
Object files can be purely episodic

Stephen R. Mitroff

Center for Cognitive Neuroscience, Department of Psychology & Neuroscience, *Duke University*

Brian J. Scholl & Nicholas S. Noles

Department of Psychology, *Yale University*

Our ability to track an object as the same persisting entity over time and motion may primarily rely on spatiotemporal representations which encode some, but not all, of an object's features. Previous research using the 'object reviewing' paradigm has demonstrated that such representations can store featural information of well-learned stimuli such as letters and words at a highly abstract level. However, it is unknown whether these representations can also store purely episodic information (ie information obtained from a single, novel encounter) that does not correspond to pre-existing type-representations in long-term memory. Here, in an object-reviewing experiment with novel face images as stimuli, observers still produced reliable *object-specific preview benefits* in dynamic displays: A preview of a novel face on a specific object speeded the recognition of that particular face at a later point when it appeared again on the same object compared to when it reappeared on a different object (beyond display-wide priming), even when all objects moved to new positions in the intervening delay. This case study demonstrates that the mid-level visual representations which keep track of persisting identity over time — e.g. 'object files', in one popular framework — can store not only abstract types from long-term memory, but also specific tokens from online visual experience.

A fundamental challenge for visual processing is not only to parse the visual world into distinct objects, but also to keep track of those objects as the same entities from one instant to the next. In contrast to long-term object matching where surface features may be the only means to identification (eg recognizing a colleague at a conference), research has suggested that this challenge is met in online visual processing via mid-level visual representations that track objects on the basis of spatiotemporal information — how and where objects move, rather than what they look like. As a result, surface features may not always be encoded at this level of processing (eg Bahrami 2003; Burt & Sperling 1981), may be overwritten on the basis of spatiotemporal factors (eg Moore & Lleras 2005), and may fail to determine object persistence even when they are encoded (eg Flombaum et al 2004; Flombaum & Scholl 2006; Mitroff & Alvarez in press; Xu & Carey 1996). At the same time, other research has shown that at least some visual features of objects are reliably encoded as a result of tracking identity over time (Kahneman, Treisman, & Gibbs 1992). Here we explore what kinds of features may be encoded in this way, and whether they can be truly visual in nature.

Mid-level visual representations have been theoretically characterized in many ways, including as object tokens (eg Kanwisher 1991) and as visual indexes (eg Pylyshyn 1989). Here, following Kahneman and Treisman, (Kahneman et al 1992), we shall often draw upon the notion of 'object files' for such representations. This characterization — and the object-reviewing paradigm that is most directly associated with it — naturally emphasizes the relationship of visual surface features to persisting object representations. In the initial object-reviewing experiments (Kahneman et al 1992), observers viewed a 'preview' display that contained two or more objects with a different letter placed in each. The letters then disappeared and the objects moved to new locations. Once the objects stopped, a single 'probe' letter then appeared in one of the objects, and the observers simply named it aloud. The probe could be one of the initial preview letters (on Match trials) or it could be novel to the trial (on No-Match trials). Further, on Match trials, the probe letter could reappear on the same object in which it had been previewed (on Congruent trials) or on a different object (on Incongruent trials). Observers are typically quicker to respond on Congruent trials than on Incongruent trials — an effect termed an *object-specific preview benefit* (OSPB). In this way, OSPBs provide an index of object persistence: manipulations that attenuate enduring object representations result in weakened OSPBs.

For helpful conversation and/or comments on earlier drafts, we thank Rob Gordon, Steve Franconeri, Andrew Hollingworth, Melody Lu, Pamela Yee, anonymous reviewers, and the members of the Scholl and Chun labs at Yale University. SRM was supported by NIMH #F32-MH66553-01. BJS was supported by NSF #BCS-0132444.

Prior research with this paradigm has demonstrated that object files are not necessarily tied to specific visual properties — for example a letter can be represented regardless of its typeface (Henderson & Anes 1994), and a concept can be represented regardless of its medium (eg the abstract identity ‘fish’ persists despite being previewed as a word and probed as a line drawing; Gordon & Irwin 2000). In the present experiment we address the degree to which such abstraction is an intrinsic part of object file processing. It is consistent with all previous studies that object files can store *only* abstract information, rather than visual detail. However, this seems unlikely. First, this would not allow object files to function when abstract type information could not be extracted, and would thus render this mid-level system helpless to support visual experience unless recognition had already occurred. Second, the previously used stimuli may have directly lent themselves to abstraction. To our knowledge, all previous object-reviewing experiments have used either letters, words, or line-drawings of highly familiar objects which already have long-term memory representations. Third, many types of previous work have demonstrated that visual details *are* stored in other contexts. To take one example, previous eye-tracking research has demonstrated that specific visual properties such the orientation of an object (which can vary without changing the underlying semantic representation) are stored in visual working memory, and can influence later perception (eg Henderson & Siefert 2001).

For these reasons, we predicted that object files can store not only object types (eg the letter ‘A’, a dog) but also purely episodic object tokens (eg the specific contours of a nonverbalizable shape, or the identity of a face). We addressed this question in the present experiment via a case-study, exploring whether it is possible for object files to be mediated by purely episodic information, when type information is either not available or not informative. Is it necessary to tap into previously stored long-term memory types or can an OSPB be guided by novel tokens that are only available in the single ‘episode’ of a given trial? To determine this, we employed face images in an object-reviewing paradigm. The faces were entirely novel to observers, such that they had no pre-existing long-term memory representations. Since the faces were similar to one another and

never to be seen again, there was no way (nor incentive) for observers to readily verbalize the identities of the faces. Moreover, because our task forced observers to respond to *particular* faces, any OSPB would have to reflect stored object tokens, and not merely types.

Method

Twenty members of the Yale University community participated for course credit or payment. The displays were presented on a Macintosh iMac computer using custom software written using the VisionShell graphics libraries (Comtois 2005). Each trial began with two square frames (4.48 deg²) presented as white outlines (.15 deg thick) on a black background. Each frame started along the vertical midline 4.98 deg from the center, one to the left and one to right (with distances calculated from the frames’ centers and all visual angles are based upon an approximate viewing distance of 50 cm). After 500 ms, a color image of a face (approximately 2.14 deg x 3.09 deg, drawn from the Max-Planck Faces Database¹) appeared centered in each frame (see figures 1 & 2). After 1500 ms, these ‘preview faces’ disappeared and the frames began moving. The motion took 1500 ms and traced either a clockwise or counterclockwise path (equally often in each direction). Regardless of direction, the frames always stopped along the horizontal midline 4.98 deg from the center, one above and one below. A single ‘probe’ face then appeared in one of the two frames (equally often in each). Observers made a speeded response, pressing one key (‘1’) to indicate that the probe face was the same as *either* of the two preview faces (‘Match’), or another key (‘2’) to indicate that it did not appear in the preview display (‘No-Match’). 50% of the trials were No-Match trials, and of the remaining Match trials, 50% were ‘Congruent’ (in which the probe face was the same as the preview face that initially appeared in that frame), and 50% were ‘Incongruent’ (in which the probe face was the same as the preview face that initially

¹ The face database was provided by the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany. The faces were created by laser-scanning heads without hair and then morphing the scans to avoid close resemblances so as to hide the identity of the individuals. The set of 200 faces consisted of 100 males and 100 females. Within any given trial in the current experiment, the faces were either all male or all female, and only front-view orientations were used.

appeared in the *other* preview frame). This is a modified version of the object-reviewing paradigm that offers a flexible and robust measure of object persistence by using a matching task rather than naming (eg Kruschke & Fragassi 1996; Mitroff & Alvarez in press; Mitroff, Scholl & Wynn 2004, 2005; Noles, Scholl & Mitroff 2005).

Each face appeared in only one trial throughout the entire experiment. Observers completed 16 practice trials before the 72 test trials (with faces from the practice trials never appearing during test). Four image sets were constructed such that, across observers, each face was equally likely to be seen during a Match and a No-Match trial, and likewise as part of a Congruent and an Incongruent Match trial. Five observers were assigned to each image set and each completed the test trials in a unique random order.

Results

Response times more than two standard deviations from an observer's mean were removed prior to subsequent analyses (*Mean* trials removed = 3.25, *SD* = 1.12). On average, observers responded correctly to the Match/No-Match question on 84.31% of the trials (*SD* = 5.35%) and there was no significant difference between Congruent (*Mean* = 86.67%, *SD* = 11.17%) and Incongruent trials (*Mean* = 83.39%, *SD* = 11.53%; $t(19)=0.89$, $p=0.387$). Only trials with an accurate response were analyzed further. Our primary question was whether there was an OSPB — a response time benefit for Congruent Match over Incongruent Match trials. Collapsing the data across motion direction and probe location, observers were faster to respond on Congruent trials (*Mean* = 987.43 ms, *SD* = 216.60 ms) than on Incongruent trials (*Mean* = 1025.62 ms, *SD* = 226.34 ms), producing a significant OSPB of 38.19 ms ($t(19)=2.83$, $p=.011$). There was no general priming effect beyond the OSPB: Responses on Incongruent trials were not significantly faster than on No-Match trials (*Mean* = 1029.64 ms, *SD* = 226.65; $t(19) = 0.169$, ns.).

Discussion

In contrast to every previous study of persisting object representations that has used the object-reviewing paradigm, the stimuli in our experiment did not correspond to pre-existing representations

in long-term memory, and could not be readily verbalized.² The fact that we still obtained a reliable and fairly large OSPB with such stimuli indicates that object files can store not only abstracted information about object types, but also specific visual features of individual object tokens. Although it is possible that the observers created new long-term memory representations ‘on the fly’ during the preview phase of this experiment, it remains clear that object files can be formed and tracked on the basis of representations there were not previously held in a long-term store. This helps to bring the ‘object file’ framework closer to other notions of mid-level object processing (eg ‘proto-objects’) wherein episodic memory traces go beyond sensory representations, yet precede long-term memory (eg Rensink, 2000; Wolfe & Bennett, 1997).

In future studies it may be possible to determine just when object files store information of each type. In particular, two aspects of the present experiment may have encouraged the storage of visual details. First, our stimuli could not be readily mapped to pre-existing semantic type representations, since the faces were all novel. Second, our task required that observers encode the initial displays into visual working memory: the faces needed to be represented in some way which captured the particular details of each token stimulus, beyond its type-identity *as* a face. Future studies with further variants of these paradigms and stimuli might thus be able to tease these factors apart, and could also explore the duration for which such representations persist (beyond single trials), and whether the resulting representations are purely visual or newly-created type representations.

What the current experiment does show, for the first time, is that object files *can* be purely episodic in at least some situations: they can be truly *visual* representations which track token objects over time and motion. In this way, such representations can serve to mediate coherent online visual experience wherein novel objects and views are constantly encountered.

² Observers may have noticed some resemblance between a face and a previous acquaintance, of course, but this could not have driven our results, since 1) it is unlikely that this occurred for more than a few faces out of the 144 preview faces seen, 2) no face was presented in more than a single trial, and 3) the faces were presented too quickly for any serious scrutiny.

References

- Bahrami B, 2003 "Object property encoding and change blindness in multiple object tracking" *Visual Cognition* **10** 949 - 953
- Burt P, Sperling G, 1981 "Time, distance, and feature trade-offs in visual apparent motion" *Psychological Review* **88** 171 - 195
- Comtois R, 2005 VisionShell PPC [Software libraries]. Cambridge, MA: author
- Flombaum J I, Kunder S, Santos L R, Scholl B J, 2004 "Dynamic object individuation in rhesus macaques: a study of the tunnel effect" *Psychological Science* **15** 795 - 800
- Flombaum J I, Scholl B J, 2006 "A temporal same-object advantage in the tunnel effect: facilitated change-detection for persisting objects" *Journal of Experimental Psychology: Human Perception and Performance* **32** 840 - 853
- Gordon R, Irwin D, 1996 "What's in an object file? Evidence from priming studies" *Perception & Psychophysics* **58** 1260 - 1277
- Gordon R, & Irwin D, 2000 "The role of physical and conceptual properties in preserving object continuity" *Journal of Experimental Psychology: Learning, Memory, & Cognition* **26** 136 - 150
- Henderson J M, 1994 "Two representational systems in dynamic visual identification" *Journal of Experimental Psychology: General* **123** 410 - 426
- Henderson J M, Anes M D, 1994 "Roles of object-file review and type priming in visual identification within and across eye fixations" *Journal of Experimental Psychology: Human Perception and Performance* **20** 826 - 839
- 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 & Review* **8** 753 - 760
- Kahneman D, Treisman A, Gibbs B J, 1992 "The reviewing of object files: Object-specific integration of information" *Cognitive Psychology* **24** 174 - 219
- Kanwisher N G, 1991 "Repetition blindness and illusory conjunctions: Errors in binding visual types with visual tokens" *Journal of Experimental Psychology: Human Perception and Performance* **17** 404 - 442
- Kruschke J K, Fragassi M M, 1996 "The perception of causality: Feature binding in interacting objects" In Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society (pp. 441 - 446) Hillsdale, NJ: Erlbaum
- Mitroff S R, Alvarez G A, in press "Space and time, not surface features, guide object persistence" *Psychonomic Bulletin & Review*
- 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, Lleras A, 2005 "On the role of object representations in substitution masking" *Journal of Experimental Psychology: Human Perception and Performance* **31** 1171 - 1180
- Noles N S, Scholl B J, Mitroff S R, 2005 "The persistence of object-file representations" *Perception & Psychophysics* **67** 324 - 334
- Pylyshyn Z W, 1989 "The role of location indexes in spatial perception: A sketch of the FINST spatial-index model" *Cognition* **32** 65 - 97
- Rensink R A, 2000 "The dynamic representation of scenes" *Visual Cognition* **7** 17 - 42
- Wolfe J M, Bennett S, 1997 "Preattentive object files: Shapeless bundles of basic features" *Vision Research* **37** 25 - 44
- Xu F, Carey S, 1996 "Infants' metaphysics: the case of numerical identity" *Cognitive Psychology* **30** 111 - 153

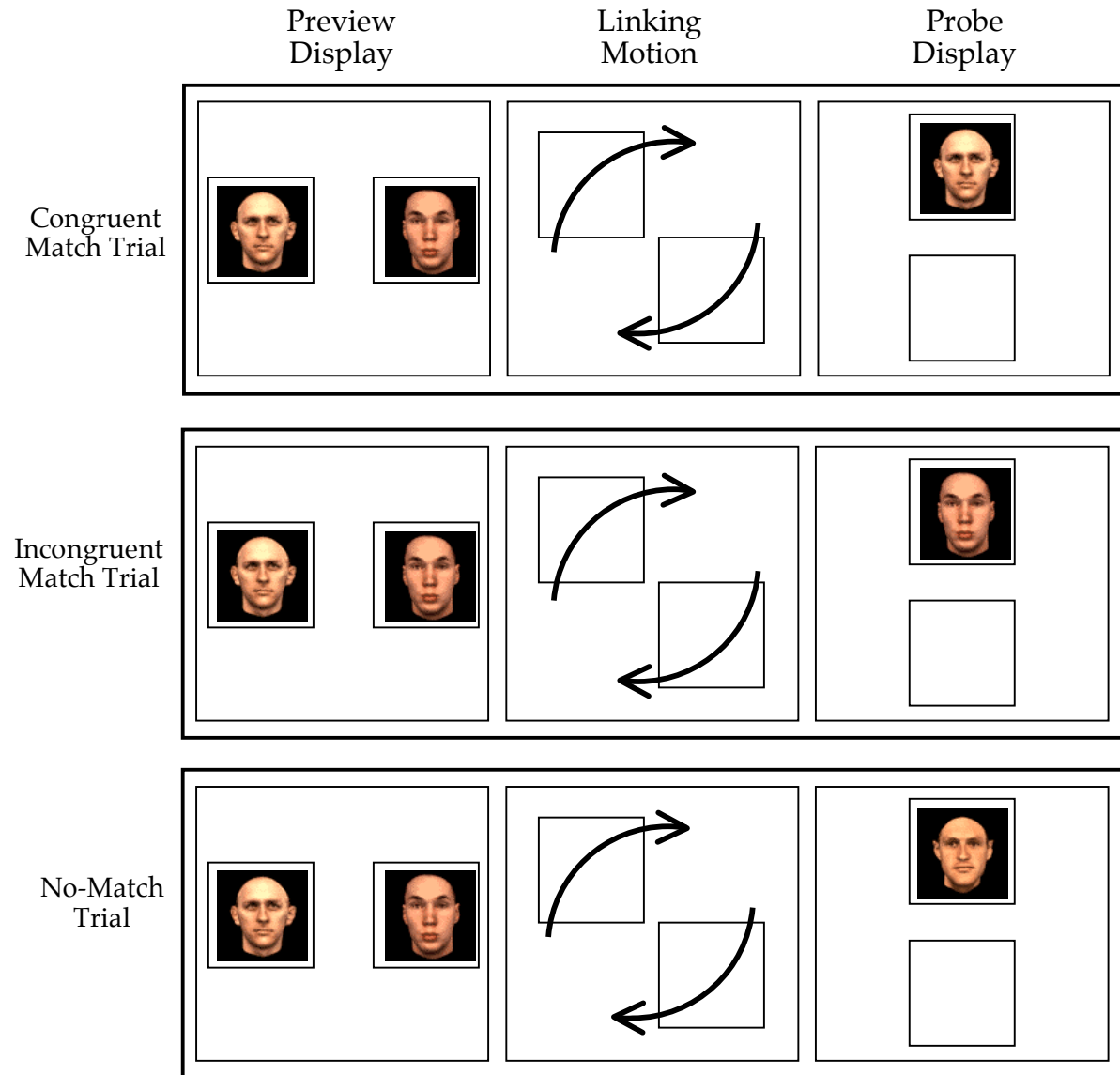
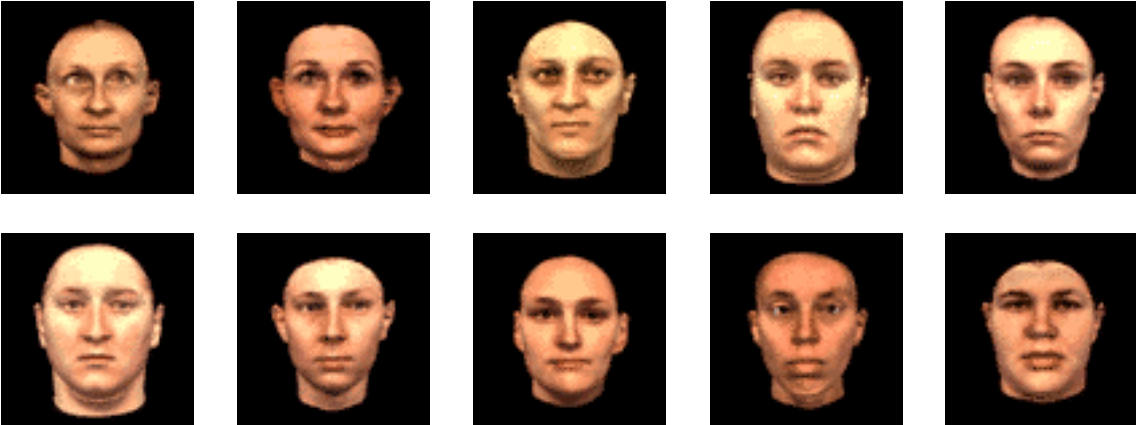


figure 1. Depiction of the object-reviewing paradigm and the face stimuli used in the current experiment. Each trial consisted of Preview, Linking, and Probe displays. The probe could be the reappearance of one of the two preview items (on Match trials) or a novel item (on No-Match trials). On Match trials, the probe could reappear in the same object in which it was previewed (on Congruent trials) or in the other object (on Incongruent trials). The observers' task was simply to report whether or not the probe item had appeared anywhere in the preview display. Note that this depiction is not drawn to scale and that the probe face could reappear in either the top or the bottom frame.

Sample Female Face Stimuli



Sample Male Face Stimuli



figure 2. A subset of the face stimuli used in our study.