017.04.020

advantage over children for well-established associations but showed no such advantage for newly acquired pairings. This suggests that the influence of semantic associations on multisensory processing does not change with age but rather these associations become more robust and, in turn, more influential.

Introduction

Events in our world provide signals to multiple senses. The ability to use these different signals in conjunction confers numerous advantages on perception. First, senses can be complementary in providing unique kinds of information; for example, only vision can reliably tell us whether we are faced with a gray squirrel or a red one. Second, senses often provide redundant information about the same property; for example, both visual size and auditory amplitude could be cues to an animal's size. Using multiple redundant cues across senses allows adults to detect stimuli more readily (Lovelace, Stein, & Wallace, 2003; Stein, London, Wilkinson, & Price, 1996) and to respond more accurately (e.g., Alais & Burr, 2004; Ernst & Banks, 2002) and more rapidly (Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994).

Studies investigating the development of multisensory perception have found mixed results. There is a body of evidence to suggest that some multisensory abilities are present during early infancy (e.g., Bremner, Slater, Johnson, Mason, & Spring, 2012; Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006; Scheier, Lewkowicz, & Shimojo, 2003). However, studies investigating the development of cross-modal cue combination suggest that in some tasks children do not combine information across senses as adults do until 8 years of age or later (e.g., Gori, Del Viva, Sandini, & Burr, 2008; Jaime, Longard, & Moore, 2014; Nardini, Bales, & Mareschal, 2016; Nardini, Bedford, & Mareschal, 2010; Nardini, Begus, & Mareschal, 2013; Nardini, Jones, Bedford, & Braddick, 2008; Petrini, Remark, Smith, & Nardini, 2014). Considering audio–visual stimuli specifically, children appear to integrate cues more frequently and less selectively than adults (Adams, 2016; Innes-Brown et al., 2011), and they also show a bias toward auditory stimuli (Nava & Pavani, 2013), which develops into an adult-like visual dominance across middle childhood. In addition, children show a diminished McGurk interference effect, which also suggests that they may be processing auditory information over visual information (e.g., Massaro, Thompson, Barron, & Laren, 1986). An early bias for auditory stimuli may partially be explained by the differential experience of the auditory and visual systems in the prenatal environment (e.g., Lecanuet & Schaal, 1996). This developmental shift in the way that audio–visual information is weighted suggests that the mechanisms underlying these processes are changing across this period.

Many studies now show that multisensory perception in adults is influenced by factors beyond simple low-level stimulus properties such as spatial and temporal coincidence. For example, adult multimodal perception is also influenced by how attention is allocated within a scene (Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010) as well as by variation in the congruency (in terms of both perceptual and semantic features) between the different sensory inputs (e.g., Heron, Whitaker, & McGraw, 2004; Jackson, 1953; Slutsky & Recanzone, 2001). There is a growing body of evidence suggesting that adults are sensitive to the semantic congruency between multisensory signals and that this influences the way in which these signals are processed, enabling more accurate and efficient recognition (e.g., Chen & Spence, 2010; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Lehmann & Murray, 2005; Senkowski, Saint-Amour, Kelly, & Foxe, 2007). This could be advantageous because it allows observers to use their previous experiences to improve their chances of making correct perceptual judgments. Semantic congruency is a particularly important factor when sensory reliability is reduced. For example, older adults (whose vision and hearing have become degraded over time) benefit substantially from semantic congruency when processing audio–visual speech (Maguinness, Setti, Burke, Kenny, & Newell, 2011).

Moreover, there is also an impressive body of research considering the development of semantic associations across the first years of life. In particular, these studies explore how children develop lexical semantic associations. Infants are capable of associating words with objects in their world by the first year of life (e.g., Schafer, 2005). By 2 years of age, infants show enhanced visual target recognition following a related word prime (Styles & Plunkett, 2009), demonstrating that they are already making semantic associations between auditory and visual stimuli. By 2 years, infants are also able to make some semantic associations between the words in their lexicon (Arias-Trejo & Plunkett, 2009), suggesting that they are beginning to build a representation of semantic associations. The development of these semantic associations is an ongoing process as children create more elaborate representations, acquire new content, and structure the existing representations accordingly (e.g., Bjorklund, 1985; Bjorklund, 1987). It is clear from this work that children are building semantic representations, across multiple sensory domains, from very early in life. Consequently, these higher-level associations have the potential to influence perception through top-down processes across all of childhood.

We know that multisensory processing changes from infancy through childhood and into adulthood (e.g., Gori et al., 2008; Innes-Brown et al., 2011; Nardini et al., 2008; Nava & Pavani, 2013; Neil et al., 2006; Scheier et al., 2003). It also seems that semantic associations across the senses influence perception throughout development. For example, Jordan and Baker (2011) found that redundant audio–visual information helped 3- to 5-year-olds' numerical matching performance compared with unisensory information, suggesting that at this age children can already benefit from cross-modal semantic associations. Here, we investigated how children (and adults) bring together concurrent complementary or conflicting sensory information presented in vision and audition. It seems that semantic associations play a role in perception from early in development; however, it is less clear how this role develops. It is possible that across development the role of semantic knowledge in multisensory perception may become more established (e.g., Murray, Lewkowicz, Amedi, & Wallace, 2016). Alternatively, it could be that we observe an apparent increase in the role of semantic associations with age, but this is driven by strengthening associations rather than by a more general change. To disentangle these two possibilities, we presented children (and adults) with audio–visual pairings with which they were familiar and also introduced new audio–visual pairings to explore their relative influence. To this end, we used a Stroop-like paradigm (Stroop, 1935) in which participants were required to attend to and identify a sound while simultaneously being presented with a visual stimulus.

Children are susceptible to various forms of Stroop interference. From around 7 years of age, children experience a large degree of interference for the classic color–word Stroop paradigm (Comalli, Wapner, & Werner, 1962). From around 3 years of age, children are susceptible to Stroop interference in tasks that do not require reading ability (e.g., Gerstadt, Hong, & Diamond, 1994; Prevor & Diamond, 2005; Wright, Waterman, Prescott, & Murdoch-Eaton, 2003). Across all these latter studies, children were slower to name an item when it was presented alongside semantically incongruent information.

In cross-modal Stroop paradigms, participants are typically asked to attend to a stimulus presented in one modality while ignoring a stimulus presented in another modality. Participants tend to be slower to respond to a stimulus presented in one modality when it is accompanied by an incongruent stimulus presented in another modality (e.g., Cowan & Barron, 1987; Vogler & Titchener, 2011). Yuval-Greenberg and Deouell (2009) investigated the influence of visual stimuli on auditory processing as well as the influence of auditory stimuli on visual processing. They presented adult participants with pictures and vocalizations of animals that had a congruent, incongruent, or neutral relation to one another. They found an asymmetry in the extent to which one modality influenced the other. Participants were faster to respond to congruent trials irrespective of the modality to which they were responding; however, this advantage was greater when participants were responding to the auditory stimulus. It seems that in adults visual stimuli confer a particular advantage for recognition of auditory stimuli. However, evidence from the developmental literature (e.g., Massaro et al., 1986; Napolitano & Sloutsky, 2004; Nava & Pavani, 2013; Robinson & Sloutsky, 2004) suggests that younger children may be less influenced by visual information and, in fact, demonstrate an auditory dominance bias.

Cross-modal Stroop has also been explored developmentally. For example, Hanauer and Brooks (2003) found that at 4 or 5 years of age children were slower to respond to a color patch when the auditory distractor was an incongruent color word rather than a noncolor adjective. The extent of this interference (difference in reaction times between conditions) decreased with age, and the effect was relatively small in adults. In a later study, Hanauer and Brooks (2005) found that between 3 and 7 years of age children were slower to respond to a line drawing when the auditory distractor was from the same semantic category rather than a different one. Again the extent of this interference reduced with age, and the extent of interference was mediated by whether the distractor item was from the same response set (e.g., animals vs. clothes). These studies suggest that cross-modal semantic associations influence children; however, the task-irrelevant information was always a word. Thus, these studies do not inform us about whether nonverbal information is automatically processed at a semantic level.

Studies of sensory dominance suggest that early in development auditory stimuli tend to dominate perception (Massaro et al., 1986; Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004;), whereas this bias tends to be reversed during adulthood (e.g., Colavita, 1974; Howard & Templeton, 1966; McGurk & MacDonald, 1976). Developmental studies have demonstrated the potential time course of this change. Constantidou, Danos, Nelson, & Baker (2011) found that children's (7–13 years of age) memory for spoken words presented with a visual stimulus was better than that for spoken words alone; however, this performance was not better than that for visual stimuli alone. This might be because memory for spoken words was enhanced by visual stimuli, or it might be due to a visual dominance effect, which appears to emerge from 7 years of age (Nava & Pavani, 2013). Heikkilä and Tiippana (2016) also demonstrated that 8- to 12-year-olds had better recall for stimuli presented in semantically congruent audio–visual pairs than for stimuli presented in nonsemantic pairs. Incongruent pairs did not interfere with recall. Thus, it seems that semantically congruent information can enhance recall during childhood. As such, it seems that children can link semantic information across audition and vision and can use this to encode a single item.

The aim of the current study was to investigate the following questions. First, do primary-school-aged children benefit from semantically congruent audio–visual information during auditory object recognition, and/or are they disadvantaged by incongruent audio–visual information? Second, with development, does knowledge of semantic associations play an increasing role in combining sensory stimuli, or is this influence constant and instead the semantic associations become more robust?

In the current study, we asked children to focus on information presented in the auditory domain while ignoring information presented in the visual domain. We chose to investigate children between 5 and 9 years of age because previous research has shown some evidence of multisensory integration and perceptual benefits (e.g., Nardini et al., 2016) across this age range but typically not at a mature level. We presented participants with audio–visual pairings with which they were familiar and audio–visual pairings that were introduced during the experiment. This allowed us to tease apart two potential trajectories for the development of semantic associations in multisensory perception, namely that (a) the role of semantic associations changes across development and (b) semantic associations become robust with development, leading to a change in their role. Given the existing literature on sensory dominance, we predicted that younger children would be less susceptible to Stroop-like interference from conflicting visual information than adults and that, in turn, this might also lead to less facilitation from redundant visual information. Throughout the experiment, participants were given the task of identifying an animal vocalization while ignoring task-irrelevant visual information. This animal vocalization was presented simultaneously with a visual stimulus that was either congruent (same animal), incongruent (different animal), or neutral (black and white pattern). An auditory prompt (the name of an animal) was then presented. If this prompt matched the previous vocalization, then participants were required to make a button press.

Method

Participants

A total of 59 children contributed data to this study (32 girls and 27 boys). Children ranged in age from 6.00 to 9.42 years. Children were divided into two age groups: 6- and 7-year-olds ($n = 26$; $M_{age} = 7.12$ years, $SD = 0.54$, range = 6.00–7.89) and 8- and 9-year-olds ($n = 33$; $M_{age} = 8.75$ years, $SD = 0.40$, range = 8.05–9.42). An additional 10 children were excluded; of these children, 8 ($M_{age} = 7.00$ years, $SD = 1.44$, range = 5.33–8.91) failed to reach 85% performance during practice trials and 2 (1 5-year-old and 1 8-year-old) asked to stop the study before completing a sufficient number of test trials. In addition, 17 adults participated (10 women and 7 men), ranging in age from 18 to 53 years ($M = 27.70$ years, $SD = 7.88$). An additional adult was excluded because she failed to reach 85% performance during the practice trials. All participants had normal or corrected to normal vision and hearing. The age range of adults participating in the study was quite large. All but 2 of our participants ranged in age between 19 and 29 years. To check whether these older participants should be considered separately from the younger adults in our study, we removed them and reran our analyses; this produced the same pattern of results, and as such we report on the full group of adults.

Apparatus and stimuli

The current study adapted a paradigm designed by Yuval-Greenberg and Deouell (2009). The experiment was conducted using a Hewlett Packard G600 laptop computer with a resolution of 1280 × 800 and a refresh rate of 60 Hz. Stimulus programming, presentation and response collection were carried out using E-Prime (Version 2; http://www.pstnet.com). Visual stimuli presented in the test trials of the experiment consisted of colored photographs of eight animals (lion, sheep, koala, meerkat, dog, rhino, raccoon, and elephant) presented on a white background, as well as black and white checkerboard/dot patterns, all presented for 500 ms. Three images were used as exemplars for each animal, resulting in 24 different animal images (see Fig. 1). Auditory stimuli presented in the test trials consisted of eight animal vocalizations (lion, sheep, koala, meerkat, dog, rhino, raccoon, and elephant). Vocalizations were selected from an online database of naturally recorded sounds (http://ww1.freesounds.org); all of them were processed using a freely available program (Audacity; http://www.audacityteam.org) to have a sampling rate of 44,100 Hz and a resolution of 16 bits per sample, and they were normalized to their maximum amplitude. Each vocalization was represented by three exemplars, leading to a total of 24 animal vocalizations. Sounds were edited to fill the 500-ms interval using Audacity. In some cases, this meant selecting a section of the full vocalization (e.g., raccoon chattering). In other instances, this meant repeating a single vocalization (e.g., dog barking). Auditory prompts were also presented. These consisted of the names of the aforementioned animals spoken in a natural female voice. Auditory prompts were recorded using a (Yoga EM-278 microphone; http://www.maplin.co.uk/p/handheld-condenser-microphone-l97aq). Additional auditory and visual stimuli were presented during practice trials; these were sounds and images of bears, cows, frogs, cats, and deer. All auditory stimuli were presented binaurally through a set of (Sennheiser HD201 stereo headphones; https://www.amazon.co.uk/Sennheiser-Closed-Dynamic-headphones-Performance/dp/B0007XJSQC) at 55 dB. Participants responded using the space bar situated within the laptop keyboard.

Procedure

Participants were tested individually in a quiet room. They were seated at a desk with the computer positioned approximately 30 cm in front of them. The study consisted of four parts (see Fig. 2).

Familiarity check

Initially, participants were presented with one exemplar of each of the eight animal vocalizations and were asked to name an animal that they thought might make this sound. This was a free response,

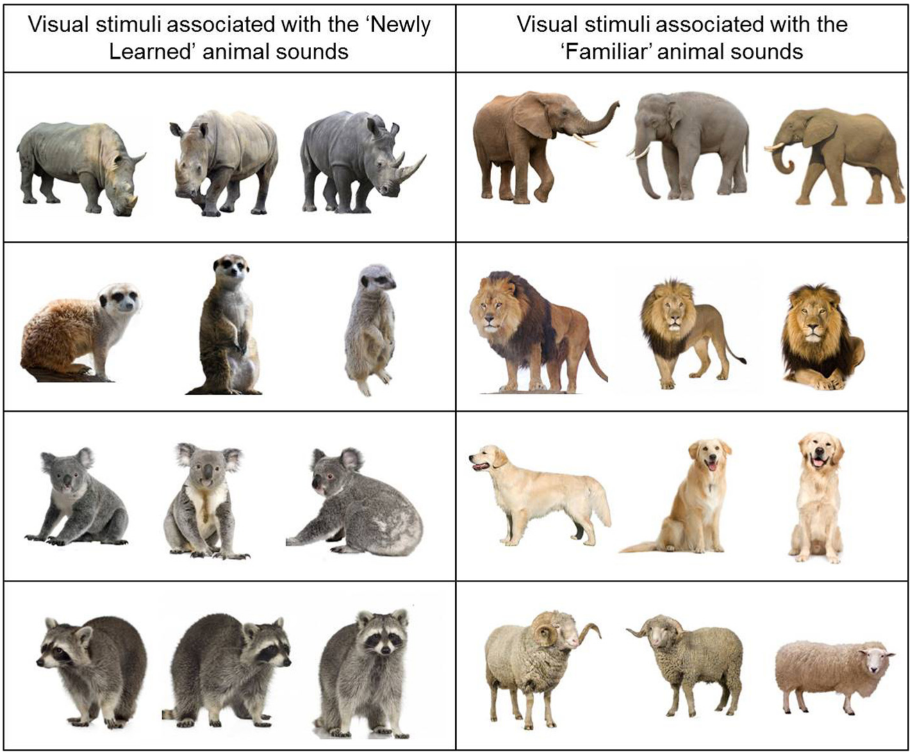
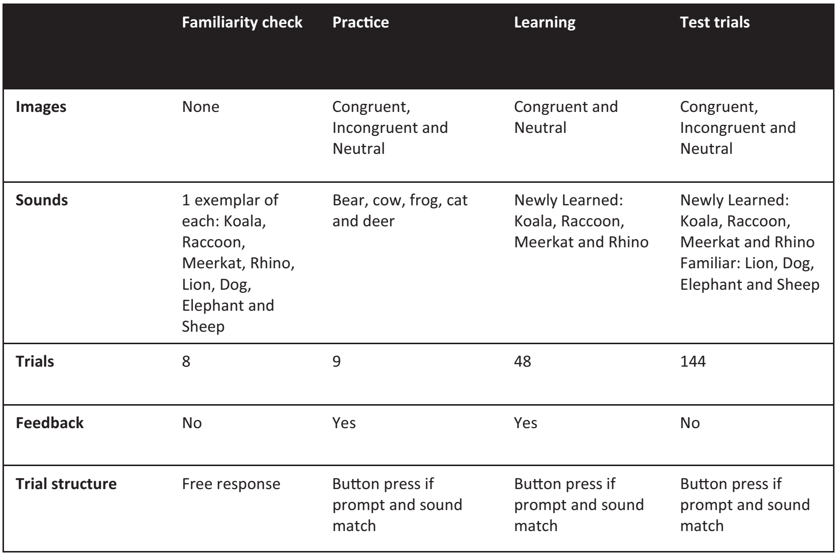


Fig. 1. Visual stimuli presented in the experiment. "Newly Learned" images were presented in the familiarization phase and in test trials. "Familiar" images were presented only during test trials.



	Familiarity check	**Practice**	**Learning**	**Test trials**
Images	None	Congruent, Incongruent and Neutral	Congruent and Neutral	Congruent, Incongruent and Neutral
Sounds	1 exemplar of each: Koala, Raccoon, Meerkat, Rhino, Lion, Dog, Elephant and Sheep	Bear, cow, frog, cat and deer	Newly Learned: Koala, Raccoon, Meerkat and Rhino	Newly Learned: Koala, Raccoon, Meerkat and Rhino Familiar: Lion, Dog, Elephant and Sheep
Trials	8	9	48	144
Feedback	No	Yes	Yes	No
Trial structure	Free response	Button press if prompt and sound match	Button press if prompt and sound match	Button press if prompt and sound match

Fig. 2. Description of each phase of the experimental procedure.

so participants could name as many animals as they wanted. This gave a measure of whether the selected auditory stimuli were truly "Familiar" or unfamiliar to each participant at the start of the study.

Participants were then told that they were going to play a computer game in which they would hear an animal sound and, at the same time, they would see a picture. This picture could be of the same animal, could be of a different animal, or might not be related to the sound at all. Participants were told that next they would see a question mark on the screen and that they would hear a lady's voice saying the name of an animal. If the lady named the animal that they had heard, then they should press the indicated key as quickly as possible. However, if the lady said the name of any other animal, then they should not make any response. We decided to use a single key for children (and adults) to record their responses because previous research (e.g., Davidson, Amso, Anderson, & Diamond, 2006) had demonstrated that participants are slower to respond to trials that require a different response site from the previous trial (even though this response is made with a different finger). Furthermore, the extent of this delay is mediated by task and age. Finally, because we excluded trials in which participants responded incorrectly, we independently examined response speed rather and accuracy.

Practice trials. Participants were informed that they would have a chance to practice the game to make sure that they understood the rules. They then completed nine practice trials, which had the same structure as the test trials (see Fig. 3 for a schematic of the sequence of events within a single trial); however, the stimuli presented in these practice trials were not subsequently repeated.

Each trial began with text on the screen that read "What animal makes this sound?" This text was read aloud to children. The experimenter initiated a trial when the participant was ready. A black fixation cross then appeared on the screen and remained there for 1000 ms. Next, an image appeared on the screen. This image showed either an animal or a black and white pattern. At the same time, an animal vocalization was presented. Either the animal vocalization and image were from the same animal (Congruent), they were from different animals (Incongruent), or the animal sound was presented with a neutral image (Neutral). During practice trials, participants were given feedback on both their

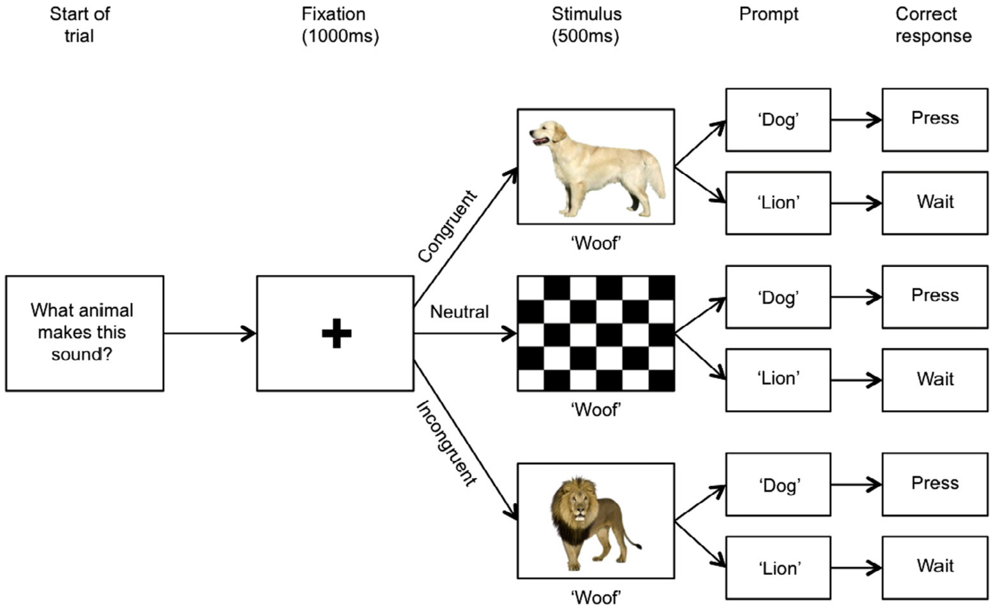


Fig. 3. Schematic of the structure of a trial.

performance (correct or incorrect) and their speed of response; this was presented as text on the screen and read aloud to children. If participants did not respond correctly to at least 85% of trials, then they were presented with a further nine practice trials. If participants did not respond correctly on at least 85% of trials after two practice sessions, then they were excluded from participating in the remainder of the study.

Learning phase. The next phase of the study gave participants experience with "Newly Learned" animal vocalizations. During this phase, participants were presented with the vocalizations of raccoons, koalas, meerkats, and rhinos. During piloting, these sounds were identified as being unfamiliar to the majority of British children and adults (participants were unable to identify which animal produced these sounds given the opportunity to give as many suggestions as they could). The structure of the trials was the same as described above (and as shown in Fig. 3); however, participants were presented with only Congruent and Neutral trials. The purpose of this phase of the experiment was to give participants experience with the animal vocalizations and their associated images. In this phase, participants were given feedback on their performance, and when they gave an incorrect response they were informed of the correct answer. Participants completed 48 training trials (24 Congruent and 24 Neutral), with half of these trials requiring a button press as a correct response. Participants were exposed to each type of Newly Learned animal sound (e.g., koala) 12 times, with each specific sound being repeated 4 times. During Neutral trials, these were presented with one of the black and white neutral stimuli described above selected at random on each trial. During Congruent trials, these were presented with one of the three possible congruent animal images selected at random on each trial. We included Neutral trials in this phase of the experiment to ensure that children were learning the auditory information presented. Prior to the experiment, children had some knowledge of the image of these animals but no knowledge of the sounds they made. Including Neutral trials meant that children needed to use purely auditory information to respond and were given feedback if this response was incorrect.

Test trials. The final phase of the experiment required participants to play the same "game" again. Trials again had the same format as described above (and as shown in Fig. 3); however, participants were not given any feedback about their responses. They completed two blocks of 72 test trials, with each block comprising 24 each of Neutral, Congruent, and Incongruent trials in a random order. Two thirds of these trials were accompanied by the correct prompt, and so a button press was the correct response. On the remaining trials, withholding a button press was the correct response. Participants were reminded to respond as quickly and accurately as possible.

Results

The categorization of Familiar and Newly Learned animal sounds held true for all participants tested. During the initial familiarity check, none of the participants was able to label a Newly Learned animal sound correctly, whereas all participants were able to name the Familiar animal sounds. Therefore, no participants were excluded based on these data.

Accuracy

Performance was high for all age groups (see Fig. 4). An analysis of variance (ANOVA) on the proportion of correct trials was performed with congruency and familiarity as within-participants factors and age group as a between-participants factor. This revealed a main effect of familiarity, $F(1, 73) = 45.297$, $p < .001$, $\eta_p^2 = .383$; overall, participants made more correct responses during Familiar trials ($M = 89.2\%$, $SE = 1.0$) than during Newly Learned trials ($M = 80.9\%$, $SE = 1.2$). This analysis also revealed main effects of age group, $F(1, 73) = 7.342$, $p < .001$, $\eta_p^2 = .167$, and congruency, $F(2, 146) = 4.647$, $p = .011$, $\eta_p^2 = .060$. A significant interaction emerged between familiarity and congruency, $F(2, 146) = 4.473$, $p = .013$, $\eta_p^2 = .058$. Neither the interaction between congruency and age group, $F(4, 146)$

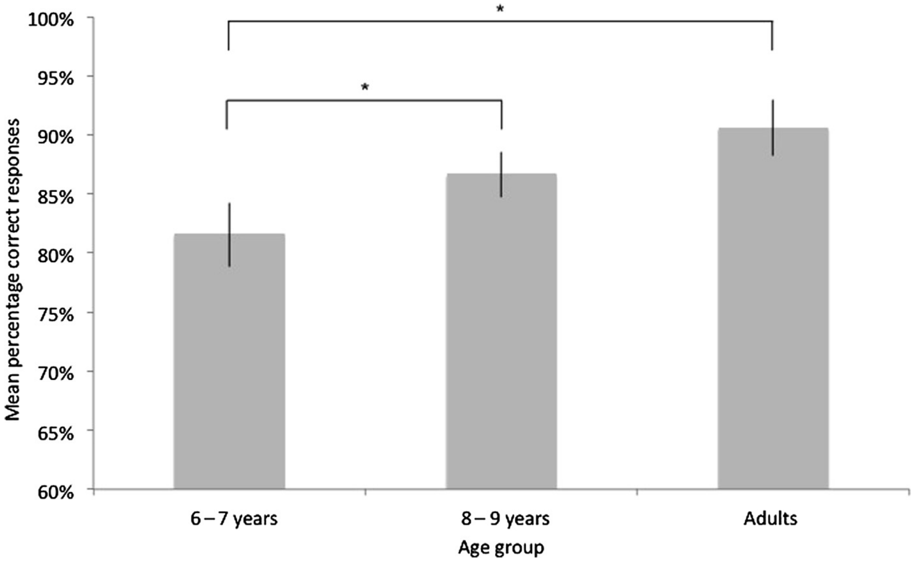


Fig. 4. Mean percentage of correct responses averaged across conditions for each age group. Error bars plot standard errors. *Indicates significant difference between Age groups ($p < 0.05$).

= 1.226, $p = .302$, $\eta_p^2 = .033$, nor the interaction between familiarity and age group, $F(2, 73) = 1.138$, $p = .326$, $\eta_p^2 = .30$, nor the three-way interaction, $F(4, 146) = 0.542$, $p = .705$, $\eta_p^2 = .015$, reached significance.

The main effect of congruency was explored using Bonferroni-corrected paired-samples *t* tests. These revealed that participants made significantly more correct responses in the Congruent condition ($M = 87.7\%$, $SE = 1.0$) than in the Incongruent condition ($M = 84.8\%$, $SE = 1.1$), $t(76) = 3.822$, $p < .001$, $d = 0.33$. No other significant differences emerged between conditions.

Bonferroni-corrected pairwise comparisons also revealed that the adults ($M = 90.5\%$ correct, $SE = 1.9$) made significantly fewer errors than the 6- and 7-year-olds ($M = 81.4\%$ correct, $SE = 1.5$, $p = .001$, $d = 1.16$) (see Fig. 4). The 8- and 9-year-olds ($M = 86.6\%$ correct, $SE = 1.4$) also made significantly fewer errors than the 6- and 7-year-olds ($p = .042$, $d = 0.64$).

To further explore the interaction between familiarity and congruency, paired-samples *t* tests were conducted comparing performance across the Familiar and Newly Learned conditions for each congruency condition. These revealed that participants made significantly fewer errors in the Familiar trials than in the Newly Learned trials for every congruency condition ($p < .001$ in all instances). To further unpack this interaction, a familiarity accuracy score was calculated by subtracting the mean percentage of correct responses made in the Newly Learned condition from the mean percentage of correct responses made in the Familiar condition for each congruency condition (see Fig. 5). Paired-samples *t* tests were then performed on these scores across congruency conditions. These revealed that the familiarity accuracy score was significantly larger in the Congruent condition ($M = 10.6\%$, $SE = 1.12$) than in the Incongruent condition ($M = 6.0\%$, $SE = 1.30$), $t(75) = 2.987$, $p = .004$, $d = 0.42$. Thus, performance was best when the auditory stimulus was Familiar and presented alongside a congruent visual stimulus. At the other extreme, performance was worst when the auditory stimulus was Newly Learned and presented alongside an incongruent visual stimulus. There were no other significant differences across congruency conditions.

In summary, analysis of accuracy across conditions revealed that the youngest children performed the worst and that performance increased with age. Participants made more errors when the auditory stimulus was paired with an incongruent visual stimulus than when it was paired with a congruent visual stimulus. This effect did not interact with age, suggesting that all age groups were similarly

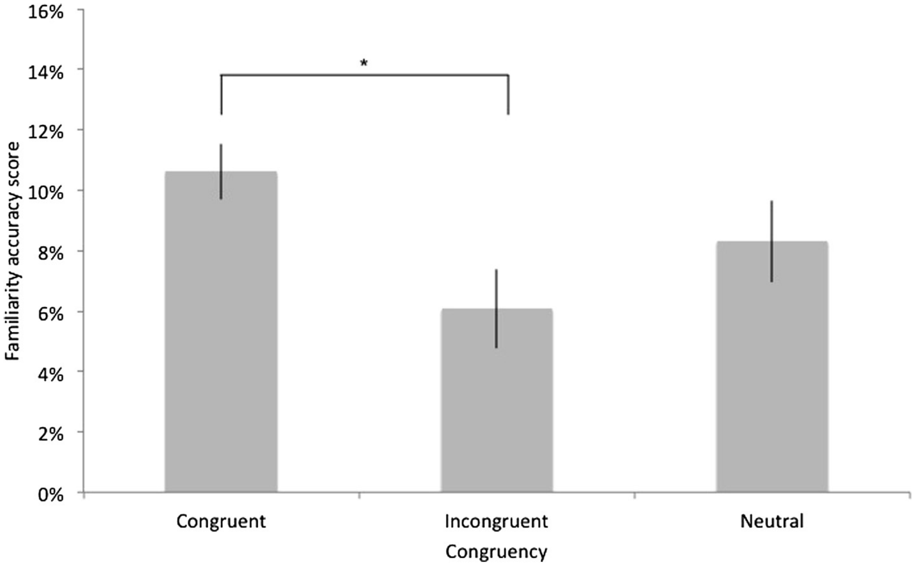


Fig. 5. Familiarity accuracy score calculated by subtracting percentage correct in Newly Learned trials from percentage correct in Familiar trials. A positive score indicates that accuracy was higher for Familiar trials compared with Newly Learned trials. Error bars plot standard errors. *Indicates significant difference between Age groups (p < 0.05).

influenced by the relationship between the auditory and visual stimuli. Participants also made more errors in the Newly Learned condition, where the auditory stimuli were unknown to participants before they were introduced during the experiment. Again, this did not interact with age group, suggesting that participants were similarly affected by the depth of knowledge they had of the pairings irrespective of age. Participants were most accurate when they were asked to recognize an auditory stimulus that was familiar to them and this was presented with a congruent visual stimulus. As might be expected, accuracy was worst when participants were asked to recognize an auditory stimulus with which they had limited experience and which was presented with an incongruent visual stimulus.

Reaction times

A third of experimental trials were excluded because the prompt was invalid (when the prompt and vocalization did not match, participants were not required to respond); this was independent of whether the trial was Congruent, Incongruent, or Neutral. Analyses were then performed only on trials that participants responded to correctly. Mean reaction times and standard deviations were calculated for each participant. Trials in which reaction times fell further than ±3 standard deviations from a participant's mean reaction time in each condition were excluded as outliers. Finally, reaction times under 150 ms were also excluded because any response under this time was assumed to be too fast for processing of the stimuli and so was likely to be preemptive. This resulted in 2.9% of responses being excluded for the 6- and 7-year-olds, 3.4% for the 8- and 9-year-olds, and 1.8% for the adults. Estimates of button press reaction times in adults range from around 200 to 250 ms (e.g., Eckner, Kutcher, & Richardson, 2010), so 150 ms is a conservative lower cutoff that should not exclude any genuine rapid responses. The remaining reaction time data were normally distributed in all age groups (as determined by the Kolmogorov–Smirnov test for normality).

Reaction times were analyzed in a mixed-design ANOVA with familiarity (Familiar or Newly Learned) and congruency (Congruent, Incongruent or Neutral) as within-participants factors, and age group (6- and 7-year-olds, 8- and 9-year-olds, or adults) as a between-participants factor. The dependent variable was the mean reaction time to make a button press during valid trials. The ANOVA revealed main effects of familiarity, $F(1, 73) = 267.198$, $p < .001$, $\eta_p^2 = .785$; overall, participants responded faster during Familiar trials ($M = 795$ ms, $SE = 19$) than during Newly Learned trials ($M = 1024$ ms, $SE = 19$). This analysis also revealed main effects of congruency, $F(2, 146) = 123.441$, $p < .001$, $\eta_p^2 = .628$, and age group, $F(4, 73) = 6.563$, $p = .002$, $\eta_p^2 = .152$. These main effects were qualified by significant interactions of familiarity with congruency, $F(2, 146) = 4.27$, $p = .017$, $\eta_p^2 = .055$, and familiarity with age group, $F(2, 73) = 4.967$, $p = .009$, $\eta_p^2 = .120$. The interaction of congruency with age group approached but did not reach significance, $F(4, 146) = 2.206$, $p = .08$, $\eta_p^2 = .196$. The three-way interaction among familiarity, congruency, and age group did not reach significance, $F(4, 146) = 1.395$, $p = .240$, $\eta_p^2 = .037$.

Paired-samples t tests were used to investigate the main effect of congruency. Participants were significantly faster during Congruent trials ($M = 803$ ms, $SE = 18$) than during Neutral trials ($M = 953$ ms, $SE = 21$), $t(75) = 13.15$, $p < .001$, $d = 1.51$. Participants were also significantly faster during Neutral trials than during Incongruent trials ($M = 1021$ ms, $SE = 21$), $t(75) = 4.34$, $p = .001$, $d = 0.49$. Finally, participants were significantly faster during Congruent trials than during Incongruent trials, $t(75) = 15.90$, $p < .001$, $d = 1.82$. Thus, participants responded most rapidly when an auditory stimulus was accompanied by a congruent visual stimulus, followed by a neutral visual stimulus, and were slowest to respond when the visual stimulus was incongruent.

The main effect of age group was examined using Bonferroni-corrected pairwise comparisons. These revealed that the adults were significantly faster than both the 6- and 7-year-olds ($p = .006$, $d = 1.15$) and the 8- and 9-year-olds ($p = .004$, $d = 1.09$). There was no significant difference in average speed of response between the two child age groups.

To explore the interaction between age group and familiarity, Bonferroni-corrected one-way ANOVAs were performed with age group as the between-participants factor independently for Familiar and Newly Learned trials (see Fig. 6). These revealed that the adults were significantly faster than the 6- and 7-year-olds ($p \leq .001$, $d = 1.21$) and the 8- and 9-year-olds ($p = .001$, $d = 1.04$) in the Familiar condition; however, no significant differences between age groups emerged in the Newly Learned

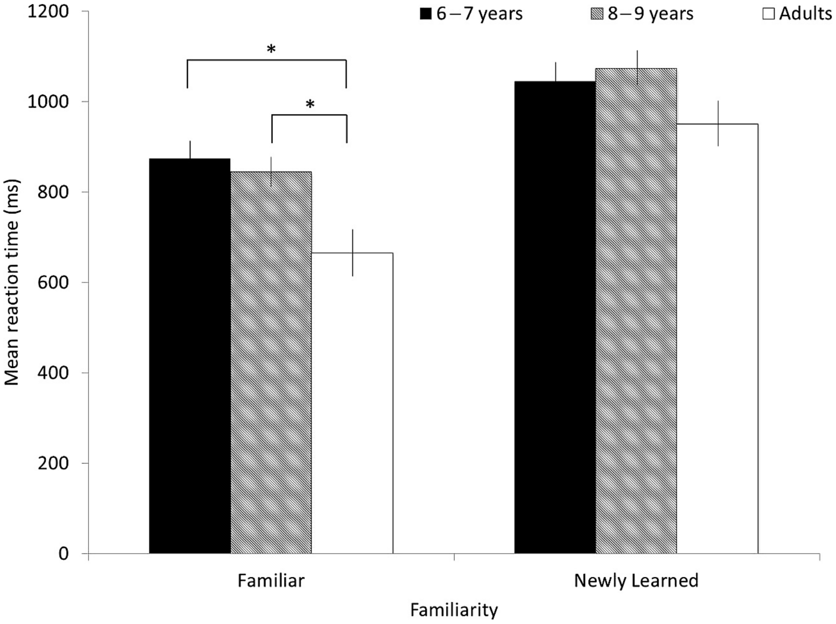


Fig. 6. Mean reaction time to respond to the auditory prompt collapsed across congruency condition for Familiar and Newly Learned animal vocalizations within each age group. Error bars plot standard errors. *Indicates significant difference between Age groups ($p < 0.05$).

condition. Thus, it seems that when the auditory stimulus was Familiar, adults had a significant advantage over children. However, when the auditory stimulus was Newly Learned, performance for adults and children was similar.

To explore the marginally significant interaction between age group and congruency (see Fig. 7), paired-samples *t* tests were performed among congruency conditions independently for each age group. These revealed significant differences in reaction times across all congruency conditions (in the same directions as reported for the main effect of congruency) for the 8- and 9-year-olds and adults ($p < .001$ in all cases). For the 6- and 7-year-olds, significant differences emerged between the Congruent and Incongruent trials ($p < .001$, $d = 1.20$) as well as between the Congruent and Neutral trials ($p < .001$, $d = 1.23$). However, the difference in reaction times to Incongruent and Neutral trials was not significant, $t(25) = 0.405$, $p = .689$, $d = 0.079$. Thus, it seems that the older age groups are advantaged by a congruent visual stimulus and are disadvantaged by an incongruent visual stimulus. In contrast, at 6 or 7 years of age, children are advantaged by a congruent visual stimulus but are not significantly disadvantaged by an incongruent visual stimulus (as compared with a neutral one). This may suggest that they are able to suppress irrelevant visual information; however, it could also be the case that the "neutral" image is equally as interfering for younger children as the semantically incongruent image.

Finally, to explore the interaction between familiarity and congruency, a Newly Learned delay was calculated (Newly Learned reaction time − Familiar reaction time) for each congruency condition (see Fig. 8). Paired-samples *t* tests were performed among congruency conditions on these scores. Newly Learned delay was significantly smaller in the Congruent condition than in the Incongruent condition, $t(75) = 2.534$, $p = .013$, $d = 0.32$. Likewise, Newly Learned delay was significantly smaller in the Congruent condition than in the Neutral condition, $t(75) = 2.752$, $p = .007$, $d = 0.39$. The difference in Newly Learned delay between the Incongruent and Neutral conditions did not reach significance, $t(75) = 0.657$, $p = .513$, $d = 0.08$. Thus, it seems that participants benefited more from the auditory stimuli being Familiar when the relationship between the auditory and visual stimuli was either incongruent or neutral. When the relationship between the auditory and visual stimuli was congruent, familiarity had less impact on speed of responses.

In summary, at all ages tested, participants responded most rapidly during trials in which an auditory stimulus was accompanied by a congruent visual stimulus. This was followed by trials in which

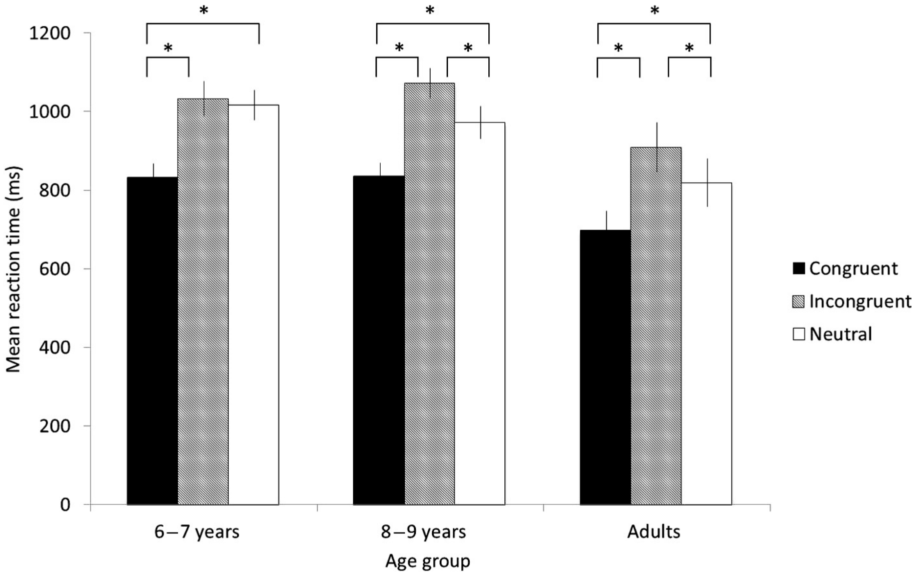


Fig. 7. Mean reaction to the auditory prompt collapsed across Familiarity conditions. Error bars plot standard errors. *Indicates significant difference between Age groups ($p < 0.05$).

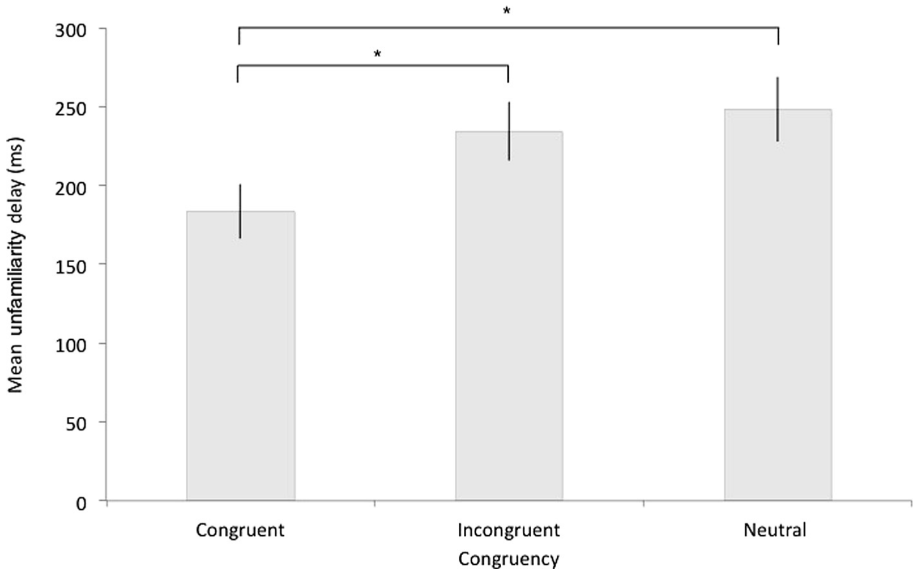


Fig. 8. Mean Newly Learned disadvantage score (Newly Learned–Familiar reaction time). A positive value indicates that participants were faster to respond to Familiar trials than to Newly Learned trials. Error bars plot standard errors. *Indicates significant difference between Age groups (p < 0.05).

the accompanying visual stimulus had no relation to the auditory stimulus. The youngest participants responded at a similar speed when the auditory stimulus was accompanied by either a neutral or incongruent visual stimulus. By 8 or 9 years of age, as well as during adulthood, participants were significantly slower to respond to trials where the auditory stimulus was accompanied by an incongruent visual stimulus.

The same order of reaction times (Congruent, Neutral, and Incongruent) emerged in both the Familiar and Newly Learned conditions, suggesting that participants had rapidly learned the new image and vocalization pairings and that this association influenced their subsequent perception. It seems that even relatively limited experience with these pairings was enough to enhance perception when congruent stimuli were presented and to interfere with perception when incongruent stimuli were presented; however, the extent of this influence was mediated by familiarity with the pairings.

When both children and adults knew the semantic association between the auditory and visual stimuli prior to the experiment, they responded significantly more rapidly than when they had learned these associations during the study. Adults (with the opportunity to acquire extensive experience with the auditory stimuli across their lifetime) outperformed children at all ages when the auditory stimuli were Familiar. However, when all groups had equivalent experience (when the auditory stimuli were unfamiliar at the start of the experiment), this advantage was no longer evident.

Discussion

In the current study, we investigated how children and adults bring together current complementary or conflicting sensory information presented in vision and audition. Multisensory perception in adults seems to go beyond low-level stimulus properties and is also influenced by factors such as the congruency (in terms of both perceptual and semantic features) between the different sensory inputs (e.g., Heron et al., 2004; Jackson, 1953; Slutsky & Recanzone, 2001). Here, we used a Stroop-like paradigm (Stroop, 1935) in which participants were required to attend to and identify a sound while simultaneously being presented with a visual stimulus. We presented participants with audio–visual pairings with which they were familiar and audio–visual pairings that had been introduced during the experiment, allowing us to tease apart two potential trajectories for the development of semantic associations in multisensory perception, namely that (a) the role of semantic associations changes across development and (b) semantic associations become robust with development, leading to a change in their role. Across middle childhood, children appear to undergo a significant shift in their processing of audio–visual information, changing from an auditory bias to an adult-like visual bias (e.g., Nava & Pavani, 2013). This is also the period of time during which children begin to integrate information across the senses in a mature adult-like manner (e.g., Gori et al., 2008; Jaime et al., 2014; Nardini et al., 2008, 2010, 2013, 2016; Petrini et al., 2014). As such, this is a particularly important period of development for us to explore the role of semantic associations across the senses.

We set out to address two key questions.

(1) Do school-aged children benefit from semantically congruent audio–visual information during auditory object recognition, and/or are they disadvantaged by incongruent audio–visual information?

We found that stimuli presented in the task-irrelevant modality (vision) influenced processing of stimuli in the task-relevant modality (audition) from 6 years of age. This influence was demonstrated by differential reaction times across congruency conditions. Interestingly the pattern of reaction times across congruency conditions changed with age. From 8 years, participants showed a significant facilitation effect when presented with a congruent stimulus in the task-irrelevant modality and showed a significant interference effect when presented with an incongruent stimulus in the task-irrelevant modality. At 6 years of age, participants showed a significant facilitation effect, with significantly faster responses in the Congruent condition compared with both the Neutral and Incongruent conditions. However, at this age participants did not demonstrate a classic interference effect; reaction times were very similar in the Neutral and Incongruent conditions. It appears that in this task facilitation effects emerge earlier in development than interference effects (relative to neutral), which might suggest that the way in which semantic information is used changes with development. However, there are a number of possible explanations for the pattern of results that emerged in the youngest age group. One possibility is that the youngest children found the "neutral" visual stimuli and the incongruent visual stimuli to be equally distracting. This would suggest that at this age the influence of a visual distractor is not limited by its semantic association with the auditory stimulus. The fact that children were faster for Congruent trials suggests that they were taking the visual information into account. An alternative explanation of this finding could be that at this age children may process the visual stimuli faster than the auditory stimuli and that it is this difference that allows them to be speeded in the Congruent condition. If this were the case, then performance in the other conditions might reflect children's speed of processing auditory information. Although the auditory dominance literature might predict (although primarily in younger ages) that auditory input is processed more rapidly due to its dynamic and transient nature, there are also studies suggesting that vision is processed preferentially (e.g., Colavita, 1974) or that speed of processing of stimuli is moderated by the stimuli being attended to (for a review, see Spence & Parise, 2010). In light of these varied hypotheses and our finding that "neutral" visual stimuli were not processed differently from incongruent stimuli by the 6-year-olds, future research should include a unisensory baseline. Such a condition would give us further insight into the interaction between the senses and would allow for easier comparisons between this research and the literature investigating sensory dominance. Interestingly, the increase in reaction times from Congruent to Neutral is similar in absolute terms across the age groups. If the youngest children were processing the visual information followed by the auditory information in serial order, then one might expect that the interference effects they experienced would be enhanced, which was not the case. In addition, children in this age group do not seem to be preferentially processing auditory information over visual information (as might be predicted by the auditory overshadowing literature; e.g., Massaro et al., 1986; Napolitano & Sloutsky, 2004; Nava & Pavani, 2013; Robinson & Sloutsky, 2004) because this should lead to similar reaction times irrespective of congruency condition.

(2) With development, does knowledge of semantic associations play an increasing role in combining sensory stimuli, or is this influence constant and instead the semantic associations become more robust?

The same order of reaction times across congruency conditions (Congruent, Neutral, and Incongruent) was observed for semantic associations introduced during the course of the experiment as well as for previously established associations. It appears that these associations were readily acquired throughout the course of the experiment and had an immediate impact on processing of the task-relevant stimulus.

The extent of experience of semantic associations had a main effect on average reaction time, with faster reactions being demonstrated for previously acquired (Familiar) associations. In addition, the effect of extent of experience interacted with age. Adults (with extensive experience of previously acquired associations) were able to significantly outperform children in terms of reaction times when the auditory stimulus was Familiar to them. However, when adults and children had equivalent experience of the auditory stimulus, this speed advantage was negated. Given that developmental changes were observed only for Familiar audio–visual pairings, it appears that these changes are caused by experience-driven increases in the strength of associations rather than by maturational change in the role of semantic associations in multisensory processing. This finding is important to consider when comparing multisensory performance across development with performance during adulthood because it suggests that children could underperform not just because their perceptual system may be immature but also because they have less experience with the stimulus pairings. This finding also suggests that, despite children and adults being able to rapidly form associations presented across their senses, such associations continue to mature over a protracted period. The fact that age interacted with familiarity suggests that the age effects we found cannot be explained away solely by more broad developmental changes such as general improvements in inhibitory skills. As such, it seems that the influence of semantic associations on multisensory processing remains constant across the tested age range, but the associations themselves become more robust.

Because the three-way interaction did not approach significance, we did not follow this up in the Results section; however, here we believe that it is relevant to consider whether adults' speed in the Familiar trials could be attributed to certain congruency conditions. Within Familiar trials, adults were faster than children across all congruency conditions; this might suggest that adults are speeded due to their experience with the auditory information given that congruency of visual stimulus did not influence the effect.

In conclusion, as has been suggested in previous studies, visual information can have a potent effect on processing in other domains (e.g., Yuval-Greenberg & Deouell, 2009) even if this information is in a task-irrelevant modality and is semantically incongruent. In addition, visual information can have a facilitatory effect on processing of auditory stimuli even if the semantic association between these signals is newly acquired. We found this pattern of results from 8 years of age. It appears that the influence of visual information on auditory object recognition might change between 6 and 8 years.

Task-irrelevant visual information influenced the processing of auditory stimuli from 8 years of age. This influence occurred for both Newly Learned and Familiar audio–visual pairings. At this age, children experienced significant facilitatory and inhibitory influences from vision on audition, suggesting that they were not able to modify the use of visual information, depending on its semantic relation to the auditory stimulus. Similar patterns of reaction times were evident for 8-year-olds and adults, suggesting that despite ongoing changes in multisensory processing across this age range (e.g., Gori et al., 2008; Nardini et al., 2008), there is some continuity in audio–visual interactions between 8 years of age and adulthood. The mechanisms behind these effects may vary across this age range, but the result in terms of relative speed of processing is comparable.

Acknowledgements

This work was funded by an ESRC PhD studentship awarded to RLT and ESRC grant RES-062-23-0819. We would like to thank all the children, parents and teachers who made this work possible.

References

Adams, W. J. (2016). The development of audio–visual integration for temporal judgements. *PLoS Computational Biology, 12*(4), e1004865.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262.

Arias-Trejo, N., & Plunkett, K. (2009). Lexical–semantic priming effects during infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 3633–3647.

Bjorklund, D. F. (1987). How age changes in knowledge base contribute to the development of children's memory: An interpretative review. *Developmental Review, 7*(2), 93–130.

Bjorklund, D. F. (1985). The role of conceptual knowledge in the development of organization in children's memory. In C. J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development* (pp. 103–142). New York: Springer.

Bremner, J. G., Slater, A. M., Johnson, S. P., Mason, U., & Spring, J. (2012). The effects of auditory information on 4-month-old infants' perception of trajectory continuity. *Child Development, 83*, 954–964.

Chen, Y. C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition, 114*, 389–404.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*, 409–412.

Comalli, P., Jr., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color–word test in childhood, adulthood, and aging. *Journal of Genetic Psychology: Research and Theory on Human Development, 100*, 47–53.

Constantidou, F., Danos, M., Nelson, D., & Baker, S. (2011). Effects of modality presentation of working memory in school-age children: Evidence for the pictorial superiority hypothesis. *Child Neuropsychology, 17*, 173–196.

Cowan, N., & Barron, A. (1987). Cross-modal, auditory–visual Stroop interference and possible implications for speech memory. *Perception & Psychophysics, 41*, 393–401.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4–13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037–2078.

Eckner, J. T., Kutcher, J. S., & Richardson, J. K. (2010). Pilot evaluation of a novel clinical test of reaction time in National Collegiate Athletic Association Division I football players. *Journal of Athletic Training, 45*, 327–332.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Gerstadt, C., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like day–night test. *Cognition, 53*, 129–153.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology, 18*, 694–698.

Hanauer, J., & Brooks, P. (2003). Developmental change in the cross-modal Stroop effect. *Perception & Psychophysics, 65*, 359–366.

Hanauer, J., & Brooks, P. (2005). Contributions of response set and semantic relatedness to cross-modal Stroop-like picture–word interference in children and adults. *Journal of Experimental Child Psychology, 90*, 21–47.

Heikkilä, J., & Tiippana, K. (2016). School-aged children can benefit from audiovisual semantic congruency during memory encoding. *Experimental Brain Research, 234*, 1199–1207.

Heron, J., Whitaker, D., & McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio–visual interaction. *Vision Research, 44*, 2875–2884.

Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. New York: John Wiley.

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual–auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology. Human Perception and Performance, 20*, 131–153.

Innes-Brown, H., Barutchu, A., Shivdasani, M. N., Crewther, D. P., Grayden, D. B., & Paolini, A. G. (2011). Susceptibility to the flash-beep illusion is increased in children compared to adults. *Developmental Science, 14*, 1089–1099.

Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology, 5*, 52–65.

Jaime, M., Longard, J., & Moore, C. (2014). Developmental changes in the visual-proprioceptive integration threshold of children. *Journal of Experimental Child Psychology, 125*, 1–12.

Jordan, K. E., & Baker, J. (2011). Multisensory information boosts numerical matching abilities in young children. *Developmental Science, 14*, 205–213.

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioural performance. *Experimental Brain Research, 158*, 405–414.

Lecanuet, J. P., & Schaal, B. (1996). Fetal sensory competencies. *European Journal of Obstetrics, Gynecology and Reproductive Biology, 68*, 1–23.

Lehmann, S., & Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research, 24*, 326–334.

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

Maguinness, C., Setti, A., Burke, K. E., Kenny, R. A., & Newell, F. N. (2011). The effect of combined sensory and semantic components on audio–visual speech perception in older adults. *Frontiers in Aging Neuroscience, 3*. http://dx.doi.org/10.3389/fnagi.2011.00019.

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology, 41*, 93–113.

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

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory processes: A balancing act across the lifespan. *Trends in Neurosciences, 39*, 567–579.

Napolitano, A. C., & Sloutsky, V. M. (2004). Is a picture worth a thousand words? The flexible nature of modality dominance in young children. *Child Development, 75*, 1850–1870.

Nardini, M., Bales, J., & Mareschal, D. (2016). Integration of audio–visual information for spatial decisions in children and adults. *Developmental Science, 19*, 803–816.

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of American, 107*, 17041–17046.

Nardini, M., Begus, K., & Mareschal, D. (2013). Multisensory uncertainty reduction for hand localization in children and adults. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 773–787.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology, 18*, 689–693.

Nava, E., & Pavani, F. (2013). Changes in sensory dominance during childhood: Converging evidence from the Colavita effect and the sound-induced flash illusion. *Child Development, 84*, 604–616.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science, 9*, 454–464.

Petrini, K., Remark, A., Smith, L., & Nardini, M. (2014). When vision is not an option: Children's integration of auditory and haptic information is suboptimal. *Developmental Science, 17*, 376–387.

Prevor, M., & Diamond, A. (2005). Color–object interference in young children: A Stroop effect in children 3½–6½ years old. *Cognitive Development, 20*, 256–278.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development, 75*, 1387–1401.

Schafer, G. (2005). Infants can learn decontextualized words before their first birthday. *Child Development, 76*, 87–96.

Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Science, 6*, 233–241.

Senkowski, D., Saint-Amour, D., Kelly, S. P., & Foxe, J. J. (2007). Multisensory processing of naturalistic objects in motion: A high-density electrical mapping and source estimation study. *NeuroImage, 36*, 877–888.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport, 12*(1), 7–10.

Spence, C., & Parise, C. (2010). Prior-entry: A review. *Consciousness and Cognition, 19*, 364–379.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience, 8*, 497–506.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Styles, S. J., & Plunkett, K. (2009). How do infants build a semantic system? *Language and Cognition, 1*(1), 1–24.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*, 400–410.

Vogler, J., & Titchener, K. (2011). Cross-modal conflicts in object recognition: Determining the influence of object category. *Experimental Brain Research, 214*, 597–605.

Wright, I., Waterman, M., Prescott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function: Typical developmental trends. *Journal of Child Psychology and Psychiatry, 44*, 561–575.

Yuval-Greenberg, S., & Deouell, L. (2009). The dog's meow: Asymmetrical interaction in cross-modal object recognition. *Experimental Brain Research, 193*, 603–614.