

Tracking attention with the focus-window technique: The information filter must be calibrated

MICHAEL N. JONES and D. J. K. MEWHORT
Queen's University, Kingston, Ontario, Canada

Many researchers have recommended the focus-window tracking (FWT) technique to study the course of information acquisition and visual attention. In FWT, the stimulus is blurred, and the subject scans for information by moving a clear window with a computer mouse. We show that poor calibration of the blur filter will bias search behavior in FWT because diagnostically useful information may leak through a filter even though the subjects are unaware of it. To provide a valid index of information acquisition, FWT requires an information filter that is calibrated to the task.

Many authorities recommend a movable focus-window technique as a method of tracking visual information acquisition and attention (e.g., Jansen, Blackwell, & Marriott, 2003). With the focus-window tracking (FWT) technique, an information filter blurs the image except for a small movable section that remains unblurred. The clear section is called the *focus window*, and the subject moves it around on the stimulus image using the computer's mouse. To infer how the subject uses visual information to make judgments, the experimenter records the location of the focus window on the visual display. The assumption is that location of the focus window indicates what information is used to make the judgment.

FWT is related to eye tracking, but there are some important differences. Eye tracking traces subjects' exploration of the stimulus but, by itself, does not indicate where subjects extract particular information; to supplement simple eye tracking, experimenters have adopted the gaze-contingent technique and similar techniques (e.g., Reingold & Loschky, 2002). FWT does not perform an online tracing of subjects' visual exploration because they may move their eyes to inspect the image without shifting the window. However, because the filter blocks information from outside of the focus window, the technique indicates where the subjects seek useful information from the display and traces the location on the image from which the subjects use information to make judgments.

Figure 1 illustrates FWT in action. Panel A shows part of a face in clear view. Panel B shows the same image after a Gaussian blur filter has been applied to it. Panel C illustrates a focus window during search; in the example, the focus window uncovers a portion of the right eye of the face. Panel D shows recorded coordinates from the center of the focus window across a typical search trial. Presumably, the information that the subject has used can be inferred from the focus window coordinates. The record of successive coordinates can be used to infer the subject's acquisition strategy.

FWT has been advanced as a cheap and portable alternative to eye tracking without the inconveniences, such as calibration of the eye's position in the image space (Jansen et al., 2003). Our experiments using FWT suggest, however, that the technique is not without its own inconveniences. Considerable effort must be put into objective calibration of the information filter because filters that have been calibrated subjectively sometimes allow diagnostic information to leak through. Jansen et al. caution experimenters that "the blurred image . . . should reveal enough information to allow a subject to successfully navigate from one region of interest to another [but that] subjects should not be able to identify different syntactic elements without moving the focus window over them" (p. 59). We concur with their caution and will present a method of calibrating the filter.

A crucial feature of FWT is that the subject's eyes are free to wander independently of the recorded focus-window coordinates; indeed, this is how the subject navigates and selects areas of the image to scan with the focus window. Visual attention and recorded focus-window coordinates need not point to the same location. Hence, it is important to ensure that the subject cannot perform the experimental task using information outside the focus window.

A stimulus image (such as a face, a scene, or an equation) contains both navigation and diagnostic informa-

This research was supported by a grant to D.J.K.M. from the Natural Sciences and Engineering Research Council of Canada (NSERC) and by an AEG grant from Sun Microsystems of Canada. M.N.J. was supported by a postgraduate research scholarship from NSERC. The work was presented at the meeting of the Society for Computers in Psychology, Vancouver, BC, in November 2003; M.N.J.'s contribution received the John Castellan Student Paper Award for that year. We thank Randy Jamieson for comments on a draft of the paper. Correspondence concerning this article should be addressed to M. N. Jones, Department of Psychology, Queen's University, Kingston, ON, K7L 3N6 Canada (e-mail: mike@psyc.queensu.ca or mewhortd@post.queensu.ca).

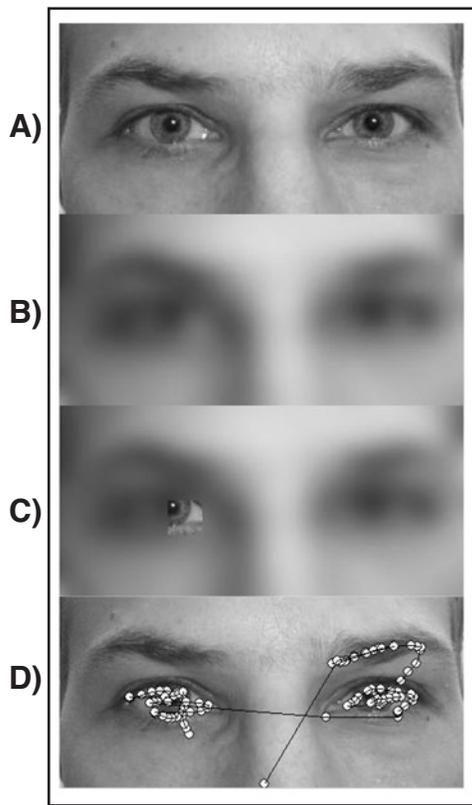


Figure 1. An example of the focus-window technique. (A) the original image; (B) the image with a Gaussian blur filter applied; (C) the filtered image with a focus window; (D) search coordinates recorded from the focus window across a trial.

tion. By diagnostic information, we mean detailed information needed to perform the experimental task. Navigation information, on the other hand, is general information about the layout of the stimulus and location of the diagnostic information. For example, consider a gender judgment task for our face from Figure 1. The filter in panel B blocks high-resolution diagnostic information needed to make that judgment but leaves the navigation information—the location of eyes, eyebrows, and nose. By definition, diagnostic information is task dependent.

An ideal information filter should pass a maximum amount of navigation information while completely blocking diagnostic information. If the filter blocks diagnostic information, the scan pattern provides valid data about what information has been used. If the filter leaks diagnostic information, however, the situation is less clear. On the one hand, the subjects may spend less time scanning the image but not alter their scan pattern. On the other hand, the subjects may alter their scan pattern because they have used leaked diagnostic information rather than information collected through the focus window. If the leak changes the scan pattern, the proportion of time in particular regions of the image would not accurately reflect information used because attention had been paid to a diagnostic region without scanning it with the focus win-

dow. Unfortunately, the experimenter often has no idea whether the leaky filter has produced an invalid measure of information used because the features that are truly diagnostic for the task are usually unknown; discovering them is often the purpose of the experiment.

In summary, a scan pattern cannot be interpreted without blocking diagnostic information (determined by the task) and passing navigation information. Hence, the information filter must be calibrated for diagnostic and navigation information in terms of the experimental task before scan patterns can be reliably interpreted. The following experiment illustrates the filter-specific nature of diagnostic information by showing an example of bias with a leaky filter.

To demonstrate such bias, we designed artificial stimuli for which we know a priori what regions of the image carry diagnostic information for different judgments. If the FWT data suggest that the diagnostic region of the stimulus was not used, but accuracy is high, we know that conclusions about focus of attention drawn from the FWT data must be incorrect.

METHOD

Subjects

Thirty-six undergraduate students from Queen's University were each paid \$10 to participate in the experiment. The subjects were between the ages of 18 and 26 (median = 20 years), and all reported normal or corrected-to-normal vision.

Stimuli

We adapted to our task the *Greebles* stimuli designed by Gauthier and Tarr (1997). Forty-eight Greebles were generated in total, with an even split between genders and among four families (Osmit, Radok, Samar, and Tasio). Greeble gender was determined by orientation of appendages—those pointing upward were males, and those pointing downward were females. Greeble family was determined by texture of the body midsection (skin, fur, feathers, or scales). Examples of Greebles from the four families are shown in Figure 2; the top two Greebles are females, and the bottom two are males. All Greebles were rendered in grayscale and presented on a dark background. Orientation of appendages on Greebles was correlated, but because the appendages differed in position and length, they may not have offered equally salient cues to gender.

For each image of a Greeble, edge-detection software was used to locate and classify each pixel from the body or appendage regions. Any pixel not classified as body or appendage was assigned to a catchall “other” category, including parts of the display extraneous to the Greeble. Hence, for each image we had an objectively defined record of pixel coordinates for all regions.

Procedure

Training. The subjects were told that the Greebles were an endangered nocturnal species living on the planet Xaelon, and that the subjects had been hired to help track Greeble migration patterns to better understand the species.

During the training phase, the subjects were shown a picture of a Greeble above a probe word labeling either a gender (e.g., female) or a family (e.g., Osmit). The subjects' task was to judge (yes/no) whether the Greeble displayed was labeled correctly, and they received feedback. If the response was correct, a message confirming their choice was shown: “Correct, it's an Osmit” (printed in green). If incorrect, they were told “Incorrect, it's a Samar” (printed in red).

Across training, the probability of a family or gender judgment was set to .5. The probability that the presented label was valid (cor-

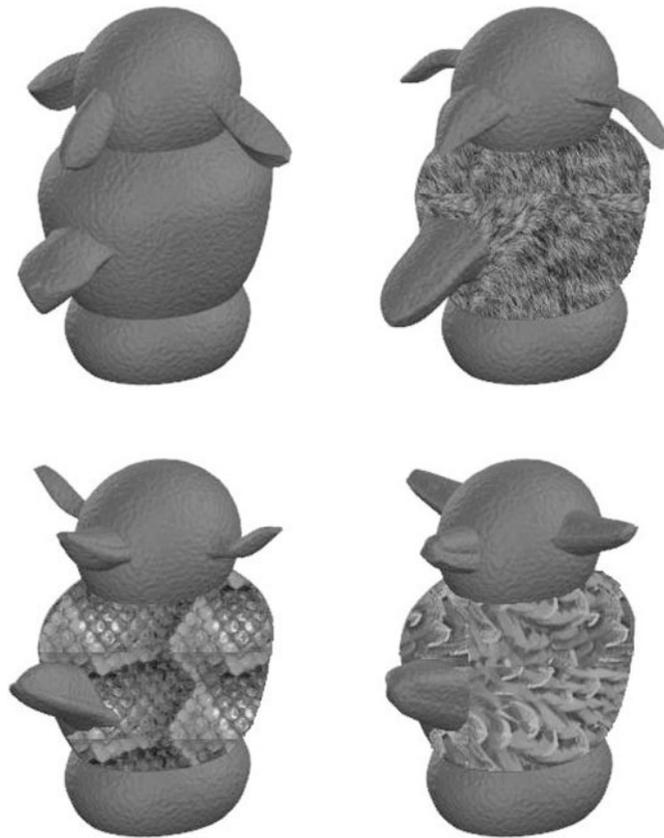


Figure 2. Examples of Greebles from the four families. Family was determined by texture of body midsection (skin, fur, feathers, or scales, clockwise from top left), and gender was determined by orientation of appendages (the top panels are both females, and the bottom panels are both males).

rect label for the presented Greeble) was also set to .5. The assignment of family names to textures was randomized independently for each subject.

Training was administered in blocks of 10 trials. The subjects continued training until they had performed perfectly on four consecutive blocks, whereupon they were declared to be “Greeble Experts.”

Test. When subjects were trained to criterion, the computer displayed a blurred target Greeble. A label was printed below the target and the subject’s job was to confirm (yes/no) the label. All 48 Greebles were presented once to each subject; presentation order was randomized for each subject. To justify the blurred image, it was explained that Greebles are nocturnal, and, thus, that the test must be run at night with reduced visibility.

Half the subjects performed the task using only the blurred images (control group), and the other half were shown how to use the focus window (described as a “night vision scope”), manipulated by the computer mouse, to help classify the blurry Greebles. The display occupied $11.25 \text{ cm} \times 10 \text{ cm}$ on the computer screen; and each Greeble was presented in the center of the display. At a viewing distance of 70 cm, each display (with the Greeble centered inside it) subtended a visual angle of 9.18° vertically and 8.18° horizontally. The focus window was 35×35 pixels in size (1 cm^2). Location of the focus window on the stimulus image was recorded every 100 msec.

Three subgroups within the window and control groups were tested under three information filter conditions (heavy, medium,

and light). The filters were generated using an RLE Gaussian blur algorithm; filter density was reduced by decreasing the radius of the blur (50-, 40-, and 30-pixel radii for heavy, medium, and light, respectively). As the radius was decreased, more information was allowed through the filter (i.e., 0-pixel radius is the original image). Examples of an image under each level of blur filter are shown in Figure 3.

RESULTS

Figure 4 illustrates typical scan patterns, one for a family judgment and one for a gender judgment. As is shown in the left panel, the subject concentrated on the body when making a family judgment. By contrast, the subject concentrated on the appendages when making a gender judgment (right panel).

Accuracy

Figure 5 (left panel) shows performance, for both types of judgments, as a function of blur filter density separately for window and control conditions. Subjects using the focus window performed near ceiling under all filter

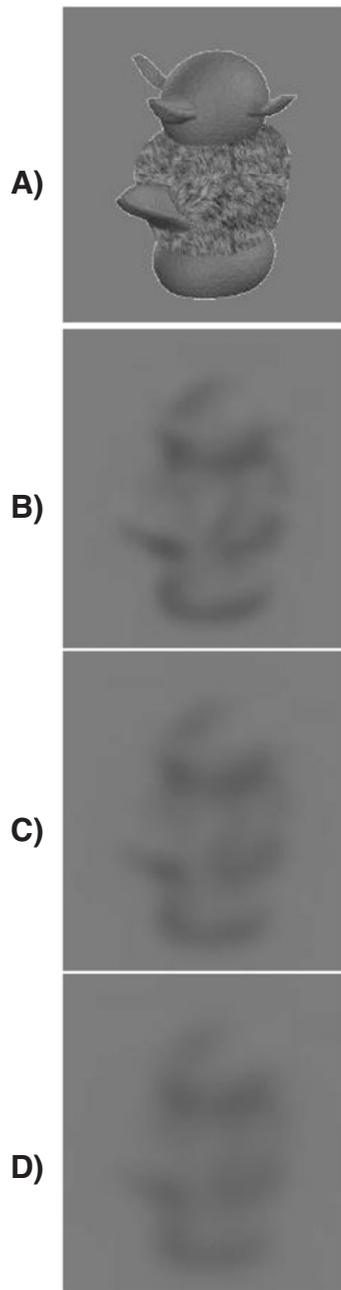


Figure 3. Example of one Greeble under (A) 0-pixel radius blur (the original image), (B) 30-pixel radius blur (light filter), (C) 40-pixel radius blur (medium filter), and (D) 50-pixel radius blur (heavy filter).

conditions. Subjects working without the focus window performed at chance in the heavy (50-pixel radius) condition for both family and gender judgments. For family judgments, accuracy remained at chance under all filters. For gender judgments, by contrast, accuracy increased linearly as the blur filter was lightened. Orthogonal contrasts confirmed the impression from the figure: There was a three-factor interaction of group (window/control) \times judgment (family/gender) \times filter density

[$F(1,60) = 10.18, p < .01$]. The pattern of results for the control group shows that diagnostic information progressively leaked through the filter as it was lightened for gender judgments, but not for family judgments.

Interestingly, many subjects in the no-window condition reported that they were guessing for all filter conditions and that no useful information was available to them. They performed significantly better than chance under the medium and light filters for gender judgments, which indicates, however, that diagnostic information leaked as the filter density was decreased.

Scan Time

Figure 5 (right panel) shows scan time for the window group as a function of filter density. As is clear in the figure, the time spent in scanning the stimuli decreased linearly as blur density was reduced [$F(1,15) = 10.70, p < .01$].

Although accuracy was at ceiling for the window group under all filters, the window subjects spent progressively less time collecting information through the focus window as the filter density was reduced. The pattern indicates that the subjects, to perform accurately, did not need to collect as much information using the focus window. In other words, the filter leaked diagnostic information about gender judgments (a point clear from the control group's accuracy), and the subjects responded by taking less time to scan the image while maintaining accuracy at ceiling.

The data presented so far leave it unclear whether or not the subjects in the window group altered their scan patterns. To address this question, we exploited the fact that diagnostic information was available from particular regions of the stimuli, and that we knew the regions a priori; family information could be obtained only from the body, whereas gender information could be obtained only from the appendages.

Scan Pattern

Figure 6 shows time spent scanning each region as a function of filter density for family and gender judgments separately. For family judgments (left panel), lightening the information filter increased the navigation information available to subjects, a point indicated by the decreased time associated with nondiagnostic areas of the image (i.e., appendages and other). These filters, however, did not leak texture information diagnostic to family judgments; the subjects spent the same amount of time collecting information from the body region across all filters. Orthogonal contrasts confirmed a linear interaction of time spent scanning the body region versus nondiagnostic regions (appendages and other combined) [$F(1,15) = 6.24, p < .05$].

The interaction of blur with body part for family judgments indicates that diagnostic information was collected through the focus window at all filter levels. The lightest filter (30-pixel radius) was ideal for family judgments because it maximized the amount of navigation information without allowing diagnostic information to leak through.

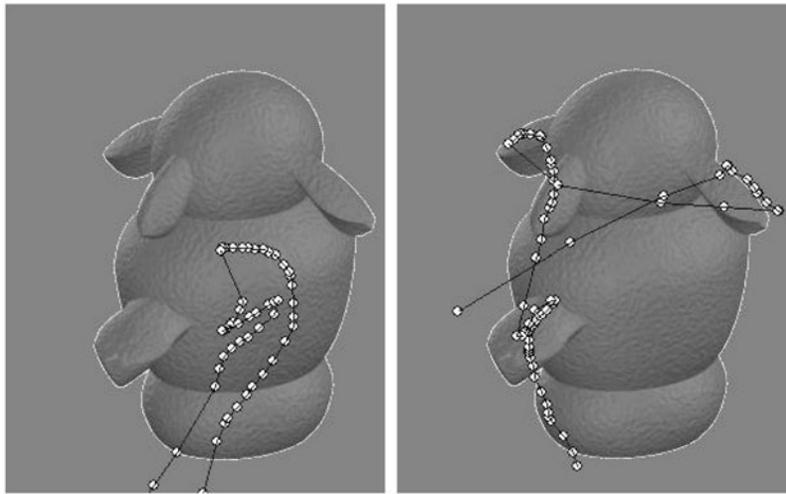


Figure 4. Typical scan patterns for a correct family judgment (left) and gender judgment (right) made on the same Greeble.

By contrast, Figure 6 (right panel) displays time spent scanning each region as a function of filter density for gender judgments. Scan time decreased linearly for all regions as the blur filter was lightened [$F(1,15) = 35.57$, $p < .001$], indicating that diagnostic shape information leaked through the filter as density decreased. The linear decrease in time scanning appendages was steeper than the corresponding decrease in the nondiagnostic regions (body and other combined) [$F(1,15) = 17.50$, $p < .01$].

For family judgments, diagnostic information did not leak through any of the filter settings, whereas the same filters allowed gender-diagnostic information to leak. For gender judgments, the time spent scanning under the heavy filter accurately reflected the allocation of visual attention (i.e., the appendages were diagnostic). As the filter was lightened, however, shape information leaked through, and the time spent scanning appendages dropped drastically, even though accuracy remained at ceiling (Figure 5, left panel).

DISCUSSION

The pattern of results illustrated in Figures 5 and 6 show some difficulty in using FWT. If we did not know that the diagnostic information for gender appeared only in the appendages, the pattern of scan times for the light filter would mislead us into concluding that the body region contains gender-diagnostic information and that the appendages do not. For example, the time spent scanning with the lightest filter in the gender task line up with the appendages receiving the least and the body receiving the most. This pattern suggests that the body contains information diagnostic to gender. That conclusion is obviously incorrect, however, because the stimuli contain gender-diagnostic information *only* in the appendages.

To provide valid data, FWT requires an information filter that is calibrated to the task. In our demonstration, the lighter filters did not pass information diagnostic to

family judgments but did allow information diagnostic to gender judgments to leak. The more the information that was leaked, the less the time that was needed to scan the stimulus while maintaining accuracy at ceiling. Moreover, for gender judgments, the subjects' scan patterns were altered because they used the leaked information.

Our demonstration shows that an information filter must be calibrated to a particular experimental task; a filter calibrated to one task (in our example, family judgments) may not block information diagnostic to a different task (in our example, gender judgments) even when the same stimuli are used. The calibration cannot be conducted subjectively; it requires a behavioral measurement with an appropriate control. Our control subjects believed that they were guessing under all filter conditions, but their accuracy data clearly show that they were not. As a consequence, we now define a calibrated information filter as one that does not allow better-than-chance performance without benefit of the focus window. We acknowledge, however, that for some stimuli, it may not be possible to calibrate an information filter to match our criterion. Blocking diagnostic information may, for example, also cripple navigation information.

The stimuli used in our demonstration were designed so that we knew which regions carried diagnostic information for each judgment, and the regions did not overlap. Often experimenters may not know a priori which stimulus regions are diagnostic. Indeed, in many cases, experimenters intend to use the FWT coordinates to reveal what information is diagnostic for the judgment at hand. As we have illustrated, however, without proper calibration it is all too easy to draw the wrong conclusion from the scan pattern about what information was used to perform the judgment. The fundamental problem is that useful information can leak through the filter.

Jansen et al. (2003) note that FWT and eye tracking are different techniques that address different questions. If the experimenter is interested in how a scene is scanned

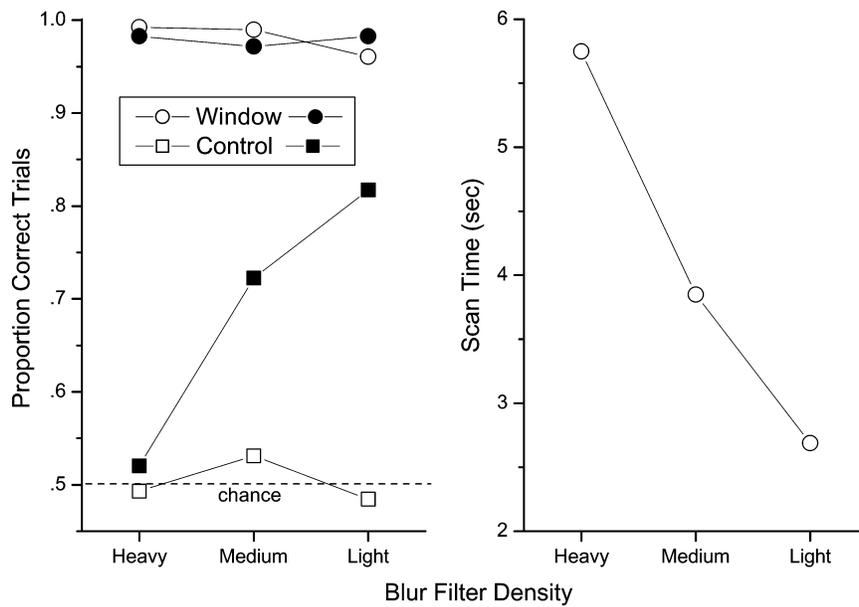


Figure 5. The left panel shows the proportion of correct trials for the window and control conditions as a function of filter density. The open symbols are family judgments, and the solid symbols are gender judgments. The right panel shows mean time spent scanning under the window condition as a function of filter density.

naturally, eye tracking with an unblurred periphery may be preferred because peripheral information may draw foveal attention. If the question is one of what information is used to make a judgment (and information acquisi-

tion is relevant), both eye tracking and FWT are appropriate. But a properly calibrated information filter must be used to ensure that unwanted information does not leak into the judgment. If the question is what visual informa-

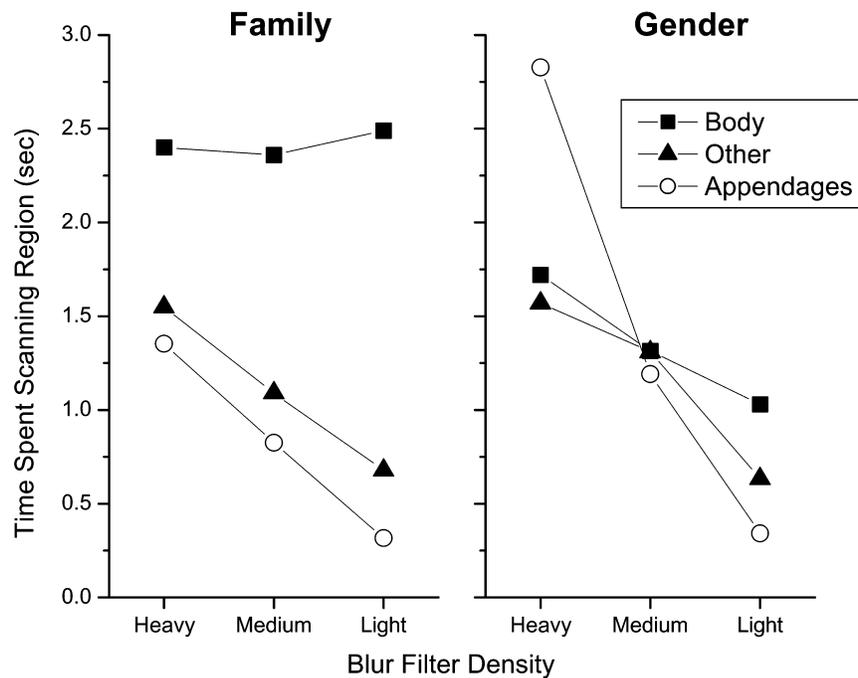


Figure 6. Time spent scanning each region as a function of filter density for family and gender judgments. The diagnostic region was *body* for family judgments, and *appendages* for gender judgments.

tion is diagnostic for a judgment (with no concern for search behavior), we recommend using a psychophysical technique designed for such a purpose (e.g., "Bubbles," Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002; see also Murray, Bennett, & Sekuler, 2002).

FWT has been recommended because it allows experimenters to avoid calibration of the eye's position required by eye-tracking techniques. Even so, it is clear from our results that FWT requires filter calibration to control information available to the subject from the display at large. A leaky filter may produce biased search data, leading to an incorrect conclusion about the use of visual information.

REFERENCES

- GAUTHIER, I., & TARR, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*, 1673-1682.
- GOSSELIN, F., & SCHYNS, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*, 2261-2271.
- JANSEN, A. R., BLACKWELL, A. F., & MARRIOTT, K. (2003). A tool for tracking visual attention: The restricted focus viewer. *Behavior Research Methods, Instruments, & Computers*, *35*, 57-69.
- MURRAY, R. F., BENNETT, P. J., & SEKULER, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision*, *2*, 79-104.
- REINGOLD, E. M., & LOSCHKY, L. C. (2002). Saliency of peripheral targets in gaze-contingent multiresolutional displays. *Behavior Research Methods, Instruments, & Computers*, *34*, 491-499.
- SCHYNS, P. G., BONNAR, L., & GOSSELIN, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*, 402-409.

(Manuscript received September 30, 2003;
revision accepted for publication March 22, 2004.)