

Semantic Congruency Across Development

Susan Geffen (sgeffen@oxy.edu)

Department of Psychology, Occidental College
1600 Campus Rd., Los Angeles CA 90041, USA

Taylor Beck (tbeck@oxy.edu)

Department of Psychology, Occidental College
1600 Campus Rd., Los Angeles CA 90041, USA

Ainsley Shelsta (shelsta@oxy.edu)

Department of Cognitive Science, Occidental College
1600 Campus Rd., Los Angeles CA 90041, USA

Christopher W. Robinson (robinson.777@osu.edu)

Department of Psychology, The Ohio State University at Newark
1179 University Dr., Newark, OH 43056, USA

Abstract

Multisensory processing often results in facilitation and/or interference effects, yet the mechanisms remain unclear. The two reported experiments used a Stroop-like task to examine how congruent, incongruent, and irrelevant information presented in one sensory modality (e.g., visual) affects processing and responding in a different modality (e.g., auditory). Across two experiments, adults (E1) and 5-year-olds (E2) were presented with pictures and sounds, and they had to determine if what they saw or heard was an animal or vehicle. Experiment 1 with adults showed evidence of both facilitation on congruent trials and interference on incongruent trials, with unattended visual stimuli having a larger effect on auditory processing than vice versa. Results in 5-year-olds were slightly more symmetrical than adults but there was no evidence that auditory input dominated visual processing. Possible mechanisms underlying these effects are discussed.

Keywords: Crossmodal Processing; Multisensory Integration; Modality Dominance; Semantic Congruency

Introduction

Many of our daily tasks require the processing of both auditory and visual stimuli. For example, when you encounter a stimulus, like a dog in the environment, you not only see it, but you may hear it as well. Much of the early research on perception and attention initially focused on only single sensory domains. However, within the last 40 years, researchers have started examining multisensory processing with a specific focus on the intersection of auditory and visual domains (see for example Calvert et al. 2004; Colavita, 1974; Driver & Spence, 2004). While many studies have documented evidence of cross-modal facilitation and interference, mechanisms underlying semantic congruency (facilitation) and modality dominance (interference) remain unclear.

Semantic congruency, or harmony between stimuli (i.e., a train paired with a train whistle) can facilitate processing (e.g., Laurienti et al., 2004; Thomas et al., 2017), whereas semantic incongruency (e.g., a dog paired with a train whistle) can interfere with processing (e.g., Wright et al., 2003). For example, participants demonstrate faster (visual) recognition and responding (Boutonnet & Lupyan, 2015; Lupyan & Thompson-Schill, 2012), and greater accuracy when responding to congruent versus incongruent stimuli (Thomas et al., 2017). These findings show that congruency matters, however, proper controls are needed in behavioral studies to determine if differences stem from congruent multisensory stimuli being encoded more efficiently or from responses being faster on congruent trials because both modalities are associated with the same response. The same issues are presented when examining interference, and it is unclear if these effects stem from one modality disrupting encoding in the other modality or from post-perceptual factors with slower decisions because the auditory stimulus elicits one response and the visual stimulus elicits a different response.

In addition to not fully understanding the underlying mechanisms, multisensory interactions appear to change across age, with younger children being more influenced by auditory information (Nava & Pavani, 2013; Robinson & Sloutsky, 2004; Sloutsky & Napotlitano, 2003) and adults being more influenced by visual information (Colavita, 1974; Sinnett, Spence, & Soto-Faraco, 2007; see also Spence, Parise, & Chen, 2012; Robinson & Sloutsky, 2010, for reviews). Developmental shifts from auditory to visual modalities may stem from the slow protracted development of the visual system, changes in top-down control of attention, or increased focus on the visual modality due to a focus on reading (see Robinson & Sloutsky, 2004, for a review). However, it is important to note that, under some conditions, adults also show evidence of auditory dominance

(Dunifon, Rivera, & Robinson, 2016; Robinson, Chandra, & Sinnott, 2016), thus, it is important to study which testing contexts show the auditory to visual shift across development and which contexts show consistent auditory (or visual) dominance across the lifespan.

The current experiments used a Stroop-like task (Stroop, 1935) to better understand mechanisms underlying modality dominance (interference on incongruent trials) and semantic congruency effects (more efficient processing on congruent trials). Participants were presented with multisensory information that was either congruent, incongruent, or irrelevant, and they had to quickly determine if the auditory or visual information was an animal or a vehicle (see Figure 1 for trial types). In contrast to previous research, the current studies examined both the effects of auditory stimuli on visual processing and visual stimuli on auditory processing. Thus, these experiments examined which modality (auditory or visual) was more dominant at various points in development.

Based on semantic congruency effects, responding to congruent stimuli should be faster and more accurate than responding to incongruent stimuli. Comparisons to unimodal controls (just picture or sound) will determine if these effects stem from congruent stimuli being processed more efficiently or from incongruent stimuli hindering processing. Comparing semantic and response congruent trials to response congruent trials will determine if semantic congruency has any additional effects over both stimuli being associated with the same response. Regarding modality dominance effects, stronger auditory effects would be expected if auditory stimuli automatically engage attention and delay visual processing (Robinson & Sloutsky, 2010), whereas, stronger visual effects would be consistent with visual dominance research in adults, with sensory systems being inhibitory in nature and approximately 50% of the brain dedicated to vision (Desimone & Duncan, 1995; Duncan, 1996; Spence et al., 2012). Finally, comparing adults (Experiment 1) and 5-year-olds (Experiment 2) will determine if semantic congruency tasks show the auditory to visual shifts across development or relatively stable effects with age.

Experiment 1

Method

Participants and Design Forty-five adults completed the study. Data were excluded if participants missed at least half of the attention checks during the auditory block (n=4) or experimenter error (n=5), leaving 36 participants in the final sample. All participants were recruited from Occidental College or The Ohio State University at Newark (OSU), all self-reported normal hearing and normal or corrected to normal vision, and all were over 18. Participants received course credit and provided consent before the start of the experiment.

Apparatus For students who completed the study at OSU, the experiment was conducted on a 22" Dell PXL 2230 MW monitor with 1920 x 1080 resolution and Dell Optiplex 7040 systems with Intel Core i5 processors. Bose QuietComfort 25 Noise Cancelling headphones were used for auditory stimulus presentation. At Occidental, students completed the study on a personal device. At both schools, stimulus timing and reaction time/accuracy data were collected using Pavlovia software (<https://pavlovia.org/>).

Materials Visual stimuli for the experiment consisted of images of common animals (e.g., dog, elephant), vehicles (e.g., train, motorcycle), household objects (e.g., broom, banana), as well as green circles which served as a visual attention check (see Figure 1 for examples). Visual stimuli were presented centrally on a computer monitor at approximately 600 x 600 resolution. Auditory stimuli consisted of short clips of the corresponding animal/vehicle/everyday sounds that were taken from the Marcell sound library (Marcell et al., 2020) and edited to be approximately 500 ms. The sounds were presented via headphones at approximately 65 dB. Auditory and visual stimuli had the same onset and offset. There were also six attention checks per block (12 total) which consisted of a green circle that appeared in the center of the screen with no accompanying audio. Participants had to hit this spacebar on these trials. This ensured that participants stayed focused on the screen throughout the entire experiment

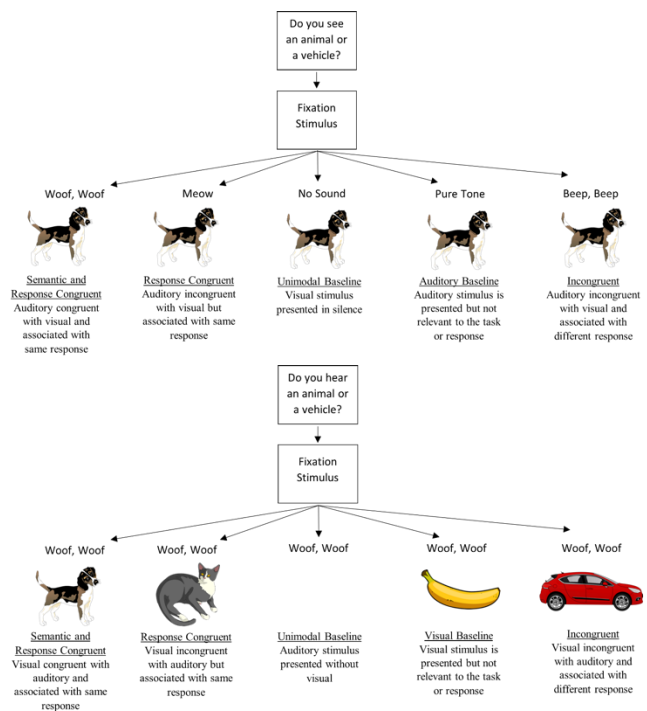


Figure 1. Examples of Attend to Visual trials (top) and Attend to Auditory trials (bottom). The correct response in all examples is animal.

Design The experiment consisted of two factors: attended modality (auditory, visual) and congruency (semantic and response congruent, response congruent, unimodal baseline, irrelevant baseline, incongruent), with both factors being manipulated within subjects. Attended modality was blocked and the order was counter-balanced, and the different congruency trials were randomized within each block. There were 48 trials for each congruency type with 240 attend to auditory trials and 240 attend to visual trials.

Procedure All participants completed the study at either Occidental College or The Ohio State University at Newark. Participants completed the study in a quiet room with minimal distractions and read through a series of instructions independently prior to the first trial of each block. Participants were instructed to press the p or q buttons on the keyboard to indicate if the stimulus was an animal or a vehicle. Button pairings were counterbalanced across subjects. Participants were given the option to take a break halfway through the experiment between the two blocks. The whole procedure lasted approximately 30 minutes.

Results and Discussion

Participants detected 96% of the attention check trials in the auditory report condition and 88% in the visual report condition, thus, participants in both conditions were attending to the target and non-target modalities. Accuracies (Figure 2A) and Median Response Times (RT) on correct trials (Figure 2B) were submitted to two separate Modality (2) x Trial Type (5) repeated-measures ANOVAs.

Regarding accuracies, the Modality x Trial Type repeated-measures ANOVA revealed a main effect of Trial Type, $F(4, 140) = 16.24, p < .001$, and post-hocs with Bonferroni adjustments revealed that accuracy on Incongruent trials ($M = .86, SE = .02$), was significantly lower than all of the other Trial Types ($M's > .92$), $p's < .001$. The analysis also revealed a Modality x Trial type interaction, $F(4, 140) = 3.90, p < .001$. As shown in Figure 2A, simple main effects revealed that the different auditory trial types had no effect on visual processing, $p = .206$ (see right side of Figure 2A). In contrast, the different visual trials had an effect on auditory processing, $p = .206$ (see left side of Figure 2A), with Incongruent trials differing from the other trial types.

Analyses examining median RTs point to the same processing asymmetry, with visual stimuli having a larger effect on auditory processing than vice versa. The Modality x Trial Type repeated-measures ANOVA revealed main effects of Modality, $F(1, 35) = 198.56, p < .001$, Trial Type, $F(4, 140) = 22.56, p < .001$, and a Modality x Trial Type interaction, $F(4, 140) = 17.65, p < .001$. As shown in Figure 2B, visual RTs ($M = 510.67, SE = 18.06$) were faster than auditory RTs ($M = 790.88, SE = 22.90$). Post hoc tests with Bonferroni adjustments revealed that Semantic and Response Congruent trials differed from the other trial types and Response Congruent trials also differed from the other trial types, $p's < .051$.

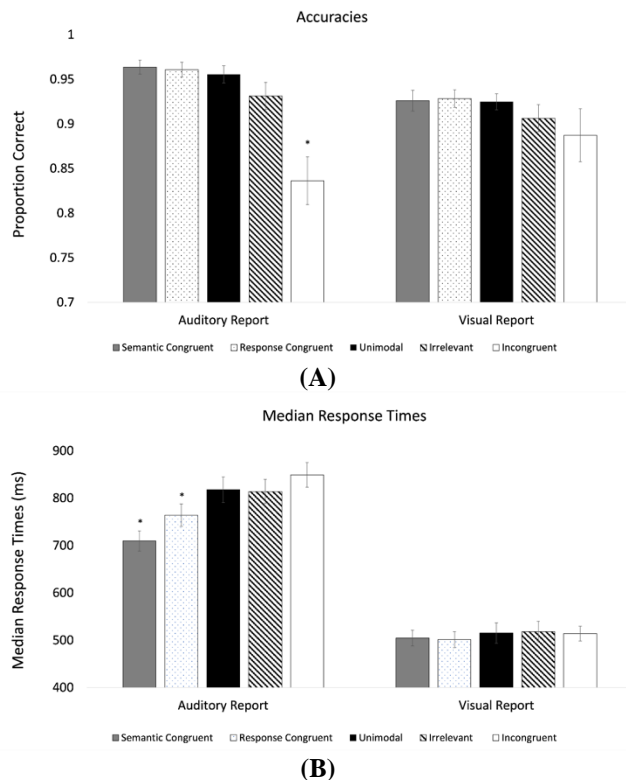


Figure 2. Proportion of correct responses (A) and median response times (B) across trial types and conditions in Experiment 1. Solid bars are associated with congruent, unimodal, and incongruent trials and patterned bars are controls. Error bars denote standard errors, and "*" denotes a significant difference from unimodal trials, $p < .001$.

In summary, the data point to visual dominance with unattended visual stimuli having a larger effect on auditory processing than vice versa. Congruent visual stimuli sped up responses while incongruent visual stimuli decreased accuracy. It is important to note that the response congruent trials also differed from the unimodal baseline, suggesting that some of the semantic congruency effects stem from cross-modal stimuli sharing the same response and semantic congruency adding to these effects.

Experiment 2

The purpose of Experiment 2 was to examine semantic congruency and incongruency effects in young children. Based on the developmental research using a variety of tasks (Nava & Pavani, 2013; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003), it was hypothesized that auditory stimuli will have larger effects in young children.

Method

Participants, Materials, and Procedure Twenty-three children ($M = 5.05$ years, Females = 11) participated in the current study. Children were recruited through Children

Helping Science (<https://childrenhelpingscience.com>) and word of mouth, they completed the study online using Pavlovia, and received a \$10 gift card for their participation. An additional seven children were tested but not included because they were not at least 75% accurate on unimodal baseline trials.

To simplify the task, the number of trials were reduced to 12 trials for each congruency type for a total of 120 trials (60 attend to auditory and 60 attend to visual). The experiment was also conducted over Zoom, so the experimenter could make sure children were focused on the screen throughout the procedure. Thus, the green circle trials were also removed to simplify the task. Finally, in contrast to Experiment 1 where participants responded by pressing p or q on a keyboard, children made animal and vehicle responses by touching on-screen buttons (touch screen devices) or by using a mouse to click on the buttons (non-touch screen devices). While the overall procedure was comparable for children and adults, there were also enough differences to avoid direct comparisons.

Results and Discussion

As in Experiment 1, we used two separate Modality (2) x Trial Type (5) repeated-measures ANOVAs to examine how modality and congruency affect accuracy and response times. See Figure 3A for means and Standard Errors. The analysis revealed a main effect of Trial Type, $F(4, 88) = 15.28, p < .001$, and post-hocs with Bonferroni adjustments revealed that accuracy on Incongruent trials ($M = .75, SE = .04$), was significantly lower than all of the other Trial Types (M 's $> .87$), p 's $< .002$. The analysis also revealed a marginally significant effect of modality, $F(1, 22) = 3.58, p = .072$, with visual accuracy ($M = .91, SE = .02$) being slightly greater than auditory accuracy ($M = .86, SE = .02$). The Modality x Trial type interaction did not reach significance, $p = .108$.

The analysis examining median response times revealed a main effect of Trial Type, $F(4, 88) = 2.72, p = .035$, and post-hocs with Bonferroni adjustments revealed that response times on Incongruent trials ($M = 2821$ ms, $SE = 340$), was slower than Semantic and Response Congruent trials ($M = 2143$ ms, $SE = 149$), $p = .048$. The Modality x Trial type interaction was marginally significant, $F(4, 88) = 2.15, p = .081$.

In summary, while effects of incongruent auditory stimuli on visual processing appeared to have a stronger effect in children than adults (see Figures 2A & 3A), there appeared to be no evidence that auditory stimuli dominated visual processing.

General Discussion

Many situations require the simultaneous processing of auditory and visual information. Under some conditions, multisensory presentation can facilitate processing, and in other conditions, it can interfere with processing. The current study used a modified Stroop task to examine these effects in

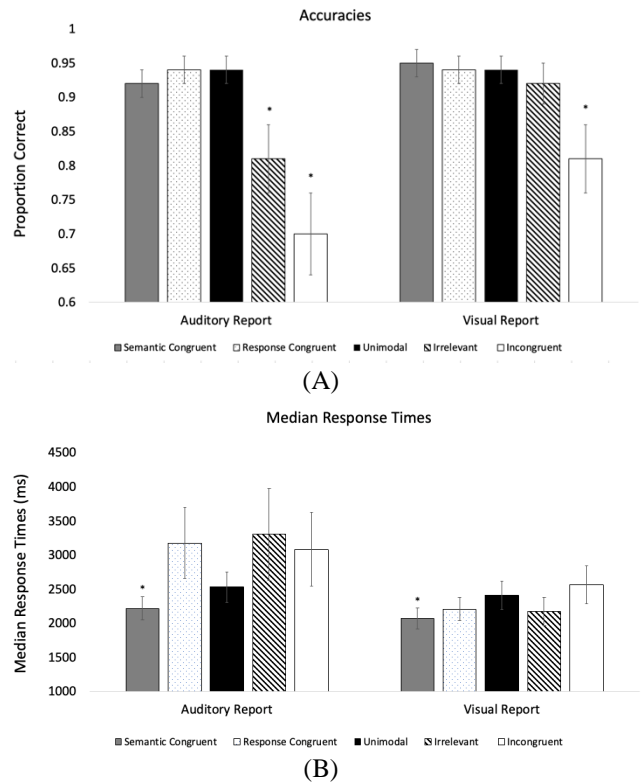


Figure 3. Proportion of correct responses (A) and median response times (B) across trial types and conditions in Experiment 2. Solid bars are associated with congruent, unimodal, and incongruent trials and patterned bars are controls. Error bars denote standard errors, and "*" denotes a significant difference from unimodal trials, $p < .05$ (one-tailed).

children and adults. The accuracy and response time data in adults show that visual input had a larger effect on auditory processing, than vice versa. Five-year-old data were slightly more symmetrical in nature, but they showed a similar pattern as adults.

One important finding is that semantically congruent stimuli sped up response times, and the response congruent trials also sped up responding compared to the unimodal baseline (see Figure 2A). This finding suggests that some of the faster responses on semantic congruent trials stem from auditory and visual stimuli being associated with the same response. However, semantic congruency also had an effect. Note that response times to semantic and response congruent stimuli were also faster than the unimodal baseline and they also differed from the response congruent trials, which suggests that shared response and semantic congruency both affect processing speed.

The current findings also challenge the notion that auditory stimuli automatically engage attention and delay visual processing, a proposed mechanism underlying auditory dominance (Robinson & Sloutsky, 2010). There are several issues that will need to be addressed in future research. First,

familiarity seems to play a role in children's modality dominance effects (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004). When stimuli are familiar, visual stimuli often have larger effects on auditory processing, whereas, auditory dominance effects are more likely to occur when presented with novel stimuli. Thus, future research will need to use a wider range of categories while systematically manipulating familiarity to determine if these effects stem from general processing factors or from more specific factors associated with the stimuli used in the current research.

The current study was also conceptual in nature, with no need to process the details of the stimuli. In contrast, many of the studies supporting auditory dominance in children and adults require the processing of stimulus details (Dunifon, Rivera, & Robinson, 2016). Thus, it is possible that visual stimuli dominate tasks that are more top-down in nature and auditory dominance effects are more likely to occur in perceptual tasks where processing of stimulus details is important. Future research will need to expand on the semantic congruency task to see if the same pattern of results will be found using low-level perceptual categories.

Finally, one other factor needs to be addressed to better understand developmental changes in semantic congruency tasks. The current study compared 5-year-olds with adults, whereas, studies have found significant changes across a wider age range (e.g., Nava & Pavani, 2013). Thus, including younger and older children would be important to capture the developmental trend.

The current studies expanded on multisensory processing by examining the impact of modality dominance and semantic congruence at various points in development. Preliminary results suggest that congruency has a fairly stable effect across the lifespan, decreasing response times relative to unimodal baseline, and suggesting congruent stimuli are processed more efficiently. These experiments also expanded on previous research by examining both the effects of auditory stimuli on visual processing and visual stimuli on auditory processing. The visual dominance demonstrated by adults paired with the somewhat symmetrical data for 5-year-olds suggests that, contrary to our hypothesis, there are relatively stable effects of visual dominance with age. While multiple factors may influence multisensory processing, these results suggest that these influences take root early and remain consistent across the lifespan.

References

- Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience*, 35(25), 9329-9335.
- Calvert, G., Spence, C., & Stein, B. (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.
- Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, 16, 409-412.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193-222.
- Driver, J., & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention*. Oxford, UK: Oxford University Press.
- Duncan, J. (1996). Cooperating brain systems in selective perception and action. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 549 – 578). Cambridge, MA: The MIT Press.
- Dunifon, C. M., Rivera, S., & Robinson, C. W. (2016). Auditory stimuli automatically grab attention: Evidence from eye tracking and attentional manipulations. *Journal of Experimental Psychology: Human Perception and Performance*, 42(12), 1947.
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental brain research*, 158, 405-414.
- Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, 141(1), 170.
- Marcell, M. M., Borella, D., Greene, M., Kerr, E., & Rogers, S. (2000). Confrontation Naming of Environmental Sounds. *Journal of Clinical and Experimental Neuropsychology*, 22, 830 - 864.
- 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(2), 604-616.
- Robinson, C.W., Chandra, M. & Sinnett, S. (2016). Existence of competing modality dominances. *Attention, Perception, & Psychophysics*, 78, 1104–1114 . <https://doi.org/10.3758/s13414-016-1061-3>
- Robinson, C. W., & Sloutsky, V. M. (2010). Development of cross-modal processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 135-141.
- Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, 75, 1387-1401.
- Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. *Open Mind*, 1(1), 4-14. https://doi.org/10.1162/OPMI_a_00002
- Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual Fdominance and attention: Revisiting the Colavita effect. *Perception & Psychophysics*, 69, 673–686.
- Sloutsky, V.M., & Napolitano, A. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, 74(3), 822-833.
- Spence, C., Parise, C., & Chen, Y. C. (2012). The Colavita visual dominance effect. In M.M. Murray, & M.T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (pp. 529-556). Boca Raton, FL: CRC Press.

- Stroop J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–62.
- Thomas, R.L., Nardini, M., & Mareschal, D. (2017) The impact of semantically congruent and incongruent visual information on auditory object recognition across development. *Journal of Experimental Child Psychology*, *162*, 72-88.
- 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.