RUNNING HEAD: REVERSING THE COLAVITA EFFECT

Reversing the Colavita Visual Dominance Effect

Mary Kim Ngo[1], Michelle L. Cadieux[2], Scott Sinnett[3],

Salvador Soto-Faraco[4,5], & Charles Spence[1]

[1]Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, Oxford, UK

[2]Department of Psychology, Neuroscience & Behaviour, McMaster University, Ontario, Canada

[3]Department of Psychology, University of Hawaii at Manoa, Hawaii, USA

[4]Multisensory Research Group, Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra, Barcelona, Spain

[5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

RESUBMITTED TO: *Experimental Brain Research* (JULY, 2011)

CORRESPONDENCE TO: Mary Kim Ngo, Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, South Parks Road, OX1 3UD, Oxford, UK. Tel.: +44 (01865) 271307. E-mail: thuan.ngo@psy.ox.ac.uk.

ABSTRACT

Many researchers have taken the Colavita effect to represent a paradigm case of visual dominance. Broadly defined, the effect occurs when people fail to respond to an auditory target if they also have to respond to a visual target presented at the same time. Previous studies have revealed the remarkable resilience of this effect to various manipulations. In fact, a reversal of the Colavita visual dominance effect (i.e., auditory dominance) has never been reported. Here, we present a series of experiments designed to investigate whether it is possible to reverse the Colavita effect when the target stimuli consist of repetitions embedded in simultaneously-presented auditory and visual streams of stimuli. In line with previous findings, the Colavita effect was still observed for an immediate repetition task but, when an N-1 repetition detection task was used, a reversal of visual dominance was demonstrated. These results suggest that masking from intervening stimuli between N-1 repetition targets was responsible for the elimination and reversal of the Colavita visual dominance effect. They further suggest that varying the presence of a mask (pattern, conceptual, or absent) in the repetition detection task gives rise to different patterns of sensory dominance (i.e., visual dominance, an elimination of the Colavita effect, or even auditory dominance).

Reversing the Colavita visual dominance effect

Colavita (1974) has been popularly credited with first reporting the phenomenon whereby people frequently fail to respond to an auditory stimulus if they have to respond to a simultaneously presented visual stimulus. This phenomenon is now commonly referred to as the Colavita visual dominance effect. Researchers studying the Colavita effect have expanded the original demonstration from simple stimuli (beeps and flashes) to include streams of common sounds and pictures, and, for more than three decades attempted to eliminate, or reverse, this particular example of visual dominance (i.e., and show auditory dominance; see Spence, 2009, for a review)[1]. Attempts to do this have involved increasing the intensity of the auditory stimulus (Colavita, 1974), increasing the probability of bimodal targets from 20%, as is typical of most earlier studies on the Colavita effect, to 90% (Koppen & Spence, 2007a), varying the stimulus onset asynchrony (SOA) between the visual and auditory stimuli (Koppen & Spence, 2007b), manipulating the spatial coincidence between the visual and auditory stimuli (Koppen & Spence, 2007d), and shifting the relative focus of attention to one modality versus the other (Sinnett, Spence, & Soto-Faraco, 2007). Despite these numerous attempts, only a few studies have been able to eliminate the Colavita effect, while a *reversal* of the effect has yet to be observed[2].

Sinnett et al. (2007) reported that when the frequency of auditory targets (60%), was greater than visual or bimodal targets (both 20%) in a pre-specified target detection task, participants made just as many auditory as visual responses (i.e., errors) to bimodal targets. That is, the Colavita visual dominance effect was eliminated. Sinnett et al. attributed this result to the introduction of a bias towards attending and responding to the auditory stimuli. Similarly, Moro and Steeves (2010) recently reported that unilaterally enucleated patients (i.e., those individuals who had had early unilateral surgical eye removal due to cancer of the retina), who presumably rely less on vision than other normally-sighted individuals, made just as many unimodal auditory responses as visual responses to bimodal targets in a task requiring the participants to respond to pre-specified targets, again demonstrating an elimination of the Colavita visual dominance effect. Assuming that such patients would rely less on vision and more on audition (Lessard, Pare, Lepore, & Lassonde, 1998), Moro and Steeves concluded that in order to eliminate the Colavita effect there must be some form of degradation in the reliability of the visual information presented to participants. Under such conditions, the enucleated patients may have biased their attention, and in turn their responding, toward the auditory modality. That is, while the percentage of auditory-only responses was approximately equal between both groups, the enucleated patients made significantly fewer visual-only responses to bimodal targets as compared to the normally-sighted participants.

Ngo, Sinnett, Soto-Faraco, and Spence (2010) recently examined the Colavita effect in the context of a repetition detection task. People typically find it easier to detect auditory repetitions than to detect visual repetitions (e.g., Soto-Faraco & Spence, 2002). Moreover, the modality appropriateness hypothesis (Welch & Warren, 1980, 1986) contends that audition is better suited for temporally demanding tasks such as detecting repetitions. As such, Ngo et al. hypothesized that the temporal demands of the repetition detection task should have led to a bias toward responding to the auditory rather than the visual information. This may ultimately have resulted in the elimination, or even the reversal, of the Colavita effect. In their study, participants had to monitor a stream of simultaneously-presented visual and auditory stimuli and respond whenever they saw or heard the immediate repetition of a picture, a sound, or both. Note that while all of the previously-mentioned studies have been based on the simple detection of predefined targets much like Colavita's (1974) original study[3], the repetition detection task relies on a more abstract, rule-based level of representation. That is, the task required participants to hold the representation of an object in working memory (see

Baddeley, 1992, 2000, for a review) and to make a comparison with a subsequently-presented object. As expected on the basis of the previous repetition detection literature, auditory superiority was observed on unimodal target trials. That is, participants missed significantly more unimodal visual than unimodal auditory repetitions. However, despite this unimodal auditory dominance, a strong visual dominance effect for bimodal targets, with participants making more visual-only responses (i.e., errors) than auditory responses on these bimodal trials, was nevertheless still observed. It would therefore appear that the temporal demands of the repetition detection task used in their study were insufficient to bring about a bias toward responding preferentially to stimuli presented in the auditory modality.

One known advantage of auditory processing over visual processing is that short-term auditory (echoic) memory is longer lasting than short-term visual (iconic) memory. According to some estimates, echoic memory lasts upwards of 2000 ms (see Cowan, 1984), whereas iconic lasts for only up to a maximum of 1000 ms (Sperling, 1960), with some studies suggesting conservative estimates of only 200-300 ms (Di Lollo, 1977; Long, 1980). Note that Ngo et al. (2010) used an immediate repetition task in which a potential target repetition required participants to retain the visual information for only about 250 ms. It is possible then that the visual dominance effect reported by Ngo et al. may have been related to the persistence of the visual stimulus in participants' iconic memory store. This is likely given that the average response latency to repeated targets in Ngo et al.'s task were still well-within the 1000 ms upper limit of iconic memory.

Rather than detecting immediate repetitions, as in Ngo et al.'s (2010) study, we required participants to detect non-adjacent (n-1) repetitions in Experiment 1. This potentially induced a sufficient degradation of the internal memory of the visual signal by way of the rapidly decaying iconic memory trace (Di Lollo, 1977; Long, 1980; Sperling, 1960). As the onset of the repeated n-1 target started at least 979 ms after the first item of the repeated pair, it would be unlikely that any decision would be based on a simple iconic memory trace. Additionally, introducing an intervening stimulus between target n-1 repetitions may also have given rise to potential masking effects (see Enns & Di Lollo, 2000; Intraub, 1984; Loftus & Ginn, 1984; Potter, 1976, 1999). Specifically, an intervening stimulus might mask potential n-1 targets by competing with and making the first part of the target difficult to access from working memory.

The participants monitored a stream of simultaneously-presented visual and auditory stimuli and responded as soon as they saw or heard the n-1 repetition of a picture (i.e., separated by an interleaved different picture and sound), a sound, or both. If either the temporal decay of iconic memory, or the masking effects of the intervening stimulus, imposes a limit on visual dominance, then by extending presentations beyond this temporal limit, we would expect to observe an elimination of the Colavita effect as well, or perhaps a reversal (i.e., auditory dominance).

EXPERIMENT 1

Methods

Participants. Twenty participants from the University of Oxford (12 female) ranging in age from 20-55 years (mean age of 28 years) took part in this experiment. All of the participants reported normal or corrected-to-normal vision, and normal hearing. All but two were right-handed by self-report. The experiment took approximately 20 minutes to complete. The participants received a £5 (UK Sterling) gift voucher in return for taking part in the study. The experiment was conducted in accordance with the Declaration of Helsinki. Written informed consent was given by participants prior to the start of the experiment.

Apparatus and materials. Fifty line drawings of common objects chosen from the Snodgrass and Vanderwart (1980) database were used as the visual stimuli. Fifty sounds (identical to those used in Ngo et al., 2010) were selected from a database of 103 sounds on the basis of their clarity and familiarity, as rated by three judges. The sounds were presented through closed-ear headphones (Beyer Dynamic DT 531) at 70 dB (A)[4]. All of the sounds were adjusted so as to have a duration between 400-500 ms (average of 479 ms). The pictures were presented on a 17" Philips 107E visually-flat CRT computer monitor (screen refresh rate = 75 Hz) for the same duration as the temporally corresponding sounds. The pictures and sounds were presented synchronously and their order of presentation was pseudo-randomized using the DMDX software (downloaded from http://www.u.arizona.edu/~jforster/dmdx.htm, on 12/01/2009). The audiovisual (picture-sound) stimulus pairs were separated by a 250 ms silent/blank interval, giving rise to an SOA of 650-750 ms. Care was taken to ensure that none of the pairs of sounds and pictures had the same or related semantic meaning.

INSERT FIGURE 1 ABOUT HERE

Procedure. The participants sat approximately 60 cm from a computer screen in a dimly-lit experimental room. They were instructed to monitor the streams of auditory and visual stimuli and to respond as rapidly as possible whenever they saw or heard the repetition of a picture, sound, or both separated by an interleaved different picture and sound (i.e., an n-1 repetition task; see Figure 1). The participants used their right hand to press the "J", "K", or "L" key on a standard keyboard in response to the targets (the response keys were counterbalanced across participants for each response type: visual, auditory, or bimodal).

There were 10 blocks, each containing a 100 element sequence thus giving rise to a total of 1000 trials of which 250 trials involved the presentation of an n-1 target repetition (25 per block). Of the 250 target trials, 40% were visual, 40% were auditory, and 20% were bimodal (note that similar proportions have been used in the majority of previous studies of the Colavita effect; e.g., see Koppen & Spence, 2007b, c, d). In each block of 100 trials, every picture and sound was presented twice, with 25 target repetition trials per block (10 visual, 10 auditory and 5 bimodal), with a maximum of 5 target trials being presented in a row. It is important to note that while the target trials could be presented consecutively, they were never interleaved. That is, the first half of a repetition never preceded the second half of the previous repetition. The bimodal stimuli were constructed so that a picture and sound were never combined in the same way across different trials within the same block. That is, the picture of a chair only appeared with the sound of a cat once, unless, of course, the item constituted a bimodal target (5% of all trials). Note that the auditory and visual targets on the bimodal target trials were not semantically related. The participants were given training with the task using a stimulus sequence containing six targets — two visual, two auditory, and two bimodal — before starting the main experimental session. The training session was repeated until the participant felt comfortable with the task.

Results

The reaction times (RT) for correct responses, the percentage of correct responses (see Appendix), the percentage of missed targets (trials on which participants failed to make a response to a target within the allotted 1250 ms), and the distribution of errors to the bimodal targets were calculated for each participant and condition. Of particular interest was the distribution of errors on bimodal targets, where a Colavita effect might be observed. The accuracy data (percentage of correct responses, misses, and errors) for all experiments are shown in Figure 2. Analyses of the RT data for all experiments were non-significant ($F<1$ for all experiments except Experiment 2, where $F(2,32)=1.49$, $p=.24$). These data are presented for completeness in Table 1. The errors made by participants in response to the bimodal targets were classified into one of three categories: unimodal visual responses, unimodal auditory responses, or misses. Thus, the participants may have failed to respond to the auditory component of the target, the visual component of the target, or failed to make any response at all.

A repeated-measures analysis of variance (ANOVA) with target type (visual, auditory, or bimodal) as the within-participants factor was conducted on the percentage of correct responses. This analysis highlighted significant differences among the three target types, $F(2,38)=17.42$, $p<.001$ (see Figure 2A). Separate pairwise comparisons revealed that participants responded more accurately to unimodal auditory targets ($M=56.0\%$, $SE=4.0\%$) than to either unimodal visual ($M=35.4\%$, $SE=2.8\%$, $t(19)=-5.88$, $p<.001$) or to bimodal ($M=34.4\%$, $SE=4.7\%$, $t(19)=4.64$, $p<.001$) targets. The difference between the latter two conditions was not statistically significant, $p=.81$.

An ANOVA conducted on the percentage of misses (i.e., where no response was made to a target) revealed a significant main effect of target type, $F(2,38)=88.94$, $p<.001$ (see Figure 2B). The participants missed significantly more unimodal visual targets ($M=57.0\%$, $SE=2.7\%$) than unimodal auditory ($M=38.9\%$, $SE=3.9\%$) or bimodal ($M=21.9\%$, $SE=2.8\%$) targets, $p<.001$ for both comparisons. This result, together with the previous analysis of correct responses, strongly suggests that participants found it significantly more difficult to detect n-1 visual repetitions than to detect auditory repetitions. The participants also missed significantly more unimodal auditory than bimodal target trials, $p<.001$, suggesting that the bimodal targets were the easiest to detect. This reduced number of misses in bimodal targets may have been due to a redundancy effect as the participants could have responded to any one of its constituent components (see also Spence, Parise, & Chen, 2011).

As mentioned previously, participants could have committed one of three types of errors on the bimodal target trials: A miss, a unimodal visual response, or a unimodal auditory response. Of the errors made on the bimodal trials, 21.9% of those were misses, and the remaining trials were either unimodal visual or auditory responses. A closer look at the pattern of bimodal response errors revealed that participants made significantly more auditory based (*M*=27.2%, *SE*=2.7%) than visual based unimodal responses (*M*=16.0%, *SE*=2.2%), *t*(19)=-3.48, *p*=.002 (see Figure 2C), thus demonstrating a clear reversal of the Colavita visual dominance effect!

Discussion

The results of Experiment 1 provide the first clear evidence for a reversal of the Colavita visual dominance effect when participants have to respond to targets consisting of non-adjacent repetitions. One possible account for this finding is couched in terms of the duration of iconic memory, which, as mentioned previously, typically only lasts for up to one second (Sperling, 1960; Di Lollo, 1977; Long, 1980), whereas echoic memory typically lasts upwards of two seconds (Cowan, 1984). Here, the participants would have had to wait at least 979 ms (250 ms blank interval + 479 ms intervening stimulus presentation + 250 ms blank) before being able to determine whether an n-1 repetition had occurred.

Given the known limits of iconic and echoic memory, it is likely that the target (the second instance of the repeated pair) will have been presented just at, if not beyond, the upper limit of iconic but not echoic memory. Thus, it is possible that the visual information contained in the first stimulus presentation may have decayed prior to the presentation of the potential n-1 repeat (i.e., the third stimulus presented approximately 979 ms after the first stimulus), while the auditory information would have remained relatively less degraded. This would have rendered the visual information relatively “unreliable” as compared to the auditory information in the stream.

Moro and Steeves (2010) previously suggested that degradation in the reliability of the visual information is necessary to eliminate the Colavita visual dominance effect. It would appear then that the n-1 repetition detection task brought about a sufficient degradation in vision not only to eliminate, but apparently to reverse, the Colavita effect observed in

Experiment 1. It is unclear, however, whether the visual degradation was due to the decay of the iconic memory trace or the interference from the intervening stimulus, or some combination of the two.

The goal of Experiment 2 was therefore to determine the role of the temporal limit of iconic memory in reversing the Colavita visual dominance effect. Here, we shortened both the duration of the stimulus presentations and the blank interstimulus interval. The amount of time required to detect a visual n-1 target repetition would now fall well within the limits of iconic memory. Any potential masking effects should remain intact, however, as the intervening stimuli in the n-1 task could still have competed with the target stimuli for processing resources. Should the Colavita visual dominance effect be observed once again, and auditory dominance be eliminated, on the bimodal target trials, this would suggest that the limits on sensory memory played an important role in reversing the Colavita effect in Experiment 1. However, if we were to observe auditory dominance in the next experiment this would, once again, suggest that the masking effects from the intervening stimulus presentations were responsible for the reversal of the Colavita effect.

EXPERIMENT 2

Methods

Eighteen participants (13 female; age range = 18-36 years; mean age of 27 years) took part in Experiment 2. All were right-handed by self-report. The experimental set-up was identical to that used in Experiment 1, except that the visual and auditory stimulus pairs were now presented for 200 ms and the blank interval was now 150 ms. Thus, the elapsed time between n-1 repetitions in this experiment was 500 ms, which is well within the upper limit of iconic memory.

Results

The same analyses as in Experiment 1 were conducted on the present data. An ANOVA on the percentage of correct responses revealed a significant main effect of target type, $F(2,34)=35.8$, $p<.001$ (see Figure 2D). Separate pairwise comparisons indicated that participants responded more accurately to unimodal auditory targets ($M=67.6\%$, $SE=3.9\%$) than to either unimodal visual ($M=46.6\%$, $SE=3.1\%$, $p<.001$) or bimodal ($M=34.8\%$, $SE=4.8\%$, $p<.001$) targets. Participants also responded more accurately to unimodal visual than to bimodal targets, $p=.012$.

An ANOVA conducted on the percentage of misses (i.e., where no response was made to a target) also revealed a significant main effect of target type, $F(2,34)=15.66$, $p<.001$ (see Figure 2E). The participants missed significantly more unimodal visual targets ($M=44.0\%$, $SE=3.5\%$) than unimodal auditory ($M=26.7\%$, $SE=3.8\%$), or bimodal ($M=20.8\%$, $SE=3.7\%$) targets, $p<.001$, for both comparisons. This result, together with the previous correct response analysis, suggests that the visual n-1 repetition detection task was more difficult than the auditory task. The numerical difference between the percentages of missed auditory and bimodal targets failed to reach significance, $p=.22$.

Most importantly, a paired-samples t-test revealed that participants also made significantly more erroneous auditory-only responses ($M=28.0\%$, $SE=2.4\%$) than erroneous visual-only responses ($M=20.3\%$, $SE=3.0\%$), $t(17)=-2.29$, $p=.035$ (see Figure 2F) on the bimodal target trials. Thus, a reversal of the Colavita effect (i.e., auditory dominance) was still observed in the present experiment.

Discussion

The results of Experiment 2 demonstrate two important points: First, when the minimum time in which an n-1 target repetition could be detected was shortened to be within the limits of both iconic and echoic memory, the task became slightly easier, as evidenced by an overall increase in accuracy and decrease in the percentage of missed targets as compared to the results of Experiment 1[5]. Nevertheless, the overall accuracy of participants' responses still fell well below ceiling. Second, and more importantly, a reversal of the Colavita effect, whereby more auditory rather than visual responses were made to bimodal targets, was again observed in Experiment 2. In fact, for both Experiments 1 and 2, participants made approximately 8% more auditory than visual responses to bimodal targets, even though the overall percentage of incorrect responses to bimodal targets was lower in Experiment 2. Taken together, it would appear then that the temporal limit of visual sensory memory was not responsible for, or at least played only a small role in, the auditory dominance elicited by the n-1 repetition detection task.

Rather than occurring as a result of some intrinsic temporal limitation of sensory memory, it is likely that the reversal of the Colavita visual dominance effect may instead have resulted from interference from the intervening stimulus in the n-1 target repetition. Specifically, each intervening stimulus may have functioned as a conceptual mask (see Loftus & Ginn, 1984), which influenced performance by interrupting and competing for higher-level, semantic processing (Enns & Di Lollo, 2000; Intraub, 1984; Potter, 1976, 1999). Potter (1999) argued for the existence of a conceptual short-term memory store in which conceptual information is processed rapidly, but is highly unstable, and decays within a few hundred milliseconds. Unless this conceptual information is selected for further processing, it is quickly forgotten. Similarly, Enns and Di Lollo suggested that conceptual masks do not terminate target processing, but rather they become the new focus of attention. Each stimulus presentation in Experiment 2 carried semantic content and could have been a part of a repetition. Thus, the nature of this n-1 task required participants to constantly update and match previously-presented items to current items based on their semantic content. Consistent with the previous masking literature, the intervening stimulus could easily have become the new focus of attention and replaced the preceding stimulus (the first part of the target) for further processing. It is important to note that the intervening stimulus in this experiment appears to have differentially affected information processing / retention of the target stimuli in the visual and auditory modalities. This is evidenced by the fact that the accuracy of participants' responses was higher for auditory than for visual targets.

The goal of Experiment 3 was therefore to test whether instead of having an intervening meaningful stimulus separating potential n-1 repetition targets, the presentation of an irrelevant, meaningless visual stimulus (a pattern mask) paired with a burst of noise, would serve to extinguish any potential conceptual masking effects. Should we observe visual dominance rather than auditory dominance, as had been observed in Experiments 1 and 2, on the bimodal target trials, this would suggest that the Colavita effect was indeed due to the conceptual masking of the intervening meaningful stimuli having a differential effect on vision than on audition. If the Colavita effect is still absent in Experiment 3, however, then it is possible that there is still some processing of the pattern mask giving rise to interference effects with the target stimuli.

EXPERIMENT 3

Methods

Sixteen participants (12 female; age range = 18-37 years; mean age of 24 years) took part in Experiment 3. All of the participants were right-handed by self-report. The set-up for this experiment was similar to that used in Experiment 1 except that between the presentation of each visual and auditory stimulus pair, a pattern mask was presented in the center of the screen along with a 500 ms white noise burst (see Figure 1B). The pattern mask consisted of a pattern of meaningless, overlapping squiggly lines that did not form a recognizable shape.

Results

An ANOVA on the percentage of correct responses revealed a significant main effect of target type, $F(2,30)=13.08$, $p<.001$ (see Figure 2G). Separate pairwise comparisons indicated that participants responded more accurately to unimodal auditory targets (M=84.3%, SE=4.3%) than to either unimodal visual (M=75.4%, SE=4.3%, $p<.001$) or bimodal (M=73.3%, SE=3.3%, $p=.001$) targets. There were no significant differences between the accuracy of participants' responses to unimodal visual and bimodal targets, $p=.38$.

An ANOVA conducted on the percentage of misses (i.e., where no response was made to a target) also revealed a significant main effect of target type, $F(2,30)=11.16$, $p<.001$ (see Figure 2H). The participants missed significantly more unimodal visual targets (M=21.4%, SE=4.3%) than unimodal auditory (M=13.4%, SE=4.4%, $p<.001$) or bimodal (M=14.0%, SE=3.2%, $p=.003$) targets. The difference between the percentages of missed auditory and bimodal targets was not significant, $p=.77$. These results, together with the previous correct response analysis, suggest that the visual n-1 repetition detection task was still more difficult than the auditory task.

A paired-samples t-test revealed no significant difference between the percentage of incorrect auditory-only (M=3.9%, SE=0.9%) and visual-only responses (M=5.8%, SE=0.9%)**,** $t(15)=1.15$, $p=.27$ (see Figure 2I) on bimodal target trials. Thus, the participants made similar percentages of visual and auditory based errors. This result demonstrates that the removal of the semantically-meaningful intervening stimulus within the n-1 target stream resulted in an elimination of the auditory dominance that had previously been observed in Experiments 1 and 2. Importantly, the Colavita visual dominance effect was not observed in the present experiment either.

Discussion

The results of Experiment 3 suggest that it was indeed the interference caused by the processing of the semantically-meaningful intervening stimulus within an n-1 target trial that played a crucial role in determining whether or not the Colavita visual dominance effect would be reversed. Here, the participants had to detect a repetition that was separated by stimuli that were not related to the task (i.e., did not constitute any part of any target repetition). The SOA between target repetitions was the same as in Experiment 1. This allowed us to rule out a sensory memory explanation of the effects, where the decay of the visual memory trace due to the passage of time might have influenced the reliability of the visual information. Rather than making more auditory than visual responses to bimodal target repetitions, participants now made similar percentages of visual and auditory responses, thus, demonstrating an elimination (but, importantly, not a reversal) of the Colavita effect. Note that, numerically, participants made more visual-only than auditory-only responses to bimodal targets.

It appears that the pattern mask gave rise to some masking effects (Enns & Di Lollo, 2000; Loftus & Ginn, 1984) and, consequently, some reduction in visual dominance. In contrast with the results of Experiment 2, however, it appears that the reduction in visual dominance resulting from the pattern mask was not as great as that observed when a semantically meaningful intervening stimulus was presented. Thus, although auditory dominance was no longer observed in Experiment 3 (as it had been in Experiment 2), the relatively weaker masking effect of the pattern mask (as compared to the semantic intervening stimuli) might explain why the Colavita visual dominance effect was not completely reinstated either. Next, therefore, we went on to test whether eliminating an intervening stimulus altogether, thereby requiring participants to detect immediate rather than n-1 repetitions, would again give rise to a Colavita visual dominance effect on the bimodal trials. We also maintained the long SOA between targets (1000 ms) in order to minimize the likelihood of a sensory memory explanation for the results.

EXPERIMENT 4

Methods

Twenty-four participants (20 female; age range = 18-37 years; mean age of 25 years) took part in Experiment 4. All but one of the participants were right-handed by self-report. The experimental setup was similar to that used in Experiment 1. However, instead of detecting an n-1 repetition, participants now had to detect any immediate repetition of a picture, sound, or both. Additionally, the blank interstimulus interval was now extended from 250 to 1000 ms.

Results

An ANOVA on the percentage of correct responses revealed a significant main effect of target type, $F(2,46)=11.61$, $p<.001$ (see Figure 2J). Separate pairwise comparisons indicated that participants responded more accurately to unimodal auditory targets (M=86.2%, SE=2.7%) than to either unimodal visual (M=80.5%, SE=3.6%, $p=.001$) or bimodal (M=77.7%, SE=2.8%, $p<.001$) targets. There were no significant differences between the accuracy of participants’ responses to the unimodal visual and bimodal targets, $p=.17$.

An ANOVA conducted on the percentage of misses (i.e., where no response was made to a target) also revealed a significant main effect of target type, $F(2,46)=15.32$, $p<.001$ (see Figure 2K). The participants missed significantly more unimodal visual targets (M=17.0%, SE=4.3%) than unimodal auditory (M=11.3%, SE=2.7%, $p=.001$) or bimodal (M=9.3%, SE=2.3%, $p<.001$) targets. The difference between the percentages of missed auditory and bimodal targets was marginally significant, $p=.086$. This, together with the previous correct response analysis, suggests that detecting auditory repetitions was easier than detecting visual repetitions.

A paired-samples t-test revealed that participants made significantly more visual responses (*M*=8.5%, *SE*=1.0%) than auditory responses (*M*=4.5%, *SE*=1.0%)**,** *t*(23)=2.96, *p*=.007 (see Figure 2L) on bimodal target trials, a clear demonstration of the Colavita effect.

Discussion

The results of Experiment 4 replicated and extended the findings of Ngo et al. (2010) by demonstrating that when detecting immediate repetitions of bimodal targets, participants made significantly more visual-only than auditory-only responses. Thus, despite the SOA between targets being as long as 1000 ms, we nevertheless demonstrated the Colavita visual dominance effect. This further suggests that the presentation of an intervening stimulus, be it semantically meaningful or not, rather than temporal processing limits, is the key factor in determining whether or not visual dominance will be observed.

GENERAL DISCUSSION

The results of the present study demonstrate that under the appropriate experimental conditions and task demands, the Colavita visual dominance effect can be reversed. Consistent with Ngo et al.'s (2010) recent findings, the participants in the present study detected auditory n-1 repetitions more accurately than visual or bimodal repetitions across all four experiments. Importantly, however, the participants in Experiments 1 and 2 also made significantly more unimodal auditory than unimodal visual responses to bimodal targets, thus demonstrating a clear reversal of the Colavita visual dominance effect in favour of auditory dominance. While few studies of the Colavita visual dominance effect have successfully eliminated the effect (Koppen & Spence, 2007b; Moro & Steeves, 2010; Sinnett et al., 2007), no study up until now has been able to reverse the effect and show auditory dominance instead. In fact, as far as we are aware, this is the first reported reversal of the Colavita visual dominance effect in more than three decades since the phenomenon was reported in the scientific literature.

Based on accepted constraints of the duration of iconic memory (which lasts for less than 1000 ms; see Sperling, 1960) and echoic memory (which has been estimated to last for up to 4000 ms; see Cowan, 1984), we predicted that participants would have been biased toward responding to the auditory rather than the visual targets as a default. Such a pattern of responding was thought to arise from the fact that the amount of time needed for the comparison between potential visual n-1 targets in Experiment 1 exceeded the temporal limit of the iconic memory store. While the results of Experiment 1 corroborated this prediction, with participants making significantly more auditory than visual responses to bimodal targets, the results of Experiments 2 and 3 suggest that sensory memory temporal limitations were not crucial in terms of modulating the Colavita effect.

The results of Experiment 2 demonstrated that shortening the SOA between target repetitions, so that detecting a repetition would be well-within the bounds of iconic memory, not only failed to reinstate the Colavita visual dominance effect, but it reproduced the auditory dominance first demonstrated in Experiment 1. Thus, we suspected that the pattern of results might depend on the interference by the meaningful, intervening event separating target repetitions. Because each stimulus presentation in Experiment 2 could have been a potential n-1 repetition target, we suspected that visual masking may have played a role, such that the presentation of any semantically meaningful intervening stimulus would have given rise to competition for the higher-level processing resources that are recruited during object recognition (Enns & Di Lollo, 2000; Loftus & Ginn, 1984).

In Experiment 3, we demonstrated that presenting semantically meaningless intervening stimuli instead gave rise to similar percentages of visual and auditory responses to bimodal targets. Thus, the Colavita visual dominance effect was successfully eliminated, but not reinstated. Importantly, auditory dominance was no longer observed in Experiment 3 (as it had been in Experiments 1 and 2), suggesting that changing the semantically meaningful intervening stimulus to one that carried no semantic meaning did give rise to the elimination of auditory dominance that we expected to observe. Still, it appeared that some interference/backward masking from the pattern mask, even though it carried no semantic meaning in either the visual or auditory modality, was present, given that the visual dominance effect was not reinstated as we had predicted.

Loftus and Ginn (1984) demonstrated that while semantically meaningful (e.g., picture) masks produced strong masking effects, pattern masks gave rise to significant (albeit weaker) masking effects as well. In their study, the participants had to recall as many details from a previously-presented naturalistic picture as possible. In one condition, the target picture was masked by either a different naturalistic picture or a random noise pattern mask after a delay of 300 ms. Both the presentation of the picture and pattern mask led to participants recalling significantly fewer details from the target (M=1.86 and M=2.21 details recalled, respectively) as compared to the no-mask condition (M=2.35). Importantly, the picture mask produced a greater magnitude of interference than the pattern mask. The results of Experiments 2 and 3 are therefore consistent with Loftus and Ginn's findings in demonstrating a stronger masking effect in the presence of meaningful intervening stimuli as compared to the meaningless pattern mask. When removing the intervening stimulus completely in Experiment 4, effectively making it an immediate rather than n-1 repetition detection task, participants made more visual rather than auditory responses to bimodal targets. Thus, the Colavita visual dominance effect was reinstated despite the long SOA between target repetitions.

Taken together, the results of the present study suggest that the Colavita effect can be modulated by the potential masking effects that may differentially affect the retention of visual and auditory information. Introducing an intervening stimulus to the repetition detection task seems to have effectively made the task more difficult overall. More importantly, the interference from the intervening stimulus seems to have had a more pronounced effect on vision than on audition. Not only were participants better at detecting unimodal auditory as compared to unimodal visual targets, but they also made more auditory only rather than visual only responses when it came to bimodal n-1 targets. Thus, the memory of the visual n-1 stimulus seems to have been wiped out by the intervening visual stimulus, much more than the extent to which memory of the auditory n-1 target seems to have been by the intervening auditory stimulus. In other words, the new (intervening) visual stimulus was dominant enough to wipe out the previous visual stimulus, but this effect did not occur at the same proportions in the auditory domain – yet another demonstration of just how dominant the visual system really is!

We suspect that the differential masking effects of the intervening stimuli on vision and audition may be due to the fact that masking has a different time course in vision (see Enns & Di Lollo, 2000, for a review) and audition (Elliot, 1971; Oxenham & Moore, 1994; Oxenham & Wojtczak, 2010). For both vision and audition, forward masking is much more sensitive to the introduction of any temporal delay between the target and mask as compared to backward masking. Forward masking seems to occur in both vision and audition when the mask is presented within 50 ms of the target's onset. In vision, the disruptive effects of backward masking on the low-level iconic representation of a target are present up to 50 ms after the offset of the target (Loftus & Ginn, 1984). After 300 ms of the offset of the target, Loftus and Ginn suggest that conceptual masking occurs, whereby identification of the mask *"interrupts the higher level processing that is required for long-term storage of the information corresponding to the picture [the target repetition, in our case]"* (p. 435). In audition, low-level backward masking appears to be effective within 200 ms of target offset

(Massaro, Cohen, & Idson, 1976). Unlike in vision, however, Massaro and Burke (1991) suggest that auditory masks do not work retroactively as conceptual masks. That is, they do not disrupt the processing of information that might have been obtained before the presentation of the mask.

It would appear then that visual masks are more effective than auditory masks at producing substantial conceptual interference effects, especially at longer SOAs. This notion seems to hold true in light of the findings of the present study. Here, we observed that not only was detecting n-1 repetitions in the visual modality more difficult than in the auditory modality, but also auditory-only responses dominated over visual-only responses on bimodal trials. The masking effects appear to have given rise to a sufficient degradation in the reliability of the information presented in the visual but not the auditory modality, which has been shown to be necessary for eliminating the Colavita visual dominance effect (Moro & Steeves, 2010). We further demonstrated that the pattern of visual or auditory dominance could be modulated by the presence of a mask (conceptual, pattern, or absent) in the repetition detection task. That is, in the presence of a conceptual mask, auditory dominance prevails; in the presence of a pattern mask, auditory dominance disappears and vision begins to take the lead; in the absence of any mask, visual dominance is restored on bimodal target trials. Most notably, the fact that we had to push the visual system so hard to reverse the Colavita effect and that, with just a little less favouring of the auditory modality, we still see visual dominance (see also Ngo et al., 2010) is quite impressive. Moreover, even though unimodal visual target detection was worse than unimodal auditory target detection, vision still dominated over audition in bimodal target trials. These results go to show just how resilient and robust the Colavita visual dominance effect really is!

ACKNOWLEDGEMENTS

MKN was supported by the Clarendon Fund Scholarship from Oxford University. SS-F was supported by grants SEJ 2007-64103/PSIC and CDS00012 from the Spanish Ministerio de Ciencia e Innovación, ERC StG 263145, and grant 2009SGR-292 from DURSI.

REFERENCES

Baddeley, A. D. (1992). Working memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **302**, 311-324.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, **4**, 417-423.

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

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin,* **96**, 341-370.

Di Lollo, V. (1977). Temporal characteristics of iconic memory. *Nature*, **267**, 241-243.

Elliot, L. L. (1971). Backward and forward masking. *Audiology*, **10**, 65-76.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, **4**, 345-352.

Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 115-125.

Koppen, C., Alsius, A., & Spence, C. (2008). Semantic congruency and the Colavita visual dominance effect. *Experimental Brain Research*, **184**, 533-546.

Koppen, C., & Spence, C. (2007a). Assessing the role of stimulus probability on the Colavita visual dominance effect. *Neuroscience Letters*, **418**, 266-271.

Koppen, C., & Spence, C. (2007b). Audiovisual asynchrony modulates the Colavita visual dominance effect. *Brain Research*, **1186**, 224-232.

Koppen, C., & Spence, C. (2007c). Seeing the light: Exploring the Colavita visual dominance effect. *Experimental Brain Research*, **180**, 737-754.

Koppen, C., & Spence, C. (2007d). Spatial coincidence modulates the Colavita visual dominance effect. *Neuroscience Letters,* **417**, 107-111.

Lessard, N., Pare, M., Lepore, F., & Lassonde, M. (1998). Early-blind human subjects localize sound sources better than sighted subjects. *Nature*, **395**, 278-280.

Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 435-441.

Long, G. M. (1980). Iconic memory: A review and critique of the study of short-term visual storage. *Psychological Bulletin*, **88**, 785-820.

Massaro, D. W., & Burke, D. (1991). Perceptual development and auditory backward recognition masking. *Developmental Psychology*, **27**, 85-96.

Massaro, D. W., Cohen, M. M., & Idson, W. L. (1976). Recognition masking of auditory lateralization and pitch judgments. *Journal of the Acoustical Society of America*, **59**, 434-441.

Moro, S. S., & Steeves, J. K. E. (2010). An absence of the Colavita effect: Enhanced multisensory processing in monocular blindness. Poster presented at the *11th International Multisensory Research Forum*, 16-19th June, Liverpool, UK.

Ngo, M. K., Sinnett, S., Soto-Faraco, S., & Spence, C. (2010). Repetition blindness and the Colavita effect. *Neuroscience Letters,* **480**, 186-190.

Oram, M. W., Xiao, D., Dritschel, B., & Payne, K. R. (2002). The temporal resolution of neural codes: Does response latency have a unique role? *Philosophical Transactions of the Royal Society of London B, Biological Sciences*, **357**, 987-1001.

Oxenham, A. J., & Moore, B. C. J. (1994). Modeling the additivity of nonsimultaneous masking. *Journal of the Acoustical Society of America*, **98**, 1921-1934.

Oxenham, A. J., & Wojtczak, M. (2010). Frequency selectivity and masking. In C. J. Plack (Ed.), *The Oxford handbook of auditory science* (pp. 5-44). Oxford: Oxford University Press.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, **2**, 509-522.

Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 13-46). Cambridge, MA: MIT Press.

Sinnett, S., Soto-Faraco, S., & Spence, S. (2008). The co-occurrence of multisensory competition and facilitation. *Acta Psychologica*, **128**, 153-161.

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception & Psychophysics*, **69**, 673-686.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 174-215.

Soto-Faraco, S., & Spence, C. (2002). Modality-specific auditory and visual temporal processing deficits. *Quarterly Journal of Experimental Psychology*, **55A**, 23-40.

Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progress in Brain Research*, **176**, 245-258.

Spence, C., Parise, C., & Chen, Y.-C. (2011). The Colavita visual dominance effect. In M. M. Murray & M. Wallace (Eds.), *Frontiers in the neural bases of multisensory processes* (pp. 523-550). Boca Raton, FL: CRC Press.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, **74**, 1-29.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, **381**, 520-522.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, **88**, 638-667.

Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. I. Sensory processes in perception* (pp. 1-36). New York: Wiley.

FOOTNOTES

1. It is important to note that the task used to study the Colavita visual dominance effect has evolved from Colavita's (1974) simple detection paradigm and expanded to include the recognition of predefined pictures and sounds (see, for example, Ngo et al., 2010; Sinnett et al., 2007). Furthermore, while we believe it appropriate to call visual dominance observed in such tasks the Colavita effect, others may not. For the purposes of the current investigation we consider the paradigm and stimuli used here to be an expansion of the Colavita visual dominance effect as done so by previous authors (see Spence et al., 2011, for a review).

2. It should be noted that a reversal of the Colavita visual dominance effect has never been observed in any variant of Colavita's original visual dominance task when measuring unimodal response error rates to bimodal target stimuli. However, a recent demonstration of multisensory competition using a passive oddball event-related potential (ERP) design suggests that auditory dominance might be evident when measuring the P300 latency. Specifically, Robinson, Ahmar, and Sloutsky (2010) measured visual and auditory oddball detection in a passive (i.e., no response) task in which the participants were presented with frequent (i.e., standards) and infrequent stimuli (i.e., oddballs) embedded in a stream of auditory and visual events. The main comparison involved the potential modulation of unimodal baseline ERPs under conditions of multisensory presentation, as indicated by an increased or decreased latency of the P300 component. Interestingly, compared to their respective unimodal baseline ERPs, bimodal presentations either slowed down or sped-up the P300. Specifically, the visual P300 was delayed when paired with sounds whereas the auditory P300 occurred somewhat earlier when paired with pictures. Despite these results

suggesting a reversal in visual dominance in favour of auditory dominance, it should be noted that the task utilized by Robinson et al. was markedly different from the traditional Colavita visual dominance task. Additionally, there was no measure of accuracy on bimodal trials. Therefore, a direct comparison of the results of these studies is somewhat problematic (see also Lewkowicz, 1988a; b; Robinson & Sloutsky, 2004; 2010; Sloutsky & Robinson, 2008, for similar results with infant studies).

3. Note, though, that while many of the Colavita studies still involve simple detection, the complexity of the targets used in these studies have evolved from the simple detection of beeps and lights to the recognition of predefined pictures and sounds (see, for example, Koppen, Alsius, & Spence, 2008; Ngo et al., 2010; Sinnett et al., 2007).

4. In Colavita's (1974) original study, he addressed the question of whether increasing the intensity and saliency of the auditory relative to the visual stimuli would eliminate or reverse the visual dominance effect. Colavita asked participants to subjectively increase the intensity of a tone until it was twice as intense as a light, yet this stimulus modulation had no effect on the magnitude of the Colavita effect.

5. Between-experiments comparisons (independent-samples *t*-tests) were conducted on the percentages of correct responses and missed targets data from Experiments 1 and 2. These comparisons revealed that participants responded significantly more accurately to visual targets (*mean difference*=11.2%, p=.01) and auditory targets (*mean difference*=11.6%, p=.05) in Experiment 2 as compared to Experiment 1. There was no difference in the percentage of correct responses for bimodal targets (*mean difference*=0.4%) between Experiments 1 and 2, p=.95. The participants also missed significantly fewer visual targets

(*mean difference*=13.0%; *p*=.006), auditory targets (*mean difference*=12.2%, *p*=.03), and bimodal targets (*mean difference*=12.7, *p*=.03) in Experiment 2 as compared to Experiment 1.

APPENDIX

The percentages of correct responses for all three target types (visual, auditory, and bimodal) were rather low as compared to those values observed in previous studies of the Colavita visual dominance effect (e.g., see Ngo, Sinnett, Soto-Faraco, & Spence, 2010). We suspected that this may have been due to the difficulty that participants experienced in trying to simultaneously divide their attention between the visual and auditory modalities in the present study. In order to assess this possibility, we therefore conducted an additional control experiment in which the participants (N=10) were instructed to detect only unimodal visual or auditory n-1 repetition targets in separate blocks of experimental trials. The procedure was identical to that used in the main experiment reported in the text, except for the fact that the participants completed five blocks of trials detecting each target modality. The order of presentation of each target modality was counterbalanced across participants. At the start of each block of trials, the participants were instructed to respond to n-1 repetition targets in either the auditory <u>or</u> visual modality by pressing a key on the keyboard.

The percentage of correct responses and the RTs from those trials in which the participants responded correctly for each participant and condition were calculated. Separate paired-samples t-tests revealed that participants responded significantly more accurately to auditory targets (*M*=79.6%, *SE*=4.6%) than to visual targets (*M*=67.6%, *SE*=4.5%), $t(9)=2.71$, p=.02. What is more, the participants detected the visual targets (*M*=591 ms, *SE*=29 ms) significantly faster than the auditory n-1 targets (*M*=608 ms, *SE*=34 ms), $t(9)=2.40$, p=.04. This may have been due to the fact that the auditory stimuli time-evolving, whereas the visual stimuli were immediately perceptible. That is, the information needed to identify a naturalistic sound, such as a cat's "meow" may not arrive until relatively late in the presentation of the sound. All of the information contained within a picture, on the other hand, is immediately available as soon as it is presented (and is likely processed after approximately 150 ms of stimulus onset; Oram, Xiao, Dritschel, & Payne, 2011; Thorpe, Fize, & Marlot, 1996). It is also possible that there were some speed-accuracy trade-offs that might have been present in the data (see Edwards, 1965) given that participants were faster but less accurate at detecting visual, as compared to auditory, targets.

The results of this control experiment therefore confirm that participants found the task much easier when they need only attend to one sensory modality (either vision or audition). This was evidenced by the substantially higher percentages of correct responses (68% and 80% for visual and auditory, respectively) as compared to those observed in Experiment 1 (35% and 56% for visual and auditory targets, respectively), where participants had to divide their attention between the visual and auditory modalities in order to detect either type of target n-1 repetitions in the experiments reported in the main text.

Table 1

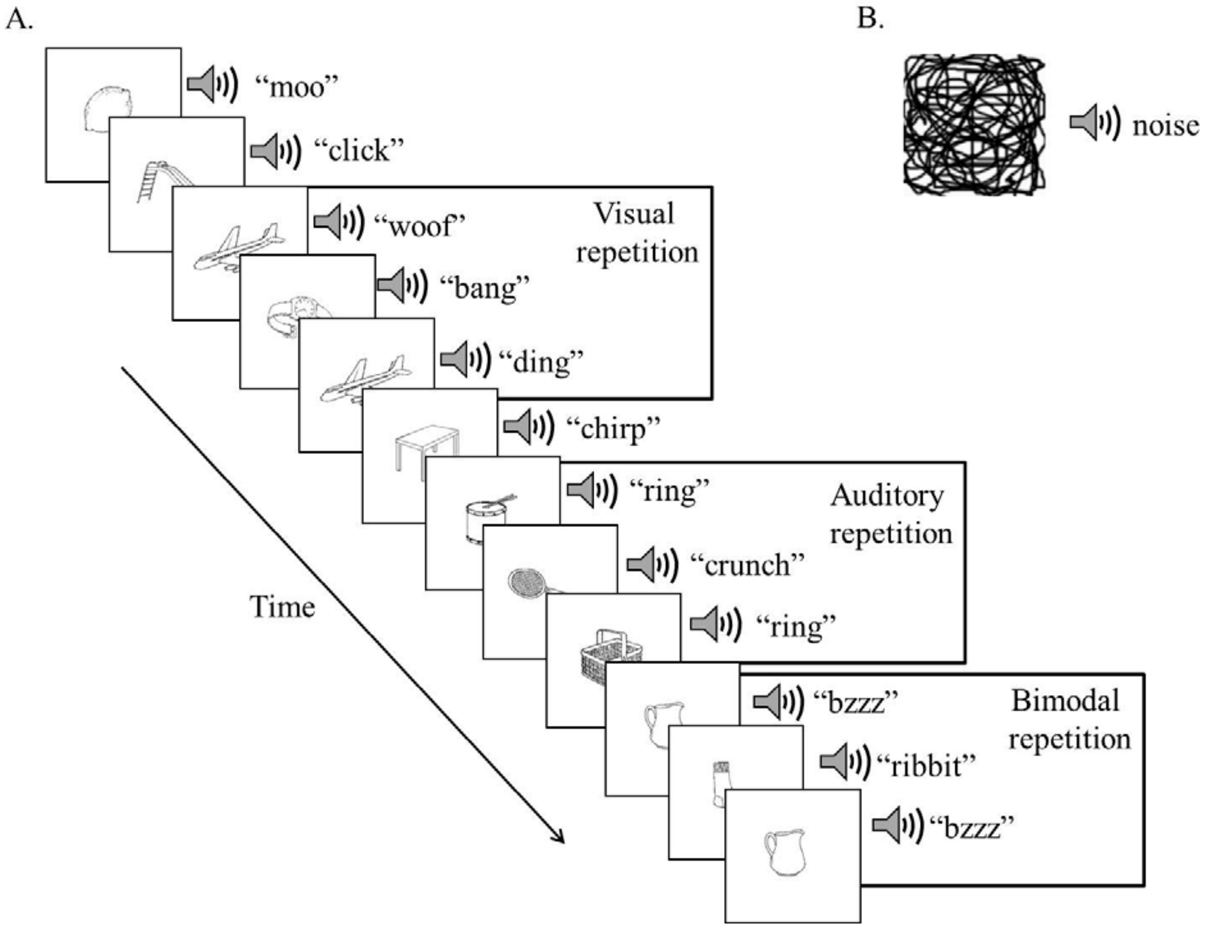
Mean RTs and SEMs (in ms) in response to visual, auditory, and audiovisual targets in Experiments 1 through 4.

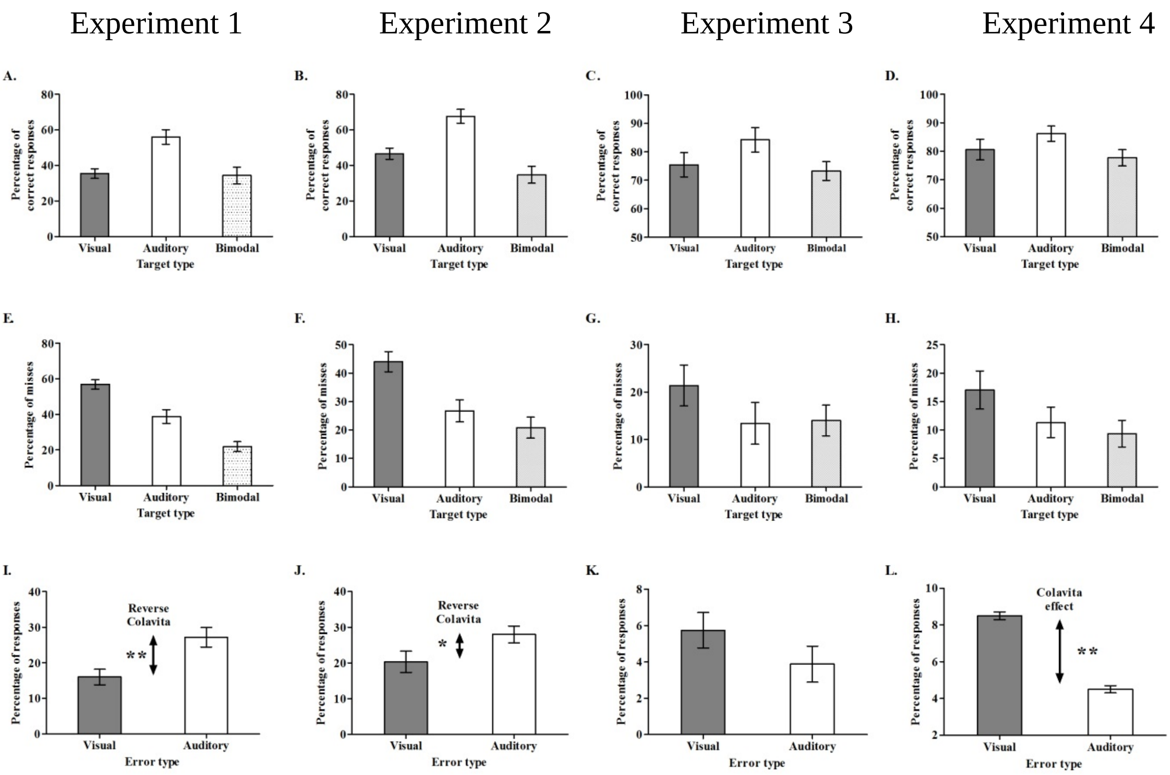
	Target		
Experiment	Visual	Auditory	Audiovisual
1	905 (17)	914 (18)	913 (22)
2	884 (28)	860 (18)	898 (30)
3	855 (30)	859 (30)	863 (24)
4	866 (21)	854 (18)	863 (16)

FIGURE CAPTIONS

Figure 1. A) Schematic illustration of the synchronous presentation of visual and auditory stimuli in Experiment 1. An example of a visual target, an auditory target, and a bimodal target are included. Each stimulus pair was presented for 400-500 ms, and separated by a 250 ms blank screen and silence making for an SOA of 650-750 ms. B) Image of visual pattern mask and noise burst between target stimuli used in Experiment 3.

Figure 2. A-D) Mean percentage of correct responses and misses for the visual, auditory, and bimodal repetition trials for Experiment 1-4. E-H) Mean percentage of misses for the visual, auditory, and bimodal repetition trials for Experiment 1-4. I-L) Mean percentage of erroneous visual-only versus auditory-only responses made on bimodal repetition trials for Experiment 1-4. The error bars represent the standard errors of the means.


A.
"moo"
"click"
"woof"
Visual repetition
"bang"
"ding"
"chirp"
"ring"
Auditory repetition
"crunch"
"ring"
"bzzz"
Bimodal repetition
"ribbit"
"bzzz"
Time
B.
noise



Experiment 1
Experiment 2
Experiment 3
Experiment 4
A.
B.
C.
D.
Percentage of correct responses
Visual
Auditory
Bimodal
Target type
E.
F.
G.
H.
Percentage of misses
I.
J.
K.
L.
Percentage of responses
Reverse Colavita
**
*
Colavita effect
Error type