A tool for tracking visual attention:
The Restricted Focus Viewer

ANTHONY R. JANSEN
Monash University, Victoria, Australia

ALAN F. BLACKWELL
University of Cambridge, Cambridge, England

and

KIM MARRIOTT
Monash University, Victoria, Australia

Eye-tracking equipment has proven useful in examining the cognitive processes people use when understanding and reasoning with visual stimuli. However, eye-tracking has several drawbacks: accurate eye-tracking equipment is expensive, it is often awkward for participants, it requires frequent recalibration, and the data can be difficult to interpret. We introduce an alternative tool: the *Restricted Focus Viewer* (RFV). This is a computer program that takes an image, blurs it, and displays it on a computer monitor, allowing the participant to see only a small region of the image in focus at any time. The region in focus can be moved using the computer mouse. The RFV records what the participant is focusing on at any point in time. It is cheap, nonintrusive, does not require calibration, and provides accurate data about which region is being focused on. We describe this tool and also provide experimental comparisons with eye-tracking. The RFV (Version 2.1) is freely available at http://www.csse.monash.edu.au/projects/RFV/.

Tracking visual attention is an important aspect of understanding how humans reason with visual stimuli. Complex visual representations can rarely be taken in at a single glance, and, thus, following the focus of visual attention can provide important insight into the strategies used in reasoning with such representations. For this reason, eye-tracking equipment has been of great benefit to researchers in examining the processes involved in comprehending visual stimuli, such as those used in diagrammatic reasoning (Hegarty, 1992), reading (Rayner, 1998), cartography (Steinke, 1987), scene perception (Rayner & Pollatsek, 1992), and cognitive processes in general (Just & Carpenter, 1976; Yarbus, 1967).

Yet despite the benefits that traditional eye-tracking provides, it also has significant drawbacks. First is expense: Typically, the better the resolution accuracy and sampling rate, the more expensive the system is, and high quality systems are very expensive indeed. Second, many systems are awkward for participants, using head-mounted gear or requiring chinrests or bite bars to suppress head movements. Third, most systems require frequent recalibration, and often they cannot be used with participants who wear glasses. Fourth, blinks or glances away from the stimulus can cause spurious trajectories in the eye-movement data. Finally, and perhaps most important, it can be impossible to determine whether a participant is taking in a broad overview of a stimulus or focusing on a specific region. It is not surprising then that in some papers that discuss eye-tracking data, the results of some participants can not be used due to an inability to accurately record their eye fixations (e.g., see Carpenter & Shah, 1998; Schnipke & Todd, 2000).

Here, we describe an alternative computer-based tool for tracking visual attention: the Restricted Focus Viewer (RFV). This allows visual attention directed toward an image presented on a computer monitor to be tracked. This tool is a cheap system, which is easy to set up and which can be flexibly tailored to the specific elements of interest in the stimulus. It is nonintrusive and requires no calibration. It is important to understand that we do not feel that the RFV is a replacement for eye-tracking. Rather, it is a complementary technique with its own advantages and disadvantages.

The RFV uses image blurring to restrict how much of the image can be clearly seen, with only a small region in focus.

The region of focus can be moved around using a computer mouse. The idea behind using restrictions in the visual field is not new. In research on reading, for example, the number of characters that can be processed in one fixation has been examined by using visual restrictions (Osaka & Oda, 1994; Rayner & Pollatsek, 1989). Image blurring has been used to understand how people take in information from software manuals (Ummelen, 1997). Studies have also been conducted that involve the use of artificial scotomas to disrupt visual processing (Henderson, McClure, Pierce, & Schrock, 1997), and the notion of a movable window has been used before in examining visual search (Stark et al., 1992). However, the difference in this work is that we combine these features in a generic configurable tool.

The main contribution of this paper is threefold. First, we describe our implementation of the RFV in detail. It is a generic computer-based tool, explicitly designed for research into how humans reason with and comprehend visual stimuli. Features include graded blurring, motion blur, and a data replay tool. Second, we provide a qualitative and quantitative analysis of the RFV. This includes an empirical validation of the RFV, in which results obtained from a mental animation experiment using eye-tracking equipment are compared with results obtained using the RFV instead. Our third contribution is a detailed discussion of the relative advantages and disadvantages of the RFV over eye-tracking.

RESTRICTED FOCUS VIEWER

The RFV is a computer program that takes a visual stimulus, blurs it, and displays it on a computer monitor, allowing the participant to see only a small region of the diagram in focus at any time. The region in focus can be moved using the computer mouse. The RFV records what the participant is focusing on at any point in time, and the data can be played back using a replayer.

The RFV and replayer programs have been written in Java, which is a platform-independent language. As such, they can be run on many different architectures and with many different operating systems, including PCs running either Windows or Linux. We now describe the RFV and the replayer in more detail.

Description of the RFV

The human visual system can only focus on objects at the center of the visual field. The region surrounding this area of sharp focus is still perceived, but the farther from the center of the visual field an object is, the more coarse is the perception of it (Tovée, 1996). Most of the time, visual attention is directed to the center of the visual field, although this is not always the case, since it is possible to covertly attend to other locations (Coren, Ward, & Enns, 1994).

The RFV has been designed to reflect these aspects of the human visual system. The key idea is that only the part of the image under the focus window is shown clearly, with the remainder of the image out of focus. To keep most of the visual stimulus out of focus, the RFV uses image blurring. Figure 1 gives an example of this with a pulley system diagram as the stimulus. The original pulley system is shown on the left, with a blurred representation of it on the right. The blurred image still allows the general form of the diagram to be perceived, thus allowing the user to move directly from one region of the image to another. However, it does not reveal the finer details of diagram, with the individual components being indiscernible unless the user of the RFV specifically focuses on them with the focus window. (Note that the blurred images in this paper have been made slightly darker so that they are clearer when printed. Some detail has been lost due to a reduced grayscale when printing.)

It is clear from the example figure that large structural features of the diagram are suggested in the blurred image. The degree of blurring required for a specific type of vi-
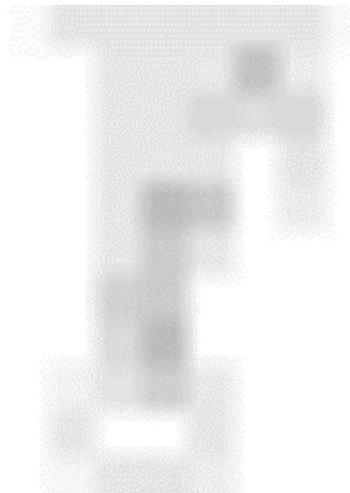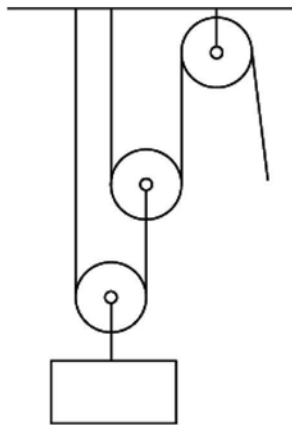

Figure 1. Example of a visual stimulus and its corresponding blurred image.

sual stimulus depends on the size and visual characteristics of the stimulus elements of interest. This is discussed in more detail later in this section.

The *focus window* of the RFV is the region in which the stimulus is visible in full detail. Two important issues concerning the focus window for the RFV are (1) how "natural" it will look and (2) how big it should be.

Initial experience with the RFV suggested that, with regard to the focus window, it is not sufficient to simply have a box on the screen in which the stimulus is in focus, while the rest of the image is blurred. The boundary between the two regions is too distinct, leading to a very unnatural effect. Also, a sharp cut-off will enable the participant to guess about neighboring parts of the image from Gestalt continuation effects. A graded blurring effect, such that the transition from blurred to focus appears smooth and seamless, is needed to prevent this.

A graded blurring effect is achieved by the technique illustrated in Figure 2. The outer rectangle defines the stimulus area that is fully blurred. The innermost box is the region of focus. Surrounding this focus region are three transition regions. Each transition region is slightly more blurred than the last, so that there is only a subtle difference between neighboring regions. The overall result is the appearance of a smooth transition from the region of the image in focus to the region that is fully blurred. Using the mouse to move the focus window therefore moves not only the focus region but also the transition regions.

Figure 3 gives two examples of the focus window positioned over different regions of the stimulus shown in Figure 1. The size of the focus window is determined not only by the dimensions of the focus region but also by those of the three transition regions. However, the two most important sets of values to be considered are the size of the focus region box and the size of outermost transition region box. For the experiments discussed in this paper, the focus region and transition region boxes were all square and were centered around the current mouse coordinates, although this need not be the case. At the center of the focus window, there was also a small gray dot, to allow the users to keep track of the focus window location when it was placed on an empty region of the image.

The degree of blurring and the size of the focus region need to be adjusted in accordance with the size of syntactic elements in the stimulus. In particular, there are two key factors that should be considered when adjusting these parameters. First, the blurred image of the stimulus should

reveal enough information to allow a participant to successfully navigate from one region of interest to another. If this is not possible, participants may spend more time getting their bearings than actually performing the task being investigated. However, an important second consideration is that participants should not be able to identify different syntactic elements without moving the focus window over them. If this is not the case, participants would not need to use the focus window at all, and, thus, it would not be possible to track where their attention is being directed. In our experience with the RFV to date, the guidelines described in Table 1 have proven useful in determining appropriate settings for the RFV parameters. Note that, although these parameters can be adjusted easily for a given experiment, the RFV tool does not allow them to be changed during an individual trial. This is to prevent inconsistencies in the way that the participant views the stimulus.

Another feature that was implemented so that the RFV would more accurately mimic the way humans perceive visual stimuli is *motion blur*. This is based on the fact that, during saccadic eye movements, visual information is not processed (Rayner & Pollatsek, 1989). Thus, if the user of the RFV moves the mouse at high speed (i.e., over a large distance on the screen in a small amount of time), the focus window will not achieve full focus. Once the user reduces the speed of the mouse motion back to below a certain threshold or stops moving the mouse completely, full focus in the focus window will return.

When the focus window is stationary or moving slowly, all of the regions listed in Figure 2 are present. During motion blur, however, only the outermost transition region is present. Because this region has less blurring than the rest of the image, the user is still able to track the location of the focus window on the stimulus. However, it is not possible to determine the finer details of that location without slowing or stopping the mouse. Only then will full focus be available.

The location of the focus window is recorded many times each second by the RFV, resulting in a large number of data points for each trial. The motion blur feature allows these numerous data points to be easily separated into groups that represent movements and fixations. A sequence of data points that are motion blurred (indicating fast movement of the mouse) represents a mouse movement, whereas a sequence of data points that are not motion blurred (indicating that the mouse is either stationary
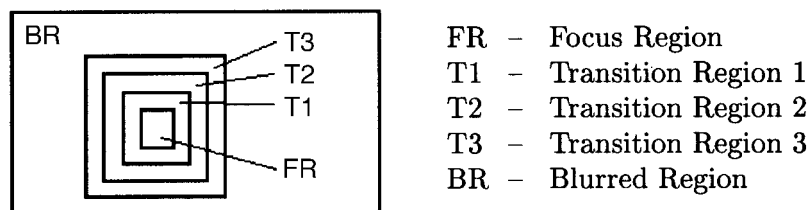

FR	–	Focus Region
T1	–	Transition Region 1
T2	–	Transition Region 2
T3	–	Transition Region 3
BR	–	Blurred Region

Figure 2. Regions of the stimulus used to achieve the graded blurring effect.
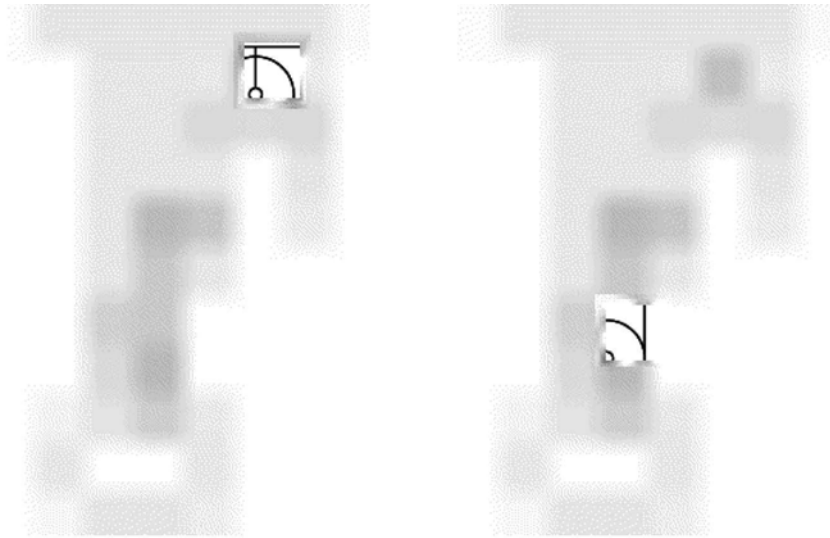
Figure 3. Two examples of the focus window on different regions of the stimulus.

or moving slowly) represents a fixation. The speed at which the onset of motion blur occurs can be controlled by the experimenter. Table 1 describes guidelines for appropriate motion blur settings.

The RFV outputs a data file that records the motion of the focus window. For each trial, information about the trial is specified, followed by lines that contain details about each updated mouse movement. The lines are composed of five data items. First is the stimulus number (more than one stimulus may be presented in a single trial). Second is the time elapsed (in milliseconds) since the RFV was initiated for that particular trial. The next two values are the x and y coordinates of the center of the focus window with respect to the top left corner of the stimulus

image. The final piece of information is a flag to indicate whether or not the focus window was motion blurred. These data allow the experimenter to exactly replicate the state of the RFV while the participant was performing the task. For a more complete discussion of user input and the output produced by the RFV, see Jansen (2001).

Data Replayer

The data replayer is a companion tool to the RFV that can read in a data file generated by the RFV and replay the way that the RFV participant moved the focus window over the stimulus. This has two important benefits.

First, it can be used as an experimental analysis tool by the experimenters when reviewing participant actions. For

Table 1
Guidelines for Setting RFV Parameters

RFV Parameter	Setting Guidelines
Focus Region Size	
Goal	Should be slightly smaller than the bounding box of a typical stimulus element
Lower limit	Must allow recognition of any one element of the stimulus when region is centered over the element
Upper limit	Should prevent simultaneous recognition of two neighboring elements when placed between them
Transition Region Size	
Goal	Should indicate the direction of neighboring elements
Level of Blurring	
Lower limit	Should be sufficient that any two elements are indistinguishable and that overall connectivity cannot be established
Upper limit	Should allow identification of stimulus boundaries (at least the overall shape)
Motion Blur Onset	
Goal	Should allow separation between fixation and movement
Lower threshold	Should not allow "brass rubbing" strategy—that is, identification of stimulus by waving window rapidly over it
Upper threshold	Should allow slow navigation with continuous focus over a connected stimulus when the task requires it

example, it can play back the focus trace at faster than real time. The data replayer also has another function that is useful to experimenters. It can draw a scan-path line over the original stimulus on the basis of the locations of the center of the focus box. Figure 4 gives an example of this for the pulley system diagram. It can be seen that, in this instance, the person started at the free end of the rope on the right of the diagram. From there, he/she moved to the top pulley, continued to the middle pulley, and, finally, moved down to the bottom pulley. At each pulley, time was spent examining that region of the diagram.

The second important benefit is that the data replayer can be used during experimental sessions to elicit retrospective verbal reports from participants about their strategy and actions during an experimental trial. As noted by Ericsson and Simon (1993), it can be very difficult for participants to verbalize concurrently while carrying out a complex problem-solving task. We have used the data replayer to play back focus movements slower than real time, so that the participants can record a verbal protocol describing their actions after completing the main task. Ericsson and Simon report that it is difficult to report retrospectively on a problem taking longer than 10 sec to solve, whereas the use of the data replayer with the RFV reminds the participants of their actions and allows more extensive protocols than can usually be obtained.

EXPERIMENT 1
A Qualitative Analysis

In Experiment 1, we examined the performance of the RFV in contrast to eye-tracking in a qualitative manner by using both techniques to explore how participants scan
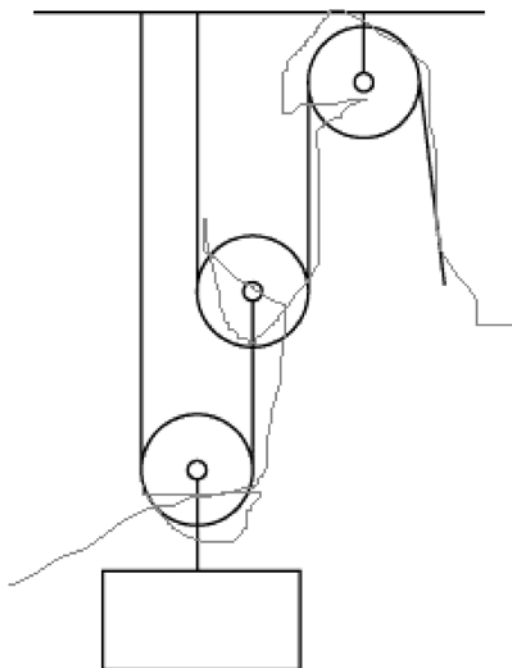

Figure 4. Example of a data replayer scan path.

over simple algebraic expressions. Algebraic expressions were chosen because they are composed of similarly sized symbols whose layout is two dimensional in nature. The aim of this experiment was to compare how well the RFV and eye-tracking equipment can indicate where in a visual stimulus attention is being directed and what problems can arise when obtaining these data. Despite the fact that we were not trying to create a replacement for eye-tracking, it was still logical to compare the performance of the RFV with eye-tracking equipment since eye-tracking is a dominant method used for tracking visual attention. The experiment was designed such that a single group of participants used both techniques, thus reducing any possible influences that could arise from individual differences. Note that we were not actually interested in how participants scan over the algebraic expressions; rather, we were interested in the relative performance of the RFV and eye-tracking equipment.

Method
Participants. Eight participants successfully completed the experiment. All were graduate or undergraduate students from the Psychology Department at the University of California, Santa Barbara. All participants were volunteers with normal or corrected-to-normal vision, and all were familiar with algebra. Data from 1 participant were excluded due to an excessive error rate.

Materials and Design. Twenty-four algebraic expressions were constructed, all consisting of exactly seven characters. The expressions contained at most one fraction, and the variable names were x and y, since these are most commonly used. For each expression, a statement about the expression was constructed that was either true or false. Half of the expressions had true statements, and the other half had false statements. The statements varied in how difficult it was to judge their correctness. Some were simple, requiring the participant to determine only whether a particular symbol was present in the expression. Others were more difficult, requiring the expression to be used in a calculation. Table 2 gives examples of the expressions that were created and their corresponding statements.

In order to monitor how the participants viewed each expression using both the RFV and eye-tracking equipment, while ensuring that the participants were presented with each expression only once to avoid practice effects, two counterbalanced versions of the experiment were constructed. For each version, there were 12 expressions assigned to the RFV and 12 expressions assigned to the eye-tracking equipment. Two additional expressions and statements were constructed for each group of 12 expressions to be used as practice items. The items for each group of expressions (RFV and eye-tracking) were presented in a different pseudorandom order for each participant.

Procedure. The participants were initially presented with written and verbal instructions, giving examples of the sort of expressions and statements involved in the experiment. Half of the participants viewed the eye-tracking expressions before the RFV expressions, and the other half performed the experiment in the reverse order. For both the RFV and eye-tracking components of the experiment, items were displayed in black on a white background on a 20-in. monitor.

For the RFV component, the monitor was running at a resolution of 1,024 × 768, with the presentation of items controlled by an IBM-compatible computer running a version of the RFV tool whose settings had been tailored to this experiment. The participants were seated comfortably at a viewing distance from the monitor of approximately 50 cm. The average width of the algebraic expressions in pixels was 92 (range 61–120), with an average height of 38 (range 24–58). The RFV focus box had an edge length of 16 pixels, and the outermost transition box had an edge length of 30 pixels. This al-

lowed one symbol to be viewed in focus at a time. The motion blur was set to a low tolerance, requiring the mouse to be stopped or to be moving very slowly in order for full focus to be maintained.

The only interface mechanism used by the participants was a standard computer mouse. Progress was self-paced, with each trial initiated by a single click with the mouse on a button at the bottom of a blank screen containing the prompt "Click button to continue." This action started the timer (to provide a reference time for the rest of that trial), and a blurred image of the expression appeared in the center of the screen. The participant could move the window of focus over different symbols in the expression by moving the mouse, allowing the entire expression to be read. The participants were initially given two practice items (allowing them to become familiar with the RFV tool), followed by the 12 experimental items.

Once the participants had read the expression, they clicked a button at the bottom of the screen using the mouse. This stopped the timer and made the blurred expression disappear, replaced by the statement about the expression (which was not blurred). The participants were required to determine whether the statement was true or false with respect to the expression that immediately preceded it. At the bottom of the screen were two buttons, one labeled TRUE and the other labeled FALSE. When the participant had decided on the validity of the statement, he/she single clicked on the appropriate button. The RFV tool recorded the response given and the time taken. No feedback was given to the participant. The participants were instructed to try to read the algebraic expressions as quickly as possible, while still ensuring that they did not to make too many errors when determining the validity of the statements.

The eye-tracking component of the experiment was conducted using an Iscan corneal-reflectance and pupil-center eye-tracker (Model RK-426). It has a resolution of less than 1° of visual angle and samples the participants' gaze every 16 msec. The participants were seated approximately 1 m from the monitor, and a headrest was fitted comfortably to restrict head movements. The participants were asked to move as little as possible during the experiment. In order to better determine which part of each expression was being fixated on at any given time, the size of the algebraic expressions for the eye-tracking component of the experiment were larger, with an average width in pixels of 368 (range 246–478) and an average height of 149 (range 92–228). The screen resolution was also lowered to 800 × 600. As a result, each symbol subtended a larger visual angle than for the RFV component of the experiment.

After the eye-tracking equipment had been calibrated, the participants were asked to fixate on an asterisk that appeared in the center of the screen and to push a button to begin each trial. The algebraic expression then appeared in the center of the screen, and the participants read the expression. Once the expression was read the button was again pushed, and the expression was replaced with a statement about the expression. After determining whether the statement was true or false, the participants responded verbally, and the experimenter recorded their answers. As with the RFV component, two practice items preceded the 12 experimental items.

The experiment took approximately 25 min to complete.

Data treatment. Only 1 participant failed to get more than 75% correct for the statements and, thus, was excluded from further analysis. While error rates were considered, the time taken to determine whether a statement was correct was not a primary interest and, therefore, was not analyzed. The main focus instead was on how the algebraic expressions were scanned by the participants.

Results and Discussion

The primary purpose of the statements was to force the participants to carefully read the expressions. There was no significant difference between the error rates for using the RFV (10.5%, $SD = 7.4$) and those for using the eye-tracking equipment (10.4%, $SD = 5.9$). However, the participants took approximately 22% longer to view each expression when using the RFV (7,814 msec, $SD = 2,856$) than when using the eye-tracking equipment (6,401 msec, $SD = 2,229$). This difference was analyzed using a two-way analysis of variance (ANOVA; versions × interface type) and was found to be statistically significant [$F(1,7) = 14.84, p < .01$]. The slower performance was likely to be a result in part of the fact that the participants interfaced with the stimulus using a computer mouse. This was due to the difference in the ballistic speed of human motor control of the arm and the eye. Also, the blurring of the stimulus with only a small region of focus means that one would expect the participants to take longer in moving from one area of the expression to another area. Since the same participants completed both the eye-tracking and RFV components of the experiment, this difference was not due to individual differences.

The main aim of this experiment was to compare how well both the RFV and eye-tracking equipment perform in tracking visual attention. Thus, in this experiment, we did not analyze fixation or gaze durations but, instead, focused on the locations in the stimulus where visual attention was directed.

The process of calibrating the eye-tracking equipment proved to be tedious for many of the participants, and it also became uncomfortable for them to sit still with their head restrained for the entire eye-tracking component of the experiment. The calibration procedure varied in difficulty between participants, and how well the calibration was maintained across the experimental items also varied between participants. In general, the first few items seen by a participant produced good scan-path data. This is not

Table 2
Examples of the Expressions Used in Experiment 1,
Along With Their Corresponding Statements
and Whether Those Statements Were Correct

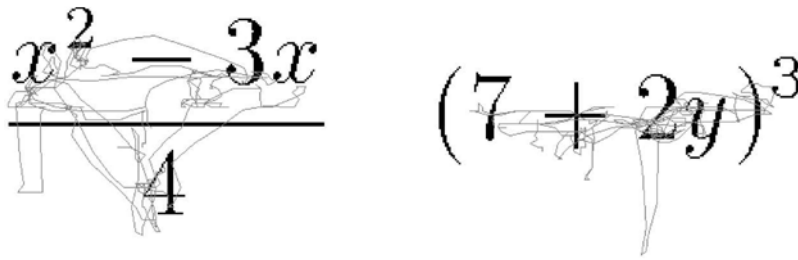| Expression	Statement	Correct
$(7 + 2y)^3$	The digit 4 appears in the expression	False
$\dfrac{x^2 - 3x}{4}$	When $x = 3$, the expression solves to equal 0	True
$4x^2 - \dfrac{x}{2}$	The expression contains the subexpression $4y^3$	False

Figure 5. Examples of eye-tracking scan-path data that show good calibration of the eye-tracking equipment.

surprising given that calibration of the equipment occurred immediately before the first items were presented. Examples of these good scan paths are shown in Figure 5. Note that even these good scan paths would need some minor adjustments before further analysis. For example, much of the scanning around the $3x$ in the left expression appeared lower than it should have.

Even when good scan-path data are obtained, eye-tracking data can still have drawbacks. The main problem is that, at any point in the data, it is impossible to determine whether the participant was focusing on a single symbol or a collection of symbols. For example, in the right expression in Figure 5, it is not clear whether the subexpression $2y$ was being focused on in a single fixation or whether the two component symbols were examined in separate fixations. (The amount of activity around that part of the expression suggests that perhaps both occurred at different times during the trial, although this only serves to further complicate data analysis.) Also, the exponent of the bracketed term in the right expression (the digit 3) was hardly scanned directly at all, yet it is an important part of the expression. It is therefore not clear from this example how much processing this symbol is receiving, since it appears that it was always being viewed in conjunction with neighboring symbols. The eye-tracking data also appear "noisy," because the eyes are constantly in motion, and the paths between fixations are not always direct.

A major problem with eye-tacking equipment is that calibration never lasts. Even though the number of expressions in each version of this experiment was quite low,

the calibration of all participants noticeably deteriorated during the task. Also, the data from several participants demonstrated that, apart from gradual deterioration, the calibration can be suddenly lost at almost any moment (sometimes, the reason is clear, such as if the participant moves too much; however, often, the reason is not clear). Examples of eye-tracking data when the calibration has effectively been lost are shown in Figure 6. These scan paths are no longer useful as data and must be discarded. Possible solutions to this problem include frequent recalibration, but, even then, there is the risk that the participant may never properly settle into the task due to the interruptions. Also, frequent recalibration really only represents a partial solution, because calibration loss can happen suddenly during any given trial.

The scan paths in Figure 5 illustrate eye-tracking data produced with good calibration, and those in Figure 6 clearly illustrate data resulting from bad calibration. However, these examples only represent extremes from a range that contains many levels. Unfortunately, this in itself presents a problem: Since calibration can vary in accuracy, when is eye-tracking data considered to be too corrupted by poor calibration to be useful? In extreme cases such as these examples, this is easy to decide. But experimenters will often be forced to make arbitrary judgments about how acceptable certain eye-tracking data are—a situation that is far from ideal. When making such arbitrary decisions, experimenters also run the risk of misinterpreting unusual scan-path data as the result of poor calibration, when in fact the data may be the result of a novel strategy used by the participant.
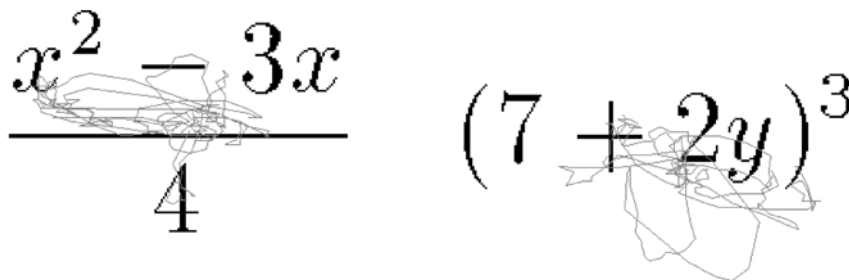

Figure 6. Examples of eye-tracking scan-path data that show bad calibration of the eye-tracking equipment.
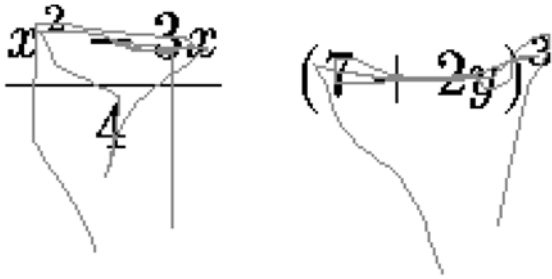
Figure 7. Examples of RFV scan-path data on the same algebraic expressions shown in Figures 5 and 6.

In contrast, the RFV does not have any problems relating to calibration. No initial calibration is required, and, by the nature of the RFV's design, it cannot get out of alignment. Figure 7 gives examples of scan paths obtained with the RFV on the same expressions shown in Figures 5 and 6. These scan paths are much more directed, since they result from conscious movements of a computer mouse. Also, since the size of the focus window corresponds to the size of a single symbol, it is always easy to determine exactly what is being focused on. A further benefit of the RFV is that the stimuli can be presented in a normal size, whereas the stimuli for the eye-tracking component had to be made unnaturally large in order to obtain accurate scan paths.

This experiment illustrated that, at least for stimuli such as algebraic expressions, the RFV has several advantages over eye-tracking equipment. The RFV has none of the problems associated with calibration, and robust data are obtainable even with participants who are inexperienced in using the RFV. The problems we encountered with eye-tracking equipment are also consistent with the findings of other researchers who, as discussed in the introduction, also had difficulty in accurately recording eye fixations. These results included stimuli ranging from graphs (Carpenter & Shah, 1998) to interactive computer software (Schnipke & Todd, 2000). Nonetheless, for certain stimuli (e.g., scene perception in which scenes contain many differently sized objects), we feel the use of eye-tracking equipment may be more appropriate.

EXPERIMENT 2
A Quantitative Analysis

The results of Experiment 1 indicate that the RFV has several advantages over eye-tracking equipment. However, a key question before we can consider using the RFV in experimental work is to determine whether it interferes with the strategies used by humans when comprehending a visual stimulus. Clearly, there is some overhead, since the participants must use a computer mouse rather than simple eye and head movements to change their focus. However, it is important to the validity of the RFV technique that this should not affect the strategies used. In Experiment 2, we tested this in a quantitative manner by replicating a classic experiment on diagram interpretation (Hegarty, 1992), originally done with eye-tracking equip-

ment, but instead using the RFV. Our hypothesis was that the participants would use the same strategy with the RFV as with the eye-tracking equipment.

Method

Participants. Eleven participants successfully completed the experiment. All were graduate or undergraduate students from the Computer Science and Psychology Departments at Monash University, Victoria, Australia. All participants were volunteers with normal or corrected-to-normal vision. Data from 1 additional participant were excluded due to excessive error rates; however, the errors were due to a misinterpretation of the instructions and were not attributable to the RFV.

Materials and Design. Two diagrams of pulley systems were constructed that were identical to those used in Experiment 1 in Hegarty (1992; see Figure 8). They each consisted of three pulleys, a weight, braces from the ceiling to support some pulleys, and sections of rope that were attached to the ceiling or weight and went over or under the pulleys. In each system, there was also a free end to the rope, and the participants were required to infer the motion of the system components when the free end of the rope was pulled. The mirror images of these pulley systems were also used, giving a total of four pulley system images.

For each pulley image, there were 12 statements, each about the motion of one of the system components. Six of these statements were true, and 6 were false. For this experiment, only kinematic statements were used (referring to the system in motion), with none of the statements addressing static aspects of the pulley system.

When the free end of the rope is pulled in any pulley system, a causal chain of inferences can be made about the motion of the components. For example, in Pulley System 1, pulling the rope causes the rope to move right over the upper pulley, turning it clockwise. From this knowledge, we can infer the motion of the middle pulley, and so on. In this way, we can define each pulley in the pulley system as being at the beginning, middle, or end of the causal chain of events. The statements about the motion of the pulley system components are equally divided among the pulleys at each of these three locations in the causal chain. In Pulley System 2, "The rope moves to the right under the lower pulley" is an example of a true statement about a kinematic event at the middle of the causal chain, and "The upper right pulley turns counterclockwise" is an example of a false statement about a kinematic event at the beginning of the causal chain.

Each statement was presented as a single line of text. A stimulus was composed of a statement appearing on the left, with the diagram
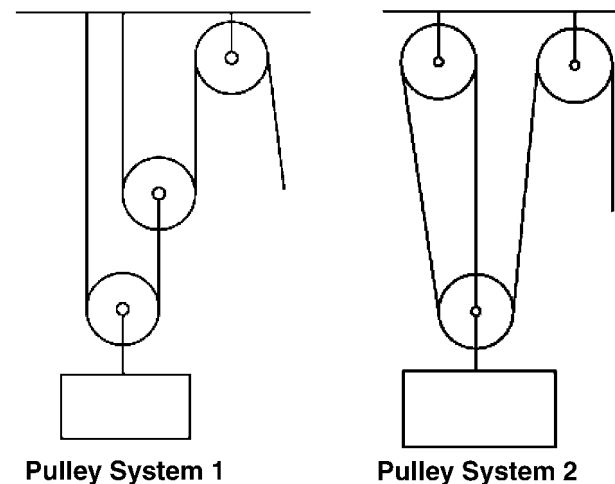

Pulley System 1   Pulley System 2

Figure 8. The two pulley systems used.

of a pulley system on the right. This gave a total of 48 stimuli. Each participant was shown all 48 stimuli twice, with a rest between the two blocks. In each block, the stimuli were presented in a different pseudorandom order.

Procedure. The participants were seated comfortably in an isolated booth. Items were displayed in black on a white background on a 17-in. monitor at a resolution of 1,024 × 768, controlled by an IBM-compatible computer running a version of the RFV tool whose settings had been tailored to this experiment. The original eye-tracking experiment was conducted using an Iscan corneal-reflectance and pupil-center eye-tracker (Model RK-426) that had a resolution of less than 1° of visual angle and sampled the participants' gaze every 16 msec.

The size of the images were 200 × 293 pixels for Pulley System 1 and 160 × 300 pixels for Pulley System 2. The text statements were on average 338 pixels across (range 244-417) and 16 pixels high. The RFV focus box had an edge length of 36 pixels, and the outermost transition box had an edge length of 50 pixels. The motion blur was set to a high tolerance, allowing for full focus to be maintained even during moderately fast movements of the mouse.

The participants were given a brief statement of instructions before the experiment began and were shown diagrams that labeled all of the pulley components referred to in the statements. They were then given some practice items involving a very simple pulley system (only two pulleys). The practice items allowed the participants to become familiar with the RFV tool.

The only interface mechanism used by the participants was a standard computer mouse. Progress was self-paced, with each trial initiated by a single click with the mouse on a button at the bottom of a blank screen containing the prompt, "Press the button to continue." This action started the timer (to provide a reference time for the rest of that trial), and a blurred image of the stimulus appeared on the screen. On the left was a statement; on the right was a diagram of a pulley system. The participant could move the window of focus over different regions of the stimulus by moving the mouse. By doing this, the participant could then read the text of the statement and look at the attributes of the diagram.

The participants were required to determine whether the statement was true or false with respect to the pulley system presented. At the bottom of the screen were two buttons: one labeled TRUE and the other labeled FALSE. When the participant had decided on the validity of the statement, he or she single clicked on the appropriate button. This stopped the timer, and the RFV tool recorded the response given and the time taken. No feedback was given to the participant.

The participants were instructed to try to respond as quickly as possible while still trying not to make too many errors. The experiment consisted of two blocks of 48 trials, with a brief rest period in between. The full set of stimuli was shown in each block, resulting in two repetitions for each item in the experiment. The experiment took approximately 40 min to complete.

Data treatment. To reduce the unwanted effects of outlying data points, an absolute upper cut-off was applied to response latencies, such that responses longer than 30 sec were excluded from the response time data analysis and were designated as errors. For consistency with the original eye-tracking experiment, the data analysis was conducted only over the items that contained true statements, with the false items acting only as fillers in the experiment.

One participant's results were excluded, due to getting no correct responses at a particular causal chain location in one of the pulley systems and thus not allowing the calculation of a mean response time at that position.

Results and Discussion

The aim of this experiment was to compare the results obtained using the RFV with eye-tracking results, to determine whether the RFV affects the strategies used by humans when comprehending visual stimuli in a manner different from the way that eye-tracking equipment might affect strategy. The main focus of this analysis was therefore to examine key data trends and significant results obtained in the original eye-tracking experiment and to see whether the RFV results concurred.

Errors. The overall error rate was only 5.3%. This was much lower than the error rate of the participants in the original eye-tracking experiment. However, this is not surprising given that the participants in the original experiment were all psychology undergraduates, and many of the participants in this experiment had more experience in dealing with technical diagrams.

The data comparisons were conducted using two-way ANOVAs (causal chain position × repetition), carried out over participant data. In the original eye-tracking experiment, the position in the causal chain of the component referred to in the statement had a significant effect on error rates. This effect was also present in the RFV data [$F(2,20) = 5.33, p < .05$]. This can be clearly seen in Figure 9, where the data from this experiment on the right is compared with the data from the original eye-tracking experiment on the left. There was also a trend for the participants to make fewer errors on the second repetition of the stimuli (3.0%, $SD = 5.4$) than on the first repetition (7.6%, $SD = 12.5$); however, this trend was not statistically significant [$F(1,10) = 4.52, p = .059$]. This trend was also apparent in the eye-tracking experiment.

Response times. Figure 10 shows the mean response times (overall height of the bars) for each pulley system, for statements referring to components at different positions in the causal chain. As with the error graphs, the data from this experiment are shown on the right, and the eye-tracking data are shown on the left for comparison. The times have been divided into the time spent reading the statement and the time spent inspecting the diagram.

Response times for the two pulley systems were analyzed separately, to allow for differences in the configurations. In the original eye-tracking experiment, the repetition caused a practice effect that resulted in the participants' responding significantly faster on the second repetition of the stimuli than on the first. The same effect was seen in the RFV data. The response advantages were 3.65 sec for Pulley System 1 [$F(1,10) = 44.40, p < .01$] and 4.05 sec for Pulley System 2 [$F(1,10) = 38.07, p < .01$]. The original experiment also showed that the position in the causal chain of the component referred to in the statement had a significant effect on response times. Again, the RFV data corresponded to the data from the original experiment, with position in the causal chain significantly influencing response times [for Pulley System 1, $F(2,20) = 18.87, p < .01$; for Pulley System 2, $F(2,20) = 24.15, p < .01$]. This effect can be seen in Figure 10.

These results are clearly in agreement with the eye-tracking experiment. However, the participants using the RFV took approximately 50% longer to respond. As was mentioned in Experiment 1, response time was likely to
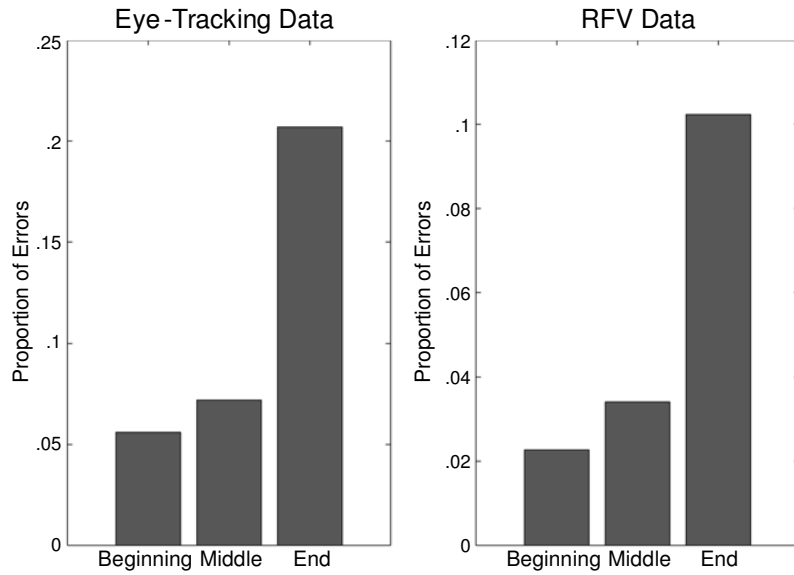
Eye-Tracking Data

RFV Data

Figure 9. Comparison of eye-tracking data from Hegarty's (1992) original experiment and RFV data from our replicated experiment, examining the proportion of errors at different causal chain positions.

result in part from the fact that the participants scanned the stimulus using a computer mouse. Despite the extra time taken, the overall trends in the data were very similar.

Further data analysis that was done in the eye-tracking experiment involved examining how long the participants inspected different components of the pulley systems. In particular, for each statement, the components in the diagram were divided into those whose motion occurred before the component referred to in the statement, the referent itself, and those components whose motion occurred
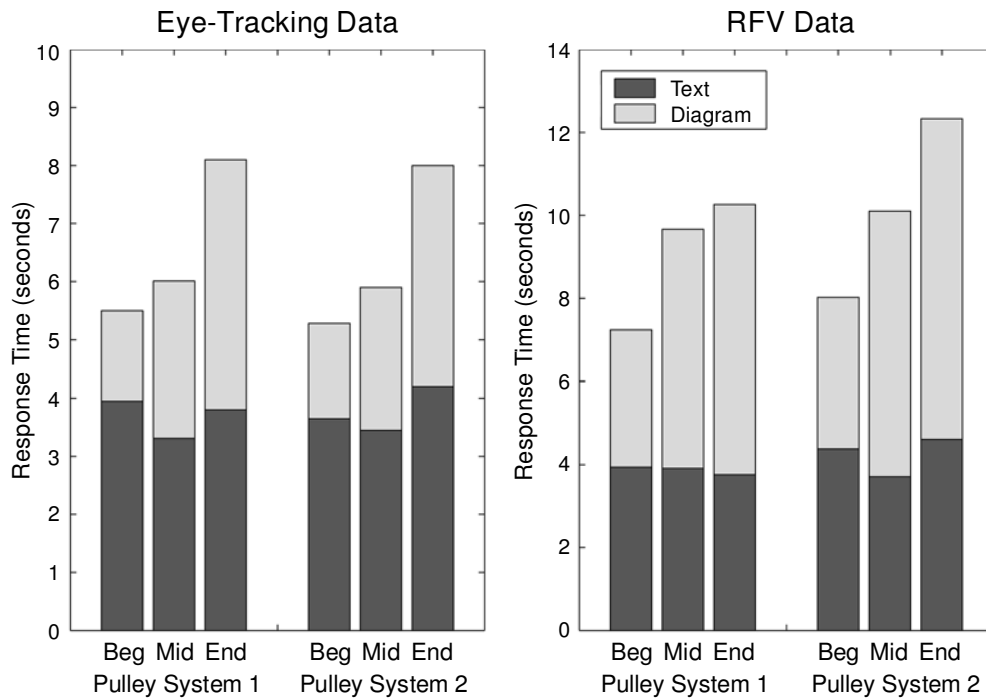
Eye-Tracking Data

RFV Data

Figure 10. Comparison of eye-tracking data from Hegarty's (1992) original experiment and RFV data from our replicated experiment, examining the mean response times for different trial types.
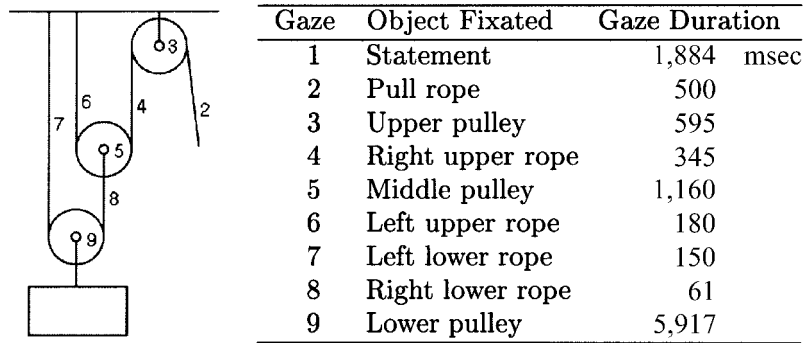
Gaze	Object Fixated	Gaze Duration	
1	Statement	1,884	msec
2	Pull rope	500	
3	Upper pulley	595	
4	Right upper rope	345	
5	Middle pulley	1,160	
6	Left upper rope	180	
7	Left lower rope	150	
8	Right lower rope	61	
9	Lower pulley	5,917	

Figure 11. Example of an RFV fixation protocol for the statement, "The lower pulley turns counterclockwise."

after the referent. This allowed for a further breakdown of response time spent viewing the diagram. Rectangular bounding boxes were used to enclose regions of the diagrams containing pulleys, rope strands, the ceiling, and the weight, just as in the original eye-tracking experiment. This allowed the order of fixations on the components of the diagrams to be determined, along with how long those fixations were. Figure 11 gives an example of a fixation protocol taken from the RFV data.

The gaze duration is defined as the total time spent fixating on components, in certain locations in the causal chain with respect to the referent in the statement. The graphs of the gaze duration data are shown in Figure 12. Again, the data from this experiment are on the right, and the eye-tracking data are on the left. The original eye-tracking experiment showed that, when looking at the pulley system, the participants spent most of their time inspecting the referent and the components whose motion preceded that of the referent in the causal chain of events. The RFV data show the same result for both Pulley System 1 (91%, $SD = 4.9$) and Pulley System 2 (92%, $SD = 4.6$).

Due to the fact that many participants in our experiment had more experience with technical diagrams than the participants in the original eye-tracking experiment, we expected some minor differences in strategy. However, the results of this experiment indicate that the response time, accuracy, and gaze duration trends obtained from the original experiment using eye-tracking techniques were the same as those obtained using the RFV. There were no key significant results from the original eye-tracking experiment that the RFV failed to obtain.

GENERAL DISCUSSION

The *Restricted Focus Viewer* (RFV) is designed to collect data about visual attention, as are eye-trackers. However, the RFV is not intended as a replacement for eye-tracking techniques. Rather, it is an alternative experimental technique and apparatus that adds to the toolbox available to cognitive scientists as they try to understand the processes involved in comprehending and reasoning with visual stimuli. It is therefore important to understand the relative advantages and disadvantages of the RFV and eye-tracking.

One primary concern is whether the RFV affects the high-level strategy used by participants. We note that this issue is not just confined to the RFV: In eye-tracking also, repeated calibrations and head gear (such as head-mounted cameras and infrared reflectance devices) may influence participant strategy. Experiment 2 suggests that high-level strategy is not changed by using the RFV instead of eye-tracking equipment. However, this depends on the task and the choice of RFV parameters. In particular, there are two important issues to consider.

The main issue is that, with the RFV, the experimenter can modify the size of the focus window and the amount of blurring, so as to ensure that the participant must explicitly focus on those components of the stimulus they are interested in. Such changes, however, may change the strategy used by the participant.

Consider again the algebraic expressions used as the visual stimulus in Experiment 1. It is preferable that the size of the expressions not be excessively large, since this would make any task involving it seem less natural to the participant. With the limited resolution accuracy of many eye-trackers, this is not always possible. Also, trying to determine the specific symbols that are being focused on was shown to be particularly problematic using eye-tracking, since each eye fixation can take in a large collection of symbols. With the RFV, the size of the focus window can be reduced so that only a few symbols or only a single symbol can be viewed at a time. This allows visual attention to be recorded at a level of detail not available using eye-tracking techniques. However, the reduction in focus window size also reduces the amount of information available to the participant at any given moment. This could affect the way a given task is approached. It appears that the more accurately one records the focus of the participants' attention (by reducing the size of the focus window), the more likely one is to affect the strategy that they would normally use. However, this is not a defect of the RFV; rather, it is an aspect that experimenters need to be aware of in experiment design and in setting the RFV parameters.

The second issue is that a computer mouse is used rather than eye and head movements to change the direction of attention. Thus, participants in the experiments need to be confident in the use of a mouse. Also, response
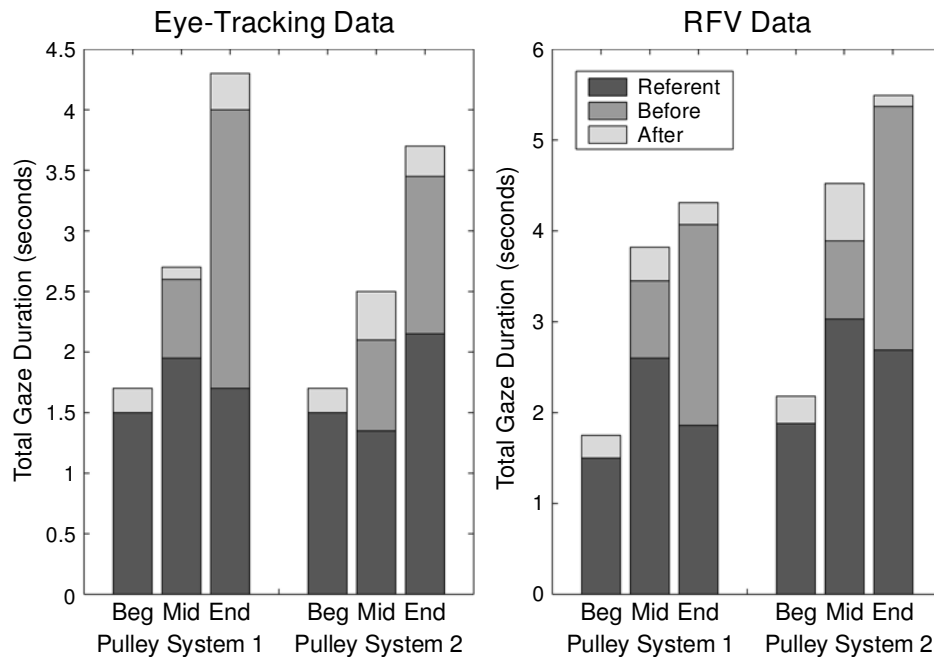
Eye-Tracking Data

RFV Data

Figure 12. Comparison of eye-tracking data from Hegarty's (1992) original experiment and RFV data from our replicated experiment, examining the breakdown of gaze duration on different components of the pulley systems.

times will be slower due to the use of the arm and the hand. For some tasks that require fast responses, this may present a problem. However, we believe that, for many tasks, this should not affect data trends or the significance levels of experimental results. One difference between the RFV and eye-tracking is what happens when the participants are not explicitly focusing on any part of the stimulus, because their attention is directed toward internal processing. At this time, their gaze may drift across the stimulus. With eye-tracking, this means that spurious fixations may appear in the data, whereas, with the RFV, the object last in focus will have a longer gaze duration.

By comparison, we have experienced far more significant difficulties when using eye-tracking equipment for experimental tasks to which we had easily applied the RFV. This was clearly seen in Experiment 1, and, although such problems will be familiar to experienced users of eye-tracking equipment, they are nevertheless significant obstacles to new researchers. Eye-trackers are expensive. They require substantial expertise in calibration and adjustment. They do not work reliably in strong daylight conditions. They can be unreliable with participants who have shiny skin, watery eyes, or contact lenses. Output data are often subject to positional drift, in addition to local nonlinear uncertainty. The vertical resolution of many eyetrackers is often poor in comparison with their horizontal resolution, making them less useful with detailed twodimensional stimuli. Analysis of data requires subjective classification of fixation and saccade thresholds. Fixations are often at a point between two display elements, leaving it unclear whether the participant is defocused or

viewing both elements as a unit. The tracker can lose stable gaze identification during the experiment, leading to invalid trials. In comparison with experiments conducted with the RFV using the same stimuli, eye-tracking results provided very little useful data.

Overall, the RFV has several advantages over traditional eye-tracking techniques. The system is cheap and easy to set up, providing accurate data about the region that is being focused on. It is nonintrusive, requiring no special gear to be worn or restrictions on the movement of participants. It does not require any calibration and can be used by participants who wear glasses. The RFV data are not corrupted by blinks or glances away from the stimulus, and the replayer provides a useful tool for immediate feedback on participant performance. Finally, the RFV has flexibility in its parameter settings, allowing it to be tailored to meet specific goals. As a result of these benefits, other researchers have also begun using the RFV (Futrelle & Rumshisky, 2001; Romero, Cox, Boulay, & Lutz, 2002).

The RFV is freely available in the public domain for other researchers to use at http://www.csse.monash.edu.au/projects/RFV/

REFERENCES

BLACKWELL, A. F., JANSEN, A. R., & MARRIOTT, K. (2000). Restricted focus viewer: A tool for tracking visual attention. In M. Anderson, P. Cheng, & V. Haarslev (Eds.), *Theory and application of diagrams: Lecture notes in artificial intelligence (LNAI) 1889* (pp. 162-177). New York: Springer-Verlag.

CARPENTER, P. A., & SHAH, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, **4**, 75-100.

Coren, S., Ward, L. M., & Enns, J. T. (1994). *Sensation and perception* (4th ed.). Harcourt Brace.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.

Futrelle, R. P., & Rumshisky, A. (2001). Discourse structure of text-graphics documents. In A. Butz, A. Krueger, P. Oliver, & M. Zhou (Eds.), *1st International Symposium on Smart Graphics*. New York: ACM Press.

Hegarty, M. (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1084-1102.

Henderson, J. M., McClure, K. K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception & Psychophysics*, **59**, 323-346.

Jansen, A. R. (2001). *Restricted Focus Viewer (RFV) Version 2.1 User's manual and tutorial* (Tech. Rep. No. 2001/87). Monash University, School of Computer Science and Software Engineering.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, **8**, 441-480.

Osaka, N., & Oda, K. (1994). Moving window generator for reading experiments. *Behavior Research Methods, Instruments, & Computers*, **26**, 49-53.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**, 372-422.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology*, **46**, 342-376.

Romero, P., Cox, R., Boulay, B. du, & Lutz, R. (2002). Visual attention and representation switching during Java program debugging: A study using the Restricted Focus Viewer. In M. Hegarty, B. Meyer, & N. H. Narayanan (Eds.), *Diagrammatic representation and inference: Lecture notes in artificial intelligence (LNAI) 2317* (pp. 221-235). New York: Springer-Verlag.

Schnipke, S. K., & Todd, M. W. (2000). Trials and tribulations of using an eye-tracking system. In *CHI 2000 extended abstracts* (pp. 273-274).

Stark, L. W., Ezumi, K., Nguyen, T., Paul, R., Tharp, G., & Yamashita, H. I. (1992). Visual search in virtual environments. *Proceedings of the SPIE: Human vision, visual processing, and digital display III*, **1666**, 577-589.

Steinke, T. R. (1987). Eye movement studies in cartography and related fields. *Cartographica*, **24**, 40-73.

Tovée, M. J. (1996). *An introduction to the visual system*. Cambridge: Cambridge University Press.

Ummelen, N. (1997). *Procedural and declarative information in software manuals*. Unpublished doctoral dissertation, Universiteit Twente, Utrecht.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.