Using eye-tracking to parse object recognition: Priming activates primarily a parts-based but also a late-emerging features-based representation

Peter Gerhardstein[1] · Sarah Olsen[2]

Abstract

Biederman and Cooper (*Cognitive Psychology, 23*, 393-419, 1991), using parts-deleted and features-deleted stimuli, presented evidence that object priming occurs at the level of the objects' parts, but not features. A control condition confirmed that some priming was accrued by the (non-visual) object concept that the stimulus represented, but all visual-level priming appeared to be at the more global level of the parts of objects, rather than the local level of the individual features (edges, vertices). This outcome has long been viewed as an important piece of supporting evidence for the existence of structural descriptions (e.g., Biederman, *Psychological Review, 94*, 115, 1987). The original report used a naming response task, and concluded that stimuli presenting half of the parts of an object primed only those parts, whereas half-features-deleted stimuli primed both themselves and the "complementary" half, containing the features deleted from the first image, equally. The current study adapts an eye-tracking approach to enable examination of the time course of priming across an exposure to both the primed image and unprimed competitors. Parts-deleted images primed themselves quickly and exclusively, replicating the finding of Biederman and Cooper (1991). Features-deleted images showed a deviation across time, however; initially a features-deleted prime attracted looking to itself and to its complement equally, but later on, looking to the target deviated upward, demonstrating an ability to distinguish between the two versions. The outcome of the present tests provide support for the primacy of a structural parts description, while also demonstrating the existence of multiple types of representations, both global and local.

Keywords Priming · Object recognition · Eye movements · Cognitive

Introduction

A traditional view of object recognition argues that the process is essentially bottom-up, building a representation of objects by detecting low-level local features (edges, vertices) and using these features to activate higher-level structural descriptions consisting of parts (Biederman, 1987; Dickinson, Pentland & Rosenfeld, 1992b; Marr, 1982; see also Hummel, 2000). This account asserts that the visual system uses a set of "volumetric primitives" (parts) as the elements from which lasting representations are constructed, and that the local features are not contained in these lasting object representations.

A seminal finding in support of this structural parts-based representation for object recognition (Biederman & Cooper, 1991; see also Hayworth & Biederman, 2006) examined priming using either parts or features (see Fig. 1), and found that while features-based priming was non-specific, parts-based priming was highly specific to the parts presented. In an object-naming task, Biederman and Cooper (1991) presented participants with contour-deleted images of common objects in which 50% of either the object's parts or features were deleted (parts- and features-deleted, respectively; see Fig. 1) and asked them to name the object. In a subsequent test phase, participants were asked to name another set of images that contained the identical images they had seen during the study phase, their 50% parts- or features-deleted counterpart (called the complement), or a different exemplar with the same object name (e.g., an upright piano and a grand piano). Both parts- and features-deleted objects showed non-visual priming, as evidenced by increased speed and accuracy for naming the

✉ Peter Gerhardstein
gerhard@binghamton.edu

[1] Department of Psychology, Binghamton University-SUNY, Binghamton, NY 13902, USA

[2] Department of Psychiatry, School of Medicine, University of Minnesota, Minneapolis, MN, USA
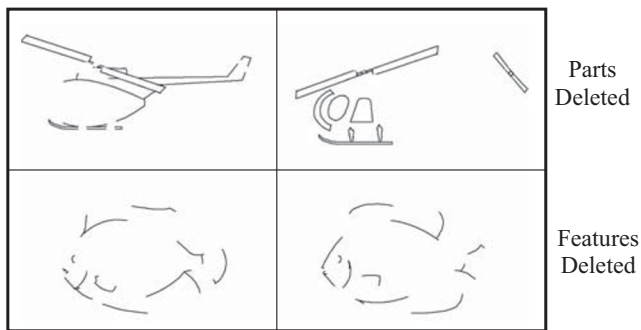
Fig. 1 Examples of stimuli from Biederman and Cooper (1991). The top row displays two complementary images deleted at the level of the object parts, while the bottom row displays two features-deleted complementary images

different exemplar at test. For features-deleted images, there was equivalent visual priming (increased speed and accuracy relative to the different exemplar) for the identical primed image and its complement. However, for parts-deleted images, only the identical primed image was visually primed, whereas the complement received no more (visual) priming than the different exemplar.

The combination of these two outcomes (sensitivity to a parts-based change, no sensitivity to a features-based change) was interpreted as strong support for the existence of a structural representation in which, following a minutes-long delay, only the parts activated by a displayed image remained active in memory. In Biederman's (1987) structural representation view, features-deleted stimuli and their complements both display incomplete information for all of the image's parts, while parts-deleted stimuli contained relatively complete information, but for only half of the image's parts. Therefore, features-deleted images would be predicted to activate all of the parts of an object, and thus both the primed image and its complement would activate the identical, and complete, object representation (all parts). Conversely, parts-deleted images offer no direct path to activation of all of the parts of the object, and therefore the representations (which the structural viewpoint argues is at the level of the parts) activated by the parts-deleted image and its complement are distinct.

Given this perspective, observers either did not ever encode the presented image at the level of the individual features or could not retrieve those individual features at test, but were clearly able to encode and retrieve the specific parts presented. Biederman and Cooper (1991), as well as other early studies, motivated the argument for a structural description such that the representation accessed by this type of object priming contains only parts, and not local features. This finding and subsequent studies (Hayworth & Biederman, 2006; Hummel, 2000; Hummel & Stankeiwicz, 1996; Lerner, Hendler, & Malach, 2002) have cemented the assertion that the visual system constructs and uses an enduring parts-based representation (see Hummel, 2000, for discussion regarding the importance of such representations).

In the years following this finding, however, the research literature has developed a more nuanced view of the nature of object representations (see Hummel, 2013 for a review, but see also Hummel, 2000), including the idea that both structural descriptions and features- or view-based representations (Edelman & Weinshall,1991; Poggio & Edelman, 1990; Riesenhuber & Poggio, 2002) are present in human object recognition. Bar and colleagues, for example (Bar, 2003, 2004; Bar, Kassam, Ghuman, et al., 2006; Kveraga, Boshyan & Bar, 2007) have proposed that top-down processes play a primary role in typical object recognition by generating an initial guess about the identity of an object. Kveraga, Boshyan, and Bar (2007) proposed that the visual system uses an initial magnocellular-centered "gist" process, involving potentially low spatial frequency information, to quickly generate initial guesses about object identity, followed by a slower parvocellular process, instantiating something similar to the more traditional idea of a bottom-up edge- or features-based analysis. This description appears to require both parts-based and features-based representations (although both might not be encoded into long-term memory). Note that this idea coincides with the finding of a global precedence effect in visual processing (Navon, 1977), in which observers are found to perceive characteristics of the global shape of objects faster than local details of the same objects. The argument by Bar and colleagues supplements the more classic bottom-up understanding of the process of object recognition, in which small segments are used to detect contours, contours are used to detect parts, and parts to represent objects in an essentially feedforward process (see Serre, Oliva, & Poggio, 2007, for more on this point). The gist-driven process proposed by Bar and colleagues would constrain the outcome of the features-driven process, and accelerate identification.

One result supporting a multiple-representations view is a finding by Schendan and Kutas (2007) in which contour-deleted images, similar to the features-deleted images used by Biederman and Cooper (1991), elicited different event-related potential (ERP) waveforms in observers. In this study, participants were shown partially deleted or intact images in a study phase, followed by the identical, primed (features-deleted) image or its complement in a subsequent test phase. Results revealed that an early (P2) waveform centered on the occipito-temporal area (possibly overlapping the lateral occipital complex; see *Discussion*), was insensitive to local changes, as was a somewhat later (N350) waveform occurring in more frontal regions. A third waveform (later still), however, demonstrated a clear sensitivity to local feature changes, supporting accounts positing multiple types of representation, both holistic and features-based.

This result by Schendan and Kutas appears to agree substantively with the sequential two-part process posited by Kveraga, Boshyan, and Bar (2007), and suggests early generation of a gist-level representation, followed by a later local-

level representation. Through the use of ERPs, which are able to provide a much more detailed glimpse into the time-course of the processes occurring during object recognition than simple accuracy or reaction time measures (such as those used in earlier studies), Schendan and Kutas were able to shed light on the more complicated nature of object representations. However, Schendan and Kutas (2007) did not explicitly manipulate parts-based information in their images, and therefore did not directly address questions regarding the nature of the higher-level representation.

Schendan and Kutas (2007), like Biederman and Cooper (1991), employed a study- and test-phase design with a delay of minutes between the two phases, which assesses the long-term memory of representations. It is possible that the nature of the representations differs at shorter timescales. For example, Ellis and Allport (1987; see also Schendan & Kutas,2003) argue for the existence of a limited-duration initial representation tied to the specific properties of the presented image (e.g., local features), along with activation of a more durable (and more global) representation (i.e., parts). Following shorter delays (100 and 500 ms), they saw effects of their featural manipulation, but following longer delays (2,000 ms), those effects were absent. Their tests, however, presented changes in viewpoint, rather than using fragmented objects. Additionally, they did not have the advantage of a measure sensitive to the time course of processes afforded by eye-tracking or EEG.

The use of eye-tracking, and in particular the use of the visual world paradigm (VWP; Allopenna, Magnuson, & Tanenhaus, 1998), has the potential to shed light on the time course of object representations at shorter timescales, as well enabling assessment of the direct competition between differing types of information (features as compared to parts). The VWP (for review, see Huettig, Rommers, & Meyer, 2011) presents a prime (verbal in the original application) accompanied (or followed) by a display of multiple images in an array. Proportion of looking is assessed using eye-tracking over a short display interval following exposure to the priming stimulus, to determine the effect of the prime on looking to the target image in comparison to a set of simultaneously presented visual competitors. This eye-tracking approach reveals the time-course of priming, allowing investigation of the relative impact of the prime on different types of related and unrelated stimuli across time as priming unfolds. Eye-tracking, used in this manner, offers insights into the priming process that do not appear to be available with other measures.

The use of the VWP approach enables a direct test of the question of whether the parts-based representation postulated by structural descriptions is formed rapidly and exclusively in the visual system. Biederman and Cooper's original finding argues for an exclusively parts-based process, while Schendan and Kutas (2007) and others suggest that a local features-based process, slower than the more global parts-based

process (and possibly informed by it, as suggested by Kveraga et al., 2007), is operating. The goal of the present tests was to address this question, and to determine whether a parts-based representation would appear on the same time course as the more global representation tested by Schendan and Kutas. A more general goal is to introduce the use of the VWP in the context of eye-tracking to demonstrate the utility of this approach for testing questions in this area.

Experiment 1: Parts and features priming across a short interstimulus interval (ISI)

The present investigation was intended to examine the processes invoked by short-term priming to determine whether evidence for both parts-based and features-based representation can be detected using the same stimuli that elicited the effects reported by Biederman and Cooper (1991). This was done by adapting the VWP to present a *visual* (rather than auditory) prime prior to the array presentation, using parts- and features-deleted stimuli as in Biederman and Cooper (1991) The array contains the prime itself (target), the complementary version of the same object (complement), a different exemplar, and an unrelated distractor. The different exemplar was presented as part of the array to enable discrimination of the priming attributable to the visual aspects of the displayed prime from non-visual (category, name-level) priming. A fully unrelated distractor was also included in the test array as a baseline control (as is commonplace in VWP studies). If, following the visual prime, observers demonstrate a differential looking pattern to the primed target as compared to the unprimed complement, this would indicate a clear distinction, which would establish the activation of a prime-specific representation. Overlapping looking proportions, on the other hand, would argue for priming of a single representation indistinguishable between target and complement at the level of the prime (features- or parts-deleted).

Critically, the structural perspective (e.g., Biederman, 1987; Dickinson, Pentland, & Rosenfeld, 1992a, b) and multiple-representations viewpoints (Kveraga et al., 2007; Schendan & Kutas, 2007) make different predictions regarding looks to the target and complement for parts- as compared to features-deleted images. Specifically, the structural viewpoint argues that observers will not perceive any difference between the two features-deleted versions of the object at test, because the features are not theorized to be encoded in the representation. The parts-deleted prime, however, would be expected to produce clearly differential priming, with the parts-based target accruing all visual perceptual priming at test, and the parts complement engendering no more looking than the different exemplar (sharing only a category name with the target). A multiple-representations view, however, would predict that the parts-deleted primes will elicit a

strongly discriminative response early in the array exposure, while features-deleted primes could also promote a discriminative response somewhat later in the probe array exposure.

Method

Participants

Binghamton University college students (N=45) served as participants. This level of participants was selected based on that used by previous VWP studies (e.g., Allopenna et al., 1998; Altmann & Kamide, 2009; Huettig & Altmann, 2005). Post hoc power analyses, while not an optimal approach (e.g., Zhang et al., 2019), indicated that this number of participants was sufficient (see Target Only and Target + Complement dwell time analysis, below). Participants received course credit for their participation. The data from six participants were discarded due to problems with the apparatus or failure to calibrate to the eye-tracker (most of these wore vision correction). Two additional participants were eliminated following track-loss analysis (see *Results* section, below), resulting in a final N=37. Data collection adhered to the ethical principles of the American Psychological Association (APA) and was conducted under a reviewed and approved protocol.

Apparatus

An SMI RED 120 Hz eye-tracker was used to record eye movements. It was placed on a table 60 cm in front of the participant and positioned to an angle of 12.5°. The experiment was presented using a 112 × 70 cm Samsung monitor at a viewing distance of approximately 180 cm from the observer. The calculated visual angle of the monitor was 32° in width and 21° in height. All participants sat on a movable seat taken from a vehicle. The electric controls from the vehicle seat allowed the researcher to tilt the car seat forwards and backwards or move the seat itself forwards or backwards by a small amount to enable the eye-tracker to clearly detect the eye. The experiment itself was constructed and presented using E-prime 2.0 (Psychology Software Tools), and integrated with the SMI hardware using the SMI SDK for E-prime. E-Prime scripts were used to implement the gaze-contingent aspects of the experiment.

Stimuli

Stimuli consisted of 16 features-deleted and 16 parts-deleted line drawings of common objects (e.g., airplane, table, rabbit, etc.) taken from Biederman and Cooper (1991). Half of the items/objects (eight) had both a features- and parts-deletion version, and the other half did not, meaning that there were 24 different objects in the test, across features and parts. The condition in which each object was presented (parts or features) was counterbalanced within subjects for those objects having both a features-deleted and a parts-deleted version; they would see one exemplar in the features-deleted condition (as prime) and the other exemplar in the parts-deleted condition. An additional 16 Target-Only trials (eight features-deleted and eight parts-deleted) were included in which there was no competitor of any type; the primed image was presented with three unrelated distractors. Each participant therefore observed a total of 48 experimental trials.

Test arrays

Each trial presented a fixation cross and then a prime image, which was followed by a test array (see Procedure, below, for details). Prime (target) images were displayed in the center of the screen and were sized to fit inside a square spanning 6.0° of visual angle. The four images in the test array were also sized to (each) fit inside a 6.0° square. The objects in the test array varied based on the type of trial (Target + Complement and Target-Only trials), but were always consistent with the prime in that all matched in terms of the type of deletion presented (parts or features). Images in the test arrays were arranged in four quadrants on the screen such that there was a distance of 12.4° horizontally and 11.9° vertically between the center of adjacent images, meaning that there was substantial space between images in the array.

Target + complement trials

Test arrays in these trials (32 per observer) displayed the primed (target) image, its 50% deleted complement, a different exemplar image, and an unrelated distractor image. If the target was a features-deleted image, the test array contained only features-deleted images; the same was true for parts-deleted trials. See Fig. 2 for an example of test array images for both features- and parts-deleted trials. Subjects saw one item from each test array as the target, resulting in 16 part- and 16 features-deleted trials (32 total experimental trials in a given list). Given that eight of the 24 objects were the same across features and parts, if one image served as the target for a features-deleted trial, one of the different exemplars images would serve as the target for the parts-deleted (counterbalanced across subjects), such that an object image would be seen as the target (prime) only once by each observer. Finally, the unrelated distractor was randomly selected from the set of unused different exemplar images.

Images were counterbalanced across eight lists such that each image in a test array served as the target, complement, and different exemplar across the lists. The location of the four items in the array was quasi-random such that the target, complement, different exemplar, and unrelated distractors appeared in each of the four quadrants an equal number of times.
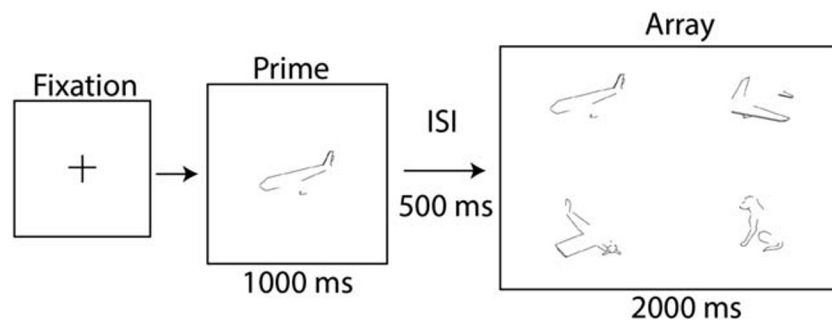
Fig. 2 Trial sequence. Note that an array from a *target + complement* trial is shown (see text for a description of the difference when a target-only trial was presented). Each trial began with an attention-attracting fixation point (a small colorful spinning orb). This fixation point was gaze-contingent, such that subjects needed to accrue 350 ms of looking time in order for the trial to progress, to ensure that all participants were looking at the middle of the screen when the prime appeared. After 2,000 ms, however, the trial progressed automatically. The fixation point was followed by a 250-ms blank screen. The prime (or target) image followed the blank screen and was presented for 1,000 ms. The prime was replaced by a black dot in the middle of the screen in order to maintain looking in the center of the screen. This dot was present for 500 ms, and was replaced by the test array. Following the array there was a 1,000-ms blank screen before the start of the next trial

Target-only trials

Each list also contained 16 target-only trials (eight features-deleted and eight parts-deleted) in which the test array contained the target image and three unrelated distractors. Targets in these trials were eight of the 16 items that had served as unrelated distractors in the experimental trials. The three unrelated distractors in each array were selected quasi-randomly from the remaining set of different exemplars and unrelated distractors from a given experimental list, so that the target and complement images from experimental trials never appeared in a target-only trial, and such that the different exemplar and unrelated distractor images from the experimental trials were only repeated once as fillers.

The 48 total trials in each list were presented in random order, such that parts- and features-deleted trials, and target-only and target + complement trials were randomly intermixed. Participants saw only one of the eight lists (lists were counterbalanced across participants).

Procedure

Prior to starting the experiment, participants provided informed consent and performed a Snellen eye test; participants with less than 20/40 vision (corrected) would be dropped (none were). Participants were then seated in front of the eye-tracker and were instructed that the system would be calibrated to their eyes, at which point the experiment would begin. They were instructed that they would see an image followed by an array of four images, one of which was the one they had just seen, and their task was to find the image they had just seen. Participants were also informed that the experiment was gaze-contingent, and that they would make their response simply by looking at the images. Following instructions, the eye-tracker was calibrated to the participants'

eyes (using a 5-point calibration routine, followed by validation of the calibration) and the experiment began.

In order to acquaint participants with the task and the pacing of the trials, the experiment began with four practice trials. The images in these trials consisted of completed (i.e., not contour-deleted), easily recognized, black and white cartoon images of animals and insects. Trials containing the contour-deleted images immediately followed the practice trials. See Fig. 2 for a visual depiction and description of the trial sequence. The array exposure was gaze-contingent such that if a participant looked at one of the four images (in one of the four quadrants) for 660 ms consecutively, the trial ended (a timeout of 2,000 ms ended trials if this did not occur; in practice, a majority of trials were terminated by gaze-contingent looking to any of the four areas for at least 660 ms). The sequence was the same for each of the four practice trials and 48 contour-deleted trials (32 target + complement and 16 target-only trials). Participants were debriefed following the experiment.

Results

Data were processed using scripts written in R-Studio (R Core Team, 2018) and using EyeTrackingR (Dink & Ferguson, 2015) in particular. Raw data were processed into the format required for EyeTrackingR using R. All observers' data was filtered to include only the first 1,600 ms following array presentation, for a total of 192 possible eye-tracking samples per trial. Four areas of interest (AOIs; one for each quadrant) were defined as squares of 7.5° visual angle centered around each of the four images in the array. Applying the EyeTrackingR "trackloss" function with an elimination criterion of 25% maximum permissible loss (with on-screen non-AOI looks included) for a trial resulted in removal of 14 trials or < 1% of trials. When classifying non-AOI looks as trackloss, 107 trials exceeded 50% loss, with 52 of those from

two participants. Those participants were excluded from the analysis, and all other trials with over 50% track loss (55, or < 1%) were also dropped, leaving 1,721 trials across all remaining participants (N=37). In all of the analyses described below, each of the four locations in the test array (see Fig. 2) was assigned as an AOI and looking samples that did not land in one of these areas were discarded (i.e., were treated as trackloss). In this way, data for analysis were filtered to remove all offscreen and off-AIO looks. Note that this resulted in apparently highly variable looking data early in each trial (generally in the first 250 ms following array onset); this is misleading, as the vast majority of looks during this time were to non-AOI locations because the dot presented between the prime and target centered looking at array onset to a non-AOI location. This required observers to make at least one saccade to fixate in any AOI. These non-AOI data, however, do not contribute to the comparisons of interest and rapidly decline in proportion across the time course of array looking, and so the looking proportion analyses below were performed with these samples excluded, as is typical in the VWP literature.

Two types of analyses are presented. For each type of target condition (Target-Only, Target + Complement), a dwell-time analysis is presented, followed by a divergence analysis (Dink & Ferguson, 2015). The dwell-time analysis uses overall looking time within each AOI; the divergence analysis is described next.

Divergence analyses were conducted on the proportion of looking in each of the AOIs across time (in 25 ms time bins). These analyses were conducted using the EyeTrackingR divergence analysis package (Dink & Ferguson, 2015; Maris & Oostenveld, 2007), which enables examination of the onset and duration of a difference between two effects across time while controlling for family-wise error using a boot-strapping routine in a cluster-based permutation analysis. This package iteratively samples the data many times from subjects in each condition, smoothing each sample, and from those samples builds a null distribution. Continuous clusters with a mean outside the 2.5–97.5% range of the null distribution are reported (see the EyeTrackingR package description and Maris & Oostenveld, 2007, for more on this approach). Results below report the summary statistic and the p-value, but do not report the null mean (which was near zero in all cases) or range (typically within -22 to +22; never exceeding 25 when using 250 iterations) unless the statistic is relatively close to the null distribution.

Target-only presentation – dwell analysis

The primary purpose of this condition was to test the amount of priming from a parts-deleted prime to itself in comparison to a features-deleted prime to itself, in the absence of any competition from a complementary image or a different exemplar. A two-factor within-subjects ANOVA was conducted using dwell time within each AOI as a percentage of overall looking time, with factors of stimulus condition (feature or part) and AOI (target, average distractor). Looking within the three distractors in these trials was averaged for purposes of analysis. The main effect of AOI was reliable; $F(1, 144) = 5413.79$, $p \ll .0001$, $\eta^2 = 0.97$, unsurprising as the overall difference in looking between the target and distractors was expected to be large in all cases. The main effect of condition was not significant; $F < 1$. This lack of an effect was mediated by a significant interaction; $F(1,144) = 11.62$, $p < .0009$, $\eta^2 < .002$, a notably small effect. Examination of the data (see Fig. 3) and follow-up contrasts showed that the interaction was due to significantly more looking to the parts-based than to the features-based target AOI [$t(144) = 3.07$, $p = .0023$], while in the average distractor AOI there was no such difference [$t(144) = 1.71$, $p > .08$]. A post hoc power analysis (R-WebPower for repeated-measures ANOVA; Zhang & Yuan, 2018) using the sample size and effect sizes (converted to Cohen's f) produced in the dwell time ANOVA as the measure of effect size and number of groups as directed by the analysis, indicated a power of 1.0 for the main effect of AOI, but only 0.05 for the interaction effect. It should be noted that a post hoc power analysis is less than optimal; the sample size was selected using typical sample sizes in the VWP literature that have previously produced reliable results, but issues with a post hoc power analysis persist regardless (see, e.g., Zhang et al., 2019), and any follow-up work should avoid this approach.

Divergence analysis

A divergence analysis using the R package t.test for time clusters (Dink & Ferguson, 2015) was used to compare target to distractor looking proportion for each type of deletion (features and parts). These analyses demonstrated that both parts- and features-deleted targets primed themselves strongly, diverging from a randomly-selected distractor AOI at 350 ms for parts (cluster sum statistic = 1281.03; $p < .0001$), and slightly later at 400 ms for features-deleted images (cluster sum statistic = 919.37; $p < .0001$). Initial inspection of the features- and parts-deleted conditions (within subjects) suggests that the two target types appear to have attracted similar priming responses (see Fig. 4), but analysis revealed a difference between 350 and 900 ms (cluster sum statistic = -55.18, $p < .0001$); the features-deleted stimuli appear to have taken longer to achieve the same level of priming (as reflected by fixation proportion) as the parts-deleted stimuli.

Target + complement trials When the prime is followed by an array displaying both the prime itself and the complementary image, plus a different exemplar and an unrelated distractor, there are several types of potential priming that may affect looking. The identity prime is assumed to be the most
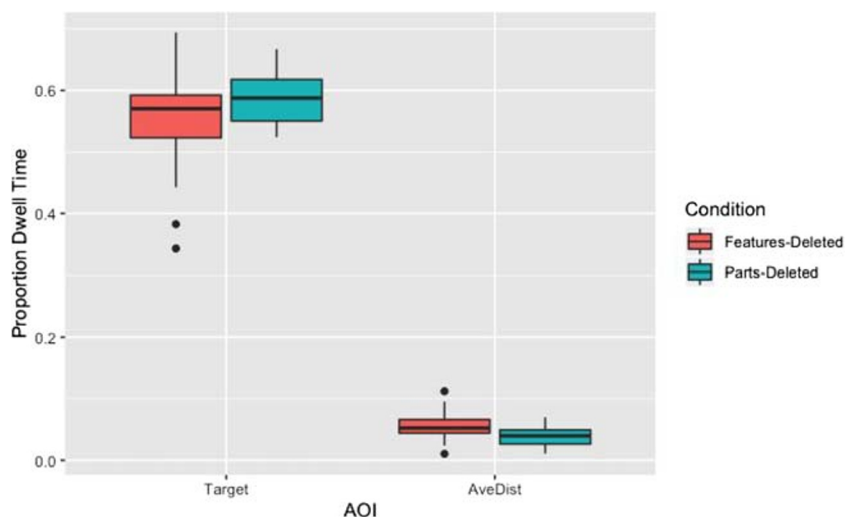
Fig. 3 Dwell time as a function of AOI and condition when the target was paired with three unrelated distractors. Targets clearly attracted more looking than distractors overall; only in the target AOI did the parts-based stimuli elicit more looking than the features-based stimuli

effective priming target (because there is no difference between the priming image and the probe image). Facilitation from the prime to the complementary image is expected to depend upon the amount of overlap between the prime and the complement at the level of a persisting representation (with different levels of activation expected on parts- as compared with features-deleted trials). Non-visual priming should occur at a lower level between the prime and the different exemplar in all cases. A baseline control (from the distractor, an unrelated object) is expected to attract relatively little looking in the array. To investigate these differences, an analysis on dwell time was conducted including both parts and features. Additionally, the same divergence analysis (Dink & Ferguson, 2015) as above was conducted using looking

proportion across time between AOIs following priming with a parts-deleted or features-deleted stimulus.

Dwell-time analysis

As above, a two-factor within-subjects ANOVA [stimulus condition (feature or part) and AOI (target, complement, different exemplar, distractor)] was conducted using dwell time as a dependent measure. The main effect of AOI was reliable; $F(3, 288) = 1417.30, p \ll .0001; \eta^2 = 0.48$. This was again an expected outcome as the overall difference in looking between the target and non-visually similar AOIs was expected to be extensive. The main effect of condition was not significant; $F(1,288) = 1.17, p > 0.28$. As in the target-only condition, this
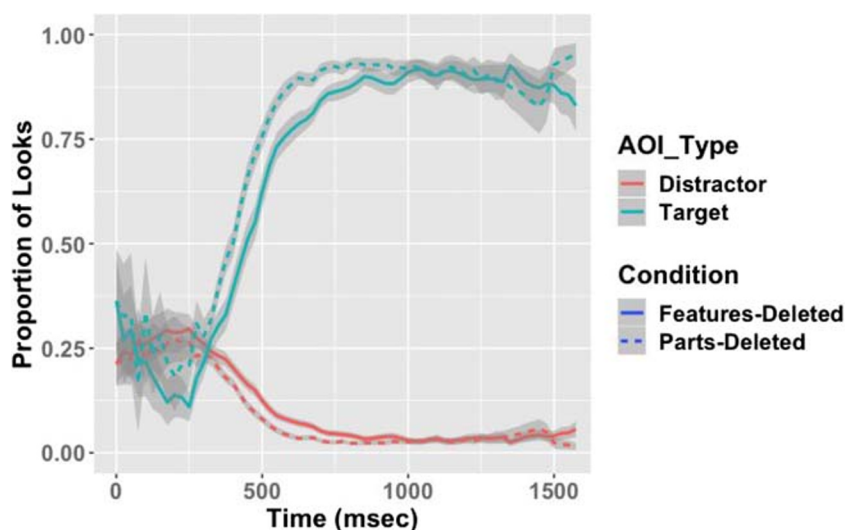

Fig. 4 Proportion of looks to the target (the prime) in comparison to a distractor in the arrays for target-only trials. Both parts-deleted and features-deleted trials are shown. Non-areas of interest (AOIs) looks are

not graphed, but are included in the calculation of data for the graph, resulting in less than 100% proportions at early time points

lack of an effect was mediated by a significant AOI × condition interaction; $F(1,288) = 174.06$ $p \ll .0001$, $\eta^2 < .06$. Visual inspection of the data (see Fig. 5) and follow-up contrasts revealed significantly more looking to the parts-deleted than to the features-deleted target AOI [$t(288) = 14.84$, $p < .0001$]; this was also true for the complement AOI [$t(288) = 17.40$, $p < .0001$]. Additional contrasts comparing target to complement AOIs in the features-deleted [$t(288) = 11.33$, $p < .0001$] and parts-deleted [$t(288) = 43.58$, $p < .0001$] conditions separately revealed that in both cases, the target AOI attracted significantly more looking than the complement AOI. A post hoc power analysis using the sample size and Cohen's f statistics converted from the analysis to estimate effect size (R-WebPower for repeated measures; Zhang & Yuan, 2018) on dwell time for Target + Complement conditions indicated a power of 0.76 to detect the main effect and low power (0.25) for the interaction effect.

Parts-deleted divergence analysis

Looking to the target AOI in comparison to the complement, different exemplar, and distractor AOIs, respectively (see Fig. 6, left panel), following priming with a parts-deleted stimulus, was examined using a divergence analysis. These analyses, using the R package t.test for time clusters, all showed reliable effects. The target and complement AOIs diverged from 350–1,600 ms (cluster sum statistic = 815.06, $p < .0001$), as did the target and different exemplar AOIs (cluster sum statistic = 925.96, $p < .0001$). The target AOI attracted more looking than the distractor from 325–1,600 ms (cluster sum statistic = 1229.95; $p < .0001$). The complement AOI, however, did not attract reliably more looking than the different exemplar AOI at any time during the array exposure. The different exemplar AOI demonstrated a clear non-visual priming

component in observers' looking in comparison to the distractor AOI, attracting more looking from 450–1,225 ms (cluster sum statistic = 148.6, $p < .0001$).

Features-deleted divergence analysis

The same tests were conducted on priming from a features-deleted stimulus (see Fig. 6, right panel). In this condition, however, the target-complement AOI comparison showed a substantially different pattern relative to the parts-deleted condition. The analysis showed no divergence in looking until 950 ms (cluster sum statistic = 82.88; $p < .0001$); this divergence remained significant through the end of the array exposure analyzed. The target and different exemplar AOIs diverged from 375–1,600 ms (cluster sum statistic = 615.47; $p < .0001$), similar to the parts-deleted condition. Looking to the target also increased relative to the distractor, beginning at 300 ms (cluster sum statistic = 845.86; $p < .001$). The complement AOI attracted more looking than the different exemplar AOI from 425–1,600 ms (cluster sum statistic = 437.02, $p < .0001$), a clear difference relative to the parts-deleted priming condition, in which this pairing showed no divergence in looking at any point. Again, there was a clear non-visual priming component in observers' looking: Different exemplars attracted more looking than distractors from 425–1,125 ms (cluster sum statistic = 127.9, $p < .0001$).

Discussion

Adult observers largely replicated the original findings of Biederman and Cooper (1991) and Schendan and Kutas (2007), and provided support for the presence of a structural representation, as well as a late-emerging features-based representation. As in the original study, parts-deleted primes
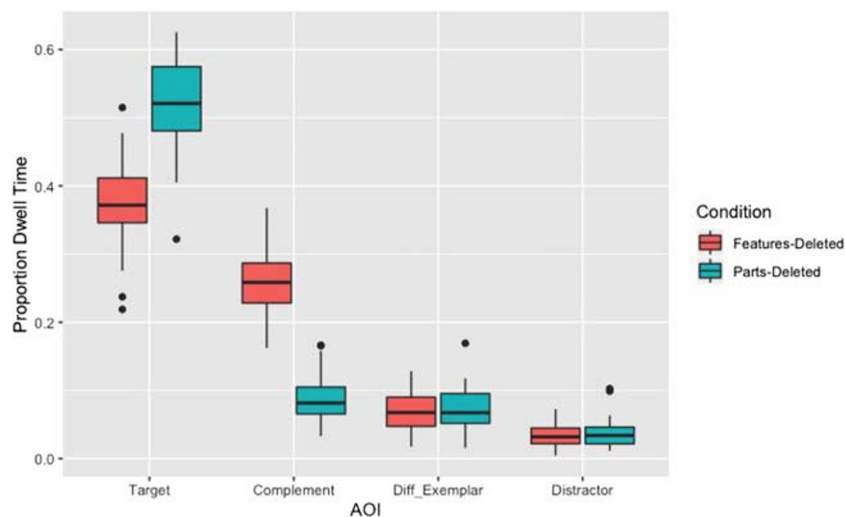

Fig. 5 Mean dwell time by stimulus type (features-deleted, parts-deleted) and area of interest (AOI). Note that while the Target-Complement difference in looking is significant in both the parts- and features-deletion conditions, the disparity is much larger in the part-deleted condition
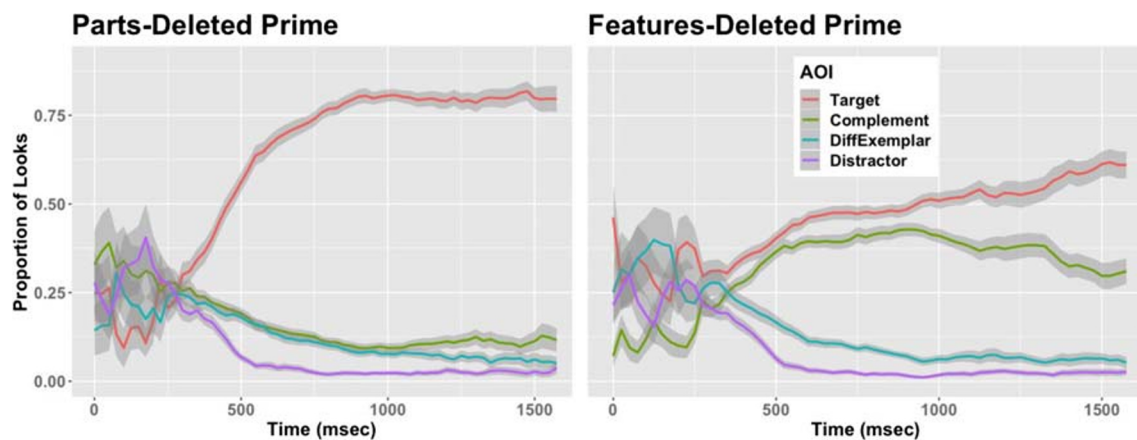
Fig. 6 (**Left**) Proportion of looking to the parts-deleted target + complement trial. (**Right**) Proportion of looking for the features-deleted trials

prompted looking to themselves from early in the array exposure to a much greater degree than to their parts-level complement. The parts-deleted complement did not attract looking at a level that was different to that of the different exemplar. This strongly supports the original finding that the parts retained all of the visual perceptual priming from the prior exposure, and that the only other significant source of priming was from the category name, which was the same between the complement and the different exemplar, and greater than for the unrelated distractor. Importantly, there is no evidence that a parts-deleted prime activated the missing parts or features. This point is in strong agreement with Biederman and Cooper's report, and the presence of a parts-based representation, even at a short ISI. The parts-based finding was not tested by Schendan and Kutas (2007), who tested features-deleted stimuli in comparison to a global (complete object) shape. The current result further illustrates the nature of object representations, suggesting that objects may be represented by their component parts from early in the process of recognition.

The features-level priming condition was somewhat more nuanced, however, as results provided evidence supporting Biederman and Cooper's (1991) original finding as well as a multiple representation view. In agreement with Biederman and Cooper, the features-deleted prime and its complement attracted equal amounts of looking from early on in the array exposure. Given that Biederman and Cooper measured reaction time, the first portion of the curve depicted in Fig. 5, in which the feature target and complement images attracted equal amounts of looking, would likely have driven responding in their test. The late-occurring divergence in looks to the features-deleted target and complement (discussed below), however, would not have been captured by their test. Additionally, both the feature target and its complement attracted greater looking in comparison with the different exemplar from early in the array exposure as well, also matching the visual priming outcome seen in Biederman and Cooper (1991).

There was also evidence in favor of a multiple-representation perspective. Specifically, looks to the features-deleted target, as compared with its complement, did eventually diverge late in the array exposure. This was also reflected in the analysis of dwell times, in that there was more overall looking to the features-deleted target than its complement. This is in line with the findings of Schendan and Kutas (2007), and shows that this features-based distinction is present even at much shorter priming ISIs. The eventual deviation in looking suggests that observers possess the capacity to differentiate between the features-deleted prime and its complement, an observation apparently at odds with Biederman and Cooper's original finding, and in line with the idea that objects are represented at multiple levels. The outcome supports the assertion of Bar and colleagues that there is an early emerging and fundamentally global representation followed by a late-emerging featural representation (Kveraga, et al., 2007), even across quite short ISI durations.

Experiment 2: Feature- and part-level priming following a longer delay

While the results of Experiment 1 point to the presence of a features-based representation as well as a parts-based structural description, it is possible that the difference seen in looks to the features-deleted target and complement was a function of the short ISI, and that the features-based representation is only a momentary phenomenon (although the results of Schendan & Kutas, 2007, strongly argue otherwise). Specifically, Ellis and Allport (1987) saw effects of their featural manipulation only at shorter ISIs, but not at longer durations. Therefore, in Experiment 2 the ISI between the prime and test array was increased from 500 to 1,400 ms, to examine whether a longer duration would alter the finding from Experiment 1 with regard to feature-level priming. The findings of Schendan and Kutas (2007) predict that the increased delay will have no

effect, as they saw the features-level effect after a much longer delay. This test is intended to either confirm this prediction or to demonstrate, if not, that Ellis and Allport's result applies at the level of features-deleted stimuli.

Method

Participants

Forty-five participants were recruited as in Experiment 1. Six participants' data were discarded due to calibration issues/ technical failures or for a participant's stated need to use eyeglasses. Three additional participants were excluded due to issues with data collection (eye makeup or other tracking issues), and two additional participants were excluded in preanalysis filtering (see Results), resulting in a final N = 34. This experiment was conducted under an IRB-approved protocol and conformed to the APA's ethical guidelines for the treatment of human subjects.

Apparatus

This was the same as in Experiment 1.

Procedure and design

Experiment 2 used the same procedure and design as in Experiment 1 except for a lengthened ISI (1,400 ms) between the end of the prime and the start of the test array.

Results

Data were processed as in Experiment 1. Trackloss analysis (elimination criterion of 25% maximum permissible loss with non-AOI looks included) for a trial resulted in removal of three trials and no participants. When excluding non-AOI looks, 113 trials exceeded 50% loss, with 32 of those from two participants. Those two participants were excluded from the analysis, and all other trials with over 50% track loss (n=82; < 1% of trials; non-AOI looks excluded) were also dropped, leaving 1,550 trials across all remaining participants (N=34). No included participant had more than six trials dropped as a result of this filtering, with the modal participant losing zero trials. Both dwell time and divergence analyses were conducted as described in Experiment 1.

Target-only trials

Dwell-time analysis:

As in Experiment 1, a two-factor within-subjects ANOVA was conducted using dwell time within each AOI as a percentage of overall looking time, with factors of stimulus condition (feature or part) and AOI (target, average distractor). Looking within the three distractors in these trials was averaged for purposes of analysis. The main effect of AOI was reliable; $F(1, 132) = 4074.60$, $p << .0001$, with a large effect size ($\eta^2 = 0.96$), as expected. The main effect of condition was not significant; $F<1$. The interaction was significant; $F(1,132) = 4.63$, $p < .04$, $\eta^2 < .01$. Examination of the data (see Fig. 7) suggests that this interaction is not informative; the effect size is quite small and concern regarding family-wise error leads to the conclusion that effects close to the set level of alpha (0.05) should be interpreted with caution in any case. Post hoc power analysis (R-WebPower; Zhang & Yuan, 2018) using dwell time data and the Cohen's f converted from the η^2 as the estimate of effect size indicated that power for the main effect was 1.0, but much less for the interaction (0.07). The same assumptions/approach as in Experiment 1 were used in this power analysis, and the same cautions apply; as before, the literature was used to inform the sample size.

Target-only trials

Divergence analysis (as in Experiment 1) comparing deletion type (features and parts) demonstrated that both parts-primed and features-primed targets effectively primed themselves, diverging from a randomly selected distractor AOI from 325–1,600 ms for parts (cluster statistic = 730.74, $p < .0001$) and slightly later, from 400–1,600 ms for features-deleted images (cluster statistic = 692.39, $p < .0001$). Features- and part-deleted primes showed similar-appearing curves (see Fig. 8), but a divergence analysis revealed a difference from 350–950 ms (cluster statistic = 81.20, $p < .0001$), reflecting, as in Experiment 1, that the features-deleted stimuli appear to have taken longer to achieve the same level of priming (as reflected by fixation proportion) as the parts-deleted stimuli.

Target + complement trials

Dwell-time analysis

As in Experiment 1, a two-factor within-subjects ANOVA was conducted over stimulus condition (feature or part) and AOI (target, complement, different exemplar, distractor). The main effect of AOI was reliable; $F(1,264) = 900.24$, $p << .0001$, $\eta^2 < 0.81$; as in Experiment 1, the target was expected to (and did) attract much more looking than the non-visually similar objects. The main effect of condition was not significant; $F<1$. This lack of an effect was mediated by a significant AOI × condition interaction; $F(1,264) = 120.19$, $p << .0001$, $\eta^2 < 0.10$. Visual inspection of the data (see Fig. 9) and follow-up contrasts revealed significantly more looking to the parts-deleted than to the features-deleted target AOI [$t(264) = 12.46$, $p < .0001$], with the reverse for the complement AOI [$t(264) = 14.33$, $p < .0001$]. Contrasts between target and
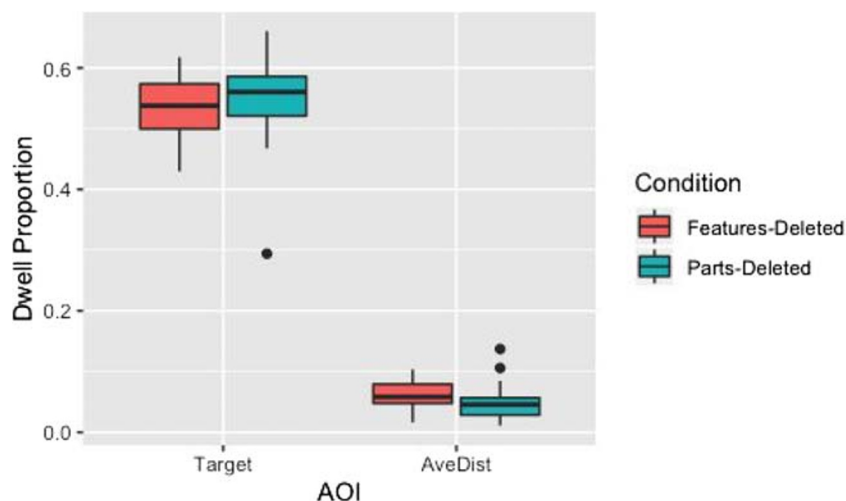
Fig. 7 Dwell time as a function of area of interest (AOI) and condition (parts- or features-deleted). Note that while the interaction is significant, the data do not suggest that the interaction is an important descriptor of the outcome

complement AOIs for the parts-deleted condition [t(264) = 33.95, p < .0001] and the features-deleted condition [t(264) = 7.16, p < .0001] were also reliable. Post hoc calculations indicated that power for the main effect was 1.0, as before, and for the interaction effect, 0.32.

Divergence analysis: Parts-deleted Following priming with a parts-deleted stimulus, looking proportion to the target AOI in comparison to the complement, different exemplar, and distractor AOIs was examined using a divergence analysis. These analyses showed a clear effect: In comparison to the complement, target fixations deviated upward from 350–1,600 ms (cluster sum statistic = 600.31; p < .0001), in comparison to the different exemplar, again, from 350–1,600 ms (cluster sum statistic = 706.62; p < .0001), and in comparison to the distractor, from 300–1,600 ms (cluster sum statistic = 954.36; p < .0001). Comparison of the complement and the

different exemplar AOIs revealed no reliable deviations. The different exemplar AOI deviated (above) from the distractor AOI from 400–1,300 ms (cluster sum statistic = 156.21, p < .0001), demonstrating the presence of non-visual (name) priming across most of the array exposure. These data are depicted in Fig. 10 (left panel).

Features-deleted

Features-deleted priming resulted in a different outcome (see Fig. 6, right panel). Looking proportion to the target AOI in comparison to the complement showed a significant effect late in the array exposure, between 1,250 and 1,600 ms (cluster sum statistic = 38.16, p = .004), echoing the finding of Experiment 1. Target looking deviated upward relative to the different exemplar AOI significantly from 425–1,600 ms (cluster sum statistic = 424.58, p < .0001). The target AOI
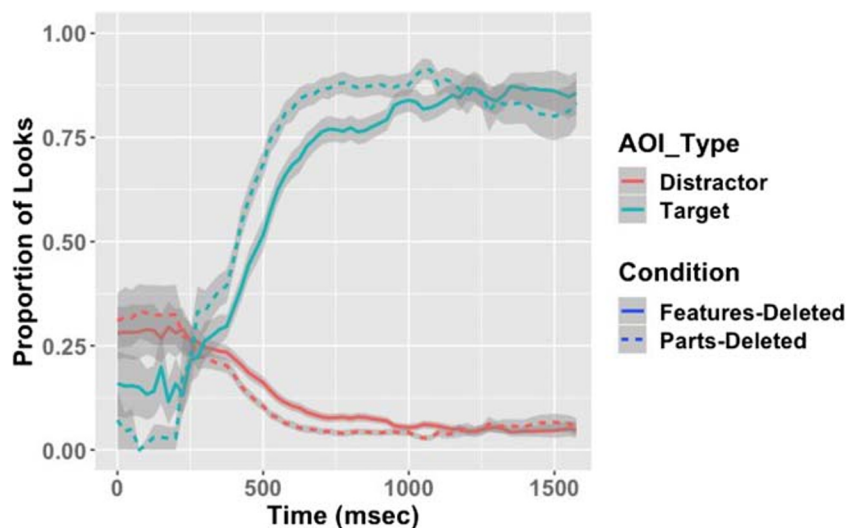

Fig. 8 Priming from the parts- and features-deleted objects to themselves in the absence of any related distractors
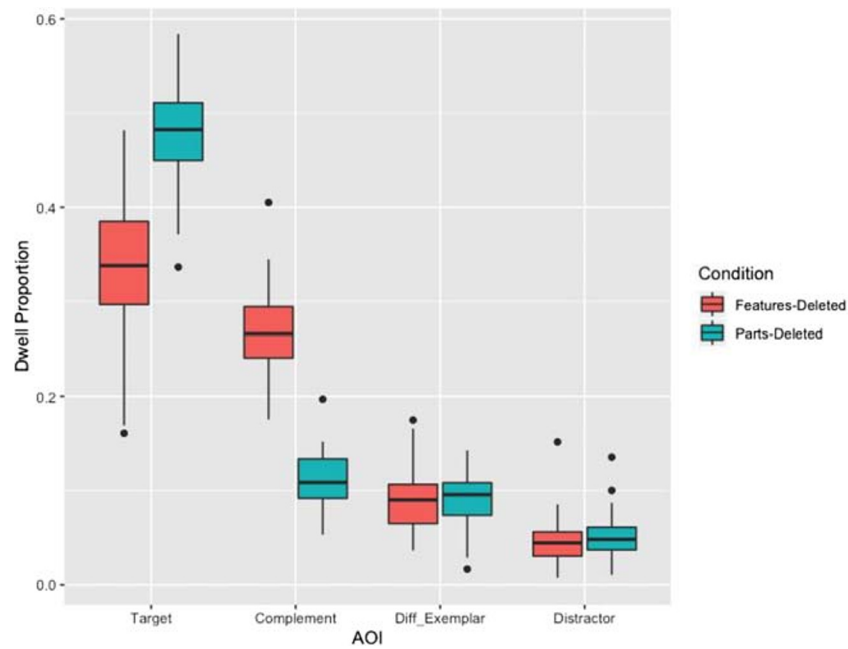
Fig. 9 Dwell time as a function of area of interest (AOI) and condition (features- or parts-deleted), showing that the difference in looking between target and complement is much larger for parts than for features

deviated significantly above distractor from 375–1,600 ms (cluster sum statistic = 634.76, $p < .0001$). The complement and different exemplar AOIs also deviated from 375–1,600 ms (cluster sum statistic = 372.01, $p < .0001$). The different exemplar AOI deviated (above) from the distractor AOI in two clusters from 450–1,600 ms with a small break at 1,150 ms (both clusters reliable; larger cluster sum statistic = 103.95, $p < .0001$), again demonstrating the presence of non-visual (name) priming across most of the array exposure.

Discussion

Overall, Experiment 2 replicated Experiment 1 in both the dwell time and the divergence analysis. Parts-deleted priming elicited the same parts-specific looking, and features-deleted primes again elicited an initial response indicating no differentiation, with looking diverging between the identity-match to the prime and its complement emerging later. This divergence, however, began later in Experiment 2 (1,250 ms) than
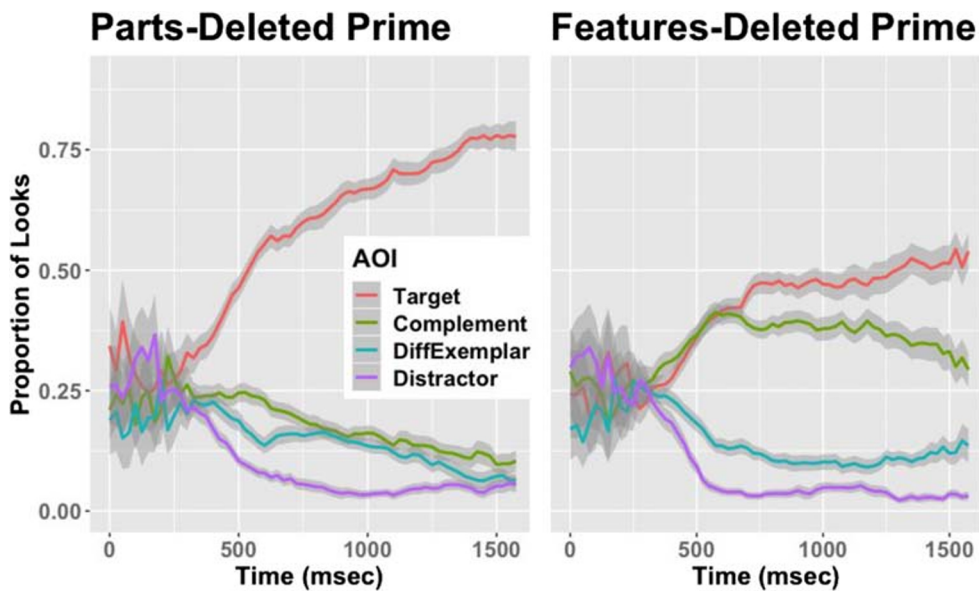

Fig. 10 Proportion of looks in Experiment 2 for parts-deleted (**left**) and features-deleted (**right**) primes, following a 1,400-ms interstimulus interval. Results largely mirror those of Experiment 1

in Experiment 1 (975 ms). The difference is likely attributable to the increase in ISI; possible reasons for this effect are reviewed in the *General discussion*. This result does not support the interpretation suggested by Ellis and Allport (1987) of a local or feature-level representation persisting over a short duration, but instead suggests that the system is constructing both a persistent global level and local level representation of objects from the incoming image, in agreement with multiple findings (Hummel, 2001; Schendan & Kutas, 2007). The choice of a test ISI for Experiment 2 intermediate between the two used by Ellis and Allport (500 ms and 2,000 ms) might have allowed some residual feature-level priming to persist. If so, a longer duration presented during the ISI might eliminate the priming of a feature-level representation, but the finding by Schendan and Kutas, with a much longer delay, clearly argues otherwise. It is clear from the present findings, however, that priming is fastest (and apparently strongest) at the level of parts in the context of shorter delays, replicating the original report of Biederman and Cooper (1991), which used a delay of minutes between the prime block and test block.

General discussion

The present investigation employed an adaptation in the VWP to investigate the time course, and nature, of our representation of objects. Participant eye movements were monitored as they first saw parts- or features-deleted primes, followed by an array of images presented after either a short (500 ms in Experiment 1) or longer (1,400 ms in Experiment 2) delay. Image arrays contained the identical primed image (target), its parts- or features-deleted complement, a different exemplar of the same category name, and an unrelated distractor. Results for the parts-deleted images support the findings of Biederman and Cooper (1991), and support an argument for the presence of a structural, parts-based representation (e.g., Biederman, 1987). Specifically, looks to the parts-deleted target indicated visual priming for that item above and beyond that for the different exemplar of the same category name, whereas looks to the parts-deleted complement did not show any priming above the non-visual priming seen for the different exemplar. As suggested by Biederman and Cooper (1991), this indicates that distinct representations were activated for the two items (target and complement), and therefore that the object representation may be, at least partially, at the level of the parts. This finding agrees with structure-based accounts, suggesting that there is an enduring parts-based, structural description-type representation.

Unlike the original test of Biederman and Cooper (1991), however, results for the features-deleted trials support the assertion that objects are represented on multiple levels (see Kveraga et al., 2007, for example). Specifically, while showing equivalent looks early on, looks to the features-deleted targets eventually won out over looks to the complement. Importantly, this features-based difference was present at both short and long ISIs, indicating that it was not simply a function of a momentary iconic image, as argued by Ellis and Allport (1987). In this way, the features-deleted results support the idea that there is an early-emerging global, or gist-based, representation, followed by a late-emerging features-based distinction. This result is directly in agreement with that of Schendan and Kutas (2007), who reported both an early and a late ERP signal. They argued that the early process was an indication of processing at the global level, while the late signal, by their interpretation, indicated the presence of a features-level process.

Does the divergence in features priming between the target and the complementary image seen later in the array exposure in the current tests represent a substantial deviation from the finding of Biederman and Cooper? Recall that Biederman and Cooper used a reaction time (naming) measure. Within the timeframe of such a response (generally about 600–800 ms), the eye-tracking procedure employed here showed no differentiation between the features-deleted target and the complementary image in either Experiment 1 or Experiment 2, while the parts-deleted target was the predominantly attended item in this timeframe, relative to its parts-deleted complement. This indicates that Biederman and Cooper's reported result would be predicted whether a late-emerging features-based representation existed or not; their test was not sufficiently sensitive to detect the second process. The parts-based prime, however, clearly accrued most or all of the visually based early priming in the present test of parts-based priming, just as in the original report. Thus, the present test – in large part – supports the original finding, in that parts-priming is clearly more impactful than features-level priming, while suggesting that, as reported by Schendan and Kutas, a features-level representation appears to be forming as well, although more slowly. The original report (Biederman & Cooper, 1991) found no support for a local-level (featural) representation, which the current findings (and others) suggest is too strong a claim. However, it is clear from the current results that parts-level priming is faster (and likely, given the results, stronger), and consequently is likely to control responding in many if not all typical tasks.

The nature of the representation underlying object recognition that remains once the retinal image is no longer available has long been a focus of theory and an issue of contention (Biederman, 1987; Edelman & Intrator, 2003; Marr, 1982), with a recent consensus appearing to emerge that several different types of representation (parts-based and features- or view-based) are likely preserved (see Hummel, 2013, for a review and argument for a combined representation). Global shape has also been suggested as a primary aspect of a persistent representation. Schendan and Kutas (2007) offer evidence

and discussion of an account positing representations of local features and global shape, giving support to a multiple-memory-systems argument; Goddard, Carlson, Dermody, and Woolgar (2016) discuss a somewhat similar set of findings involving low- and high-spatial frequency information. Recall that Bar and colleagues (Bar, 2003, 2004; Fenske, Aminoff, Gronau, & Bar, 2006; Kveraga, Boshyan, & Bar, 2007) have posited that the individual representational perspectives offer (alone) an incomplete picture of the overall process at best, and that both an initial "gist"-level (global or part-based) representation and a more detailed later representation are part of the overall process (see also Oliva & Torralba, 2006). Bar's account explains how the gist process could co-exist with and assist a more local-to-global process. Fintzi and Mahon (2014) have also argued for a global-first, followed-by-local account, using spatial frequency to define global and local levels. These multiple-representations accounts all point to a representation above the level of local features being of primary importance to the process of visual identification.

The multiple levels of representation account is also consistent with the literature on global-local processing mentioned above: Observers have been found to be biased in the direction of global-level features when challenged with hierarchical stimuli (Campana, Rebollo, Urai, Wyart, & Tallon-Baudry, 2016; Navon, 1977; Lamb & Yund, 1996). A similar bias is suggested by Gestalt principles (Koffka, 1935). Schendan and Kutas (2007) note that the timing of the two processes they identified argue for either an early global shape-based or a parts-based representation, and not a local features-based representation, in good agreement with the current findings as well as those of Biederman and Cooper (1991). Notably, the results from Shendan and Kutas and the current results offer something of an intermediate account, in that both processes may be expected to operate and produce evidence of both parts-based and features-based representations. Kimchi and Bloch (1998), examining grouping, presented findings supporting an argument that adding more global properties will affect processing more quickly than local properties, also suggesting that, even if two systems are operating, the parts-based representation will be activated first if there is sufficient priming to activate parts or other global-level factors.

The present outcome also agrees with findings from investigations of the visual pathway utilizing other methods, which also suggest that the primary process subserving object recognition is global or parts based. Kourtzi and Kanwisher (2001) report fMRI data showing that the lateral occipital complex (LOC) is tuned to parts, or global shape, not features (see also Bar, 2004; Grill-Spector, Kourtzi, & Kanwisher, 2001). Guggenmos et al. (2015) present a similar argument in support of a parts-based representation in the LOC, whether attended or not. Further, Kim, Biederman, Lescroart, and Hayworth

(2009) present evidence that the LOC is sensitive to changes specifically in shape rather than semantics (see also Kim & Biederman, 2012). Indeed, the LOC has been suggested as one location subserving primary object recognition (Kourtzi, Erb, Grodd, & Bülthoff, 2003). While there are clearly reports of representations tied to real world-aspects of a stimulus (e.g., real-world size; Konkle & Oliva, 2012; see also Kourtzi & Conner, 2011), overall, the prevailing argument is for a recognition system that concludes with an abstract representation removed from the immediate features, view, and other aspects of a particular depicted instance of an object.

In sum, the present investigation sought to extend the seminal research of Biederman and Cooper (1991) by using eye-tracking to investigate the time course of the priming of these two different types of deletion. Results support the suggestion that objects are represented at multiple levels. Specifically, the current outcome provides support for the contention that parts-based representations are present and primary in the visual system, and that they control responding in a typical behavioral response context. The present result also demonstrates, however, that a features-level representation is established, and the time-delimited nature of the current application of eye-tracking revealed that a features-level representation is established more slowly than the initial parts-based (and more global) representation. The features-level representation may be tied to more local-level uses, such as manipulation of objects in visual space or action preparation (Rizzolatti & Matelli, 2003; Schendan & Kutas, 2007), possibly through a different pathway (Almeida, Fintzi, & Mahon, 2013), and it may be the case that the early presence of a parts-based representation may aid in establishing a features-level representation, an idea also in line with the arguments of Bar and colleagues about a two-pass process (e.g., Kveraga et al., 2007).

The outcome of this pair of experiments is helpful, and the use of eye-tracking offers a complementary set of findings that support the ERP data offered by Schendan and Kutas (2007), with additional information about where observers are looking as an exposure progresses. These results are limited, however. Application of a technique such as fMRI to determine whether different areas within the visual pathway are activated by parts- as compared to features-level priming would be informative in terms of furthering understanding of the extent to which these two types of deleted stimuli activate different representations, as compared to activating the same representations at different levels. While the current approach offers informative time-delimited information, it is clearly a limitation of the approach that no data regarding the pathway location of these processes is available. Future research into this issue could apply the same types of stimuli using a method capable of providing this type of data.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439.

Almeida, J., Fintzi, A. R., & Mahon, B. Z. (2013). Tool manipulation knowledge is retrieved by way of the ventral visual object processing pathway. *Cortex, 4*, 2334–2344.

Altmann, G. T., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111(1),* 55–71.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15*, 600–609.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*, 617.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, 103*, 449–454.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*, 115.

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23*, 393–419.

Campana, F., Rebollo, I., Urai, A., Wyart, V., & Tallon-Baudry, C. (2016). Conscious vision proceeds from global to local content in goal-directed tasks and spontaneous vision. *Journal of Neuroscience, 36*, 5200–5213.

Dickinson, S. J., Pentland, A. P., & Rosenfeld, A. (1992a). 3-D shape recovery using distributed aspect matching. *IEEE Transactions on Pattern Analysis & Machine Intelligence (2)*, 174–198.

Dickinson, S. J., Pentland, A. P., & Rosenfeld, A. (1992b). From volumes to views: An approach to 3-D object recognition. *CVGIP: Image Understanding, 55*, 130–154.

Dink, J. W., & Ferguson, B. (2015). Eyetracking R: An R library for eye-tracking data analysis. Available at www.eyetracking-r.com

Edelman, S., & Intrator, N. (2003). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science, 27*, 73–109.

Edelman, S., & Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics, 64*, 209–219.

Ellis, R., & Allport, D. A. (1987). Multiple levels of representation for visual objects: A behavioural study. In *Artificial intelligence and its applications*(pp. 245–257). Wiley & Sons, Inc.

Fenske, M. J., Aminoff, E., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Progress in Brain Research, 155*, 3–21.

Fintzi, A. R., & Mahon, B. Z. (2014). A bimodal tuning curve for spatial frequency across left and right human orbital frontal cortex during object recognition. *Cerebral Cortex, 24*, 1311–1318.

Goddard, E., Carlson, T. A., Dermody, N., & Woolgar, A. (2016). Representational dynamics of object recognition: Feedforward and feedback information flows. *Neuroimage, 128*, 385–397.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*, 1409–1422.

Guggenmos, M., Thoma, V., Cichy, R. M., Haynes, J. D., Sterzer, P., & Richardson–Klavehn, A. (2015). Non–holistic coding of objects in lateral occipital complex with and without attention. *Neuroimage, 107*, 356–363.

Hayworth, K. J., & Biederman, I. (2006). Neural evidence for intermediate representations in object recognition. *Vision Research, 46*, 4024–4031.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96(1),* B23–B32.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*, 151–171.

Hummel, J. E. (2000). Where view–based theories break down: The role of structure in shape perception and object recognition. *Cognitive dynamics: Conceptual Change in Humans and Machines*, 157–185.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. Visual cognition, 8, 489–517. https://doi.org/10.1080/13506280143000214.

Hummel, J. E. (2013). Object recognition. *Oxford Handbook of Cognitive Psychology*, 32–46.

Kim, J. G., & Biederman, I. (2012). Greater sensitivity to nonaccidental than metric changes in the relations between simple shapes in the lateral occipital cortex. *NeuroImage, 63*, 1818–1826.

Kim, J. G., Biederman, I., Lescroart, M. D., & Hayworth, K. J. (2009). Adaptation to objects in the lateral occipital complex (LOC): shape or semantics? *Vision Research, 49*, 2297–2305.

Kimchi, R., & Bloch, B. (1998). Dominance of configural properties in visual form perception. *Psychonomic Bulletin & Review, 5*, 135–139.

Koffka, K. (1935). *Principles of Gestalt psychology.* Oxford, England: Harcourt, Brace.

Konkle, T., & Oliva, A. (2012). A real–world size organization of object responses in occipitotemporal cortex. *Neuron, 74*, 1114–1124.

Kourtzi, Z., & Conner, C. E. (2011). Neural representations for object perception: structure, category, and adaptive coding. *Annual Review of Neuroscience, 34*, 45–67.

Kourtzi, Z., Erb, M., Grodd, W., & Bülthoff, H. H. (2003). Representation of the perceived 3–D object shape in the human lateral occipital complex. *Cerebral Cortex, 13*, 911–920.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293* (5534), 1506–1509.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top–down facilitation in recognition. *Journal of Neuroscience, 27*(48), 13232–13240. https://doi.org/10.1523/JNEUROSCI.3481-07.2007.

Lamb, M. R., & Yund, E. W. (1996). Spatial frequency and attention: Effects of level-, target-, and locationrepetition on the processing of global and local forms. *Perception & Psychophysics, 58*, 363–373.

Lerner, Y., Hendler, T., & Malach, R. (2002). Object–completion effects in the human lateral occipital complex. *Cerebral Cortex, 12*, 163–177.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG–and MEG–data. *Journal of Neuroscience Methods, 164*, 177–190.

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. *MIT Press*.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353–383. https://doi.org/10.1016/0010-0285(77)90012-3.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*, 23–36.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three–dimensional objects. *Nature*, *343* (6255), 263–266.

R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology,12*, 162–168.

Rizzolatti, G., & Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Experimental Brain Research*, *153*, 146–157.

Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience*, *15*, 111–135.

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience,19*, 734–749.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, 104*, 6424–6429.

Stankiewicz, B. J., & Hummel, J.E. (1996). Categorical relations in shape perception. *Spatial Vision, 10*, 201–236.

Zhang, Y., Hedo, R., Rivera, A., Rull, R., Richardson, S., & Tu, X. M. (2019). Post hoc power analysis: is it an informative and meaningful analysis?. *General Psychiatry, 32(4)*.

Zhang, Z., & Yuan, K.H. (2018). Practical Statistical Power Analysis Using Webpower and R(Eds). Granger, IN: ISDSA Press.