# Two-Part Models Capture the Impact of Gain on Pointing Performance

GARTH SHOEMAKER, University of British Columbia
TAKAYUKI TSUKITANI, Osaka University
YOSHIFUMI KITAMURA, Tohoku University
KELLOGG S. BOOTH, University of British Columbia

We establish that two-part models of pointing performance (Welford's model) describe pointing on a computer display significantly better than traditional one-part models (Fitts's Law). We explore the space of pointing models and describe how independent contributions of movement amplitude and target width to pointing time can be captured in a parameter $k$. Through a reanalysis of data from related work we demonstrate that one-part formulations are fragile in describing pointing performance, and that this fragility is present for various devices and techniques. We show that this same data can be significantly better described using two-part models. Finally, we demonstrate through further analysis of previous work and new experimental data that $k$ increases linearly with gain. Our primary contribution is the demonstration that Fitts's Law is more limited in applicability than previously appreciated, and that more robust models, such as Welford's formulation, should be adopted in many cases of practical interest.

28

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Theory and methods*

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: Fitts, Welford, gain, interaction techniques, large displays, pointing

**ACM Reference Format:**
Shoemaker, G., Tsukitani, T., Kitamura, Y., and Booth, K. S. 2012. Two-part models capture the impact of gain on pointing performance. ACM Trans. Comput.-Hum. Interact. 19, 4, Article 28 (December 2012), 34 pages.
DOI = 10.1145/2395131.2395135 http://doi.acm.org/10.1145/2395131.2395135

## 1. INTRODUCTION

In this article we compare the use of one-part and two-part models of pointing performance in describing empirical data. We will consider the Fitts (Equation (1)) and Shannon (Equation (2)) one-part models, and a two-part model due to A. T. Welford (Equation (3)), as well as a new Shannon-Welford two-part model (Equation (4)) that we introduce. We will explore how the different pointing models perform at different levels of control/display gain, gain $G$ being the scaling factor between physical movement in control space and resulting motion in display space: $G = v_{display}/v_{control}$ [Casiez

Funding was provided by the Natural Sciences and Engineering Research Council of Canada through operating and equipment grants, scholarships, and the NECTAR strategic research network, by the GRAND Network of Centres of Excellence, and by Defence Research and Development Canada. Equipment and facilities were funded by the Canada Foundation for Innovation and by the Natural Sciences and Engineering Research Council of Canada.
Author's address: G. Shoemaker, Department of Computer Science, University of British Columbia, ICICS/CS Building, 201-2366 Main Mall, Vancouver, BC, V6T 1Z4, Canada; email: garths@cs.ubc.ca.

DOI 10.1145/2395131.2395135 http://doi.acm.org/10.1145/2395131.2395135

and Roussel 2011], where $v_{display}$ is the velocity of the cursor in display space and $v_{control}$ is the velocity of the hand, mouse, or other device in control space. Our emphasis will be on interaction techniques for very large displays where widely varying levels of gain are often required. We will demonstrate that two-part models can sometimes describe actual data significantly better than their one-part counterparts.

$$\text{Fitts (one-part)} \qquad MT = a + b \log_2\left(\frac{A}{W}\right) \tag{1}$$

$$\text{Shannon (one-part)} \qquad MT = a + b \log_2\left(\frac{A}{W} + 1\right) \tag{2}$$

$$\text{Welford (two-part)} \qquad MT = a + b_1 \log_2(A) - b_2 \log_2(W) \tag{3}$$

$$\text{Shannon-Welford (two-part)} \qquad MT = a + b_1 \log_2(A + W) - b_2 \log_2(W). \tag{4}$$

We first explore the space of models of pointing performance. The two axes we consider are degrees of freedom (one-part or two-part) and information-theoretic interpretation (Fitts or Shannon). We discuss our four candidate formulations (Equations (1)–(4)), and how they fit into the $2 \times 2$ matrix of pointing models. Using our different candidate formulations, we perform a reanalysis of data obtained from other researchers. This reanalysis serves to frame our own expectations and a new experiment. The reanalysis demonstrates that one-part models (Equations (1)–(2)) break down when describing pointing performance on some computing devices, and that the breakdown depends on the control/display gain. Further reanalysis demonstrates that two-part formulations (Equations (3)–(4)) can correct for the limitations of the one-part formulations. To conclude the reanalysis, we illustrate how limitations in experimental design and analysis can conceal shortcomings of one-part models. Motivated by the reanalysis, we then present results from a new experiment that investigated distance pointing on a very large wall display. Our new experimental results along with the results of the reanalysis indicate that two-part formulations are able to model pointing performance significantly better than do their corresponding one-part formulations, and that the relative contributions of movement amplitude $A$ and target width $W$ to movement time do indeed vary with gain.

The Fitts's Law model of pointing performance [Fitts 1954] has proven to be extremely robust. As Pratt et al. [2007] point out, it has been applied to physical pointing underwater [Kerr 1973], in near-zero gravity [Fowler et al. 2008], with microscopic targets [Langolf et al. 1976], and even pointing with one's feet [Hoffman 1991]. Fitts's Law has found a home in HCI, with a nearly thirty-year legacy [Soukoreff and MacKenzie 2004]. When evaluating new pointing devices or new display form factors one of the first research tasks is often to perform a Fitts's Law evaluation in order to produce a predictive model of performance and determine throughput for the particular approach. One important factor influencing pointing performance that is unique to computer interaction is control/display gain. The parameters of a Fitts pointing model are often found to vary depending on the gain level used [Arnaut and Greenstein 1990; Casiez et al. 2008; MacKenzie and Riddersma 1994; Sandfeld and Jensen 2005], in addition to other factors such as the pointing device used.

However, there are exceptions to Fitts's Law. Not long after Fitts's original paper, Welford [1971] observed that some Fitts-like tasks produce data that does not follow Fitts's model. Welford proposed an alternate two-part model to accommodate this deviation from expected performance. The new formulation allows for independent contributions of movement amplitude $A$ and target width $W$ to movement time. Rather than only considering the ratio of $A$ and $W$ captured in Fitts's "index of difficulty" as the sole independent variable, there are instead two independent variables, A and W, that determine movement time. We discuss this in detail in Section 3. In their own work identifying limitations in Fitts's Law, Pratt et al. [2007] conclude that "it now seems unlikely that a single equation will be able to accurately capture all aspects of speed-accuracy trade-offs." It is evident to us that Fitts's Law, while widely useful and validated, should not be taken as gospel. Alternate explanations should be considered when appropriate.

Very large displays, where we define "very large" to mean significantly greater than a user's physical arm span, support a class of interactions beyond those thoroughly explored in the Fitts's Law literature. These interactions can be performed in midair, using gestures or pointing, or on a surface, using a traditional device such as a mouse. Regardless of the input method, it is often necessary for the system to support control/display gain levels much higher than in smaller display systems. These gain levels are necessary in order for a user to be able to interact efficiently over the entirety of the display surface. Many researchers have found that gain can be an influencing factor on pointing performance. We thus felt it important to explore the appropriateness of different pointing models for these new display form factors over their larger ranges of relevant control/display gain values. Accurate pointing models can help us determine which devices are well-suited for use. They can also guide designers in the implementation of specific interaction techniques and inform the positioning and sizing of on-screen display elements.

We summarize here the contributions of this article in point form.

(1) We establish a matrix with two axes that defines a space of models of pointing performance. The axes are degrees of freedom (one-part vs. two-part) and information-theoretic interpretation (Fitts vs. Shannon).
(2) We introduce a new two-part formulation, $MT = a + b_1 \log_2(A + W) - b_2 \log_2(W)$ that takes inspiration from Shannon information theory, but also separates the contributions of $A$ and $W$ to movement time as with Welford's formulation.
(3) We establish the equivalence of Welford's two-part formulation to a linear version of a formulation due to Kopper, and demonstrate the importance of a constant $k$ in characterizing the relative contributions of movement amplitude $A$ and target width $W$ to movement time.
(4) In a reanalysis of data from other researchers, we show one-part formulations (i.e., Fitts and Shannon formulations) to be inadequate in modeling some pointing tasks.
(5) In the same reanalysis, we demonstrate that two-part formulations (i.e., Welford and Shannon-Welford) significantly outperform their one-part counterparts in describing some of the data, especially for mouse pointing on large displays. We also determine empirically that the constant $k$ appears to vary linearly with gain.
(6) We reveal shortcomings in how Fitts's Law has historically been applied in the HCI community, and provide recommendations on how to design and analyze the results of a pointing evaluation.
(7) In a new experiment, we investigate physical pointing on a large display. Two-part models are shown to significantly outperform their one-part counterparts, and $k$ is found to vary linearly with gain. These results are consistent with our reanalyses of data from other researchers.

The next section of this article reviews the relevant literature. The four formulations are explored in Section 3. Section 4 presents the findings from our reanalysis of previous studies and Section 5 describes a new study designed to test hypotheses that arose during the reanalysis. We present conclusions and suggestions for future research in Section 6.

Throughout this article we adopt the recommendation of Strunk and White [1979] concerning the possessive form of the proper singular noun “Fitts” when referring to work by Paul M. Fitts. Some readers may prefer nonstandard alternatives.

## 2. RELATED WORK

Relevant related work lies in the areas of large display research and Fitts’s Law pointing research. We review both of these in this section.

### 2.1 Large Display Research

The work described in this article was largely driven by our interest in very large display interaction. While large displays are increasingly the topic of research, and are beginning to be widely deployed, work in understanding pointing performance with these new form factors is ongoing. Interaction with large displays is sometimes characterized by widely varying levels of control/display gain, in order to support both accurate localized input and the ability to interact over large distances. It also frequently involves input from a distance, meaning that a user is located out of physical reach of the display. These conditions, which do not exist for most traditional computing interaction scenarios, may result in pointing performance that deviates from established empirical models.

Two categories of large display research are relevant. The first investigates the nature of large physical surfaces and how they motivate the need for, and drive the design of, large interactive surfaces. The second investigates specific interactive surface systems and techniques.

The deployment of large interactive displays is driven by the observation that large physical surfaces are themselves useful. Investigations of how large physical surfaces are used in the real world often inform the design of interactive systems [Tang et al. 2009]. Related to the importance of large physical surfaces, it has been argued that the widespread deployment of wall-sized blackboards in the 19th century, replacing smaller individual slates, was a critical advancement in classroom technology [Buxton 2008]. Other large surfaces, such as desks, also play important roles in everyday life. The organization of information on desks by individuals has been examined closely [Malone 1983], and it has been shown that spatial organization on large surfaces is important in supporting collaborative processes [Scott et al. 2004]. The properties of physical surfaces that lend themselves to certain tasks are also possessed by large interactive surfaces. Rogers and Lindley [2004] examined the advantages and disadvantages of both table and wall surfaces, and found each of them to be suitable for different tasks. They found that large wall displays support collaboration of dynamically changing groups and giving presentations to audiences.

Large electronic surfaces offer benefits over physical surfaces because of their additional computational and interactive capabilities. Examples of research systems that center around the use of large display surfaces include Flatland [Mynatt et al. 1999], the Responsive Workbench [Agrawala et al. 1997], and Dynamo [Izadi et al. 2003]. There is also a large collection of specialized interaction techniques, such as Drag-and-Pop [Baudisch et al. 2003], Frisbee [Khan et al. 2004], text input techniques [Shoemaker et al. 2009], the Vacuum [Bezerianos and Balakrishnan 2005], and body-centric interaction [Shoemaker et al. 2007, 2010], that address issues particular to large displays, and can potentially be integrated into a variety of systems.

In general, large displays have been found to be very useful in supporting specific classes of tasks, such as public sharing of media [Brignull et al. 2004], group awareness management [Huang and Mynatt 2003], and brainstorming [Cherubini et al. 2007].

### 2.2 Fitts's Law for Computer Pointing

Fitts's Law [Fitts 1954] was originally developed as a tool for modeling the performance of human physical pointing tasks, such as reciprocal tapping. It was later applied to pointing on a computer display by Card et al. [1978]. The empirically determined parameters of a Fitts model for computer pointing depend on the device used. Researchers have performed evaluations of numerous approaches, including traditional mouse, pad, and trackball devices [Epps 1986], stylus input [Forlines and Balakrishnan 2008], direct touch on tables [Forlines et al. 2007], and pointing with a laser pointer [Myers et al. 2002]. Researchers have also extended Fitts's Law to special cases. Variations have been developed for 2-D pointing [MacKenzie and Buxton 1992], 3-D pointing [Grossman and Balakrishnan 2004], pointing to expanding targets [McGuffin and Balakrishnan 2005], and pointing with snapping [Fernquist et al. 2011], among others. Researchers have even investigated such subtle points as the impact of cursor orientation on performance [Po et al. 2005], and the independence of throughput from the speed/accuracy tradeoff [MacKenzie and Isokoski 2008].

There are still open questions, however, regarding the limits of Fitts's Law in real world application. Pratt et al. [2007] discovered that allocentric information can modulate pointing performance. They found that when an array of targets is visible, the time taken to point to the last target in the array is shorter than predicted by Fitts's Law, and that pointing to the last target can be faster than pointing to closer targets in the array. This suggests that there is more to pointing than the low-level motor movement modeled by Fitts's Law. Furthermore, Keulen et al. [2002] identified multiple reference frames used by the brain in reaching tasks, where very different mental processes are used for pointing at a distance as compared to up close. This might suggest that different pointing models could be necessary to describe pointing at a distance, which is very relevant to large displays, versus pointing within physical reach, as is common on smaller displays.

Researchers regularly propose variations to Fitts's Law, or even entirely different models of pointing performance. Examples include an extension to Fitts's Law that incorporates classic control theory [Radix et al. 1999], an incorporation of gain into the Fitts's model for computer pointing [Johnsgard 1994], and a Fitts variant incorporating a tremor constant for modeling pointing to very small targets [Chapuis and Dragicevic 2011]. Older models, including those by Woodworth [1899] and Hollingworth [1909], as well as more recent models, such as by Schmidt et al. [1979], have also been proposed, but have largely been overshadowed by Fitts's Law.

## 3. ONE-PART AND TWO-PART MODELS OF POINTING PERFORMANCE

Many variants of Fitts's Law have been proposed in the literature. In this section we discuss the four formulations (Equations (1)–(4)) that we will focus on in this article, and delve into details relevant to our work.

We define one-part formulations (Equations (1)–(2)) to be those where the model depends only on the ratio of movement amplitude ($A$) and target width ($W$), but not their absolute values.[1] The best-known one-part formulation is Fitts's Law[2] (Equation (1)), which defines movement time as depending on $A$ and $W$, as well as two experimentally determined constants, $a$ and $b$. As with all of the one-part models we consider, in the Fitts formulation it is the dimensionless ratio of $A$ and $W$ that matters; the individual values of $A$ and $W$ are not in isolation important.

[1]One-part models have two degrees of freedom, because they also incorporate an additive constant in addition to a multiplicative constant.

[2]An alternate form of Fitts's Law is $MT = a + b \log_2(2A/W)$, which is functionally equivalent to Equation (1), except that applying it will result in a different $a$ constant.

Table I.

A $2 \times 2$ matrix of the pointing models considered in this article. The axes are the number of parts (one or two) in a formulation, and the information-theoretic interpretation (Fitts or Shannon). Equations in the same cell are mathematically equivalent. When $k$ is used, it is defined as $k = b_2/b_1$, as described in Section 3.2.

| | One-Part | Two-Part |
|---|---|---|
| Fitts | $a + b\,\log_2\left(\frac{A}{W}\right)$ (Eq. (1)) | $a + b_1\log_2(A) - b_2\log_2(W)$ (Eq. (3))<br>$a + b\,\log_2\left(\frac{A}{W^k}\right)$ |
| Shannon | $a + b\,\log_2\left(\frac{A}{W} + 1\right)$ (Eq. (2))<br>$a + b\,\log_2\left(\frac{A+W}{W}\right)$ | $a + b_1\log_2(A + W) - b_2\log_2(W)$ (Eq. (4))<br>$a + b\,\log_2\left(\frac{A+W}{W^k}\right)$ |

Soukoreff and MacKenzie [2004] have promoted the use of a one-part formulation (Equation (2)) based on an alternate interpretation of the information-theoretic roots of Fitts's Law. They argue that the additive constant of one within the log term of $ID$ is a more faithful transfer of Shannon's information theory, although Drewes [2010] has, perhaps controversially, questioned this claim. Regardless, when modeling real-world data, the Shannon formulation often produces better $R^2$ coefficient of determination values than does the Fitts formulation.

Debating the relative virtues of the Fitts (Equation (1)) and Shannon (Equation (2)) models is beyond the scope of this article. In our analysis we will make use of both formulations, because these are the two most commonly used one-part formulations found in the literature.

We define two-part formulations (Equations (3)–(4)) to be models that allow for separable contributions of $A$ and $W$ to movement time.[3] By separable, we mean that the individual values of $A$ and $W$ are of interest, rather than just their ratio $A/W$, and thus the contributions of $A$ and $W$ to movement time can be weighted with their own constants. Welford introduced his two-part[4] formulation (Equation (3)) to account for deviations from Fitts's Law that he observed in data collected from an improved version of Fitts's original experiments [Welford 1971, Figure 5.8 on page 158 of the "reprinted with corrections" edition of his 1968 book]. Welford's two-part formulation has been largely overlooked in the HCI literature, but there is work, including that by Wobbrock et al. [2008], that suggests a need for considering $A$ and $W$ independently in modeling performance. We think Welford's model is potentially powerful in its generality, especially when applied to interaction with large displays.

We introduce in this article a variation on Welford's two-part model, which we will refer to as the Shannon-Welford formulation. This formulation (Equation (4)) combines aspects of the one-part Shannon (Equation (2)) and the two-part Welford (Equation (3)) formulations. The mapping of noise and signal from information-theoretic origins to movement amplitude and target width are the same as with the Shannon formulation, but signal and noise are broken into independent terms with their own constants as with the Welford formulation. Along with existing formulations from the literature, this new formulation completes the space of models covering the axes of degrees of freedom (one-part vs. two-part), and information-theoretic interpretation (Fitts vs. Shannon), as we will see later in our summary of models and as outlined in Table I.

[3] Two-part models have three degrees of freedom. Two are multiplicative constants related to $A$ and $W$, and the third is captured in the additive constant.

[4] Welford also has a one-part formulation, $MT = a + b\,\log_2(A/W + 0.5)$, not to be confused with his two-part formulation. We do not consider his one-part formulation, as it is not as widely used as the other Fitts and Shannon one-part formulations.

### 3.1 The Validity of Two-Part Models in a Fitts Context

There is some disagreement within the research community regarding the validity of two-part models in the Fitts's Law context. The issue is whether or not Fitts's $ID$ (index of difficulty) can be considered in comparison to its components of $A$ and $W$. Mathematically, $ID$ is clearly calculated using $A$ and $W$, but some argue that once $ID$ has been decomposed into $A$ and $W$, we have an alternative to Fitts's Law, rather than a more general extension of it. Referring to the parameters $ID$, $A$, and $W$, Guiard [2009] states that "... it is logically unsound to inquire into three separate causal influences in a Fitts's Law experiment." He believes that $A$, $W$, and $A/W$ form a trio with only two degrees of freedom,[5] and therefore it does not make sense to investigate three different causal influences as if they were independent.

In his paper, Guiard discusses both a one-part interpretation, where $A$ and $W$ represent "just the handles we need to control the ratio," and a two-part interpretation, where $A$ and $W$ are independent factors and movement time "depends on both factors." Guiard sees this as an "inescapable choice between two mutually exclusive theoretical views of Fitts'[s] [L]aw, which disagree about something quite crucial – the number and the identity of the independent variables." We will return later to this when we consider some of the other observations made by Guiard.

Welford argues directly that separating $ID$ into $A$ and $W$ is a valid approach, and that doing so results in a model that is related to the Fitts's model, but with an extra degree of freedom. In his book, on page 156, Welford states that "two control processes or phases ought perhaps to be distinguished: a faster distance-covering phase and a slower phase of 'homing' on to the target" [1971]. We agree with Welford's stance, and our comparisons of one- and two-part models are inspired by his work. Ultimately, it may require some time before a resolution is reached between Guiard's and Welford's epistemological differences. Our investigation is more pragmatic and focuses mostly on the ability of two-part models to explain the effect of gain, something that neither Welford nor Guiard [2009] explicitly address.

### 3.2 Pointing at a Distance and the $k$ Value

Kopper et al. [2010] examined distance pointing with a laser pointer on large displays. They validated a two-part model of performance based on angular measurements $\alpha$ for movement amplitude and $\omega$ for target width (Equation (5)). We examine their model here, and relate it to existing two-part models.

$$MT = a + b\,\log_2\left(\frac{\alpha}{\omega^k} + 1\right). \tag{5}$$

Kopper et al.'s use of angular measurements is consistent with their technique, where rotation of the input device, rather than translation, results in cursor motion. The exponent $k$ they introduce allows $\alpha$ and $\omega$ to have separate degrees of impact on movement time. We call this exponent the "Kopper $k$." A formulation analogous to the Kopper angular formulation can be constructed using linear units (Equation (6)).

$$MT = a + b\,\log_2\left(\frac{A}{W^k} + 1\right). \tag{6}$$

In this version of the Kopper formulation, linear amplitude $A$ replaces angular $\alpha$, and linear width $W$ replaces angular $\omega$. A second formulation is also possible

[5]There are two senses in which the term "degrees of freedom" is used. Guiard uses it here in the sense of number of input variables that can be independently adjusted. The other sense, which we use, is the statistical notion of the number of parameters in a mathematical model that maps input variables to output values.

(Equation (7)), where we omit the Shannon-like "plus one" term. This makes it similar in form to Fitts's model, and allows us to compare the two directly.

$$MT = a + b\ \log_2\left(\frac{A}{W^k}\right). \tag{7}$$

An important observation, not noted by Kopper et al., is that the Fitts-inspired linear analog of their formulation (Equation (7)) can be derived directly from Welford's two-part formulation (Equation (3)), meaning that the two formulations are mathematically equivalent. The $k$ and $b$ values in the linear version of Kopper's formulation are equal to $b_2/b_1$ and $b_1$ from Welford's formulation, respectively. We therefore have a second means of expressing Welford's two-part formulation, with a different choice of constants. Algebraic equivalence of the two expressions is demonstrated by the straightforward derivation in Equation (8).

$$\begin{aligned} MT &= a + b_1 \log_2(A) - b_2 \log_2(W) \\ &= a + b_1 \left[ log_2(A) - \frac{b_2}{b_1} log_2(W) \right] \\ &= a + b_1 \left[ log_2(A) - log_2(W^{b_2/b_1}) \right] \\ &= a + b_1 log_2 \left( \frac{A}{W^{b_2/b_1}} \right) \\ &= a + b_1 log_2 \left( \frac{A}{W^k} \right). \end{aligned} \tag{8}$$

The significance of the linear (nonangular) Kopper formulation to our work lies in the exponent $k$. The exponent $k$ is a single parameter that conveniently encapsulates the relative magnitudes of the separable contributions of the independent variables $A$ and $W$ to the overall movement time. If experimental results determine that $k = 1$ then $b_1 = b_2$ and the model is simply Fitts's Law (without the factor of 2 multiplying $A$ that Fitts originally had, an inconsequential detail for our current discussion), and Fitts's Law will model the experimental data as well as does the two-part model. However, the more that $k$ deviates from unity, as determined by the experimental data, the worse Fitts's Law will perform at modeling results. A $k$ value greater than one ($b_2 > b_1$) indicates that the impact of target width is overshadowing the impact of movement amplitude, whereas a $k$ value less than one ($b_1 > b_2$) suggests the opposite. Thus $k$ is useful not only for gauging the relative contributions of $A$ and $W$, but is also a good indicator of the applicability of Fitts's formulation. We adopt the use of $k$ for much of the remaining discussion in this article to illuminate the separable contributions of $A$ and $W$.

It is important to note that we are not introducing a new parameter or degree of freedom with our use of $k$, merely an equivalent reformulation that replaces the Welford parameters $b_1$ and $b_2$ with the parameters $b$ and $k$.

### 3.3 Unit Dependence in Two-Part Models of Pointing

The empirically determined constants of a one-part model are independent of the choice of units (e.g., cm or mm) used in measurement. This is not the case using a two-part model such as Welford's. For example, in Equation (9) we see that the one-part Fitts formulation is not impacted by changing units used in measurement (represented by scaling $A$ and $W$ by some multiplicative constant $s$), because the scaling constants simply cancel out. The empirically derived constants $a$ and $b$ will therefore be the same regardless of units used.

$$\begin{aligned} MT &= a + b\log_2\left(\frac{sA}{sW}\right) \\ &= a + b\log_2\left(\frac{A}{W}\right). \end{aligned} \tag{9}$$

In two-part formulations, however, the units chosen will impact the derived constants. When using Welford's two-part formulation, a multiplicative constant $s$ applied to measurements will result in a change in the empirically determined additive constant $a$, as demonstrated in the derivation in Equation (10).

$$\begin{aligned} MT &= a + b_1\log_2(sA) - b_2\log_2(sW) \\ &= a + b_1(\log_2(A) + \log_2(s)) - b_2(\log_2(W) - \log_2(s)) \\ &= (a + b_1\log_2(s) - b_2\log_2(s)) + b_1\log_2(A) - b_2\log_2(W) \\ &= a' + b_1\log_2(A) - b_2\log_2(W). \end{aligned} \tag{10}$$

A multiplicative scale factor $s$ to adjust for units results in a new different constant, $a' = a + b_1\log_2 s - b_2\log_2 s$. Importantly, however, the $b_1$ and $b_2$ constants, and therefore the $k$ value, are all independent of the units chosen and thus do not depend in any way on the scale factor $s$. Thus the relative magnitude of the separable effects of $A$ and $W$ as captured in $k$ is independent of the units chosen to represent movement amplitude and target width.

This explanation glosses over an important point. Both $A$ and $W$ possess units of distance, yet appear within logarithms. This is not consistent with their intended use as dimensionless values. It is desirable to treat $\log_2 A$ and $\log_2 W$ similarly to the index of difficulty ($ID$) in Fitts's formulation, where $ID = \log_2(A/W)$ is dimensionless. Fortunately, as Graham [1996] explains, Welford anticipated this objection and postulated nominal values $A_0$ and $W_0$ that "normalize" $A$ and $W$ respectively and eliminate the problem of logarithmic dimensions by turning all of the arguments for the logarithm functions into dimensionless values (pure numbers).

$$MT = a + b_1\log_2\left(\frac{A}{A_0}\right) - b_2\log_2\left(\frac{W}{W_0}\right). \tag{11}$$

We will assume that suitable constants $A_0$ and $W_0$ are used as in Equation (11), but will simply write Equation (3) with the understanding that suitable normalization has taken place and that only the intercept value $a$ might be affected by the choice of normalization. This is consistent with much of the literature.

### 3.4 Comparing Models Statistically

Throughout this article we compare one-part and two-part models with the goal of determining which are more appropriate for application to certain data sets. Soukoreff and MacKenzie [2004] observed that two-part models will inevitably perform better (as measured by an $R^2$ coefficient of determination) than their one-part equivalents, because the two-part models incorporate an extra degree of freedom. It would be an error to select a two-part model simply because it produces a better fit than a corresponding one-part model. It is necessary, instead, to determine whether the two-part model genuinely captures aspects of the data that a one-part model cannot, rather than just providing a better fit due simply to the extra degree of freedom.

Several tests are available for comparing linear regression models, including the F-test [Draper and Smith 1998], the Akaike Information Criterion (AIC) [Akaike

1974], and the Schwarz Criterion (also known as the BIC) [Schwarz 1978]. The F-test provides a means of testing the hypothesis that a model with more degrees of freedom describes the data significantly better than a model with fewer degrees of freedom. While the AIC and BIC tests are suitable tools for model *selection* they are not suitable for *hypothesis testing*. That is, they will not determine if a higher degree-of-freedom model is significantly better at modeling the data compared to a lower degree-of-freedom model. Because of this, the F-test is the only test appropriate for our purposes.

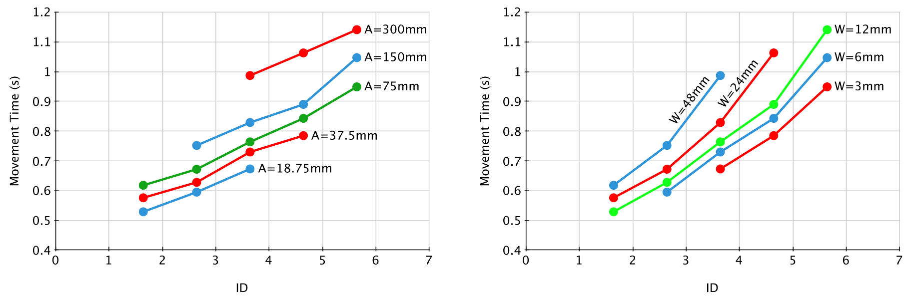
One limitation of the F-test is that it can only compare models that are nested. Models are nested if the greater degree-of-freedom model can be made equivalent to the lesser degree-of-freedom model by enforcing an additional constraint. For example, the Shannon-Welford model is equivalent to the Shannon model if we enforce the constraint that $b_1 = b_2$. Similarly, the Welford and Fitts formulations are nested using the same constraint. Thus, in our analyses we group our comparisons into Fitts-inspired models (Fitts vs. Welford) and Shannon-inspired models (Shannon vs. Shannon-Welford). The F-test that we use is $\mathrm{F}(p_2 - p_1, n - p_2)$, where $p_2$ and $p_1$ are the number of parameters in the models ($p_1$ is the number of parameters in the simpler nested model) and $n$ is the number of sample points. A $p$-value of .05 or less is considered significant when comparing the relative explanatory power of two nested models with the F-test. We are comparing two-part models with three degrees of freedom against nested one-part models with two degrees of freedom, so $p_2 = 3$ and $p_1 = 2$.

### 3.5 Summary of Pointing Models

The four models of pointing performance we focus on in this article are summarized in Table I using a $2 \times 2$ matrix. The first axis of the matrix, degrees of freedom, separates models into one-part and two-part formulations. In this article we will arrive at the conclusion that two-part models of pointing performance are sometimes necessary for modeling computer pointing. The second axis, that of information-theoretic interpretation, separates models by how the information-theoretic concepts of "signal" and "noise" have been mapped to the physical quantities of movement amplitude ($A$) and target width ($W$). Past work has argued the merits of both the Fitts and Shannon interpretations, and it has been shown that different interpretations can result in small, but sometimes important, differences in model fit ($R^2$ values). We thus consider both interpretations for completeness.

## 4. A REANALYSIS OF POINTING EXPERIMENTS

One-part models of pointing based on Fitts's formulation have garnered much more attention in the HCI literature than have two-part models such as those by Welford and Kopper et al. This may be because one-part models appear to serve us well and they are simpler and more concise than two-part models. One might well ask, is there a need for two-part models of pointing performance? In order to help answer this question we undertook a reanalysis of published experimental data. We looked at aspects of the data from three previous studies that were left unexplored by the authors of the papers we examined. We found that one-part models possess previously unappreciated limitations, and that two-part models of pointing performance frequently perform significantly better than do their one-part counterparts. We extended the reanalysis by reviewing a set of 19 papers to see whether the lack of appreciation in the community for two-part models might be due to the experimental designs and analyses that were employed.




(a) Lines connect points representing tasks with the same movement amplitude $A$.

(b) Lines connect points representing tasks with the same target width $W$.

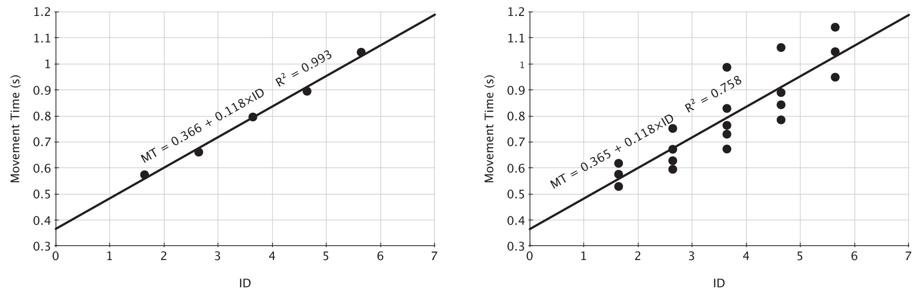
Fig. 1. Movement time results for all $A$ and $W$ combinations from the Graham data at $G = 1$. Data points are the same in both graphs. Movement times with identical $ID$ can differ, depending on $A$ and $W$. This is in violation of Fitts's Law. Graphs adapted from Graham [1996].

### 4.1 Physical Pointing on a Small Display

In his doctoral dissertation Graham described a series of experiments dealing with indirect (virtual) control of an on-display pointer using a motion-tracked finger [Graham 1996]. A subset of his results have been reported in the peer-reviewed literature [Graham and MacKenzie 1996]. His first experiment investigated pointing at three levels of control/display gain on a traditional-sized display. In analyzing his results for the lowest gain level ($G = 1$), Graham found that Fitts's formulation did not accurately model pointing because of a separable effect of $A$ and $W$ impacting performance. A visualization of Graham's results (Figure 1) reveals a pattern very similar to that shown by Welford in Figure 5.8 on page 158 of his 1971 book. Movement time increases roughly linearly with $ID$ within either a single $A$ or $W$ value, but not across changes in both $A$ and $W$ values. Graham found that Welford's two-part formulation (Equation (3)) was necessary to account for this pattern, and to accurately model movement time. The validity of applying this two-part formulation was further supported by Graham's analysis of hand velocity and acceleration profiles during pointing, which again revealed separable effects of $A$ and $W$ in different temporal segments of the motion.

The fact that both Welford and Graham found a two-part formulation to be necessary for accurately modeling performance of their respective real-world and virtual pointing tasks is intriguing. But what does this mean for other forms of computer pointing? Fitts's formulation has repeatedly been found to accurately model pointing on a computer. It would be surprising to find limitations to Fitts's Law in a computer pointing context.

Again, Graham has already provided part of the answer in his dissertation [Graham 1996]. He makes note of several facts that leave open the possibility of discovering situations where Fitts's Law may not suffice, and where two-part models may instead be appropriate. First, he notes that even for cases where there is a separable effect of $A$ and $W$, the Fitts formulation may hold for certain subsets of the data. As an example, if Graham had run his experiment either with a single $A$ and multiple $W$ values, or with a single $W$ and multiple $A$ values, Fitts's Law would have appeared to hold. This is evident from Figure 1. Second, the treatment of data during analysis can compromise one's ability to recognize a separability of $A$ and $W$. As an example, Figure 2(a) shows a regression analysis of data points from the Graham experiment that are averaged within $ID$ values, whereas Figure 2(b) shows a regression analysis of the same data, but where data points are not averaged within $ID$ values but are




(a) A regression analysis using average MT results for every $ID$ value. A very good fit of data to model is due to the averaging of data points, concealing the underlying nature of the data.

(b) A regression analysis using MT results of every combination of $A$ and $W$. The poor fit of data to model is revealed by including all data points in the analysis.

Fig. 2. An example of a possible pitfall in analyzing Fitts's Law experimental results. A regression analysis of movement times for all $A$ and $W$ combinations from the Graham data ($G$ = 1) make evident a poor fit, whereas analyzing results averaged within $ID$ values conceals the poor fit. Graphs adapted from [Graham 1996].

averaged within distinct pairs of $A$ and $W$ values. In this case averaging data points within $ID$ values conceals the poor fit of the data to a linear function, a danger that should be avoided.

The importance of careful experimental design and analysis is backed up by arguments made by Guiard [2009], who notes that standard Fitts's Law experimental designs tend to contain a confound related to the concomitant variation of $A$ and $W$ with $ID$. Including all individual pairs of $A$ and $W$ in an analysis addresses the problems identified by Graham, although Guiard offers another approach based on the notions of *form* and *scale* as alternative characterizations for the degrees of freedom inherent in pointing tasks. Unfortunately, neither approach is standard practice in HCI.

It is interesting that Guiard shows data from a disc transfer study by Fitts [Guiard 2009, Figure 3] that is very similar to our Figure 1 from Graham's data. In both cases lines of equal $A$ are roughly parallel to each other but have different slopes than the lines of equal $W$ that are themselves roughly parallel to each other (only the lines of equal $W$ are shown in the figure in Guiard's paper).

A conclusion that can be reached from Graham's discussion of pitfalls is that past experiments that found Fitts's Law to correctly model pointing may have been masking separable effects because of limitations in either design or analysis.

Graham unfortunately only analyzed the applicability of Fitts's Law at a single level of gain. We further analyzed the data from Graham's experiment.[6] Results of our linear-regression analysis using one- and two-part Fitts-inspired models are summarized in Table II. Results using one- and two-part Shannon-inspired models are summarized in Table III. The $k$ values ($k = b_2/b_1$) for the two-part models are shown for each gain level, as are the F-test results testing the hypothesis that a two-part model describes the data better than a one-part model.

There is no universally accepted threshold for a good $R^2$ value, but MacKenzie's [1992] suggestion of $R^2$ = 0.9 as a guideline when evaluating Fitts's Law results has been used in the literature, and therefore we also employ it. By this criterion, the two-part models produce a good fit to the data at all three levels of gain, whereas the one-part models only produce a good fit at the highest level of gain. The F-tests

[6]We thank Evan Graham for providing us with access to his original data and also to the team of technical staff who were able to retrieve the data from long-obsolete backup media.

Table II.

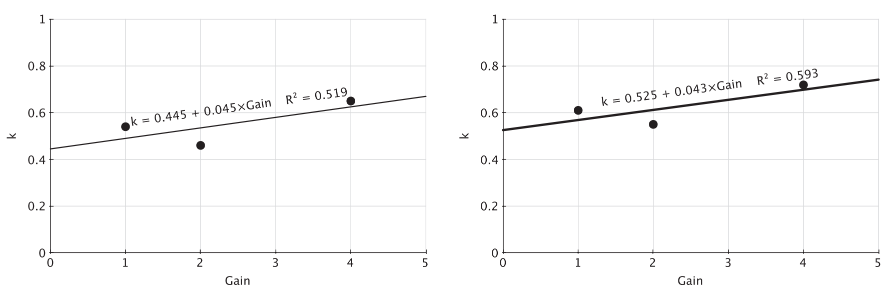
Modeling of movement time from the Graham data using the Fitts formulation and the Welford two-part formulation. For the Fitts formulation $R^2$ decreases with lower gain. For the Welford formulation $R^2$ is consistently good. An F-test indicates that the two-part model (Welford) describes the data significantly better than the one-part model (Fitts) at all levels of gain. Data is from Graham [1996].

| | Fitts | | | Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | $a$ | $b$ | $R^2$ | $a$ | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | sig? |
| 1 | 0.365 | 0.117 | 0.759 | 0.134 | 0.153 | 0.082 | 0.536 | 0.967 | 99.40 | <0.001 | yes |
| 2 | 0.346 | 0.111 | 0.789 | 0.018 | 0.151 | 0.070 | 0.464 | 0.981 | 108.66 | <0.001 | yes |
| 4 | 0.349 | 0.112 | 0.952 | 0.189 | 0.136 | 0.090 | 0.662 | 0.992 | 28.88 | 0.002 | yes |

Table III.

Modeling of movement time from the Graham data using the Shannon formulation and the Shannon-Welford formulation. For the Shannon formulation $R^2$ decreases with lower gain. For the Shannon-Welford formulation $R^2$ is consistently good. An F-test indicates that the two-part model (Shannon-Welford) describes the data significantly better than the one-part model (Shannon) at all levels of gain.

| | Shannon | | | Shannon-Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | $a$ | $b$ | $R^2$ | $a$ | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | sig? |
| 1 | 0.384 | 0.139 | 0.761 | 0.006 | 0.182 | 0.111 | 0.610 | 0.969 | 106.10 | <0.001 | yes |
| 2 | 0.364 | 0.132 | 0.793 | 0.107 | 0.180 | 0.099 | 0.550 | 0.985 | 136.24 | <0.001 | yes |
| 4 | 0.372 | 0.133 | 0.946 | 0.090 | 0.160 | 0.115 | 0.719 | 0.983 | 13.41 | 0.011 | yes |




(a) Using the Welford formulation.

(b) Using the Shannon-Welford formulation.

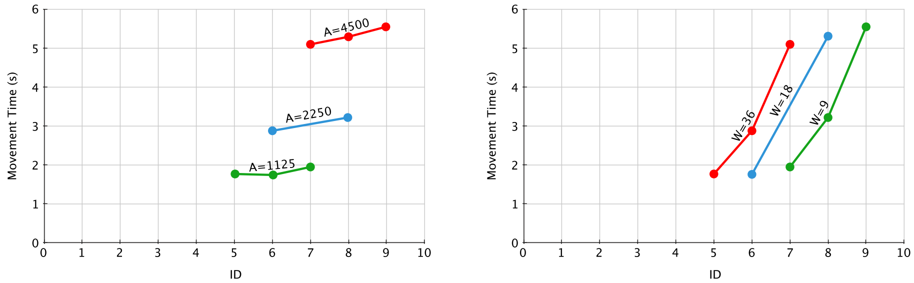
Fig. 3. The $k$ values for the Graham data as they vary with gain.

indicate that the two-part formulations model the data significantly better at all levels of gain compared to the corresponding one-part formulations, suggesting that two-part models capture aspects of the data that one-part models do not.

The relationship of $k$ to $G$ in the Graham data is visualized in Figures 3(a) and 3(b). No clear pattern is evident. A linear regression of the data points[7] for $k$ produced poor $R^2$ values of 0.519 and 0.593. It is unsurprising that no clear trend emerged, however. Just three gain values were investigated, limiting the number of data points, and only a small number of participants (six) were used in the experiment. We suspect that a more comprehensive experiment that includes a larger number and wider range of gain values might very well reveal a pattern of $k$ depending on $G$.

The reanalysis of Graham's data leaves some open questions. Most importantly, Graham's results relate to indirect (virtual) finger pointing mapped to an on-screen cursor. However, the mouse is a much more common input mechanism for computers. It is thus desirable to investigate whether we can identify a separable contribution of

[7]Performing a linear regression on three data points is questionable practice. We do so to be consistent with more robust analyses appearing later in the article.




(a) Lines connect points representing tasks with the same movement amplitude $A$.

(b) Lines connect points representing tasks with the same target width $W$.

Fig. 4. Movement time results for all $A$ and $W$ combinations for the Casiez data at $G$ = 2. Data points are the same in both graphs. This is similar to the pattern in Figure 1.

$A$ and $W$ to pointing time, similar to what Graham found, in a traditional mouse input scenario.

### 4.2 Mouse Pointing on a Large Display

Casiez et al. [2008] recently performed a very thorough analysis of mouse pointing performance that clarifies the landscape of both small and large display pointing. They evaluated pointing performance for multiple C/D gain levels as well as for different levels of pointer acceleration, on both small and large displays. They examined a wide range of $ID$ values, using a variety of $A$ and $W$ pairs for each $ID$, and included every combination of $A$ and $W$ in their analysis. In their first experiment on small display interaction they analyzed each gain level separately using the Fitts one-part formulation, and found regression $R^2$ values ranging from 0.956 to 0.984, indicating a close match between model and measurements. However, the results from their second experiment, performed on a large display, found regression $R^2$ values ranging from 0.577 to 0.959, with a roughly direct relationship between gain and $R^2$. Casiez et al. argue this relationship was due to excessive mouse clutching at low gain levels. It is possible that clutching is a contributing factor, however, based on what we learned from Kopper and Graham, we wanted to see if we could accurately model the data using a two-part formulation. Towards this goal we performed a further analysis of the large display Casiez data.[8]

We first produced two visualizations of movement times for individual $A$ and $W$ pairs for the $G$ = 2 condition, which was the condition with the worst $R^2$ in the original analysis. Figure 4 shows that when lines are drawn connecting data points with the same $A$ values, each line is approximately straight and parallel to the other lines, and when the points with the same $W$ values are connected, the lines are again approximately straight and parallel. It is clear that a single line will be unable to accurately model all data points. This is the same pattern found by both Welford and Graham in their data (and also by Guiard in Fitts's data, although Guiard did not draw explicit attention to the lines of equal $A$ values). This is evidence of a separable contribution of $A$ and $W$ to movement time. Welford and Graham both argued for a two-part formulation (Equation (3)) on this basis.

Results of a linear regression analysis of the Casiez data are shown in Table IV and Table V. Using MacKenzie's criteria of $R^2 > 0.9$ as a threshold for model fit, the one-part Shannon and Fitts models only produce good fits at gain levels 16 and 20.

[8]We thank Géry Casiez for providing us with access to his data.

Table IV.

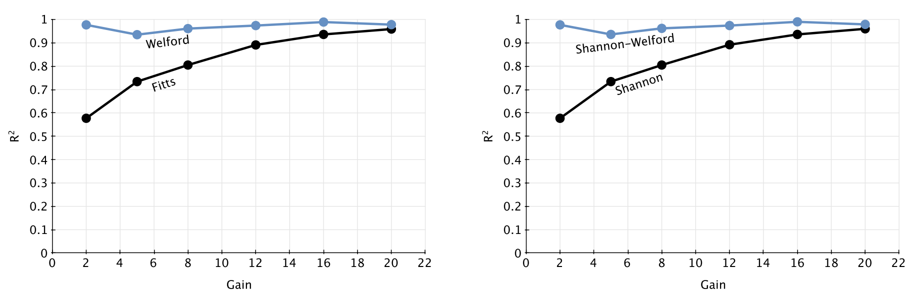
Modeling of movement time from Casiez et al. data using the Fitts and Welford formulations $R^2$ values for the Fitts formulation decrease at lower gain, but the Welford formulation is consistently good. An F-test of the models indicates that the Welford model is significantly better than the Fitts model at all levels of gain except 20. The data was provided by Casiez et al.

| | Fitts | | | Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | $a$ | $b$ | $R^2$ | $a$ | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | sig? |
| 2 | −4.125 | 0.950 | 0.577 | −15.286 | 1.742 | 0.159 | 0.091 | 0.977 | 85.79 | <0.001 | yes |
| 5 | −1.405 | 0.412 | 0.734 | −4.300 | 0.628 | 0.196 | 0.312 | 0.935 | 15.56 | 0.011 | yes |
| 8 | −0.841 | 0.308 | 0.805 | −2.608 | 0.444 | 0.172 | 0.387 | 0.961 | 19.64 | 0.007 | yes |
| 12 | −0.431 | 0.243 | 0.891 | −1.320 | 0.317 | 0.169 | 0.533 | 0.974 | 15.71 | 0.011 | yes |
| 16 | −0.393 | 0.232 | 0.936 | −1.009 | 0.287 | 0.177 | 0.632 | 0.989 | 22.50 | 0.005 | yes |
| 20 | −0.569 | 0.264 | 0.959 | −0.876 | 0.301 | 0.226 | 0.750 | 0.978 | 4.47 | 0.088 | no |

Table V.

Modeling of movement time from Casiez et al. data using the Shannon one-part and Shannon-Welford two-part formulations $R^2$ values for the Shannon formulation decrease at lower gain levels, but the Shannon-Welford formulation is consistently good. An F-test of the two models indicates that the Shannon-Welford model is significantly better than the Shannon model at all levels of gain except 20. The data was provided by Casiez et al.

| | Shannon | | | Shannon-Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | $a$ | $b$ | $R^2$ | $a$ | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | *sig?* |
| 2 | −3.259 | 0.960 | 0.577 | −15.443 | 1.760 | 0.177 | 0.100 | 0.977 | 88.47 | 0.0002 | yes |
| 5 | −1.029 | 0.417 | 0.734 | −4.358 | 0.635 | 0.203 | 0.320 | 0.936 | 15.85 | 0.011 | yes |
| 8 | −0.560 | 0.311 | 0.805 | −2.649 | 0.448 | 0.177 | 0.395 | 0.962 | 20.46 | 0.006 | yes |
| 12 | −0.209 | 0.246 | 0.892 | −1.348 | 0.320 | 0.172 | 0.538 | 0.974 | 15.48 | 0.011 | yes |
| 16 | −0.181 | 0.234 | 0.936 | −1.036 | 0.290 | 0.179 | 0.617 | 0.990 | 26.43 | 0.004 | yes |
| 20 | −0.329 | 0.266 | 0.960 | −0.904 | 0.304 | 0.229 | 0.753 | 0.979 | 4.72 | 0.082 | no |



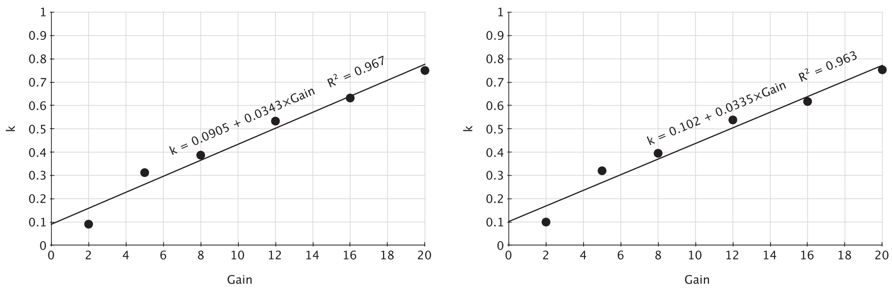

(a) Using the Fitts-inspired models (Fitts and Welford).

(b) Using the Shannon-inspired models (Shannon and Shannon-Welford).

Fig. 5. Regression $R^2$ values as they vary with gain for the Casiez data.

In sharp contrast, the two-part Welford and Shannon-Welford models produce good fits at all levels of gain. Furthermore, F-tests indicate that the two-part formulations model the data significantly better than do their corresponding one-part models at all levels of gain, except for gain 20. Figures 5(a) and 5(b) clearly show the drop-off of $R^2$ values for the one-part models at lower gains and the consistently good $R^2$ values for the two-part models at all levels of gain.

As with Graham's data, the $k$ values serve to quantify the separability of contributions of $A$ and $W$. We further analyzed these values with the goal of determining a pattern. The $k$ values for each gain level of the Casiez data are shown in Figure 6, with results shown for analyses using both the Welford and Shannon-Welford models.




(a) Using the Welford formulation. (b) Using the Shannon-Welford formulation.

Fig. 6. The $k$ values for the Casiez data.

In both cases the $k$ values increase monotonically with gain, and the dependence of $k$ on gain appears to be linear. To confirm this hypothesis we ran a linear regression and found the fit to be very good, $R^2$ = 0.967 for Welford, and $R^2$ = 0.963 for Shannon-Welford.

Our analysis of the Casiez data revealed two valuable insights that were not discussed in their paper. First, the poor modeling of pointing performance using the one-part formulations (i.e., Fitts and Shannon) in the large display condition is corrected by the use of two-part formulations (i.e., Welford or Shannon-Welford), suggesting separable impacts of $A$ and $W$ on mouse pointing performance. Second, we discovered that the $k$ values vary linearly depending on gain for all the levels of gain examined by Casiez et al.

### 4.3 Midair Pointing on a Large Display

The authors and their colleagues [Tsukitani et al. 2011] evaluated performance of users in a serial 1-D pointing task on a very large wall display. The task differed from Kopper's in that pointing was performed by translating the hand, as opposed to rotating it. This is consistent with the input techniques investigated by Graham and by Casiez. The displacement of the hand was magnified by a given control/display gain, and converted into cursor motion. The researchers' goal was to investigate the distinction between "display-space" (interaction as measured on the display) and "hand-space" (interaction as measured in user coordinates). The experimental design was such that the results provide a partial picture of the applicability of two-part formulations to modeling pointing performance on large displays.

In the Tsukitani et al. experiment the apparatus was similar to that in Figure 8. Users stood at one of three distances (1.0m, 2.5m, 3.25m) from a 5m$\times$3m wall display. The cursor was placed on the display at a location determined by a vector originating at a point 4.0m from the display, passing through the hand of the user. The vector origin was located at a height such that the cursor would appear near the middle of the display when the hand was held at waist height. Effective gain values of 1.33, 2.67, and 5.33 were computed from the projection geometry used in the experiment.

Tsukitani et al.'s original analysis using the Fitts formulation produced a good fit to the model ($R^2 > 0.9$), suggesting there was no need for a two-part formulation. We delved deeper into the data, and computed linear regression results for the one- and two-part models of both the Fitts-inspired and Shannon-inspired formulations. These results are shown in Table VI and Table VII. When the $k$ values for the three gain levels are visualized (Figure 7) a pattern of $k$ monotonically increasing with gain is evident. This is consistent with the results of our analyses of both the Graham and Casiez data sets. A possible reason for the good fit with a Fitts formulation is also

Table VI.

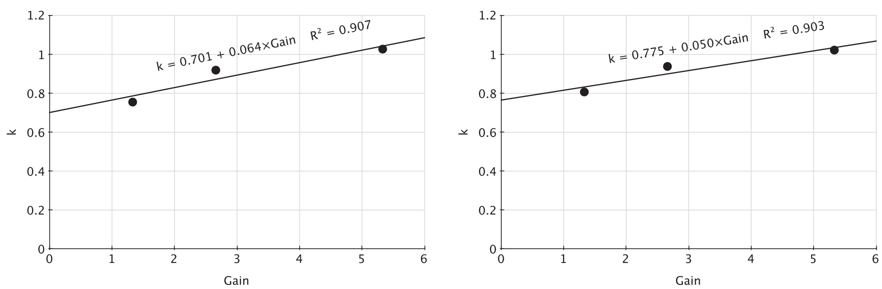
Modeling of movement time from Tsukitani et al. data using the Fitts and Welford formulations.

| | Fitts | | | Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | $a$ | $b$ | $R^2$ | $a$ | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | sig? |
| 1.33 | 0.149 | 0.247 | 0.966 | −0.045 | 0.282 | 0.213 | 0.755 | 0.985 | 7.35 | 0.035 | yes |
| 2.66 | 0.158 | 0.241 | 0.988 | 0.260 | 0.251 | 0.231 | 0.920 | 0.990 | 2.70 | 0.127 | no |
| 5.33 | 0.105 | 0.270 | 0.975 | 0.428 | 0.266 | 0.273 | 1.026 | 0.976 | 0.22 | 0.645 | no |

Table VII.

Modeling of movement time from Tsukitani et al. data using the Shannon and Shannon-Welford formulations.

| | Shannon | | | Shannon-Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | $a$ | $b$ | $R^2$ | $a$ | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | sig? |
| 1.33 | 0.146 | 0.311 | 0.971 | −0.322 | 0.354 | 0.286 | 0.808 | 0.989 | 10.08 | 0.019 | yes |
| 2.66 | 0.171 | 0.294 | 0.991 | 0.030 | 0.305 | 0.286 | 0.938 | 0.993 | 3.37 | 0.091 | no |
| 5.33 | 0.134 | 0.321 | 0.986 | 0.186 | 0.317 | 0.324 | 1.022 | 0.986 | 0.32 | 0.578 | no |




(a) Using the Welford formulation. (b) Using the Shannon-Welford formulation.

Fig. 7. The $k$ values for the Tsukitani et al. data.

evident. The Tsukitani experiment only evaluated pointing at three gain levels, and these gain levels are all close to where the regression line of best fit passes through $k = 1$. This is where Welford's formulation is effectively a one-part formulation, so not much evidence of separability of $A$ and $W$ would be present. But there was some separability, as Figure 7 demonstrates.

While the Tsukitani data reveals no strong need for a two-part formulation, we suspect that had a larger number and wider spread of gain values been evaluated such a need would have emerged. An experiment examining a range of gain levels similar to those explored by Casiez et al. would be needed in order to verify this. We describe such an experiment in Section 5.

### 4.4 Speculation about Limitations in Previous Work

Our conclusion that interactive computer pointing performance, in particular mouse pointing on a large display, does not follow Fitts's Law may be surprising. Many papers have successfully applied Fitts's Law to mouse pointing and other computer tasks. Why is it that these papers did not find an effect similar to what we found? There are several possible answers.

First, assuming that $k$ varies with gain, for some levels of gain $k$ will be close enough to unity that modeling performance using Fitts's Law produces good $R^2$ values. It is possible that the narrow range of reasonable gains for small display pointing are near the $k$-line's intercept with $k = 1$, whereas large display pointing invites the use of a wider range of gain values. Second, as Graham observed, experimental design and analysis choices can serve to partially mask or even completely obscure the separable impacts of $A$ and $W$ on performance, even when $k$ is not close to one.

In order to get a clearer picture of why the shortcomings of Fitts's Law were not previously appreciated we scrutinized the experimental design and analysis in 19 papers directly related to device pointing drawn from a review by Soukoreff and MacKenzie [2004]. We found that 15 of the 19 papers suffered from limitations, in either the experimental design phase or the data analysis phase, that could conceal a separable effect of $A$ and $W$. Of these 15 papers, seven [Akamatsu et al. 1995; Boritz et al. 1991; Guiard et al. 1999; Hornof 2001; Jones 1989, 1991; MacKenzie and Riddersma 1994] included limited combinations of $A$ and $W$ in their experimental designs. As evident in Figure 1, employing only a single $A$ value or a single $W$ value will result in a good fit, regardless of the independence of contributions of $A$ and $W$ to $MT$, concealing the need for a two-part model. In addition, eight papers [Card et al. 1978; Epps 1986; Inkpen 2001; MacKenzie and Ware 1993; Mithal and Douglas 1996; Po et al. 2004; Rutledge and Selker 1990; Zhai et al. 2003] aggregated data points in some manner based on $ID$. As previously demonstrated in Figure 2, this can result in erroneously concluding that Fitts's Law accurately models the data. Only four of the nineteen papers [Han et al. 1990; MacKenzie et al. 1991; MacKenzie and Buxton 1992, 1994] avoid both of these pitfalls and thus have potential to reveal the effect through robust experimental design and analysis.

We are not criticizing the previous research. The analyses were reasonable under the assumption that Fitts's Law is robust, which has been the functioning assumption for the last several decades. However, with the realization that a two-part formulation is sometimes more appropriate, the need to evaluate a broad range of both $A$ and $W$ values becomes clear, as does the need to include each resulting $A$ and $W$ pair in the analysis.

From our reanalysis and high-level review of related work, we offer recommendations on designing and analyzing evaluations meant to test models of pointing performance.

(1) Include multiple combinations of $A$ and $W$ values in the design, representing a realistic range of values for the interaction in question. Ideally fully cross all $A$ and $W$ values.
(2) Include multiple levels of gain. Attempt to explore the entire range of gain values practical for use with the technique being examined.
(3) Include every individual combination of $A$ and $W$ in the analysis.
(4) Perform a separate analysis for every level of gain.
(5) Consider both one-part and two-part models of pointing performance, and determine if the two-part models describe the data significantly better by using appropriate statistical tests.

### 4.5 Reanalysis Conclusions

As explained earlier in this section, we performed analyses of data from two other researchers [Casiez et al. 2008; Graham 1996], and one recent study on which the authors collaborated [Tsukitani et al. 2011]. Those analyses serve to clarify the applicability of different pointing models to different interactive scenarios.

The key insight was derived from the earlier work by Graham, who in turn was building on Welford's work, which was that $A$ and $W$ can contribute independently to the magnitude of movement time, and that a one-part Fitts-like model where only the ratio of $A$ and $W$ is considered can be inadequate for accurately describing pointing time. Instead, two-part formulations are often more appropriate for describing pointing performance on a display.

Applying this insight to other data sets helped clarify the results from those studies. We found that recent data from Casiez et al. are best modeled using two-part formulations, and that a two-part model fixes a problem of poor model fit for conditions of low gain on large displays. We further found in the Casiez et al. data that the parameter $k$ appears to vary linearly with gain. These conclusions are further supported by an analysis of results from Tsukitani et al., where we found similar patterns of $k$ varying with gain, although the results in this second analysis are weaker due to a lack of data points.

We drew the following conclusions from our reanalysis.

(1) One-part models, including Fitts's Law, do not accurately model computer pointing performance using either a finger on a small display or a mouse on a large display, except at some specific gain levels.
(2) Two-part models, including Welford's, accurately model computer pointing performance using either a finger or a mouse.
(3) The improvement of two-part models over one-part models in describing the data is frequently statistically significant.
(4) The value of $k$ appears to vary linearly with gain for mouse pointing on a large display. The slope and intercept of the $k$-line (as it varies with gain) seems to depend on the technique.

## 5. EVALUATION OF POINTING PERFORMANCE ON LARGE DISPLAYS

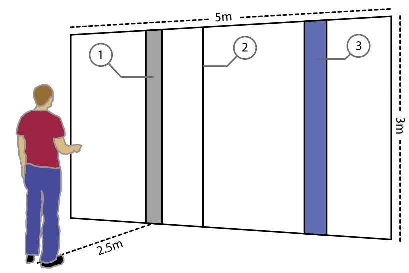
Our reanalysis of data from other researchers lent insight into physical pointing on a small display and mouse pointing on a large display. To aid in completing the picture, we undertook an evaluation of physical pointing on large displays, in order to determine if a separable contribution of $A$ and $W$ to pointing time, and as a consequence a need for two-part pointing models, could be identified. We used physical pointing in part to avoid any clutching. In our evaluation we used a wide range of gain values in addition to using multiple pairs of $A$ and $W$ values. The range of gain values considered (2–20) was the same as that investigated by Casiez et al. While gain values at the upper end of this range might seem extreme, and can be difficult to manage for some users, they are necessary in order to support interaction over the entirety of a very large display surface. The general design of our experiment is inspired by those of Casiez et al. and Tsukitani et al., which were discussed in the previous section.

### 5.1 Apparatus

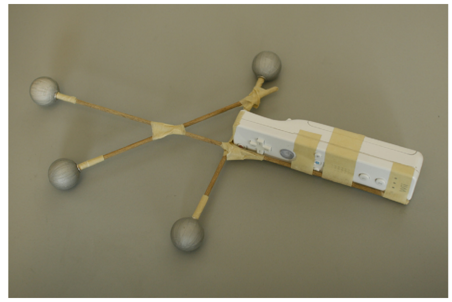
As illustrated in Figure 8(a), users stood 2.5m from a large vertical glass screen approximately 5m$\times$3m in size. The screen was rear-projected by a 4$\times$3 array of projectors, each with an 800$\times$600 pixel resolution. The images of neighbouring projectors overlapped 160 pixels with a blending function to minimize discontinuities due to possible misalignment. Overall resolution was 2720$\times$1480 pixels.

Tap events were performed using the thumb (A) button on a hand-held Nintendo Wii Remote. The Wii Remote was tracked using a Vicon motion capture system because the native Wii Remote sensing capabilities were not accurate enough for our needs. The Wii Remote was outfitted with reflective markers for this purpose, as shown in Figure 8(b).

The experimental software ran on a computer running the Windows XP operating system and was written in C# using the Microsoft XNA Game Studio library and .NET 3.5. The WiimoteLib library was used to communicate with the Wii Remote device.




(a) Layout of experimental apparatus. Labeled components are: (1) center target, currently not active. (2) cursor. (3) right target, currently active.



(b) The Wii Remote mounted with reflective Vicon tracking markers.

Fig. 8. The experimental apparatus.

The same computer ran the Vicon tracking software. Logging of events was performed in real time and stored on the machine.

Tracking accuracy is important. Casiez et al. [2008] developed a model for calculating the usable range of gain values given sensing accuracy and display resolution. With the resolution of the experimental display, and Vicon signal noise measured to be not more than 0.07mm (the capture volume was very small, making high accuracy possible), we applied Casiez's model and calculated the maximum usable gain of our apparatus to be 26. This means that every screen pixel was addressable at gain values up to 26. All values of gain investigated in the experiment (2–20) are considered "usable" under Casiez's criteria.

### 5.2 Task and Stimuli

The experimental task was a serial 1-D tapping task between two target pairs of variable movement amplitude and target width, modeled closely after tasks used by Casiez et al. [2008] and Fitts [1954]. Targets were bars both of the specified width, which spanned the entire height of the display. Lateral translation of the user's hand was magnified by a given gain level, resulting in cursor motion. The cursor was a line which similarly spanned the entire height of the display. It was decided to use a traditional 1-D task similar to what was originally used by Fitts, rather than a 2-D task such as that defined by ISO 9241-9 [Douglas et al. 1999] because we were concerned with the fundamental applicability of Fitts's Law.

For each target pair a participant first tapped the start target and then performed a sequence of eight reciprocal taps between the two targets. The active target was always blue, and the nonactive target was always grey. One target was directly in front of the participant, while the other target was to the right of the participant at the given amplitude. This arrangement was chosen to avoid any possible impact of cross-lateral inhibition, which is a difficulty in motions when the hand crosses the body's midline [Schofield 1976]. The participant was required to correctly tap the start target to initiate the trial. Time and location for the initial tap and the following eight taps were recorded by the software. After a tap the target briefly flashed green to indicate success, or red to indicate an error. There was no requirement to correct errors.

### 5.3 Participants

Nineteen participants (two female) were recruited through on-campus advertising. All were right-handed, a requirement for participation in the experiment. Ages ranged from 20 to 42, mean 26.4, SD 5.7. All participants were regular computer users (9+ hours per week). They were compensated \$10 for participating, and the half with the best performance were later compensated an extra \$10. The additional compensation was intended to be an incentive for participants to perform well.

### 5.4 Design

A within-subjects design was used. The independent variables were gain $G$ (2, 5, 8, 12, 16, 20), movement amplitude $A$ (25cm, 50cm, 100cm, 250cm), target width $W$ (5cm, 10cm, 20cm), and trial block $B$ (1, 2, 3). $A$ and $W$ combinations were fully crossed, except that the 250cm amplitude was only used at gains 12, 16 and 20, because it was not reachable at lower gains. We chose an incomplete design in the interest of investigating a wider range of $G$, $A$, and $W$ values, as demanded by our motivation. Such a partial design has drawbacks, but it does support an investigation of parameters outside of what would be possible with a full design.

Each gain level was presented during each block of the trials. Gain levels were randomly ordered within every block. For each gain level each $A$ and $W$ pair was presented in random order. Eight taps were performed for each $A$ and $W$ pair.

In summary, the experimental design was as follows.

19 participants
× 3 blocks
× (3 gains × 9 $A$ and $W$ combinations) + (3 gains × 12 $A$ and $W$ combinations)
× 8 taps
= $19 \times 3 \times (27 + 36) \times 8 = 28{,}728$ total taps.

*5.4.1 Procedure.* The experiment was performed in a single session for each participant, lasting approximately 50 minutes. Participants arrived and filled out a consent form and a prequestionnaire gathering demographic information. They were introduced to the system and the pointing task was explained. Participants were instructed to complete the task as quickly as possible with a goal of 95% accuracy. They each practiced at least five $A$ and $W$ combinations (40 taps), and were invited to practice more if they felt the need.

Participants then completed the three experimental blocks. Whenever the gain level was changed a practice $A$ and $W$ pair was presented to the participant. The purpose was to allow a participant to grow accustomed to the new gain level. The participant was not informed that an $A$ and $W$ pair was a practice pair. These were presented in the regular flow of the experiment, but the data for those pairs were not analyzed.

Between each block the participants sat at a table and played a distractor puzzle task for three minutes. They were invited to take extra time to rest, but none did so.

After all conditions were completed each participant filled out a post-questionnaire to gather qualitative feedback on aspects of the experiment.

*5.4.2 Measures.* Performance was measured as the time taken to perform each individual tap action. Timing began when the participant tapped the start target for each $A$ and $W$ pair, and ended with the eighth tap for that pair. Errors were measured as tap events that occurred outside of the current target. The location of each tap was also recorded.

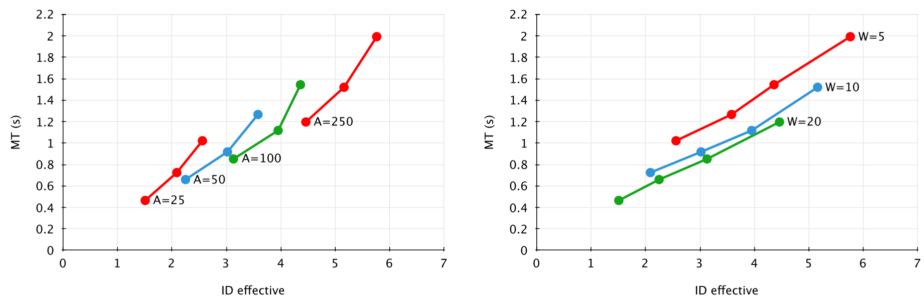
*5.4.3 Hypotheses.* We derived our hypotheses from our reanalysis of related research.

*H1*. One-part formulations (i.e., Fitts and Shannon) will not accurately model pointing performance at all gain levels.

Table VIII.

Significant ANOVA results for movement time. An asterisk (*) indicates that a Greenhouse-Geisser correction was applied because the sphericity assumption was violated.

| Factor | F-ratio | Significance | Partial $\eta^2$ |
|---|---|---|---|
| $G$ | $F_{3.0,53.8} = 36.9^*$ | $p < 0.001$ | 0.672 |
| $A$ | $F_{1.1,20.4} = 778.5^*$ | $p < 0.001$ | 0.977 |
| $W$ | $F_{1.1,20.1} = 408.3^*$ | $p < 0.001$ | 0.958 |
| $G \times A$ | $F_{10,180} = 4.3$ | $p < 0.001$ | 0.194 |
| $G \times W$ | $F_{4.3,77.6} = 15.0^*$ | $p < 0.001$ | 0.455 |
| $A \times W$ | $F_{2.5,45.8} = 18.5^*$ | $p < 0.001$ | 0.506 |




(a) Lines connect points representing tasks with the same movement amplitude $A$.

(b) Lines connect points representing tasks with the same target width $W$.

Fig. 9. Movement time results for all $A$ and $W$ combinations from the experimental data at $G = 20$. Data points in both graphs are the same. Effective target width was used in computing $ID$. An independent contribution of $A$ and $W$ to movement time is apparent, suggesting the need for a two-part linear model.

*H2*. Two-part formulations (i.e., Welford and Shannon-Welford) will accurately model pointing performance at each individual gain level.
*H3a*. (*weak*) The exponent $k$ will vary monotonically with gain.
*H3b*. (*strong*) The exponent $k$ will vary linearly with gain.

### 5.5 Results

We were concerned with the possible impact of learning effects. Before our main analysis we performed a repeated measures ANOVA to determine if there was an effect of block. We found no effect of block on either movement time ($F_{2,36} = 0.943$, $p = 0.399$) or error rate ($F_{1.382,24.873} = 0.117$, $p = 0.814$, with a Greenhouse-Geisser correction for violation of sphericity). We therefore included all blocks in our analysis.

*5.5.1 Movement Time.* Significant main effects of $G$, $A$, $W$ were found. Significant interactions of $G \times A$, $G \times W$, and $A \times W$ were also found. Results are summarized in Table VIII. A visualization of movement time data for the gain $G = 20$ condition is shown in Figure 9.

We performed a linear regression using data from all participants. Regression constants calculated using the two Fitts-inspired models are shown in Table IX, whereas results calculated using the two Shannon-inspired models are shown in Table X. In order to adjust for accuracy we performed a second analysis of the results, this time using effective target width ($W_e$), in the manner suggested by Soukoreff and MacKenzie [2004], and further explored by Zhai et al. [2004]. These results

Table IX.

Modeling of movement time using the Fitts and Welford formulations, and actual width ($W$). $R^2$ values for the Fitts formulation decrease at higher gains, but the Welford formulation is consistently good. F-tests indicate that the Welford model is significantly better than the Fitts model at gain levels 16 and 20.

| | Fitts | | | Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | $a$ | $b$ | $R^2$ | $a$ | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | sig? |
| 2 | 0.070 | 0.233 | 0.991 | 0.269 | 0.236 | 0.229 | 0.97 | 0.992 | 0.18 | 0.684 | no |
| 5 | 0.089 | 0.204 | 0.989 | 0.336 | 0.200 | 0.209 | 1.05 | 0.989 | 0.29 | 0.608 | no |
| 8 | 0.076 | 0.209 | 0.982 | 0.389 | 0.198 | 0.221 | 1.12 | 0.985 | 1.21 | 0.314 | no |
| 12 | 0.032 | 0.242 | 0.980 | 0.402 | 0.232 | 0.263 | 1.13 | 0.983 | 1.16 | 0.219 | no |
| 16 | 0.022 | 0.256 | 0.972 | 0.517 | 0.239 | 0.296 | 1.24 | 0.982 | 5.13 | 0.049 | yes |
| 20 | 0.035 | 0.275 | 0.964 | 0.653 | 0.250 | 0.331 | 1.32 | 0.983 | 9.32 | 0.014 | yes |

Table X.

Modeling of movement time using the Shannon and Shannon-Welford formulations, and actual width ($W$). $R^2$ values for the Shannon formulation decrease at higher gains, but the Shannon-Welford formulation is consistently good. F-tests indicate that the Shannon-Welford model is significantly better than the Shannon model at gain levels 16 and 20.

| | Shannon | | | Shannon-Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | $a$ | $b$ | $R^2$ | $a$ | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | sig? |
| 2 | 0.084 | 0.286 | 0.998 | 0.046 | 0.291 | 0.283 | 0.97 | 0.998 | 0.69 | 0.436 | no |
| 5 | 0.100 | 0.252 | 0.998 | 0.144 | 0.247 | 0.255 | 1.04 | 0.998 | 1.62 | 0.250 | no |
| 8 | 0.087 | 0.258 | 0.994 | 0.195 | 0.245 | 0.267 | 1.09 | 0.996 | 4.26 | 0.085 | no |
| 12 | 0.082 | 0.282 | 0.990 | 0.228 | 0.271 | 0.304 | 1.12 | 0.994 | 3.77 | 0.041 | yes |
| 16 | 0.075 | 0.299 | 0.982 | 0.336 | 0.279 | 0.338 | 1.21 | 0.993 | 15.26 | 0.004 | yes |
| 20 | 0.092 | 0.321 | 0.974 | 0.464 | 0.291 | 0.376 | 1.29 | 0.993 | 25.36 | 0.0007 | yes |

Table XI.

Modeling of movement time using the Fitts and Welford formulations, and effective width ($W_e$). $R^2$ values for the Fitts formulation decrease at higher gains, but the Welford formulation is consistently good. F-tests indicate that the Welford model is significantly better than the Fitts model at gain levels 16 and 20.

| | Fitts | | | Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | a | b | $R^2$ | a | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | sig? |
| 2 | 0.133 | 0.223 | 0.989 | 0.214 | 0.240 | 0.210 | 0.875 | 0.993 | 4.19 | 0.087 | no |
| 5 | 0.041 | 0.228 | 0.982 | 0.274 | 0.228 | 0.229 | 1.00 | 0.982 | 0.002 | 0.966 | no |
| 8 | −0.002 | 0.241 | 0.973 | 0.358 | 0.230 | 0.260 | 1.14 | 0.978 | 1.09 | 0.336 | no |
| 12 | −0.051 | 0.278 | 0.937 | 0.614 | 0.265 | 0.365 | 1.38 | 0.961 | 3.66 | 0.044 | yes |
| 16 | −0.028 | 0.292 | 0.910 | 0.985 | 0.273 | 0.457 | 1.67 | 0.970 | 18.23 | 0.002 | yes |
| 20 | 0.013 | 0.314 | 0.891 | 1.275 | 0.290 | 0.529 | 1.82 | 0.975 | 29.80 | 0.0004 | yes |

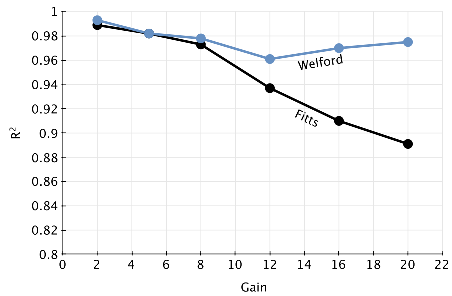
are presented in Table XI and Table XII. The tables also present results of F-tests comparing corresponding two-part and one-part models at each level of gain. In cases where a significant difference was found, the two-part model was deemed to characterize the data significantly better than did the one-part model. A visualization of $R^2$ values as they vary with gain is shown in Figure 10.

To test the hypothesis that the $k$ values varied based on gain we performed a linear regression analysis on the $k$ values computed using the Welford formulation for each gain level (Figure 11). The linear function of best fit using actual width $W$ was found to be $k = 0.946 + 0.018 \times G$, with a fit of $R^2 = 0.970$. The linear function of best fit using effective width $W_e$ was found to be $k = 0.736 + 0.055 \times G$, with a fit of $R^2 = 0.992$.

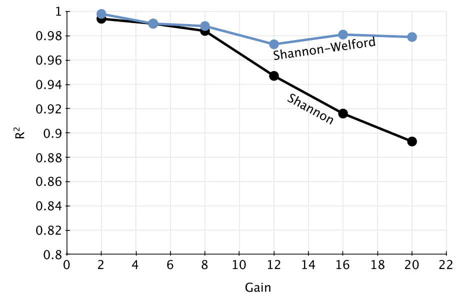
Table XII.

Modeling of movement time using the Shannon and Shannon-Welford formulations, and effective width ($W_e$). $R^2$ values for the Shannon formulation decrease at higher gains, but the Shannon-Welford formulation is consistently good. F-tests indicate that the Shannon-Welford model is significantly better than the Shannon model at gain levels 2, 12, 16 and 20.

| | Shannon | | | Shannon-Welford | | | | | F-test | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Gain | $a$ | $b$ | $R^2$ | $a$ | $b_1$ | $b_2$ | $k$ | $R^2$ | $F$ | $p$ | sig? |
| 2 | 0.132 | 0.280 | 0.994 | −0.018 | 0.299 | 0.270 | 0.90 | 0.998 | 12.95 | 0.011 | yes |
| 5 | 0.049 | 0.285 | 0.990 | 0.059 | 0.284 | 0.286 | 1.01 | 0.990 | 0.020 | 0.904 | no |
| 8 | 0.009 | 0.300 | 0.984 | 0.143 | 0.285 | 0.314 | 1.10 | 0.988 | 2.20 | 0.188 | no |
| 12 | 0.000 | 0.328 | 0.947 | 0.414 | 0.313 | 0.418 | 1.34 | 0.973 | 5.89 | 0.016 | yes |
| 16 | 0.021 | 0.348 | 0.916 | 0.780 | 0.325 | 0.515 | 1.58 | 0.981 | 31.10 | 0.0003 | yes |
| 20 | 0.065 | 0.373 | 0.893 | 1.039 | 0.346 | 0.586 | 1.69 | 0.979 | 36.29 | 0.0002 | yes |

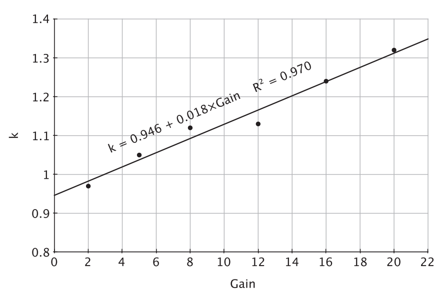



(a) Using the Fitts-inspired models (Fitts and Welford).

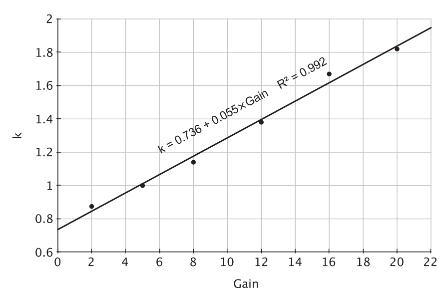



(b) Using the Shannon-inspired models (Shannon and Shannon-Welford).

Fig. 10. Regression $R^2$ values as they vary with gain for the experimental data, using $W_e$.




(a) Actual movement amplitude $A$ and actual target width $W$.




(b) Actual movement amplitude $A$ and effective target width $W_e$.

Fig. 11. The $k$ values for the large display physical pointing experiment, calculated using the Welford formulation.

*5.5.2 Error Rate.* Mean error rates were 7.8%. An ANOVA found significant main effects of $G$, $A$, and $W$. The interaction of $G \times W$ was also significant. Results are summarized in Table XIII.

*5.5.3 Subjective Measures.* A summary of results from participants' subjective ratings of the difficulty of the task is shown in Figure 12. A Friedman test comparing ratings for low (2, 5), medium (8, 12) and high gain levels (16, 20) showed a significant effect of gain on difficulty ($\chi^2_{(2,N=19)} = 30.958$, $p < 0.001$). Pairwise comparisons using a Wilcoxon Signed Ranks Test showed significant differences between

Table XIII.

Significant ANOVA results for error rate. An asterisk (*) indicates a Greenhouse-Geisser correction was applied because the sphericity assumption was violated.

| Factor | F-ratio | Significance | Partial $\eta^2$ |
|---|---|---|---|
| G | $F_{2.9,52.8} = 28.5^*$ | $p < 0.001$ | 0.613 |
| A | $F_{2,36} = 13.6$ | $p < 0.001$ | 0.431 |
| W | $F_{1.4,24.3} = 96.0^*$ | $p < 0.001$ | 0.842 |
| G×W | $F_{4.2,75.8} = 16.6^*$ | $p < 0.001$ | 0.480 |

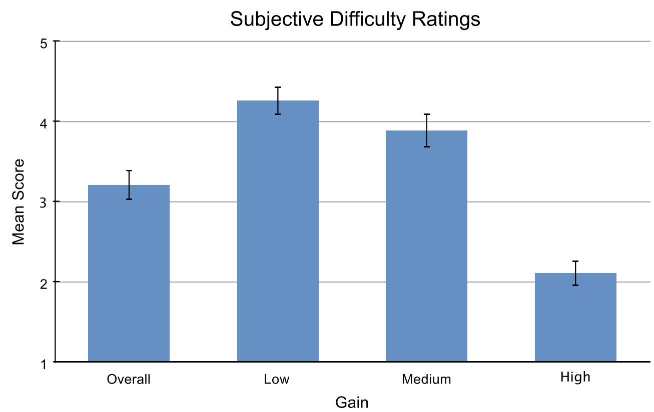



Fig. 12. Mean scores of task difficulty overall, at low (2, 5), medium (8, 12) and high (16, 20) gain levels, with standard error. Ratings on a scale of one (impossible) to five (easy). N = 19.

high and low gains ($z = -3.882$, $p < 0.001$) and between high and medium gains ($z = -3.882$, $p < 0.001$).

### 5.6 Discussion

Before considering our core hypotheses, it is useful to first discuss the results of the ANOVA for the movement time data.

The finding that there were significant effects of both $A$ and $W$ on movement time is not surprising. The fact that both movement amplitude and target width impact movement time is fundamental to any discussion of pointing performance.

The finding of a significant impact of $G$ is also unsurprising because similar effects have been found by other researchers [Casiez et al. 2008; Koester et al. 2005; MacKenzie and Riddersma 1994].

The interactions of $G \times A$ and $G \times W$ were unexpected, but are not surprising, as the role of $G$ in influencing $MT$, and how it may interact with other variables, is still not entirely understood.

We summarize the results according to our hypotheses, then discuss each in more depth.

— *H1*. One-part formulations (i.e., Fitts and Shannon) will not accurately model pointing performance at all gain levels. *Somewhat supported.*
— *H2*. Two-part formulations (i.e., Welford and Shannon-Welford) will accurately model pointing performance at each individual gain level. *Supported.*
— *H3a*. *(weak)* The exponent $k$ will vary monotonically with gain. *Supported.*
— *H3b*. *(strong)* The exponent $k$ will vary linearly with gain. *Supported.*

*5.6.1 H1.* One-part models of pointing performance had mixed success in characterizing movement time. Using actual width $W$ values, Fitts's formulation gave linear fits ranging in accuracy from a low of $R^2 = 0.964$ to $R^2 = 0.991$ at different levels of gain. For the levels of gain examined these $R^2$ values are good, surpassing MacKenzie's somewhat arbitrary 0.9 threshold. However, it is clear that the $R^2$ values are decreasing as gain increases. The Shannon formulation fares better at modeling the data. It provides very good $R^2$ values, but it is clear that the quality of fit is again decreasing at higher levels of gain.

Using effective width $W_e$, Fitts's Law is less successful. Using the Fitts formulation, linear quality of fit in this case ranges from a low of $R^2 = 0.891$ to a high of $R^2 = 0.989$ at different levels of gain, failing to produce an acceptable linear fit at one level of gain. As was the case for actual $W$ values, the Shannon formulation fares better than the Fitts formulation, but performance again drops off at higher gains. Effective width results are more relevant, due to the nature of $W_e$ as an accurate representation of the task. The reason for the difference in quality of fit between actual width $W$ and effective width $W_e$ is evident from examining the $k$ values from the two-part models as visualized in Figure 11. It is clear that $k$ varies much more for $W_e$ results than it does for $W$ results. As we have discussed, the more $k$ deviates from 1, the less well a one-part model will be able to describe the data.

We thus conclude that hypothesis *H1* is somewhat supported. The results behaved as expected, however, the contributions of $A$ and $W$ to performance did not differ enough to result in very poor $R^2$ values using one-part formulations, at least in the range of gains examined. The $R^2$ values only failed to reach the 0.9 level at the highest level of gain using the effective width data sets.

*5.6.2 H2.* Both the Welford and Shannon-Welford two-part models of pointing performance produced a good fit at every level of gain for both actual widths and effective widths data sets. The Welford formulation applied to actual width $W$ data produced linear regression fits ranging from a low of $R^2 = 0.982$ to a high of $R^2 = 0.992$. Applied to effective width $W_e$ data, linear regression fits ranged from a low of $R^2 = 0.961$ to a high of $R^2 = 0.993$. The Shannon-Welford formulation applied to actual width $W$ data produced linear regression fits ranging from a low of 0.993 to a high of 0.998. For effective width $W_e$, linear regression fits ranged from a low of 0.973 to a high of 0.998.

More importantly, the two-part models statistically described the data significantly better than did the one-part models at some gain levels. The Welford model was superior to the Fitts model at the gain levels of 16 and 20 for actual width results, and gain levels 12, 16 and 20 for effective width results. The Shannon-Welford model was superior to the Shannon model at gains 12, 16 and 20 for actual width data, and gains 2, 12, 16, and 20 for effective width data.

Thus, hypothesis *H2* is supported.

*5.6.3 H3.* The $k$ values were found to vary monotonically and linearly, according to gain, supporting hypotheses *H3a* and *H3b*. For actual width data, the $k$ values followed a linear model quite closely ($R^2 = 0.970$). For the effective width results, the $k$ values were even more accurately modeled using a linear function ($R^2 = 0.992$). Interestingly, the slopes for the two sets of results were noticeably different, with $k$ varying more in the effective width set of data. The intercept of the slope at $G = 0$ was also noticeably different, although $G = 0$ is meaningless in an interactive setting, suggesting that the intercept may not be of much significance.

While we have thoroughly examined how $k$ changes with gain, it will be useful to understand how its components, $b_1$ and $b_2$, vary as well. In the case of our experimental data, it is interesting to note that $b_2$ increases quite consistently with gain, while $b_1$ stays relatively constant. This would suggest that at higher gains the effect of $W$ to decrease movement time tends to dominate over the effect of $A$ to increase movement times. This is an issue to investigate in depth in future work.

## 6. CONCLUSIONS

Fitts's Law has been widely used as a tool for characterizing the performance of pointing tasks on computer systems, both for forming predictive models and for determining performance as characterized by throughput. It possesses the beneficial attribute of being concise and simple to understand, as well as being well grounded in information theory. However, we must remember that neither simplicity nor adherence to a particular theory of information is in isolation enough to justify the application of a particular model. A theoretical model is only as good as its ability to describe empirical data. If we find shortcomings in an existing model, or discover a new model that performs significantly better, we must resist the urge to cling to the old model. As advised in a quote often attributed to Albert Einstein, we should "Make things as simple as possible, but not simpler."

Over the years Fitts's Law has become so entrenched that researchers rarely ever question the fundamental assumptions underlying its use in HCI. We have presented here compelling evidence that Fitts's Law and its one-part variants do indeed have limitations. In some cases (Graham and Casiez) one-part models fall well short of adequately describing the data. In other cases (Tsukitani and our experiment) they perform significantly less well than corresponding two-part models such as Welford or Shannon-Welford.

To help structure our work we organized pointing formulations into a $2\times2$ matrix. The first axis of the matrix, degrees of freedom, categorizes formulations into one-part and two-part models. The second axis, information-theoretic interpretation, categorizes formulations into models with Fitts-inspired roots and those with Shannon-inspired roots. The first axis was chosen because we are primarily concerned with the relative abilities of one-part and two-part models to describe data. The second axis was chosen for completeness, even though the issue of information-theoretic interpretation is beyond the scope of our current work. We explored this space of models, where we made several contributions. First, we demonstrated the mathematical equivalence of the Welford two-part model to a linear version of Kopper et al.'s model for ray pointing. Second, we posited a two-part model combining elements of Welford's two-part model and Shannon's one-part model. Third, we described how, in two-part models, the relative contributions of movement amplitude $A$ and target width $W$ to movement time can be captured in a constant $k$ that is the ratio of the two multiplicative coefficients in Welford's formulation.

Conclusions relating to the relative merits of the different models fall into two categories. First, a reanalysis of data from other researchers provided insight into the abilities of one-part and two-part formulations to model pointing performance for mouse and physical pointing on large and small displays. Second, results from a new experiment provided additional insights into the application of different models to midair physical pointing on a large wall display.

### 6.1 Reanalysis of Third-Party Data: Rethinking the Application of Fitts's Law

Graham [1996] originally observed that Fitts's Law does not accurately model physical pointing with a tracked finger at a single gain on small displays. Our subsequent reanalysis of his data revealed that neither the Fitts nor Shannon formulations adequately model performance at two of the three gain levels that Graham examined, whereas two-part models accurately model every level of gain. Furthermore, the two-part models describe the data significantly better than do their corresponding one-part models at all three levels of gain. We also demonstrated the variation of the constant $k$ with gain. A similar analysis of data from Casiez et al. [2008] revealed that one-part models also fail to accurately model mouse pointing at constant gain on a large display, with particular shortcomings at lower gain levels. In this case as well, it was found that the two-part formulations accurately model pointing performance at all gain levels. Furthermore, a linear dependence of $k$ on gain was observed. It is important to note that the Casiez data (unlike the other data sets we analyzed) incorporated substantial amounts of clutching in the movement times, meaning that two-part models are also capable of capturing clutching. Our analysis of data from Tsukitani et al. [2011] showed a pattern consistent with those found in the other analyses, although the linear regression of $k$ showed different slopes and intercepts for the different data sets. Throughout our reanalysis, we found that two-part models regularly outperformed their one-part model counterparts in a statistically significant manner.

We investigated why the effects we observed might not have been detected by other researchers. We concluded that careful experimental design is required to identify the need for two-part models, and that past experiments in HCI have often lacked one or more elements necessary to detect these effects. We arrived at a set of recommendations to be applied to the design of experiments meant to develop models of pointing performance.

The conclusions reached in our reanalysis are as follows. One-part pointing models cannot necessarily be relied on to accurately model pointing performance in computing systems, especially at widely varying levels of gain. Two-part formulations of pointing performance can correct for the discovered shortcomings of one-part formulations. Movement amplitude $A$ and target width $W$ are both of significance in modeling performance, and should be considered as separable, independent variables when developing models of pointing performance. The relative contributions of $A$ and $W$, as captured by $k$, appear to vary linearly with gain, although the intercept of the line, and possibly the slope, vary with interaction technique.

### 6.2 A Model for Interaction with Very Large Wall Displays

As noted by Guiard [2009], “Currently we are witnessing a spectacular development of drastically *miniaturized* as well as *enlarged* interfaces (e.g., handheld devices, interactive wall displays), and so in HCI research we urgently need a better understanding of what happens in pointing at nonoptimal scales.” He suggests that in this case (when the scale of the interaction is, for example, on a very large display) it could be appropriate to separate the effects of $A$ and $W$, a question he leaves for future research. One of the contributions of our article is taking a first step in this direction.

Towards the goal of completing the exploration of the space of interactive techniques analyzed, we performed an experiment to develop a performance model for midair physical pointing on a very large wall display. Although the absolute differences in quality of fit between the models were small, our experimental results demonstrate that two-part models often significantly outperformed one-part models in describing the data. Furthermore, consistent with our analyses of other data sets, we again found $k$ to vary linearly with gain. In isolation our experimental results might not be quite convincing, due to the small absolute differences between model fits, but when considered in the context of our reanalysis of others’ data (especially the Casiez data), the results are more compelling (Figure 13).

The two most valuable analyses of the article were those of the Casiez data and of our new experimental data. These two data sets provide a good coverage of $A$

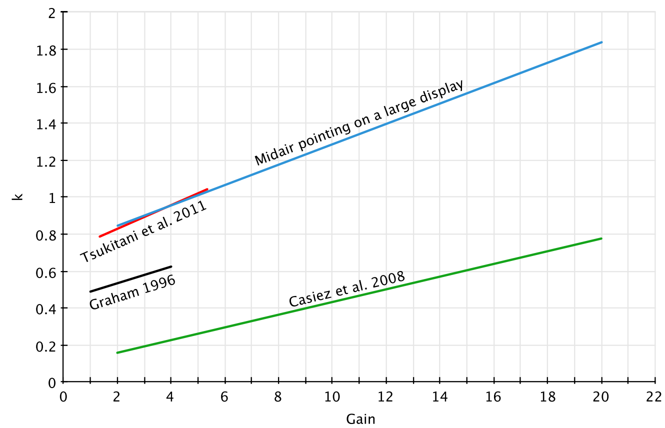



Fig. 13. The $k$-lines for the four data sets we analyzed, as determined by linear regression. Extent of lines indicate the range of gain values considered in each data set. The lines consistently have positive slopes, but their intercepts are mostly very different. Note that the $k$-lines of the Tsukitani results and our new results are similar, but this is expected as the interaction techniques in the two experiments were very similar. This figure is an aggregation of Figures 3(a), 6(a), 7(a), and 11(b).

and $W$ pairings over a wide range of gains, with a suitable number of participants. The findings of these two experiments were consistent: two-part models consistently out-perform one-part models, sometimes dramatically so. Furthermore, the separability of contributions of $A$ and $W$ to movement time, as characterized by $k$, increases linearly with gain, although the intercept and possibly the slope of the $k$-lines vary. While it is true that two-part models outperformed one-part models in opposite ranges of gains (low gains for Casiez, high gains for our data), this is possibly explained by the fact that their $k$-slope intercepts were different (Figure 13). The data from both the Graham and Tsukitani analyses lend some support, although the deficit of data points, and possibly the small number of participants, limit their contributions.

### 6.3 Future Work

With the realization of the shortcomings of Fitts's Law and other one-part models, and the applicability of two-part models of pointing performance, it is clear that there is much work still to be done. First, researchers may need to reevaluate conclusions that were long thought to be sound. Existing models of pointing performance for some input device and display combinations may be faulty, due to limitations in either experimental design or analysis.

Our understanding of the nature of $k$ is only partially complete. We believe that $k$ varies linearly with gain, but the function differs by input technique. For example, at $G = 2$ the Casiez et al. $k$ was at a very low value of 0.091, whereas for our large display midair pointing task, $k$ was at 0.875 for the same gain level. It is important that we examine the $k$ dependence for different devices and display types. We should also strive to develop a deeper understanding of the influencing factors on $k$. For example, Casiez hypothesized that mouse clutching was responsible for the deviation of their data from Fitts's Law. We may find that clutching is a variable that influences the $k$-line slope and intercept. There are also almost certainly other factors, because none of the other pointing methods we investigated allowed clutching, yet the $k$ formulation was still found to be relevant.

Even more fundamental than an understanding of $k$ is an understanding of $b_2$ and $b_1$ ($k = b_2/b_1$). We have found that $b_2$ and $b_1$ vary with gain, but we do not know why.

One interpretation of the Fitts formulation is that the reciprocal $1/b$ is the "throughput" of the human sensorimotor system as it performs physical pointing tasks, and the index of difficulty is the number of bits of information the system must process. It is possible that $b_1$ and $b_2$ are associated with throughput values for their own respective processes related to movement amplitude $A$ and target width $W$. We might imagine (and both Welford and Graham have suggested this) that there could be two distinct processes within the sensorimotor system, one responsible for the initial ballistic movement that depends on $A$ and another for the "fine tuning" during deceleration that depends on $W$. If those two processes take place at different rates, they would have different coefficients representing their respective throughput values. Movement time would be governed by the additive constant $a$ plus the multiplicative constant $b_1$ times the number of bits required for $A$ to be dealt with by the first process but mediated by $b_2$ reducing the time when the second process has determined that the obtained precision is adequate.

To see how a two-part model might reflect the theory that there are two independent processes in play, we can reformulate Equation (3) as Equation (12), where we "flip" the argument for the second logarithmic term (i.e., we use the reciprocal of the argument) and thus add rather than subtract the second logarithmic term from the first logarithmic term. From this formulation we see that movement time increases logarithmically as $A$ increases and also as $W$ decreases.

$$\begin{aligned} MT &= a + b_1 \log_2 A - b_2 \log_2 W \\ &= a + b_1 \log_2 A + b_2 \log_2 \frac{1}{W}. \end{aligned} \tag{12}$$

There is a problem with this if $W$ gets too large because the last term turns negative, but if we recall from Equation (11) that the normalizing factors suggested by Welford should be used, we actually have Equation (13). It would make sense to assume that when $W$ becomes larger than $W_0$ (which is when the logarithm goes negative) we can ignore this term because target width probably no longer has an impact on movement time because precision is no longer an issue. This also assumes that movement is still directed more or less to the center of the target so that $A$ is still a valid estimate for movement amplitude.

$$MT = a + b_1 \log_2 \frac{A}{A_0} + b_2 \log_2 \frac{W_0}{W}. \tag{13}$$

Graham's dissertation [Graham 1996] discusses some of the evidence for there being two separate processes, but the seeds for this can be found in Welford's book [1971], again on page 156, where, after describing "a faster distance-covering phase and a slower phase of 'homing' on to the target," Welford speculates that "It seems reasonable to regard the former as similar in speed to that of a ballistic movement of a given amplitude and the latter as implying an additional process of visual control." Full support for this theory will require more research, but it seems clear from our work that there are two, not one, throughput parameters, one for $A$ and another for $W$. Welford recognized this based on his analysis of the data he collected. What is different today, more than four decades after Welford wrote his book, is the existence of techniques such as functional MRI that might be used to confirm the existence of and the throughput rates for two independent sensorimotor processes.

Related to our particular interest in large wall displays, there is again much work to be done. We limited our examination to constant control/display gains. Variable gains, such as the pointer acceleration investigated by Casiez et al. [2008], are important to consider. This may allow us to expand the use of two-part models and the $k$ relationship to include pointer acceleration.

## ACKNOWLEDGMENTS

We thank Christine MacKenzie and Jocelyn Keillor for their valuable comments on our work. We thank Michiel van de Panne for providing us with access to a Vicon tracking system, and we especially thank both Evan Graham and Géry Casiez and his colleagues for providing the raw data from their experiments for use in our reanalysis. We thank Meijiao Guan and Harry Joe for helpful advice on statistical methods.

## REFERENCES

AGRAWALA, M., BEERS, A. C., MCDOWALL, I., FROHLICH, B., BOLAS, M., AND HANRAHAN, P. 1997. The two-user responsive workbench: Support for collaboration through individual views of shared space. In *Proceedings of the ACM SIGGRAPH International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'97)*. 327–332.

AKAIKE, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control 19*, 6, 716–723.

AKAMATSU, M., MACKENZIE, I. S., AND HASBROUQ, T. 1995. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics 38*, 816–827.

ARNAUT, L. Y. AND GREENSTEIN, J. S. 1990. Is display/control gain a useful metric for optimizing an interface? *Human Factors 13*, 651–663.

BAUDISCH, P., CUTRELL, E., ROBBINS, D., CZERWINSKI, M., TANDLER, P., BEDERSON, B., AND ZIERLINGER, A. 2003. Drag-and-pop and drag-and-pick: Techniques for accessing remote screen content on touch- and pen-operated systems. In *Proceedings of International Conference on Human-Computer Interaction*. 57–64.

BEZERIANOS, A. AND BALAKRISHNAN, R. 2005. The Vacuum: Facilitating the manipulation of distant objects. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'05)*. 361–370.

BORITZ, J., BOOTH, K. S., AND COWAN, W. B. 1991. Fitts' law studies of directional mouse movement. In *Proceedings of Graphics Interface '91*. 216–223.

BRIGNULL, H., IZADI, S., FITZPATRICK, G., ROGERS, Y., AND RODDEN, T. 2004. The introduction of a shared interactive surface into a communal space. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'04)*. ACM, New York, NY, 49–58.

BUXTON, B. 2008. Surface and tangible computing, and the "small" matter of people and design. In *Proceedings of the IEEE International Solid-State Circuits Conference Digest of Technical Papers*. 24–29.

CARD, S. K., ENGLISH, W. K., AND BURR, B. J. 1978. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for selection on a CRT. *Ergonomics 21*, 601–613.

CASIEZ, G. AND ROUSSEL, N. 2011. No more bricolage! methods and tools to characterize, replicate and compare pointing transfer functions. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology (UIST'11)*.

CASIEZ, G., VOGEL, D., BALAKRISHNAN, R., AND COCKBURN, A. 2008. The impact of control-display gain on user performance in pointing tasks. *Hum.-Comput. Interact. 23*, 3, 215–250.

CHAPUIS, O. AND DRAGICEVIC, P. 2011. Effects of motor scale, visual scale, and quantization on small target acquisition difficulty. *ACM Trans. Comput.-Hum. Interact. 18*, 13:1–13:32.

CHERUBINI, M., VENOLIA, G., DELINE, R., AND KO, A. J. 2007. Let's go to the whiteboard: How and why software developers use drawings. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'07)*. 557–566.

DOUGLAS, S., KIRKPATRICK, A., AND MACKENZIE, I. 1999. Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'99)*. 215–222.

DRAPER, N. R. AND SMITH, H. 1998. *Applied Regression Analysis* 3rd Ed. Wiley-Interscience.

DREWES, H. 2010. Only one Fitts' law formula please! In *Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI'10)*. 2813–2822.

EPPS, B. W. 1986. Comparison of six cursor control devices based on Fitts' law models. In *Proceedings of the 30th Annual Meeting of the Human Factors Society*. 327–331.

FERNQUIST, J., SHOEMAKER, G., AND BOOTH, K. S. 2011. "Oh Snap" – Helping users align digital objects on touch interfaces. In *Proceedings of the International Conference on Human-Computer Interaction*. 338–355.

FITTS, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psych. 47*, 6, 381–391.

FORLINES, C. AND BALAKRISHNAN, R. 2008. Evaluating tactile feedback and direct vs. indirect stylus input in pointing and crossing selection tasks. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'08)*. 1563–1572.

FORLINES, C., WIGDOR, D., SHEN, C., AND BALAKRISHNAN, R. 2007. Direct-touch vs. mouse input for tabletop displays. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'07)*. 647–656.

FOWLER, B., MEEHAN, S., AND SINGHAL, A. 2008. Perceptual-motor performance and associated kinematics in space. *Human Factors 50*, 6, 879–892.

GRAHAM, E. D. 1996. Pointing on a computer display. Doctoral dissertation. Simon Fraser University.

GRAHAM, E. D. AND MACKENZIE, C. L. 1996. Physical versus virtual pointing. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'96)*. 292–299.

GROSSMAN, T. AND BALAKRISHNAN, R. 2004. Pointing at trivariate targets in 3D environments. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'04)*. 447–454.

GUIARD, Y. 2009. The problem of consistency in the design of Fitts' law experiments: Consider either target distance and width or movement form and scale. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'09)*. 1809–1818.

GUIARD, Y., BEAUDOUIN-LAFON, M., AND MOTTET, D. 1999. Navigation as multiscale pointing: Extending Fitts' model to very high precision tasks. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'99)*. 450–457.

HAN, S. H., JORNA, G. C., MILLER, R. H., AND TAN, K. C. 1990. A comparison of four input devices for the macintosh interface. In *Proceedings of the Human Factors Society 34th Annual Meeting*. 267–271.

HOFFMAN, E. 1991. A comparison of hand and foot movement times. *Ergonomics 34*, 397–406.

HOLLINGWORTH, H. L. 1909. The inaccuracy of movement. *Arch. Psych. 13*, 1–87.

HORNOF, A. J. 2001. Visual search and mouse pointing in labeled versus unlabeled two-dimensional visual hierarchies. *ACM Trans. Comput.-Hum. Interact. 8*, 3, 171–197.

HUANG, E. AND MYNATT, E. D. 2003. Semi-public displays for small, co-located groups. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'03)*. 49–56.

INKPEN, K. M. 2001. Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Trans. Comput.-Hum. Interact. 8*, 1, 1–33.

IZADI, S., BRIGNULL, H., RODDEN, T., ROGERS, Y., AND UNDERWOOD, M. 2003. Dynamo: a public interactive surface supporting the cooperative sharing and exchange of media. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST'03)*. 159–168.

JOHNSGARD, T. 1994. Fitts' law with a virtual reality glove and a mouse: Effects of gain. In *Proceedings of Graphics Interface '94*. 8–15.

JONES, T. 1989. Psychology of computer use: XVI. Effect of computer-pointing devices on children's processing rate. *Perceptual and Motor Skills 69*, 1259–1263.

JONES, T. 1991. An empirical study of children's use of computer pointing devices. *J. Edu. Comput. Resear. 7*, 1, 61–76.

KERR, R. 1973. Movement time in an underwater environment. *J. Motor Behav. 5*, 175–178.

KEULEN, R. F., ADAM, J. J., FISCHER, M. H., KUIPERS, H., AND JOLLES, J. 2002. Selective reaching: Evidence for multiple frames of reference. *J. Exp. Psych. Hum. Percept. Perform. 28*, 3, 515–526.

KHAN, A., FITZMAURICE, G., ALMEIDA, D., BURTNYK, N., AND KURTENBACH, G. 2004. A remote control interface for large displays. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology (UIST'04)*. 127–136.

KOESTER, H. H., LOPRESTI, E., AND SIMPSON, R. C. 2005. Toward Goldilocks' pointing device: Determining a "just right" gain setting for users with physical impairments. In *Proceedings of the ACM SIGACCESS Conference on Computer and Accessibility (Assets'05)*. 84–89.

KOPPER, R., BOWMAN, D. A., SILVA, M. G., AND MCMAHAN, R. P. 2010. A human motor behavior model for distal pointing tasks. *Int. J. Hum. Comput. Stud. 68*, 603–615.

LANGOLF, G. D., CHAFFIN, D. B., AND FOULKE, J. A. 1976. An investigation of Fitts' law using a wide range of movement amplitudes. *J. Motor Behav. 8*, 113–128.

MACKENZIE, I. S. 1992. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact. 7*, 91–139.

MACKENZIE, I. S. AND BUXTON, W. 1992. Extending Fitts' law to two-dimensional tasks. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'92)*. 219–226.

MACKENZIE, I. S. AND BUXTON, W. 1994. Prediction of pointing and dragging times in graphical user interfaces. *Interact. Comput. 6*, 2, 213–227.

MACKENZIE, I. S. AND ISOKOSKI, P. 2008. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'08)*. 1633–1636.

MACKENZIE, I. S. AND RIDDERSMA, S. 1994. Effects of output display and control-display gain on human performance in interactive systems. *Behav. Inf. Techn. 13*, 328–337.

MACKENZIE, I. S. AND WARE, C. 1993. Lag as a determinant on human performance in interactive systems. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'93)*. 488–493.

MACKENZIE, I. S., SELLEN, A., AND BUXTON, W. 1991. A comparison of input devices in elemental pointing and dragging tasks. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'91)*. 161–166.

MALONE, T. W. 1983. How do people organize their desks?: Implications for the design of office information systems. *ACM Trans. Inf. Syst. 1*, 1, 99–112.

MCGUFFIN, M. J. AND BALAKRISHNAN, R. 2005. Fitts' law and expanding targets: Experimental studies and designs for user interfaces. *ACM Trans. Comput.-Hum. Interact. 12*, 4, 388–422.

MITHAL, A. K. AND DOUGLAS, S. A. 1996. Differences in movement microstructures of the mouse and the finger-controlled isometric joystick. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'96)*. 300–307.

MYERS, B. A., BHATNAGAR, R., NICHOLS, J., PECK, C. H., KONG, D., MILLER, R., AND LONG, A. C. 2002. Interacting at a distance: Measuring the performance of laser pointers and other devices. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'02)*. 33–40.

MYNATT, E. D., IGARASHI, T., EDWARDS, W. K., AND LAMARCA, A. 1999. Flatland: New dimensions in office whiteboards. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'99)*. 346–353.

PO, B. A., FISHER, B. D., AND BOOTH, K. S. 2004. Mouse and touchscreen selection in the upper and lower visual fields. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'04)*. 359–366.

PO, B. A., FISHER, B. D., AND BOOTH, K. S. 2005. Comparing cursor orientations for mouse, pointer, and pen interaction. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'05)*. 291–300.

PRATT, J., ADAM, J. J., AND FISCHER, M. H. 2007. Visual layout modulates Fitts law: The importance of first and last positions. *Psychonomic Bull. Rev. 14*, 2, 350–355.

RADIX, C., ROBINSON, P., AND NURSE, C. 1999. Extension to fitts' law to modeling motion performance in man-machine interfaces. *Trans. Syst. Man Cybern. 29*, 2.

ROGERS, Y. AND LINDLEY, S. 2004. Collaborating around vertical and horizontal large interactive displays: Which is best? *Interact. Comput. 16*, 1133–1152.

RUTLEDGE, J. D. AND SELKER, T. 1990. Force-to-motion functions for pointing. In *Proceedings of the International Conference on Human-Computer Interaction*. 701–706.

SANDFELD, J. AND JENSEN, B. R. 2005. Effect of computer mouse gain and visual demand on mouse clicking performance and muscle activation in a young and elderly group of experienced computer users. *Appl. Ergon. 36*, 547–555.

SCHMIDT, R. A., ZELAZNIK, H., HAWKINS, B., FRANK, J., AND QUINN JR., J. T. 1979. Motor-output variability: A theory for the accuracy of rapid motor acts. *Psych. Rev. 86*, 5, 415–451.

SCHOFIELD, W. N. 1976. Do children find movements which cross the body midline difficult? *The Quarterly Journal of Experimental Psychology 28*, 4, 571–582.

SCHWARZ, G. 1978. Estimating the dimension of a model. *Ann. Statist. 6*, 2, 461–468.

SCOTT, S. D., CARPENDALE, M. S. T., AND INKPEN, K. M. 2004. Territoriality in collaborative tabletop workspaces. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'04)*. 294–303.

SHOEMAKER, G., TANG, A., AND BOOTH, K. S. 2007. Shadow Reaching: A new perspective on interaction for large displays. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology (UIST'07)*. 53–56.

SHOEMAKER, G., FINDLATER, L., DAWSON, J. Q., AND BOOTH, K. S. 2009. Mid-air text input techniques for very large wall displays. In *Proceedings of Graphics Interface '09*. 231–238.

SHOEMAKER, G., TSUKITANI, T., KITAMURA, Y., AND BOOTH, K. S. 2010. Body-centric interaction techniques for very large wall displays. In *Proceedings of the Nordic Conference on Human-Computer Interaction (NordiCHI'10)*. 463–472.

SOUKOREFF, R. W. AND MACKENZIE, I. S. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *Int. J. Hum.-Comput. Stud. 61*, 751–789.

STRUNK, W., JR. AND WHITE, E. B. 1979. *The Elements of Style* 3rd Ed. Macmillan, New York.

TANG, A., LANIR, J., GREENBERG, S., AND FELS, S. 2009. Supporting transitions in work: Informing large display application design by understanding whiteboard use. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work (GROUP'09)*. 149–158.

TSUKITANI, T., SHOEMAKER, G., BOOTH, K. S., TAKASHIMA, K., ITOH, Y., KITAMURA, Y., AND KISHINO, F. 2011. A Fitts' law analysis of shadow metaphor mid-air pointing on a very large wall display. *IPSJ Journal 52*, 4, 1495–1503.

WELFORD, A. T. 1971. *Fundamentals of Skill*. Methuen, London.

WOBBROCK, J. O., CUTRELL, E., HARADA, S., AND MACKENZIE, I. S. 2008. An error model for pointing based on Fitts' law. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'08)*. 1613–1622.

WOODWORTH, R. S. 1899. The accuracy of voluntary movement. *Psych. Rev. 3*, 1–114.

ZHAI, S., CONVERSY, S., BEAUDOUIN-LAFON, M., AND GUIARD, Y. 2003. Human on-line response to target expansion. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'03)*. 177–184.

ZHAI, S., KONG, J., AND REN, X. 2004. Speed-accuracy tradeoff in Fitts' law tasks-on the equivalency of actual and nominal pointing precision. *Int. J. Hum.-Comput. Studies 61*, 6, 823–856.

Received May 2011; revised August 2012; accepted August 2012