Modeling Zooming and Scrolling Behavior in Very Large Documents

Michael Glueck\textsuperscript{1,2}, Tovi Grossman\textsuperscript{2}, and Daniel Wigdor\textsuperscript{1}

\textsuperscript{1}Dept. of Computer Science, University of Toronto
\{mglueck\|dwigdor\}@dgp.toronto.edu
\textsuperscript{2}Autodesk Research, Toronto, Canada
\{firstname.lastname\}@autodesk.com

ABSTRACT
Navigation of very large documents has not been previously well explored. Commonly employed aggregated measures of performance do not provide insight into the choices implicit in a navigation task, nor into how these choices ultimately affect performance. Thus, they support understanding of the aggregates, but not the constituent elements. We present a descriptive performance model of zooming and scrolling navigation, parameterized to account for speed-accuracy trade-offs, using standard mouse-based interaction techniques. We demonstrate the use of this model in framing the discussion of empirical results. Further, it is used to provide a basis for the design guidelines and heuristics key to achieving more performant interaction techniques for navigation of very large documents.

Author Keywords
zooming; scrolling; strategy model; very large document;

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces – Interaction Styles.

General Terms
Human Factors.

INTRODUCTION
As the size of digital documents and datasets continues to grow, so does the need for efficient ways to navigate them. While traditional scrolling UI includes some utility to support large documents, research has found that changes in viewing position beyond just a few screen widths can be significantly more efficiently carried-out by zooming out, and them back-in to the desired position in the document (c.f. Figure 1). Thus, for sufficiently large documents, some degree of zooming is essential to allowing navigation through the content. Many applications now offer some degree of zooming interaction, in addition to scrolling. Productivity software, such as Microsoft Office, readers, such as Adobe Reader, and even web browsers all provide zooming and scrolling controls to aid navigation of large documents.

Moreover, interactive visualizations of huge datasets are no longer found exclusively in the realm of trained expert users. Tools such as Google Maps [16] and Google Finance [15] provide users of all experience levels access to very large datasets through zooming and scrolling interfaces. Our domain of interest lies with the visualization very large single dimensional data, similar to the latter.

Scrolling\textsuperscript{1} and zooming are fundamental and ubiquitous interaction elements in desktop computing contexts, and, individually, have long been the focus of research in human-computer interaction. However, relatively little research has investigated user interactions in zooming interfaces, where both zooming and scrolling are used to navigate. While Guiard and Beaudouin-Lafon investigated target acquisition in zooming interfaces and developed a holistic performance model [17], their model does not elaborate on the contributions of individual zooming and scrolling operations to the whole, and thus is insufficient to fully model the details of user behavior.

Navigation in zooming interfaces is a complex interplay of sequential zooming and scrolling operations, each dependent on the previous. Parameterizing this sequence will foster a systematic exploration and comparison of navigation strategy as more than just the sum of its parts. Until now, optimal strategies have been postulated, but never empirically validated [14] (c.f. Figure 1).

In this paper, we explore navigation strategies in the context of mouse-based desktop interaction with very large documents. We theorize a user’s navigation strategy can be characterized as a sequence of interactions with trade-offs that can be parameterized; thus this work offers four key contributions. First, we present an algorithmic model highlighting the decisions contributing to these trade-offs, where sub-optimal choices will negatively impact overall performance. Second, we develop a parameterized model that describes the zooming and scrolling components of navigation. We derive costing functions for the model

- SCROLLING
- "NAIVE OPTIMAL" ZOOMING
- "OPTIMAL" ZOOMING

Figure 1. Diagrams of “optimal” strategies posited in [14]. A near target requires no zooming, only scrolling (left), “naive optimal” zooming until the target is visible (middle), and “optimal” zooming until the target is near, then scrolling (right).

\textsuperscript{1} As in [17], we do not distinguish between scrolling and panning. For the purposes of this paper, scrolling will be used to refer to both techniques.
based on empirical results, and evaluate proposed optimal strategies. Third, we investigate the effects of document familiarity on user navigation strategies. Finally, we present design guidelines and heuristics to guide the development of scrolling and zooming interface navigation techniques for very large documents.

RELATED WORK

Panning and zooming have been a part of UI design since the early 1960’s [28]. “Infinitely” zooming interfaces were first introduced with the Window [23] and Pad++ [5] systems, supporting interaction with multi-scale documents. We focus our review here on works which allow zooming and scrolling of content (‘documents’), rather than of user interface controls.

Furnas and Bederson presented scale-space diagrams as a means of understanding zooming and scrolling interfaces [14]. Based on a visual information complexity metric, they presented intuition towards an optimal zooming and panning trajectory, but these were not empirically validated. A method developed by van Wijk and Nuij calculates the optimal zooming and panning trajectory, while maintaining visual context, to smoothly animate transitions between different views within a zooming interface [29]. In contrast, we analyze users’ interactively-defined trajectories to identify where sub-optimal decisions are made and to quantify deviations from the optimal trajectory.

Text document navigation patterns were characterized in a log study of Microsoft Word and Adobe Reader [1]. In map interfaces, navigation patterns have also been analyzed to interpret task completion time differences between conditions [21] and the percentage of time spent zooming and scrolling has been compared between different tasks [23]. Our work seeks to explicitly parameterize the components of these navigation patterns and to quantify deviations from the optimal strategy.

Speed-dependent automatic zooming (SDAZ) [22] is a technique that automatically changes the zoom level with respect to the speed of scrolling, controlled through velocity-based input. This technique has been shown to be more efficient than scrolling with manual zooming [27] and traditional scroll, pan, and zoom operations [8]. Though informative, our focus is on modeling traditional zooming and scrolling interactions. Zooming and scrolling control have been combined in novel navigation techniques for position-based input; OrthoZoom Scroller [3] is a recent example. Empirical evaluations have shown that OrthoZoom Scroller outperforms SDAZ, however, this technique uses the entire document space as input, which is not suited to traditional applications where direct interactions occur over the document viewport. There is also a clear upper-bound on the size of documents which can be supported, whereas we are interested in documents of all sizes, including very large documents.

Performance models for scrolling have been presented and shown to follow Fitts’ Law [12] for scrolling to targets of known location [7,20]. Scrolling to targets of unknown location has been shown to follow either a linear model using scrollbars [2] or Fitts’ Law in a peephole pointing task [7]. The combination of zooming and scrolling has been modeled and also shown to follow Fitts’ Law [15].

While early studies [20,30] used traditional, unconstrained input, subsequent studies have mapped mouse position directly to scroll position [2,9,15], or have used bimanual input techniques [15,27]. Our work differs in several ways. First, Fitts’ Law-based models provide us with intuition of the total time we can expect a user to take to complete a task: an aggregate of all steps involved. It does not, however, provide insight into what those individual steps are, nor the sequence and use of each. Second, since we are interested in studying both zooming and scrolling when controlled by a user, we want to include the time it takes to switch from one widget to another. Third, we are interested in typical mouse-based desktop situations where bimanual interaction techniques of the type described in prior work are uncommon.

In summary, while overall performance has been modeled, a deeper understanding of zooming and scrolling behavior is lacking. In particular, we explore how users conceptualize the navigation process, which strategies they employ, and what factors impact their choice of strategy and performance. Moreover, in developing generalized performance models, highly abstracted tasks were evaluated with very controlled input. In the present work we attempt to bridge this gap and validate general theoretical models by using an abstracted navigation task with more ecologically valid input techniques. To the knowledge of the authors, no subsequent work has investigated mouse-wheel use since the evaluation by Hinckley et al. [20], nor has there been an investigation of mouse-wheel zooming, or the combination of mouse-wheel with direct manipulation techniques.

TASK CONSIDERATIONS

In this section, we summarize factors which have been shown to be important to zooming and scrolling tasks that have guided our model and experimental design.

Task Environment

We consider an abstracted scrolling and zooming task environment that could generalize to most real-world interfaces. The environment consists of a horizontally-oriented document viewport, with a scroll bar underneath, and zoom slider to the right (c.f. Figure 2).

To investigate zooming interface navigation strategies, we developed an abstract one-dimensional document. The document contains a red target line that remains the same width at all zoom levels. A light-blue goal zone appears in screen-space, fixed to the center of the viewport. To mitigate desert fog, where a user can become lost without sufficient critical zones or cues when zooming and scrolling [24], our document contains evenly spaced lines that fade-in and fade-out at different zoom levels. In addition, scroll bar and zoom slider controls were always visible to provide feedback of document position and zoom level. The
Figure 2. Screen shot of the zooming and scrolling interface, consisting of a document viewport, scroll bar underneath, and zoom slider to the right. Goal shown in blue, target in red.

zoom percentage is also displayed above the viewport, and the dark grey bars bounding the upper and lower edges of the document change color to signal when the document is zoomed back to 100%.

This interface supports a view-pointing task [18], which has been previously applied to model zooming and scrolling performance [17]. The goal is to find the target line and position it within the goal zone at a specified zoom level. To successfully find and position the target line in a very large document, one must zoom-out to find the target, and then perform a combination of interleaved scrolling and zoom-in to position it correctly.

The need to interleave scrolling and zooming-in stems from a phenomenon we call zoom-in drift – the appearance that the document is expanding from the zoom pivot, the point around which zooming is calculated. When zooming-out, document space is compressed in screen-space, resulting in binning, many document-space pixels into a single screen-space pixel. This process is reversed when zooming-in. Even if a target is positioned exactly at the zoom pivot prior to zooming-in, it will drift away due to de-binning of pixels.

**Input Device:** Our work focuses on single-handed, mouse-based input. We also include mouse-wheel input for increased external validity.

**Viewport Sizes:** viewports larger than 40px have minimal impact on performance in zooming and scrolling tasks [15]. For this reason, in our study, we use a fixed viewport size.

**Viewport Orientation:** Based on our interest in large data visualization interfaces, we use a horizontal document orientation, similar to Google Finance [15].

**Interaction Techniques:** We focus on 3 traditional techniques for zooming and scrolling: click-and-drag, mouse-wheel scroll, and mouse-wheel zoom (with CTRL modifier). These are commonly found in reader software and productivity software.

**Familiarity with Document:** Our work considers navigation within both familiar and unfamiliar documents. This will impact a user’s knowledge of the target location during the task.

**Target Saliency:** For simplicity, we assume the target is equally discernible across all zoom levels, and so display the target with a fixed width in screen-space pixels. In addition, there are no distractors. Although these decisions introduce an artificiality which could impact the ecological validity of our results, we preferred to minimize the impact of visual search complexity within our costing functions by making it as easy as possible to find the target.

**Document Bounds and Zoom Pivot:** In addition to interaction, the scroll bar visually communicates the relationship between the total document size and displayed portion in the viewport. To ensure accurate scroll bar feedback, we limited scrolling to the document bounds, and to control for individual variations in spatial ability, we used a viewport-centered zoom pivot. The efficiency of zooming interfaces is reduced when document-bounded scrolling is paired with a fixed zoom pivot: zooming-in towards document edges is encumbered since the magnitude of corrective scrolls is limited by the bounds of the document. To avoid this pitfall, we chose trial conditions avoiding document edges.

**A MODEL OF OPTIMAL PERFORMANCE**

Furnas and Bederson posited that a “naïve optimal” strategy is to zoom-out until the target is visible before zooming in [14]. This was further refined by considering separately nearby targets from distant ones. They calculated that the optimal strategy for near targets is to scroll without zooming. By extension, the optimal strategy for distant targets was to zoom-out until the target was near the center of the zoom (rather than visible on-screen), then scroll to position the center of the zoom pivot onto the target before zooming back in. “Near” was estimated to be within 1-3 viewport widths (c.f. Figure 1).

Naturally, such a model presumes that the difficulty in identifying the target is consistent across zoom levels. This is true to a greater or lesser extent depending on the domain of the document. In the present work, we begin with the same assumption, but ultimately account for target salience in our strategy model.

While empirical results suggest users prefer scrolling without zooming when targets are near [17], the key insight is that there appears to be a crossover point where zooming-out is faster than scrolling: users make this decision based on their familiarity with the document, spatial reasoning ability, and skills in performing the navigation operations.

**An Algorithmic Decision Model**

We developed a decision model to conceptualize the iterative process a user engages in while searching for a target in a zooming interface (see Figure 3). Our decision model accounts for both choices and interactions – whether or not they are optimal. At each choice (diamond), a user may misjudge the best decision and at every interaction (rectangles), a user may navigate sub-optimally. This model clearly compartmentalizes the components of navigation strategy, and provides a framework to support the discussion of user choices and interactions.
We highlight three regions, Region A, Region B, and Region C, which represent zooming-out, scrolling, and zooming-in components of navigation, respectively. Note that a user is not required to engage in any zooming. If the user chooses not to zoom-out in Region A of the model, they will not reach Region C, and the model simply reduces to describe traditional scrolling in Region B.

Region A encapsulates the zooming-out phase of interaction. The critical decision, Is target close enough?, is aided by knowledge of the target location. A misjudgment leading to an error, for example failing to zoom-out far enough, would require restarting the decision process. Without any knowledge of target location, only the “naïve” optimal strategy can be employed. However, whether or not a target is salient at all zoom levels affects a user’s ability to follow the “naïve” optimal strategy. For example, content in multi-scale documents may only be visible at a particular zoom level. This leads us to consider a target saliency threshold, an upper bound on how far a user can zoom-out given document content, which can undermine the “naïve” optimal strategy. Thus, we posit that both document familiarity and target saliency threshold impact the accuracy of choices made in Region A.

Region B and C, on the other hand, encapsulate the scrolling and zooming-in phase of interaction. Judgment of the critical decision, Has target shifted enough?, is a complex trade-off in screen-space pixels between speed, accuracy, and magnitude of scrolling, additionally complicated by zoom-in drift, since there is a limit to how far one can zoom-in before having to reposition the target at the zoom pivot. For example, if a user chooses a low threshold for “shifted enough”, more zoom-in loops would be required. Conversely, choosing a higher threshold would result in fewer zoom-in loops, but at the increased risk of losing the target through desert fog. A performance balance must exist between “many small inaccurate scrolls and short zoom-ins” and “fewer larger accurate scrolls and long zoom-ins”. Thus, we posit that a user’s spatial reasoning ability and motor skills impact the accuracy of choices made in Region C.

Sub-optimal navigation is accounted for within the decision process. However, major errors, such as getting lost in the document or completely losing track of the target, require restarting the decision process altogether. We feel this model accurately reflects how users approach navigation problems in the real world and account for errors. For example, if a user zooms-in too fast, such that the target is out of range, a user will need to decide whether the target is close enough to scroll to, or if they will zoom-out once again to relocate it.

Not only does this model underscore the complex iterative trade-offs involved in zooming and scrolling navigation, it also provides intuition for improving the choices a user makes at each step. For example, providing feedback about target location and designing interaction techniques that supplement both a user’s motor and spatial abilities will help prevent interaction errors and improve overall performance.

### A Zooming-Scrolling Strategy Model

While the decision model considers user choices and interactions, the strategy model defines parameterizations for user navigation trade-offs and global contextual limitations, such as the target saliency threshold. Based on costing functions, which will account for both visual search and interaction time, this model will calculate the expected task completion time. By iteratively searching the space of task completion times using different parameter values, optimal task completion times and parameter values can be established. In this section, we describe the logic behind these parameterizations and costing functions. The values of both will be empirically derived in our first experiment.

Without loss of generality, our model assumes a viewpoint task in a one-dimensional horizontal document that can be zoomed and scrolled. For simplicity, zooming

---

**Table 1. Summarizing meaning of strategy model parameters.**

<table>
<thead>
<tr>
<th>Param</th>
<th>Name</th>
<th>Measures</th>
<th>Eq #</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>Target distance</td>
<td>Distance to target in document pixels</td>
<td>--</td>
</tr>
<tr>
<td>$v$</td>
<td>Viewport size</td>
<td>Viewport size (width) in screen-space pixels</td>
<td>--</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Actual zoom-out</td>
<td>Number of levels zoomed-out</td>
<td>--</td>
</tr>
<tr>
<td>$z'$</td>
<td>“Naïve optimal”</td>
<td>The “naïve optimal” zoom-out level</td>
<td>(3)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Scroll accuracy</td>
<td>Average screen-pixel distance to zoom pivot at end of a zoom-in step</td>
<td>--</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Scroll magnitude</td>
<td>Average screen-pixels scrolled for each zoom-in step</td>
<td>--</td>
</tr>
<tr>
<td>$\Delta z$</td>
<td>Zoom-in delta</td>
<td>Number of levels zoomed-in at zoom-in step</td>
<td>(5)</td>
</tr>
<tr>
<td>$n$</td>
<td>Zoom-in steps</td>
<td>Average number of zoom-in steps</td>
<td>(6)</td>
</tr>
<tr>
<td>$m$</td>
<td>First scroll ratio</td>
<td>The ratio of the viewport scrolled after zooming-out</td>
<td>(7)</td>
</tr>
</tbody>
</table>
operations function with a viewport-centered zoom pivot, and the target is to be placed in the center of the viewport. The underlying logic of the model can be easily extended for arbitrary zoom pivots and goal positions.

The model calculates, for a particular navigation strategy, the task completion time, $T$, in milliseconds (ms). A navigation strategy is which is defined as the set of 7 parameters, explained in Table 1. The derivation of $T$ is:

$$T = \text{Cost}_{\text{Zoomout}}(\alpha) + \text{Cost}_{\text{Scroll}}(V, m) + \text{Cost}_{\text{Zoomin}}(\Delta z, n) + \text{Cost}_{\text{CorrScroll}}(\rho, n)$$

(1)

We start by explaining the derivation of these parameters, and follow with a discussion of the costing functions.

**Zooming-Out**

Our model is based on the relative position of the viewport and the target at the start of a given task. We define $D$ as the distance to the target from the zoom pivot, in document pixels. In our case, the zoom pivot is coincident with the center of the viewport (c.f. Figure 5).

![Diagram illustrating the distance to the target, D, based on the initial the viewport position, in document pixels.](image)

We define $V$ as the viewport size (horizontal) in screen pixels and $z$ as the zoom level. The function $R(z)$ defines the number of document pixels displayed in the viewport in screen-space pixels for a given zoom level $z$ (c.f. Figure 6).

![Diagrams (a) and (b) illustrate the viewport size, V, and the range function, R(z), in screen-space pixels. Diagram (b) also illustrates the relationship described in equation (3).](image)

$$R(z) = V \cdot 2^z$$

(2)

The document is viewed at “actual size” (zoomed to 100%) when $z = 0$. At this zoom level, screen-space pixels and document pixels are equivalent, so $R(0) = V$ (see Figure 6a). The document is decreased to 50% when $z = 1$, and so on. In general, zooming-out causes $z$ to increase and zooming-in causes $z$ to decrease.

We define $z'$ as the “naïve optimal” zoom level [10], the minimum zoom level whereby the target is first visible in the viewport. This occurs when $R(z') = D$ (c.f. Figure 6b).

Thus, substituting (2) and solving for $z'$ yields:

$$\frac{V \cdot 2^{z'}}{2} = D$$

$$\log_2 V + z' - 1 = \log_2 D$$

$$z' = \log_2 D - \log_2 V + 1$$

(3)

We define $\alpha$ as the actual zoom level achieved by the user during the zoom-out operation, given that users may either undershoot the zoom (perhaps “optimally”, as posited [10], or stop at or below the target saliency threshold) or overshoot the zoom.

**Zooming-In**

Zooming-in is an interleaved process of zooming-in and corrective scrolling due to zoom-in drift. When the target appears “closest” to the zoom pivot, it is unlikely that they are coincident. In the worst case, the range between the target and the zoom pivot is at most 1px in screen-space.

A user may not position the target as close as possible to the zoom pivot at every zoom-in step. We define $\sigma$ as the average range in screen-space pixels from the zoom pivot to the target over all zoom-in steps, where $\sigma \geq 1.0$. Then, $Q(\Delta z, \sigma)$ is the zoom-in drift of the target in screen-space pixels, away from the zoom pivot, given a change of $\Delta z$ zoom levels (c.f. Figure 4).

$$Q(\Delta z, \sigma) = 2^{\Delta z} \cdot \sigma$$

(4)

We define $\rho$ as the average distance in screen-space pixels that the target is scrolled over all zoom-in steps. Intuitively, some users may prefer that the target always remain on screen, thus ensuring that $\rho \leq \frac{V}{2}$. Others may be comfortable with a larger $\rho$, confident in their ability to re-locate the target despite shifting off-screen. If a user consistently stops zooming-in when the target is at the edge of the viewport and scrolls the target exactly to the center of viewport, then $\sigma = 1$ and $\rho = \frac{V}{2}$.

Then, the average change in zoom level, $\Delta z$, that maintains the balance between $\sigma$ and $\rho$ occurs when $Q(\Delta z, \sigma) = \rho$. Thus, substituting (4) and solving for $\Delta z$ yields:

$$2^{\Delta z} \cdot \sigma = \rho$$

$$\Delta z + \log_2(\sigma) = \log_2(\rho)$$

$$\Delta z = \log_2(\rho) - \log_2(\sigma)$$

(5)

Since the task we are modeling requires the user to return to the same zoom level at which they started, we derive the total number of zoom-in iterations, $n$, from $\alpha$ (actual zoom level) and $\Delta z$ (delta of zoom-in steps):

$$n = \frac{\alpha}{\Delta z}$$

(6)

The parameters $\sigma, \rho, \Delta z$, and $n$ quantify the impact of the speed-accuracy trade-off of the Has target shifted enough? (see Figure 3, Region C). For example, a user may choose...
to be less precise when scrolling, with a larger \( \sigma \) and smaller \( \rho \), but at the cost of a decreased \( \Delta z \), and as a result an increase in \( \pi \).

One corrective scroll is performed after zooming-out, to re-center the target prior to zooming-in. We handle this operation uniquely to account the scroll distance relative to the actual zoom-out level \( (\alpha) \). Since our model is defined relative to the initial distance to the target, we define \( m \) as the ratio of target distance, \( D \), to half the zoomed viewport size, \( R(\alpha) \):

\[
m = \frac{D}{R(\alpha)/2}
\]

The distance of this first corrective scroll in screen-space pixels is then \( m \cdot \frac{\sqrt{3}}{2} \). For example, if a user zooms-out precisely to the “naïve optimal” zoom level, \( z' \), then \( m = 1 \), and the distance to scroll is \( \frac{\sqrt{3}}{2} \).

In summary, these parameters capture both variations in initial zoom-out amplitude, as well as the speed-accuracy trade-off while zooming-in. Figure 7 illustrates all of the parameters of this model.

**Costing Functions**

To account for the time it takes the user to perform each navigation operation, we introduce costing functions with the following form:

\[
Cost_{op}(d, r) = r \cdot (A_{op} \cdot d + B_{op} + C_{op})
\]

where, the \( Cost_{op}(d, r) \) calculates the average time (ms) to complete \( r \) repetitions of distance \( d \) for the operation \( op \).

The time will be calculated with empirically derived values for \( A, B, \) and \( C \). \( A \) and \( B \) are parameters of a linear fit for physical operation performance and \( C \) represents the time requires to acquire/initiate the operation and visual search.

Unfortunately, visual search is not well understood and there are no accepted models for performance [13]. For the purposes of this model, we assume a minimized cognitive load through pre-attentive search – a uniform target saliency across zoom levels, and that the absence or presence of the target is immediately distinguishable. However, our costing functions can later be extended with appropriate models of visual search performance as they are developed.

**EXPERIMENT 1 – MODEL QUANTIFICATION**

In our first experiment, we estimate the costing function parameters \( (A, B, \) and \( C) \) for scrolling and zooming by analyzing the magnitude and duration of individual navigation operations. We recorded task completion time and all navigation operations (scroll/zoom; wheel scroll/wheel zoom/drag; magnitude; duration; position; zoom level).

**Index of Difficulty**

Since combined zooming and scrolling has been shown to follow Fitts’ Law [15], we base our selection of distance to target \( (D) \) and size of goal \( (W) \) upon the index of difficulty (ID) [12], using the common derivation \( ID = \log_2 \left( \frac{D}{W} + 1 \right) \), measured in bits. To provide perspective on the difficulty of tasks, the threshold for a target being immediately visible is \( ID = \log_2 \left( \frac{400}{8} + 1 \right) = 5.67 \), and at \( ID = \log_2 \left( \frac{2400}{8} + 1 \right) = 8.23 \), the target is beyond 3 viewport widths away.

**Participants and Apparatus**

We recruited 12 paid ($10) volunteer participants (7 female; no left-handed) with mean age 24 (min: 19, max: 30). Mean self-reported computer usage was 24 hours per week. Self-reported use frequency of zooming and scrolling interaction techniques varied: 83% of participants reported using the mouse-wheel to scroll “often” or “always” when available. For input device, 50% of participants reported primarily utilizing wheel mice, 42% touchpads, and 8% touch mice.

Participants performed the study in a private study room using a desktop computer configuration, with a 21-inch LCD monitor displaying a resolution of 1280x1024 pixels and a standard wheel mouse. Participants were seated 20 inches from the monitor. The interface was setup with a fixed-sized viewport (800x450px), displaying a document of 2 Gpx wide.

**Methods and Design**

Participants performed a view-pointing task. Trials began with the document zoomed to 100%. The goal was to find the target line, position it within the goal zone, and reset the zoom scale to 100%. We used 5 evenly-spaced target distances, based on IDs where zooming would be required, starting at 10 up to 26 [15]. The goal (8px) and target (1px) appeared with fixed-sizes in screen-space. Based on a pilot study, we found no significant effects of target direction, and simplify our design by limiting to right-directional navigation. Only three interaction techniques were available: mouse-wheel scrolling, mouse-wheel zooming, and click-and-drag scrolling to control for potential performance variations resulting from differing combinations of interaction techniques.

A repeated measures within-participant design was used, with the independent variable difficulty (10, 14, 18, 22, 26 bits). The study was divided into 3 blocks, with 4
randomized repetitions of each condition per block. One practice repetition was administered before the trials started, and discarded. Each participant completed the experiment in a single session of approximately 30 minutes.

**Analysis**

Since our focus was on parameterizing our model, we discarded trials where users commit major errors, as defined by our decision model (scrolling prior to first zoom-out and corrective zoom-outs during zoom-in). In total, 19.4% of trials were discarded (140/720). We parsed the interaction logs to extract uses of each interaction technique, using direction reversal and operation type (zoom or scroll) to delimit groups of sequential operations. We also calculated the transition time between interactions.

**Costing Functions**

We extracted magnitude and duration for each operation type to estimate parameters for our costing functions. Table 2 summarizes the costing function parameters. Parameters $A$ and $B$ are derived from fitting a line to the scatterplot data ($R^2$ value reported). The value of $C$ is the average transition time between an operation and the preceding operation.

**Discussion: Participant Strategies**

Through analysis of the interaction logs, we can quantify the parameters ($\alpha$, $\sigma$, $\rho$, $\Delta z$, and $n$) of our zooming and scrolling strategy model for each trial. The averages across participants are outlined in Table 3.

The actual zoom-out level, $\alpha$, and the number of zoom-in iterations, $n$, increase with ID, which follows intuition, since conditions with larger IDs have farther targets, and require more navigation overall.

The average zoom-in delta, $\Delta z$, fluctuates minimally as ID increases, suggesting that participants are consistent in how far they zoom-in at each step. The speculated trade-off during zoom-in iterations is supported in the data, between parameters $\sigma$, $\rho$, $\Delta z$, and $n$. In particular, the dramatic increase of $\sigma$ with ID26 suggests a shift in navigation strategy; that participants sacrificed accuracy for larger scroll distances ($\rho$).

**Estimated Optimal Strategies**

Having posited our model and populated it with costing functions reflective of our apparatus, we are now in a position to estimate both an expected completion time, based on the calculated parameters for each difficulty, as well as an optimal completion time, based only on distance to target ($D$). The results are shown in Figure 8, left.

For Experiment 1, the expected completion time lies within the 95% confidence interval of the observed data, in all but the ID10 condition, indicating that our model is able to predict task completion time. On average, our expected times are 9% faster than the observed times, and estimated optimal completion time 32% faster than observed.

Based on our model, the optimal values for each ID is the parameterization $\alpha = z' + 2$, $\sigma = 1$, and $\rho = 400$. This contradicts the “optimal” strategy posited by Furnas and Bederson, which suggests it is optimal to undershoot $z'$ [14]. We elaborate on two possible explanations for this difference. First, in Experiment 1, we simulated an unfamiliar document, where this “optimal” strategy was impossible, since target location was unknown. It has been suggested that human performance has a very low upper bound based on perceptual bandwidth [17]. This may have resulted in reduced ballistic movements by users, who were constantly engaged in the pre-attentive search task. Second, the precise metric used to determine the “optimal” strategy is unclear. Since that publication, advances in input devices, such as the mouse wheel, has made multi-modal interaction easier, in particular in-place zooming activation. In contrast, mouse pointing interaction has changed little. By drastically reducing the cost of engaging a zooming operation, the favor may have shifted from more scrolling to more zooming, as predicted by our model.

Thus, investing the time to minimize $\sigma$ and maximize $\rho$ pays dividends, as it maximized the zoom-in delta reduces the overall number of zoom-in iterations ($n$). Our model also highlights that scrolling is the major bottleneck in zooming and scrolling interaction, lending support for the

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Time</th>
<th>$\alpha$</th>
<th>$\sigma$</th>
<th>$\rho$</th>
<th>$\Delta z$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID10</td>
<td>6.6s</td>
<td>7.3</td>
<td>13.2</td>
<td>148.7</td>
<td>3.8</td>
<td>2.4</td>
</tr>
<tr>
<td>ID14</td>
<td>8.5s</td>
<td>10.6</td>
<td>16.5</td>
<td>192.0</td>
<td>3.9</td>
<td>3.4</td>
</tr>
<tr>
<td>ID18</td>
<td>11.3s</td>
<td>14.4</td>
<td>11.8</td>
<td>223.6</td>
<td>4.4</td>
<td>4.1</td>
</tr>
<tr>
<td>ID22</td>
<td>13.5s</td>
<td>16.7</td>
<td>15.2</td>
<td>253.6</td>
<td>4.3</td>
<td>5.1</td>
</tr>
<tr>
<td>ID26</td>
<td>17.9s</td>
<td>21.2</td>
<td>25.1</td>
<td>329.14</td>
<td>3.8</td>
<td>6.6</td>
</tr>
</tbody>
</table>

Table 3. Summarizing model parameters calculated from the empirical results of Experiment 1.
use of user-controlled zoom pivots, whereby $\sigma'$ is intrinsically minimized.

**EXPERIMENT 2 – DOCUMENT FAMILIARITY**

Having modeled optimal performance and analyzed strategy when target location is unknown, we wished to understand how users choose to perform document navigation, when knowledgeable about target position. Thus, in our second experiment, we explored user navigation strategies when varying the degree of (simulated) familiarity with the document. We recorded task completion time and calculated strategy model parameters for each trial.

**Document Familiarity**

We consider two classes of document familiarity: local and global. Local familiarity entails leveraging landmark features of a document to cue the relative distance of a target (e.g., wishing to view a chart of stock performance for September while the screen is currently centered on January). In contrast, global familiarity provides coarse awareness of the overall location of the target (e.g., a section located approximately half-way through the document). We simulate the former by using dynamically updating directional arrows [3,20] with magnitude indicators (c.f. Figure 9) and the latter with a document overview [9] (c.f. Figure 10).

**Participants and Apparatus**

We recruited 24 paid ($20) volunteer participants (11 female, 2 left-handed) with a mean age of 23 (min: 19, max: 29). Mean self-reported computer usage was 24 hours per week. Self-reported use frequency of zooming and scrolling interaction techniques varied: 71% of participants reported using the mouse-wheel to scroll “often” or “always” when available. For input device, 50% of participants reported primarily utilizing touchpads, 42% wheel mice, and 8% standard mice without wheels.

We used the same apparatus as in the first experiment.

**Methods and Design**

The 3 conditions for document familiarity were: none, local, and global. We tested 3 evenly spaced target distances, based on IDs, from 7 to 27. This range differed from Experiment 1 to cover the case when the local aid might prompt a user to scroll to a target within 3 viewport widths. A repeated measures within-participant design was used, with independent variables familiarity (none, local, global) and difficulty (7, 17, 27 bits). A fully-crossed design resulted in 9 combinations of variables. Indicator was counterbalanced. The study was divided into 2 blocks, with 3 randomized repetitions of each condition within each block. One practice repetition was administered before each change of conditions, and discarded. Each participant completed the experiment in a single session of approximately 60 minutes.

**Hypotheses**

We predicted that document familiarity would produce a positive impact on performance; that both local and global familiarity would outperform the lack of document knowledge (none) due to optimization of strategy (H1). Similarly, we predicted that both local and global familiarity would alter user strategies when compared to the no knowledge condition (none). In particular, that actual zoom level ($\alpha$) would be higher when familiar, leveraging document information to optimize strategy (H2). Finally, that the zoom-in trade-off ($\sigma$ and $\rho$) would not be affected by familiarity, since these are based more on spatial and motor abilities than document content (H3). Finally, we expect at least some participants to recognize that local familiarity can be leveraged to skip zooming altogether, and instead scroll to the nearest target (H4).

**Results**

A repeated measures analysis of variance was performed on the dependent variables task completion time and three model parameters: $\alpha$, $\sigma$, and $\rho$; Greenhouse-Geisser corrected for violations of sphericity. Descriptive statistics are summarized in Table 4.

As expected, a main effect was found between difficulty and time ($p<.001$, $F_{2,22}$=643.47), $\alpha$ ($p<.001$, $F_{2,22}$=4879.45), and $\rho$ ($p<.001$, $F_{2,22}$=43.76). Additionally, a main effects were found for familiarity for each of $\alpha$ ($p<.05$, $F_{2,22}$=4.71) and $\rho$ ($p<0.5$, $F_{2,22}$=4.92). No significant effect was found between time and familiarity, suggesting that any differences in strategy

<table>
<thead>
<tr>
<th>Fam.</th>
<th>Diff.</th>
<th>Time (s)</th>
<th>$\alpha$</th>
<th>$\sigma$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>ID7</td>
<td>5.3 (0.5)</td>
<td>4.3 (0.7)</td>
<td>23.2 (16.9)</td>
<td>125.4 (89.5)</td>
</tr>
<tr>
<td></td>
<td>ID17</td>
<td>11.4 (1.1)</td>
<td>13.3 (0.4)</td>
<td>13.2 (5.4)</td>
<td>228.8 (66.4)</td>
</tr>
<tr>
<td></td>
<td>ID27</td>
<td>17.5 (13.0)</td>
<td>21.3 (0.2)</td>
<td>17.2 (5.8)</td>
<td>296.2 (50.1)</td>
</tr>
<tr>
<td>Local</td>
<td>ID7</td>
<td>5.2 (0.8)</td>
<td>3.5 (0.8)</td>
<td>15.9 (14.4)</td>
<td>94.4 (61.4)</td>
</tr>
<tr>
<td></td>
<td>ID17</td>
<td>11.1 (1.0)</td>
<td>12.5 (0.5)</td>
<td>10.5 (4.7)</td>
<td>235.7 (62.3)</td>
</tr>
<tr>
<td></td>
<td>ID27</td>
<td>17.1 (1.4)</td>
<td>21.1 (0.3)</td>
<td>15.8 (4.9)</td>
<td>322.3 (102.6)</td>
</tr>
<tr>
<td>Global</td>
<td>ID7</td>
<td>5.4 (0.7)</td>
<td>3.9 (0.5)</td>
<td>8.2 (5.8)</td>
<td>64.0 (31.0)</td>
</tr>
<tr>
<td></td>
<td>ID17</td>
<td>11.5 (1.4)</td>
<td>12.9 (0.4)</td>
<td>11.9 (5.6)</td>
<td>194.2 (49.3)</td>
</tr>
<tr>
<td></td>
<td>ID27</td>
<td>18.1 (1.8)</td>
<td>21.4 (0.2)</td>
<td>17.2 (5.4)</td>
<td>253.0 (50.5)</td>
</tr>
</tbody>
</table>

Table 4. Summarizing descriptive statistics of Experiment 2. 95% confidence intervals in brackets.
related to familiarity did not significantly optimize performance time. There were no significant interactions. Bonferroni-adjusted pairwise comparisons revealed $\alpha$ was significant smaller in local than global conditions ($p<.05$). The lack of significant difference from none suggests that familiarity lead to more consistent zooming-out, however, this was not an optimal choice. Bonferroni-adjusted pairwise comparisons revealed $\rho$ was significantly smaller in the global condition over both none ($p<.05$) and local ($p<.05$). This suggests a sub-optimal reduction in scrolling magnitude when global familiarity was simulated.

Model Fit
Similar to Experiment 1, we calculated expected and optimal task completion times based on the parameters of Experiment 2 (c.f. Figure 8, right). The expected completion time only lies within the 95% confidence interval of the observed data in the ID27 condition. On average, our expected times are 19% faster than the observed times, and estimated optimal completion time 27% faster than observed.

Discussion
Contrary to our expectation (H1), it appears that participants were not able to leverage familiarity to optimize their strategy in a significant way. Based on the analysis of the parameters, it appears that participants failed to optimize their strategy by increasing $\alpha$, instead decreasing the value (H2). It would appear that participants may not know what the optimal strategy is, or how to best optimize their strategy, choosing to instead follow the “naïve optimal” strategy.

Perhaps more interesting is the lack of any significant effects for $\sigma$, suggesting this parameter is robust over difficulty and familiarity; regardless of the conditions of the task, users do not significantly vary their accuracy. This lends some support to (H3). However, the differences in $\rho$ suggest that global knowledge led to more constrained and suboptimal interaction, contrary to (H3). We hypothesize that perhaps participants were distracted by global contextual information while zooming-in; a kind of information overload leading to suboptimal performance.

We visualize participant strategies using a space-scale inspired diagram, which reveal an interesting systematic deviation from the “naïve optimal” strategy (c.f. Figure 11). Instead of zooming to reveal the target, participants used the local navigation aid to decide when to scroll to it. This was an optimal decision, as it was faster to scroll than zoom for the easiest ID. This was leveraged in 23% of trials; this lends support to (H4).

In summary, document familiarity can lead to changes in a user’s strategy, however, whether the altered decisions are better than the naïve, uninformed ones is not guaranteed. Users must be taught, or otherwise pointed towards the “better” decision, so that they can leverage additional knowledge to make better, decisions of navigation strategy.

LIMITATIONS
The accuracy of the strategy model is reflective of the costing functions used. We based our costing functions on a typical desktop user context, and have demonstrated that it is robust to this context and to document familiarity. However, these costs may need to be reassessed when applying the model non-desktop contexts, or non-mouse interaction paradigms, such as touchpads, track-points, or direct touch physical manipulation. We used generic operating system drivers with default settings. Indeed, different drivers have been shown to have different scrolling transfer functions [26]. Real-world users may have specialized drivers or customized settings which would impact the costing functions of the model.

For simplicity, our costing functions do not account for clutching in mouse drags. We also approximate the speed-accuracy trade-off using a linear model. More complex costing functions should be examined, that more accurately represent both the speed-accuracy trade-off, as well as the impact of input device clutching over longer distances.

We simulated document familiarity using proxy aids. These cannot fully account for the nuances of information provided by real world documents, nor for familiarity users may have with the content.

Finally, in our study, the task required the user to engage in a “pre-attentive” visual search, where the judgement of presence or absence of a target could be made very quickly. In real world documents, distractors may be present, or the search target may not be uniformly salient across zoom levels, both factors which would increase the time required to engage in the visual search task.

DESIGN GUIDELINES AND HEURISTICS
Simulate Document Familiarity. Techniques such as overview+detail and focus+context (see [10] for a review) or off-screen target representations are effective in supporting users as they provide context and additional information about target location [4,19]. However, not all of these techniques are directly scalable to very large documents, or even to 1-dimensional documents, and, further, would require the system to have knowledge allowing it to reduce the universe of possible targets.

Maximize target saliency threshold. Our model shows the benefit of maximizing the actual zoom-out level ($\alpha$). Techniques like semantic zooming [25] and visualization aggregation [11] can help to promote salience of targets at all levels of zoom. However, the task domain must be carefully considered when aggregating data, such that
condensed representations do not lack the detail required to complete a task. Lack of detail may be continuous, such as details simply blending together as the user zooms, or discrete, such as when zooming-out of a map application results in street labels being removed.

Enhance motor and spatial skills. Optimal navigation benefits from being able to correctly scroll quickly and accurately (minimize $\sigma$, maximize $\rho$). Techniques such as snapping, semantic pointing [6], and control-display gain schemes [26] can augment a user’s natural abilities. Especially when designing user interfaces for non-traditional input devices, such as touch or pen, which are more demanding on a user’s motor and spatial skills, it is important to ensure that both precise and ballistic navigation operations are easily controlled.

Minimize scrolling. While scrolling speed and accuracy should be well supported, another option is to factor scrolling out of the equation. For example, supporting a user controlled zoom pivot, such as mouse-cursor based zooming, will help remove the scrolling bottleneck. Moreover, if scrolling must be document-bounded, it is critical to support a user-controlled zoom pivot.

Minimize visual search. Viewport size matters less than the human perceptual bandwidth ceiling [17]. Make potential targets as easy to discern as possible – both identification of presence and confirmation of absence.

FUTURE WORK AND CONCLUSION

Perhaps the most impactful determinant to the efficiency of navigation strategy lies in the quality of a user’s decisions. We have presented a decision model that highlights two key decisions for zooming and scrolling navigation (Region A and Region C). Empirical evidence supports the intuition that better speed-accuracy trade-offs align with the optimal navigation strategy. However, results also indicate that users may not be aware of what “better” is. Our design guidelines and heuristics are intended to help designers to create interfaces which promote better decision making. Much of the work we have reviewed aims towards these goals, but has not yet been optimized to scale to the complexities and magnitude of larger datasets. This is a clear area for future research.

Our model is intended to provide a generalized approach to provide an understanding of zooming and scrolling across different contexts. The generality of the particular costing functions we have supplied is an open question; it is up to the user of our model to determine the costing functions for their context.

There is a performance trade-off in precision of scrolls in the zoom-in sequence. We found that users fail to optimize the speed/accuracy trade-off for this repetitive action. Given that performing slower, more precise and consistent corrective scrolls will improve performance, design of interaction techniques could be adjusted to promote this ‘better’ behavior.

In our work, we have shown the benefit of a systematic analysis of the components of a complex navigation strategy. It is our hope that the models described in this paper will promote further discussion and comparison of the impact of user decisions on navigation strategy. Navigation strategy is more than just the sum of its parts.

REFERENCES