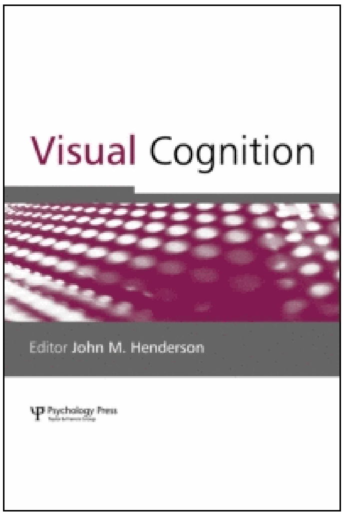

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683696

See an object, hear an object file: Object correspondence transcends sensory modality

Kerry E. Jordan [a]; Kait Clark [b];Stephen R. Mitroff [c]
[a] Department of Psychology, Utah State University, Logan, UT, USA [b] Center for Cognitive Neuroscience, Duke University, Durham, NC, USA [c] Center for Cognitive Neuroscience, and Department of Psychology & Neuroscience, Duke University, Durham, NC, USA

First published on: 01 March 2010

To cite this Article Jordan, Kerry E. , Clark, Kait andMitroff, Stephen R.(2010) 'See an object, hear an object file: Object correspondence transcends sensory modality', Visual Cognition, 18: 4, 492 — 503, First published on: 01 March 2010 (iFirst)

To link to this Article: DOI: 10.1080/13506280903338911

URL: http://dx.doi.org/10.1080/13506280903338911

VISUAL COGNITION, 2010, 18 (4), 492–503

Psychology Press
Taylor & Francis Group

See an object, hear an object file: Object correspondence transcends sensory modality

Kerry E. Jordan
Department of Psychology, Utah State University, Logan, UT, USA

Kait Clark
Center for Cognitive Neuroscience, Duke University, Durham, NC, USA

Stephen R. Mitroff
Center for Cognitive Neuroscience, and Department of Psychology & Neuroscience, Duke University, Durham, NC, USA

An important task of perceptual processing is to parse incoming information into distinct units and to keep track of those units over time as the same, persisting representations. Within the study of visual perception, maintaining such persisting object representations is helped by "object files"—episodic representations that store (and update) information about objects' properties and track objects over time and motion via spatiotemporal information. Although object files are typically discussed as visual, here we demonstrate that object–file correspondence can be computed across sensory modalities. An object file can be initially formed with visual input and later accessed with corresponding auditory information, suggesting that object files may be able to operate at a multimodal level of perceptual processing.

Keywords: Auditory; Cognition; Multisensory; Object file; Visual.

Perception is a complicated process with impressive feats seemingly accomplished with ease. For example, immense amounts of undifferentiated information are smoothly transformed into meaningful units; photons of

Please address all correspondence to Kerry E. Jordan, Department of Psychology, Utah State University, 487 EDUC Building, 2810 Old Main Hill, Logan, UT 84322-2810, USA. E-mail: kerry.jordan@usu.edu

For helpful conversation, we thank Melissa Bulkin, Sarah Donohue, Mat Fleck, Brittany James, Lujing Liu, and Yoni Mazuz. We also thank Jason Arita and Melissa Bulkin for assistance with data collection. This research was supported by Grant R03 MH080849 (SRM) from the National Institute on Mental Health.

http://www.psypress.com/viscog DOI: 10.1080/13506280903338911

light become recognizable faces, sound waves are parsed into familiar voices, and wafts of odour are identified as individuated smells. Moreover, discrete units that are parsed out of the morass of perceptual input are then tracked over time and motion as the same, persisting entities. An important goal of cognitive science is to understand such processes.

For *visual* perception, the object–file theory (Kahneman, Treisman, & Gibbs, 1992) has provided an important tool for theorizing about how visual information is tracked over time and motion as a single, persisting entity. An "object file" is an episodic representation that stores (and updates) information about an object's properties and tracks the object via spatio-temporal information (Kahneman et al., 1992). Studies have explored what is required to form and maintain object files (e.g., Mitroff, Scholl, & Wynn, 2004; Moore, Mordkoff, & Enns, 2007) and what comprises the nature of their contents, demonstrating that object files store abstracted information that is not tied to specific visual details (e.g., Gordon & Irwin, 1996; Henderson, 1994; Henderson & Siefert, 2001; Mitroff, Scholl, & Noles, 2007). Further studies have examined how object files operate, showing that their maintenance is tied to the limits of visual working memory (Noles, Scholl, & Mitroff, 2005), they might rely upon spatiotemporal continuity (Mitroff & Alvarez, 2007; but see Hollingworth, Richard, & Luck, 2008), they can diverge from conscious percepts (Mitroff, Scholl, & Wynn, 2005), they are sensitive to contextual information (Mitroff, Arita, & Fleck, 2009), their robustness can vary depending on the quality of the representation (e.g., Gordon, Vollmer, & Frankl, 2008; Henderson & Seifert, 2001), and they can be constrained by principles such as cohesion, boundedness, and containment (Cheries, Mitroff, Wynn, & Scholl, 2008; Mitroff et al., 2004, 2009).

EVIDENCE FOR MULTIMODAL INTEGRATION

Infant and adult object–file research has primarily focused on visual perception with relatively few studies exploring nonvisual representations (but see Hauser, Dehaene, Dehaene-Lambertz, & Patalano, 2002; Kubovy & van Valkenburg, 2001; Murray et al., 2004; Shen & Mondor, 2008; Shinn-Cunningham, 2008). Much can be learned by examining processing in multiple modalities and between modalities. The interaction between vision and audition has long been studied, with numerous demonstrations in both humans and nonhumans of how visual and auditory information can work in tandem to drive overall perception (e.g., Bahrick, Lickliter, & Flom, 2004; Jordan, Suanda, & Brannon, 2008; Lewkowicz & Kraebel, 2004; Lovelace, Stein, & Wallace, 2003; Melara, 1989; Stein, Huneycutt, & Meredith, 1988). For example, the McGurk effect is evidence that we automatically make use of information from both the auditory and visual modalities in language comprehension, integrating visual articulatory information with discrepant auditory syllables into a unified perception of speech (McGurk & MacDonald, 1976). A further example of audiovisual interaction in the domain of language is the "ventriloquism effect", which is produced by a spatial localization conflict between auditory and visual inputs (e.g., Alais & Burr, 2004; Bertelson & Radeau, 1981; Spence & Driver, 2000). Moreover, neurological evidence has suggested that paired visual and auditory stimuli can become associated in early processing stages, within visual object recognition areas (Murray et al., 2004). Despite such evidence that we automatically integrate auditory and visual information, to our knowledge, no studies have explored the interplay between auditory and visual information in regards to object–file representations (see Hommel, 2004, for visual–action intermodal integration in "event files").

CURRENT STUDY

Much has been learned about *visual* object files, yet much remains unknown about object files more broadly. Prior research suggests object files operate over abstract visual information (e.g., Gordon & Irwin, 1996; Henderson, 1994), but do they rely exclusively on *visual* memory and processing? It is not known whether object–file correspondence can be accomplished via auditory information; here, we explore whether object files store information from a given modality or whether their contents can be "amodal". Can auditory and visual information work in tandem to underlie object–file correspondence in the tracking and updating of persisting object representations?

EXPERIMENT 1: VISUAL-TO-AUDITORY OBJECT–FILE CORRESPONDENCE

To address the possibility that object–file correspondence can occur via an interplay between visual and auditory information, we employ a variant of the "object-reviewing" paradigm (Kahneman et al., 1992). In the modified version (see Noles et al., 2005), participants are typically presented with two simple objects (e.g., outlined frames). Information (e.g., a letter or picture) is briefly presented in each of the frames during a "preview" display (see Figure 1A). The preview information is then removed and the objects move about the display during a "linking" phase to decouple the objects from the specific spatial locations. After the motion, in a "target" display, information is presented in one of the frames and participants are to report, as quickly as possible, whether that information had been present in the preview display. For example, if the preview displays consisted of a picture of a phone in the top frame and a picture of a dog in the bottom frame,

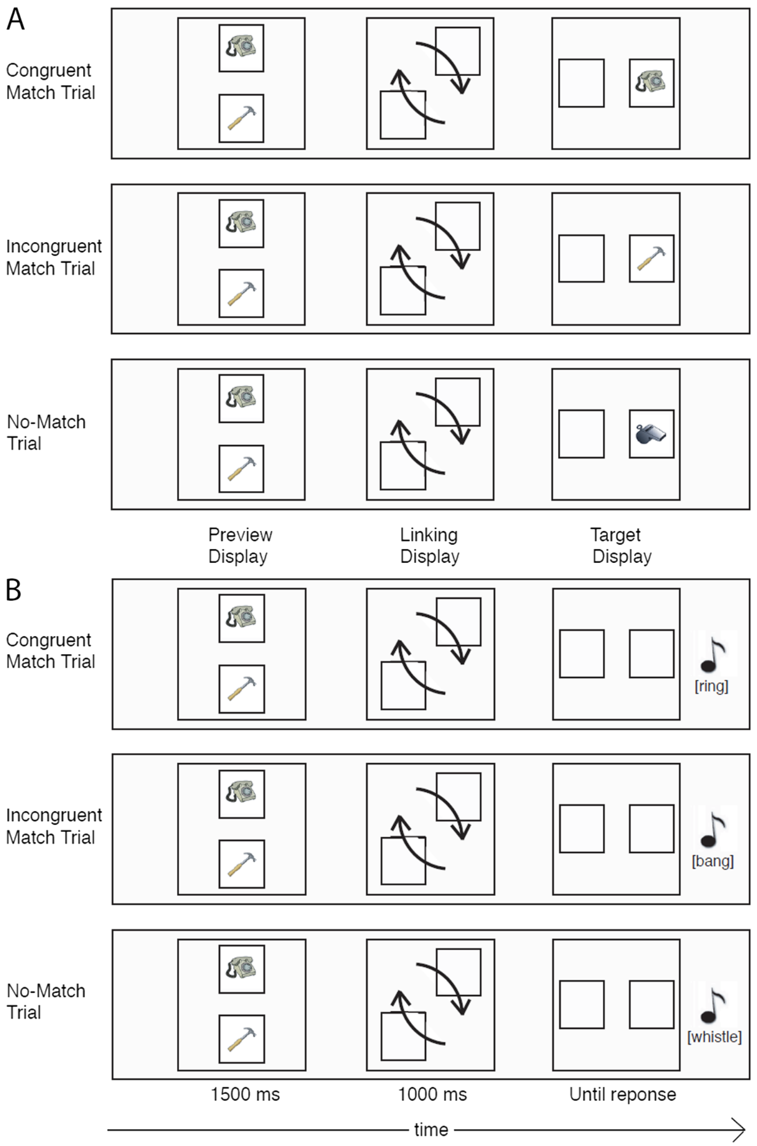

Figure 1. Depiction of congruent match, incongruent match, and no-match trial types of (A) the standard visual-to-visual object-reviewing paradigm and (B) the currently employed visual-to-auditory paradigm. This figure represents a simplified version wherein it only depicts one motion direction (the frames can move clockwise or counterclockwise), one target display location (the final visual or auditory stimulus can occur on the left or right), and does not depict a 500 ms empty frame period prior to the preview phase. Auditory stimuli are presented through speakers to the left and right of the visual display. To view this figure in colour, please see the online issue of the Journal.

participants would report "match"; if a phone or dog reappears in either frame in the target display, they would report "no match" if a novel picture to that trial, for instance, a train, appears in either frame. Typically, participants respond more quickly when the target picture is in the preview display than when it is novel, a form of general priming. More importantly, participants are quicker if the target picture reappears in the corresponding frame to where it appeared in the preview display than when it reappears in the "wrong" frame; evidence for *object-based* rather than spatial-based processing. This response time difference has been termed the "object-specific preview benefit" (OSPB) and serves as an operational measure of object files (Kahneman et al., 1992). By having previously stored specific information with a specific object (i.e., "phone" stored in the top object during the preview phase), subsequent processing is facilitated when that relationship remains constant as opposed to when it changes. That is, a "review" of an object file's contents benefits from continuity, and thus, response times are speeded (Kahneman et al., 1992).

In the current version of the paradigm, visual line-drawn pictures are presented in the preview display and then corresponding, naturally related auditory sounds occur with the target display (Figure 1B). The sounds are localized to the left and right of the display to correspond to the end locations of the two objects, and the critical question is: Can object files store visual information (line-drawn pictures) and then be reaccessed via auditory information (sounds)?

Methods

Participants. Twenty-one members of the Duke University community participated for course credit or monetary compensation. Two participants were more than two standard deviations below the mean in accuracy and were removed from all analyses.

Apparatus and stimuli. The experiment was run on a G4 Macintosh with a 19-inch CRT monitor in a dimly lit room. Participants sat approximately 45 cm from the display without head restraint. Stimulus presentation and data collection were conducted via customized software written with the VisionShell Graphics Library (Comtois, 2007). The "preview" information presented within the object frames consisted of six individual pictures (coloured line drawings with a mean size of $3.46^\circ \times 2.40^\circ$) of a dog, whistle, train, hammer, piano, and phone. Targets consisted of corresponding sounds: Dog bark, whistle blow, train horn, hammer bang, piano note, and phone ring. Preview pictures were assigned randomly, without replacement, on each trial and the target sound either matched one of the preview pictures (50% of the trials) or did not match either. The sounds were 1000 ms in duration and were presented at a constant volume via speakers attached to the left and right sides of the monitor at the approximate height of the participant's head.

Procedures. Participants began each trial by pressing the spacebar and the "preview" display would then appear with two objects vertically aligned, with their centres 10.42° above and below centre (Figure 1B). The objects were square frames with a thickness of 0.20°, measuring 6.55° on each side. After 500 ms, a preview picture appeared in each object. The pictures were removed after 1500 ms and the objects then moved in a circular path (50% of the trials clockwise) for 1000 ms until they were horizontally aligned, with their centres 10.42° to the left and right of centre. Immediately after the motion terminated, the target sound was presented for approximately 1000 ms via either the left or right speaker (50% of the time on each speaker). If the target sound corresponded with *either* of the preview pictures (no matter the location), participants were to report "match". If the sound did not correspond to either of the preview pictures, they were to report "no match". Participants indicated both responses with the numberpad of a standard keyboard, pressing "1" for match and "2" for no match. Response times were calculated from the offset of the target sound to the participant's keypress. Fifty per cent of trials were match trials and, of those, 50% were "congruent" (in which the target sound occurred at the spatial location to which the corresponding object had moved) and 50% were incongruent (in which the target sound occurred at the spatial location corresponding to the other preview picture; see Figure 1B). Participants completed 20 practice and 288 test trials.

Results

To remove extraneous data, trials were excluded if the response time was greater than 3 s (*Mean removed* = 0.24%, *SD* = 0.48%) or if they were more than two standard deviations greater than the individual participant's own average response time (*Mean removed* = 4.15%, *SD* = 1.30%). There were no significant effects of final target location or motion direction, for either this experiment or Experiment 2, and thus all analyses are collapsed over these factors. Participants were highly accurate on the match/no-match task (*Mean accuracy* = 96.48%, *SD* = 1.95%) and there was a small difference in accuracy between congruent ($M = 95.10\%$, $SD = 2.73\%$) and incongruent ($M = 96.51\%$, $SD = 2.28\%$) match trials, $t(18) = 2.16$, $p = .04$. Only trials with an accurate response were used for response time analyses.

The primary measure of interest was the object-specific preview benefit (OSPB): A response time advantage for congruent over incongruent trials.

The OSPB represents the processing advantage of information reappearing on the same object (congruent) compared to reappearing on a different object (incongruent; Kahneman et al., 1992). Participants were faster on congruent trials ($M = 711.08$ ms, $SD = 136.16$) than on incongruent trials ($M = 729.54$ ms, $SD = 144.69$), producing a significant OSPB of 18.45 ms, $t(18) = 2.19$, $p = .04$. Participants were significantly quicker to respond on incongruent match than on no-match trials ($M = 783.65$ ms, $SD = 137.99$), $t(18) = 5.54$, $p < .001$, revealing a benefit from recent exposure to the visual counterpart of the target sound (a form of general priming). Comparing the accuracy and response time data, it is tempting to infer a speed–accuracy tradeoff, since participants were faster and less accurate on the congruent than the incongruent trials. However, this seems unlikely, although not impossible, given (a) few prior object–file studies have found accuracy effects, (b) Experiment 2 finds no accuracy effect, (c) participants were highly accurate overall, and (d) only trials with an accurate response were included in the response time analyses.

Discussion

This experiment presents the first evidence for visual and auditory information working in tandem to underlie object–file correspondence. Visual information (e.g., picture of a phone or a dog) is tied to specific objects in a preview phase and then the objects are reaccessed via corresponding auditory information (e.g., a dog bark). A standard, robust OSPB is found, suggesting object files operate over abstract information that is not tied to specific perceptual details or modalities. However, before further discussing the implications of these findings, we account for a possible confound in Experiment 2.

EXPERIMENT 2: CONTROLLING FOR VERBAL ENCODING

Experiment 1 reveals crossmodal processing in which visual and auditory information work together to underlie object–file correspondence. However, alternatively, participants may see the preview pictures and encode them verbally (e.g., after seeing a phone and dog, they verbally rehearse, “phone, dog, phone, dog ...”). Then, when a corresponding sound is played, they also convert this information into a verbal code (e.g., a bark is encoded verbally as “dog”). If this were the case, the object–file correspondence may not occur crossmodally, but rather verbally. We account for this possibility here by reducing the opportunity to verbally rehearse.

Methods

Participants. Twenty members of the Duke University community participated for course credit or monetary compensation. One participant was more than two standard deviations below the mean in accuracy and their data were removed from all analyses.

Apparatus and stimuli. The experimental paradigm was identical to Experiment 1 except for the addition of a secondary, verbal interference task; participants were to verbally rehearse four digits for the duration of each trial.

Procedures. At the start of each trial, four digits from 0 to 9 were randomly selected, without replacement, and presented in the centre of the screen for 500 ms. Participants were instructed to memorize the digits in the order in which they were presented and continually rehearse them in silence for the duration of the trial. For instance, if the digits "5 1 7 2" were presented, participants would rehearse "five one seven two, five one seven two, ..." throughout the trial. At the end of each trial (all else being identical to Experiment 1), participants responded to the match/no-match task and then were prompted to enter the rehearsed digits using the number pad on the right of the keyboard. Participants were informed that response time for this secondary task was not recorded and that they should view this as a separate task. They were instructed to respond "match" or "no-match" as quickly and accurately as possible *before* completing the secondary task. The response keys for "match" and "no-match" were changed to "A" and "Z", respectively, to eliminate confusion with the digit response. Participants completed 32 practice and 288 test trials.

Results and discussion

The same data exclusion criteria as Experiment 1 were employed, removing trials with a response time greater than 3 s ($M = 0.44\%$, $SD = 0.66\%$) and a response time greater than two standard deviations of each participant's own mean response time ($M = 3.09\%$, $SD = 1.58\%$). In the remaining trials, participants performed well on the match/no-match task ($M = 96.90\%$, $SD = 3.10\%$) and the digit-reporting task ($M = 92.89\%$, $SD = 6.02\%$).

Trial congruency had no effect on accuracy for the match/no-match task (congruent $M = 96.67\%$, $SD = 2.95\%$; incongruent $M = 96.17\%$, $SD = 3.20\%$), $t(18) = 0.72$, $p = .48$, or the digit-reporting task (congruent $M = 91.81\%$, $SD = 7.59\%$; incongruent $M = 92.54\%$, $SD = 5.03\%$), $t(18) = 0.55$, $p = .59$. All remaining analyses were conducted only on the trials in which both tasks were performed correctly. Participants responded faster on

TABLE 1
Response times, by condition, for Experiments 1 and 2, with standard deviations in parentheses

	Experiment 1	*Experiment 2*
No match	783.65 ms	791.82 ms
	(137.99)	(185.10)
Incongruent match	729.54 ms	764.68 ms
	(144.69)	(198.34)
Congruent match	711.08 ms	745.97 ms
	(136.16)	(196.89)
Object-specific preview benefit (incongruent–congruent)	18.45 ms	18.71 ms
	$t(18) = 2.19$, $p = .04$	$t(18) = 2.31$, $p = .03$

congruent trials ($M = 745.97$ ms, $SD = 196.89$) than on incongruent trials ($M = 764.68$ ms, $SD = 198.34$), producing a significant OSPB of 18.71 ms, $t(18) = 2.31$, $p = .03$. As in Experiment 1, participants were also quicker to respond on incongruent match trials than on no-match trials ($M = 791.82$ ms, $SD = 185.10$), resulting in a general priming effect of 27.15 ms ($SD = 41.30$), $t(18) = 2.86$, $p = .01$ (Table 1). There was thus no evidence of a speed–accuracy tradeoff in Experiment 2; participants were equally accurate on congruent and incongruent trials in the match/no-match task, $t(18) = 0.72$, $p = .48$, while exhibiting faster response times on congruent trials, $t(18) = 2.31$, $p = .03$.

Results suggest the secondary task was effective in engaging verbal processes. Participants were accurate at the task (showing they were completing it) but not at ceiling (suggesting it was taxing). Thus, with the opportunity for verbal encoding reduced in this experiment, the results replicate those of Experiment 1, suggesting that verbal encoding does not mediate the crossmodal matching operations.

GENERAL DISCUSSION

Understanding how entities are internally represented as the same, persisting objects over time and motion is an important step towards a deeper understanding of perception and cognition. Previous research has demonstrated that object files are abstract in that they are not bound to specific visual details (e.g., Gordon & Irwin, 1996; Mitroff et al., 2007); the current experiments go further in showing that object files are not even necessarily tied to *vision*. Instead, they might store object-related information in an amodal format that can be flexibly accessed across senses.

That features from multiple sensory modalities are integrated into coherent object files should be somewhat expected, given the highly interactive nature of the senses in our everyday environments. A child's cry coming from the next room immediately after a sprinting 4-year-old has departed from sight will likely signal the continued existence of a single, persisting (and perhaps bruised) child. Although this may be an intuitive concept, these are the first data to explicitly demonstrate object files can operate *across* visual and auditory modalities.

An open question is by what means object–file correspondence can operate across multiple sensory modalities. One possibility is that object files may be purely visual, but that they can be accessed via auditory information. That is, object files may exist solely as visual representations, but the object file "reviewing operation" (Kahneman et al., 1992) may be able to operate over abstracted input. Alternatively, object–file representations may not be intimately tied to any particular sensory modality. In this sense, object files should not be conceived of as visual or auditory, but rather as abstract amodal representations. Although no evidence to date can conclusively tease apart these alternatives, the existence of nonvisual object processing (e.g., auditory objects, or audible sources; Kubovy & van Valkenburg, 2001) may support the later hypothesis. Such multisensory information could be bound in working memory via the episodic buffer's linking of visual and verbal material (e.g., Baddeley, 2000). In sum, these experiments provide the first demonstration that the object file representation can be probed by both visual and auditory inputs.

In conclusion, the current experiments offer a case study example of how object–file correspondence can be decoupled from specific sensory modalities. These data offer clear evidence for crossmodal object–file processing, but much remains unknown about the role of sensory modality information in the calculation of persisting object representations. For example, can object files transcend sensory modality in any situation, or is auditory information only incorporated into object files in the absence of competing visual information? Likewise, do object files operate across sensory modality throughout the lifespan, or might this ability only develop in adulthood? Are the OSPBs derived from visual objects alone identical in magnitude to the OSPBs derived from the same objects presented in different modalities? It also remains unknown how crossmodal object files are established in a more cluttered environment than that tested in the present experiments. Natural environments often contain many potential pairings of visual and auditory sources, and it would likely be challenging to establish crossmodal object correspondences based on purely perceptual information alone if there are many competing sounds and sights. This is especially true given the relatively poor spatial resolution of auditory perception as compared with visual perception. In such environments, semantic, top-down knowledge of which sights and sounds could potentially be paired together, along with perceptual information regarding spatiotemporal contiguity, might play a large role. The exact nature of crossmodal object files remains an exciting open question, providing a ripe area for future research.

REFERENCES

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

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

Bahrick, L., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception and cognition in infancy. *Current Directions in Psychological Science*, *13*, 99–102.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception and Psychophysics*, *29*, 578–584.

Cheries, E. W., Mitroff, S. R., Wynn, K., & Scholl, B. J. (2008). Cohesion as a constraint on object persistence in infancy. *Developmental Science*, *11*, 427–432.

Comtois, R. (2007). VisionShell PPC [Software libraries]. Cambridge, MA: Author.

Gordon, R., & Irwin, D. (1996). What's in an object file? Evidence from priming studies. *Perception and Psychophysics*, *58*, 1260–1277.

Gordon, R. D., Vollmer, S. D., & Frankl, M. L. (2008). Object continuity and the transsaccadic representation of form. *Perception and Psychophysics*, *70*, 667–679.

Hauser, M. D., Dehaene, S., Dehaene-Lambertz, G., & Patalano, A. L. (2002). Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (*Saguinus Oedipus*). *Cognition*, *86*, B23–B32.

Henderson, J. (1994). Two representational systems in dynamic visual identification. *Journal of Experimental Psychology: General*, *123*, 410–426.

Henderson, J. M., & Siefert, A. B. C. (2001). Types and tokens in transsaccadic object identification: Effects of spatial position and left-right orientation. *Psychonomic Bulletin and Review*, *8*, 753–760.

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, *137*, 163–181.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*, 494–500.

Jordan, K. E., Suanda, S., & Brannon, E. M. (2008). Intersensory redundancy accelerates preverbal numerical competence. *Cognition*, *108*, 210–221.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object specific integration of information. *Cognitive Psychology*, *24*, 174–219.

Kubovy, M., & van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*, 97–126.

Lewkowicz, D., & Kraebel, K. (2004). The value of multisensory redundancy in the development of intersensory perception. In G. Calvert, C. Spence, & B. Stein (Eds.), *The handbook of multisensory processes* (pp. 655–678). Cambridge, MA: MIT Press.

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.

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

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

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, underlie object persistence. *Psychonomic Bulletin and Review*, *14*, 1199–1204.

Mitroff, S. R., Arita, J. T., & Fleck, M. S. (2009). Staying in bounds: Contextual constraints on object file coherence. *Visual Cognition*, *17*, 195–211.

Mitroff, S. R., Scholl, B. J., & Noles, N. S. (2007). Object files can be purely episodic. *Perception*, *36*, 1730–1735.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2004). Divide and conquer: How object files adapt when a persisting object splits into two. *Psychological Science*, *15*, 420–425.

Mitroff, S. R., Scholl, B. J., & Wynn, K. (2005). The relationship between object files and conscious perception. *Cognition*, *96*, 67–92.

Moore, C. M., Mordkoff, J. T., & Enns, J. T. (2007). The path of least persistence: Object status mediates visual updating. *Vision Research*, *47*, 1624–1630.

Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage*, *21*, 125–135.

Noles, N. S., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object-file representations. *Perception and Psychophysics*, *67*, 324–334.

Shen, D., & Mondor, T. A. (2008). Object file continuity and the auditory attentional blink. *Perception and Psychophysics*, *70*, 896–915.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*, 182–186.

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: Reflexive cross-modal orienting and ventriloquism. *NeuroReport*, *11*, 2057–2061.

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: The same rules of multisensory integration apply. *Brain Research*, *448*, 355–358.

Manuscript received May 2009
Manuscript accepted August 2009
First published online March 2010