

# The Effects of Task Dimensionality, Endpoint Deviation, Throughput Calculation, and Experiment Design on Pointing Measures and Models

Jacob O. Wobbrock, Kristen Shinohara and Alex Jansen

The Information School  
DUB Group  
University of Washington  
Seattle, WA 98195 USA  
{wobbrock, kshino, ajansen7}@uw.edu

## ABSTRACT

Fitts' law (1954) characterizes pointing speed-accuracy performance as *throughput*, whose invariance to target distances ( $A$ ) and sizes ( $W$ ) is known. However, it is unknown whether throughput and Fitts' law models in general are invariant to task dimensionality (1-D vs. 2-D), whether univariate ( $SD_x$ ) or bivariate ( $SD_{x,y}$ ) endpoint deviation is used, whether throughput is calculated using the mean-of-means approach or the slope-inverse approach, or whether Guiard's (2009) *Form*  $\times$  *Scale* experiment design is used instead of fully crossed  $A \times W$  factors. We empirically investigate the confluence of these issues, finding that Fitts' law is largely invariant across 1-D and 2-D, provided that univariate endpoint deviation ( $SD_x$ ) is used in both, but that for 2-D pointing data, bivariate endpoint deviation ( $SD_{x,y}$ ) results in better Fitts' law models. Also, the mean-of-means throughput calculation exhibits lower variance across subjects and dimensionalities than the slope-inverse calculation. In light of these and other findings, we offer recommendations for pointing evaluations, especially in 2-D. We also offer an evaluation tool called *FittsStudy* to facilitate comparisons.

**Author Keywords:** Fitts' law, dimensionality, endpoint deviation, effective target width, throughput, FittsStudy.

**ACM Classification Keywords:** H.5.2. [Information interfaces and presentation]: User interfaces—*theory and methods, evaluation/methodology*.

**General Terms:** Experimentation, Measurement, Theory.

## INTRODUCTION

New pointing techniques regularly emerge, and often, these pointing techniques are evaluated using Fitts' law [7] to measure rapid aimed pointing performance independent of target distance ( $A$ ) and size ( $W$ ). Fitts' law's measure, which combines speed and accuracy, is throughput, whose invariance to the specific values of  $A$  and  $W$  is well-known. However, it is unknown whether throughput is also invariant to task dimensionality (1-D vs. 2-D). Although

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$10.00.

Fitts' original task used vertical ribbons and focused on 1-D horizontal pointing, the ISO 9241-9 standard [13], in recognizing the relevance of pointing to 2-D interfaces, has also prescribed a ring-of-circles target arrangement. In addition, researchers often create their own custom pointing evaluations involving 2-D targets [5], sometimes laid out in a random field [6].

Researchers have created many predictive movement time models for 2-D pointing [2,9,10,16,20]—but despite this work, the literature is almost silent on the issue of throughput calculation in 2-D. For example, it is unknown whether or not applying 1-D throughput calculations to 2-D pointing data results in equivalent throughput outcomes. It is also unknown whether univariate ( $SD_x$ ) or bivariate ( $SD_{x,y}$ ) endpoint deviation creates better models of 2-D pointing. Even the calculation of throughput itself requires further investigation, as the mean-of-means approach [26] insists on averaging indexes of difficulty and movement times, while the slope-inverse approach [30] insists on using the reciprocal of Fitts' regression slope.<sup>1</sup> How each calculation fares across dimensionalities and endpoint deviation schemes is unknown. Additionally, the traditional  $A \times W$  experiment design has recently been exposed by Guiard [11] as introducing possible factor confounds, but the effects of the foregoing issues as they relate to his *Form*  $\times$  *Scale* design are unknown. In short, there are many unanswered and interrelated questions regarding the use of Fitts' law in two-dimensional pointing studies. This paper investigates the confluence of these issues at once.

Endpoint deviation is crucial to the application of Fitts' law. Quantifying endpoint deviation is necessary for employing Crossman's correction [4] for normalizing subjects' personal speed-accuracy biases so that throughputs for a “fast but reckless” performer and a “slow and careful” performer are comparable. The ISO 9241-9 standard [13] and most prior literature [5,15,19,26,27] define the endpoint deviation, or spread of hits, as a univariate quantity computed as the standard deviation of endpoints'  $x$ -coordinates in a horizontal pointing task ( $SD_x$ ). But this definition is questionable in 2-D, as it ignores any deviation in the second dimension ( $\Delta y$ ). The comparative effects on

<sup>1</sup> Both approaches are mentioned, as if interchangeable, in the ISO standard [13].

throughput of using a bivariate standard deviation calculation ( $SD_{x,y}$ ) are unknown. Given Fitts' law's origins as a 1-D model, both definitions of endpoint variability in 2-D tasks seem defensible (*cf.* [5,19]). What, then, should serve as bases for addressing these issues? In this paper, we consider invariance across dimensionalities, improved model fits, and theoretical soundness as possible answers.

This paper contributes a case for, and description of, bivariate endpoint deviation, which easily generalizes to  $N$  dimensions. It also presents results from a study of 21 subjects performing the two ISO 9241-9 pointing tasks in 1-D and 2-D layouts. Throughputs are computed using both the mean-of-means approach [26] and the slope-inverse approach [30]. The effects of univariate and bivariate endpoint deviations on throughput are examined, as are the effects of Guiard's [11] *Form*  $\times$  *Scale* experiment design as an alternative to the traditional  $A \times W$  experiment design.

Our key findings include that Fitts' law is largely invariant across 1-D and 2-D, provided that univariate endpoint deviation ( $SD_x$ ) is used in both; but for 2-D pointing data, bivariate endpoint deviation ( $SD_{x,y}$ ) creates better Fitts' law models than  $SD_x$ . We also find that the mean-of-means throughput calculation exhibits lower variance across subjects and dimensionalities for the same pointing data than the slope-inverse calculation. And Guiard's *Form*  $\times$  *Scale* experiment design produces equivalent results with just one of three levels of  $A$  as the full  $A \times W$  design, resulting in the possibility of major time savings when conducting experiments. We offer a downloadable pointing evaluation tool called *FittsStudy* that provides extensive measures and a visualization tool for exploring submovement pointing profiles.

## RELATED WORK

In this section, we highlight pieces related to endpoint deviation, dimensionality, and throughput.

### Endpoint Deviation

Motor psychologists (*e.g.*, [12]) have distinguished between two sources of error in rapid aimed movements: variable error ( $VE$ ) and constant error ( $CE$ ). Simply put,  $VE$  reflects the spread of hits, or endpoint deviation. By contrast,  $CE$  is the mean distance of endpoints from the target center. Thus,  $VE$  is independent of target location while  $CE$  is not. Crossman's correction [4] relies only on  $VE$ , not  $CE$ .

The ISO 9241-9 standard [13] and many published studies treat pointing in 2-D as having only univariate endpoint deviation ( $SD_x$ ). However, Douglas *et al.* [5] provides an exception, defining 2-D endpoint deviation as a planar distance from a centroid. Unfortunately, they do not explore or discuss this choice as we do here. Murata [21] defined 2-D effective target width ( $W_e$ ) directly using a bivariate joint probability distribution rather than calculating  $W_e$  using bivariate endpoint deviation ( $SD_{x,y}$ ), which is much simpler and which has not been examined until now. Unfortunately, Murata did not compare model fits using his 2-D  $W_e$  to those using a traditional 1-D  $W_e$ , leaving the

comparative utility of his technique unknown. In general, there is still no consensus as to how endpoint deviation or effective target width should be calculated in 2-D (or 3-D).

### Predictive Models of 2-D Pointing

Two-dimensional Fitts' law studies have been conducted for some time (*e.g.*, [14]). Researchers have attempted to predict movement times ( $MTs$ ) to 2-D targets for almost as long, relying on target dimensions [2,16,20] or probability distributions [9,10,24] to inform predictions. Importantly, these efforts were focused on Fitts' law's ability to predict movement times, not its related but distinct measurement of throughput. Therefore, these efforts did not deal with endpoint deviation, speed-accuracy bias correction, or throughput calculation, topics of central importance here.

### Throughput Calculation

Throughput has been fundamental to Fitts' law from the beginning. Fitts himself [7] initially defined it using the mean-of-means approach. Twelve years later, as Zhai points out [30], Fitts and Radford [8] seem to switch gears, referring to the "slope constant" when discussing performance. Card *et al.* [3], who brought Fitts' law into HCI, defined throughput using the slope-inverse approach, *i.e.*,  $1/b$ , where  $b$  is the slope of a regression line through ( $ID$ ,  $MT$ ) points. MacKenzie [15] refers to both the mean-of-means and slope-inverse approaches. Twenty-five years after Card *et al.* [3], MacKenzie and Soukoreff [18] revised Card's calculations using the mean-of-means approach, finding significantly reduced throughputs. In general, Soukoreff and MacKenzie [26] have advocated for the mean-of-means approach, while Zhai [30] has advocated for the slope-inverse approach. The present study brings additional empirical evidence to this debate, especially for 2-D pointing tasks, which have thus far been mostly ignored. Although each throughput calculation results in a bits per second measure, the calculations consider different things. A reasonable approach, then, is to report both throughputs rather than adhere to one or the other.

Only using one  $ID$ , MacKenzie and Isokoski [19] showed that throughput remains constant across cognitive sets [8], provided that Crossman's correction [4] is used. Zhai *et al.* [31], however, has shown that Crossman's correction does not always fully correct divergent speed-accuracy biases. Nonetheless, the correction remains standard for reducing bias and equitably comparing subjects [8,15,19,26,27].

Finally, Guiard [11] has recently argued that the classic crossing of factors  $A \times W$  in Fitts' law experiments introduces possible confounds, and that experiments should use *Form*  $\times$  *Scale* designs instead, which amount to crossing  $ID \times [A \times \text{or} W]$ . Our experiment is among the first to provide results comparing both  $A \times W$  and  $F \times S$  designs.

### ENDPOINT DEVIATION IN MULTIPLE DIMENSIONS

Humans point with different speed-accuracy biases [31] despite being issued a guideline to point as fast as achieving a 4% error rate allows [15,26]. Necessary for correcting these biases is Crossman's correction [4], which enables

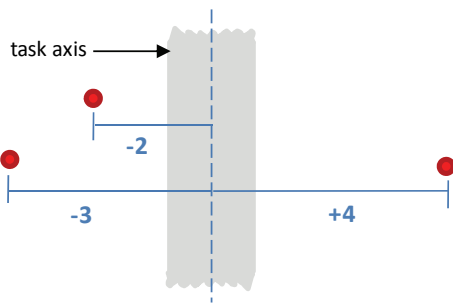
Fitts' throughput, or pointing efficiency, to be properly measured. Whether using the mean-of-means throughput calculation [26] or the slope-inverse calculation [30], throughput depends on this corrected, or "effective," index of difficulty  $ID_e$  defined as

$$ID_e = \log_2(A_e/W_e + 1). \tag{1}$$

Using  $ID_e$  allows subjects making different speed-accuracy tradeoffs to be equitably compared. The faster/slower a subject points, the lower/higher his movement time but the wider/narrower his spread of hits. In Eq. 1,  $A_e$  is the average actual movement distance [26] and  $W_e$  reflects the standard deviation of endpoints as

$$W_e = 4.133 \times SD_x. \tag{2}$$

The constant term arises from the entropy of the standard normal distribution [15,27], but it is the calculation of endpoint deviation ( $SD_x$ ) that interests us. As described elsewhere [13,15,19,26],  $SD_x$  is the standard deviation of endpoint coordinates along the axis of motion—by convention, the "x" indicating horizontal pointing to vertical ribbons. If we record each endpoint's x-coordinate such that negative values are undershoots and positive values are overshoots [26] (footnote 9), we would label endpoints as shown in Figure 1.



**Figure 1.** Vertical ribbon target with two undershoots and one overshoot.

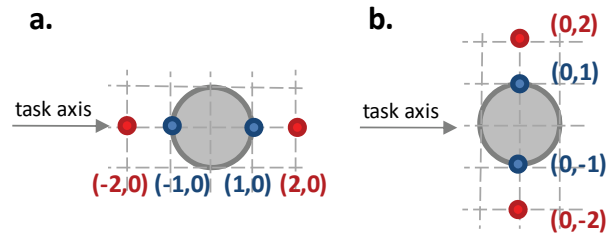
Using the data from Figure 1 as an example, the calculation for  $SD_x$  would be:

$$SD_x = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}, \tag{3}$$

$$= \sqrt{\frac{(-3 + \frac{1}{3})^2 + (-2 + \frac{1}{3})^2 + (4 + \frac{1}{3})^2}{2}} \approx 3.79.$$

The above calculation relies on the mean  $\bar{x}$ , which depends on having *signed* distances from a target center. However, in 2-D (or 3-D) pointing tasks, signed distances to target centers are untenable unless we are to discard all deviation along axes perpendicular to the task axis, turning a multidimensional task into a unidimensional one. Owing to Fitts' law's origins as a 1-D model of rapid aimed movement, it is reasonable to argue that even in 2-D pointing tasks, endpoint deviation should be regarded as univariate [19]. However, this argument forces us to ignore

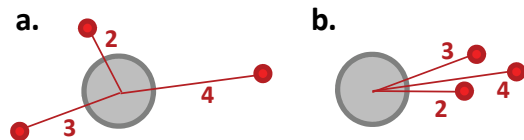
all deviation in one of two task dimensions, sometimes with rather absurd consequences as in Figure 2.



**Figure 2.** Given a horizontal approach, univariate endpoint deviation ( $SD_x$ ) would be (a) rather high and (b) zero even though the 2-D spread of points in both cases is the same.

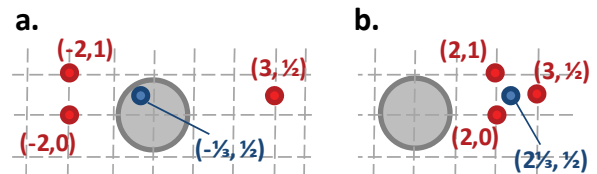
To date, despite years of Fitts' law research, the ramifications of the endpoint deviation calculation are unknown. Intuitively, it seems that endpoint deviation should be accounted for in all relevant task dimensions; after all, pointing to a circular target in 2-D (or spherical target in 3-D) clearly requires motor control in more than one dimension. But how, then, shall we quantify endpoint deviation in 2-D (or 3-D)? And will it produce comparable throughputs to those we observe in 1-D?

Consider Figure 3a, where using signed distances to a target's center is not feasible as it was in Figure 1. How can endpoint deviation be generalized to cover Figure 3? Clearly, *unsigned* distances from a target center do not suffice, lest the endpoint deviation of Figure 3b is considered equal to that of Figure 3a.



**Figure 3.** Endpoints at the same distances from target centers showing (a) a wide spread of hits, (b) a narrow spread of hits. The unsigned distances in (a) are the same as in (b), but clearly (a) exhibits much more endpoint deviation than (b).

Our answer lies in regarding endpoint deviation not as signed or unsigned distances from a target center, but as deviations around a center of mass. Mathematically, this notion is equally suited to 1, 2, 3, and  $N$  dimensions. As before, in Figure 4a and 4b, distances to the target center are the same, but the spread of hits differs. Now, the centroids ( $\bar{x}, \bar{y}$ ) are also marked.



**Figure 4.** Three-point spreads (red) around two targets, and their centroids (blue).

The center of mass, or centroid, serves as a multidimensional mean in the  $SD$  calculation, now given by Eq. 4, where the deviation is the spread of hits around the

centroid. In Eq. 4, we simply have substituted a 2-D distance-from-centroid formula for the 1-D distance-from-mean formula inside the parentheses of Eq. 3:

$$SD_{x,y} = \sqrt{\frac{\sum_{i=1}^N \left( \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \right)^2}{N-1}}. \quad (4)$$

A simplified version of Eq. 4 appears in prior work by Douglas *et al.* [5] (see Eq. 6). Applying Eq. 4 to Figures 2a and 2b yields 1.83 for both. Applying Eq. 4 to Figures 4a and 4b yields 2.93 and 0.76, respectively.

We emphasize that a spread of hits around a centroid is *not* the spread of distances from a centroid, which is a univariate deviation of 2-D distances, not a bivariate deviation of 2-D points. This incorrect definition is given for comparison as  $SD_{wrong}$  in Eq. 5.

$$SD_{wrong} = \sqrt{\frac{\sum_{i=1}^N (d_i - \bar{d})^2}{N-1}}, \text{ where } d_i \text{ is} \quad (5)$$

$$d_i = \sqrt{(x_i - \bar{x})^2} \text{ in 1-D, and} \quad (5a)^2$$

$$d_i = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \text{ in 2-D.} \quad (5b)$$

If Eq. 5b were used for Figure 2a or 2b, we would get a deviation of 0.58. If it were used for Figures 4a and 4b, we would get 0.92 and 0.04, respectively. These values represent the spread of distances from the centroid, not the spread of endpoints around the centroid.

Eq. 4 is easily increased to 3-D or decreased to 1-D:

$$SD_{x,y,z} = \sqrt{\frac{\sum_{i=1}^N \left( \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + (z_i - \bar{z})^2} \right)^2}{N-1}}, \quad (6)$$

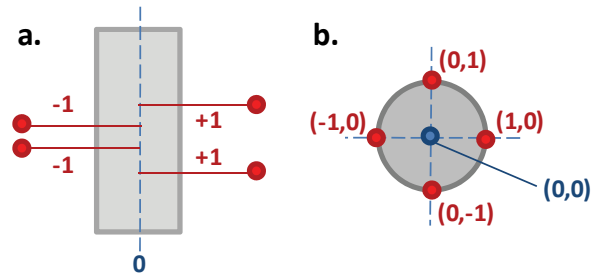
$$SD_x = \sqrt{\frac{\sum_{i=1}^N \left( \sqrt{(x_i - \bar{x})^2} \right)^2}{N-1}} = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}. \quad (7)$$

Eq. 7 shows how the deviation of 1-D coordinates around their centroid is equivalent to a standard deviation of scalars (Eq. 3). With the centroid approach, signed distances from target centers become unnecessary, and instead, 1-D coordinates can be used directly, like 2-D coordinates in Eq. 4, or 3-D coordinates in Eq. 6.

To illustrate the consistency of the centroid formulation for endpoint deviation across dimensions, consider two “equivalent” endpoint distributions in Figure 5.

In Figure 5a, univariate deviation (Eq. 7) produces 1.15. In Figure 5b, bivariate deviation (Eq. 4) also produces 1.15. (By contrast, the deviation-of-distances in Figure 5b using

Eq. 5b produces 0.00 because all points are equally 1 unit from their centroid, so there is no deviation present.)



**Figure 5.** Two endpoint distributions for (a) a 1-D task, (b) a 2-D task. Each totals four units from their respective target center.

Thus, computing endpoint deviation as the spread of points around a center of mass unifies our calculation of deviation regardless of dimensionality. Although empirical endpoint distributions show that deviation tends to be greater along the task axis than orthogonal to it [9,21], for our purposes in quantifying *amount* of deviation, we do not care where the deviation occurs, just *that* it occurs. Endpoints 2 units up from a target center or 2 units right of a target center are each equally 2 units off the mark. Importantly, all approach angles to circular (or spherical) targets are equivalent, removing concern over target shape and approach angle in the ISO 9241-9 standard tasks [13].

Although the above formulation of endpoint deviation is consistent across dimensionalities, it remains to be seen how throughputs from 2-D tasks compare to those from 1-D tasks. Also, it is unknown whether  $SD_x$  or  $SD_{x,y}$  results in better model fits for the same 2-D pointing data, how these issues interact with the two approaches to calculating throughput [26,30], or with Guiard’s contention [11] that  $A \times W$  factorial designs can be confounded. These questions are addressed in our experiment, described next.

## METHOD

### Subjects

Twenty-one subjects participated in our study. Seven were female. All were right-handed. Their average age was 29.3 years ( $SD=6.9$ ).

### Apparatus

All testing was conducted at a lab containing a 21" Samsung SyncMaster 214T flat panel monitor set to 1600×1200 resolution and a Logitech optical mouse. The computer was a Xeon CPU running Windows 7 at 3 GHz with 2 GB RAM.

We created a full-screen application called *FittsStudy* in C# to administer conditions, log data, parse files, visualize trials, and calculate results. *FittsStudy* was configured to administer 18  $A \times W$  conditions defined by 3 levels of  $A$  {256, 384, 512 pixels} crossed with 6 levels of  $W$  {8, 16, 32, 64, 96, 128 pixels} yielding 13 unique *IDs* ranging from 1.59 – 6.02 bits. For the 1-D task,  $A$  was measured from one vertical ribbon’s center to the other, and  $W$  was the width of a ribbon. For the 2-D task, the ISO 9241-9 circular

<sup>2</sup>Note that Eq. 5a *will* work using signed distances from the centroid, *i.e.*,  $d_i = (x_i - \bar{x})$ , because in 1-D, distance is univariate. However, this calculation does not generalize to 2+ dimensions where a distance is not describable by a 1-D value.

arrangement was used [13,26], with  $0.5 \cdot A$  being the radius of the target ring, and  $W$  the diameter of one circular target.

### Procedure

In each of the 18  $A \times W$  conditions per session, a subject performed 23 trials, where each trial was a single attempt to click a target. In each condition, the first three trials were logged as practice and ignored, leaving 20 test trials per condition. There was no delineation between the last practice trial and the first test trial. Presentation order of the 18  $A \times W$  conditions was randomized first by  $A$  and then by  $W$ . Subjects were instructed to move as quickly as possible while maintaining an approximate 4% error rate [15]. Cumulative and condition-level error rates were displayed after each condition. Each trial only contained one click; if a target was missed, it flashed red, audibly dinged, and the next target was immediately highlighted. With each of 21 subjects completing 18  $A \times W$  conditions consisting of 20 test trials in 2 dimensionalities, 15,120 test trials were produced as part of this study.

### Data Analysis

All trials were mathematically rotated to horizontal ( $0^\circ$ ) prior to analysis. Each subject's 18  $A \times W$  conditions resulted in 18 data points plotted as  $(ID_e, MT)$ , where  $MT$  is average movement time and  $ID_e$  is the effective index of difficulty (Eq. 1), normalized with Crossman's correction [4] using  $W_e$  (Eq. 2), which uses endpoint deviation. For the 1-D task,  $SD_x$  was computed using Eq. 7. For the 2-D task, univariate  $SD_x$  was computed using Eq. 7 and bivariate  $SD_{x,y}$  was computed using Eq. 4. Each subject's 18  $(ID_e, MT)$  points were regressed upon to produce a Fitts' law model of the form

$$MT = a + b \cdot ID_e \quad (8)$$

and Pearson  $r$  correlation coefficient.

A subject's throughput in bits per second (bps) was calculated using both the mean-of-means approach [26], which we will designate  $TP_{avg}$ , and the slope-inverse approach [30], which we will designate  $TP_{inv}$ .

$$TP_{avg} = \frac{1}{N} \sum_{i=1}^N \left( \frac{ID_{e_i}}{MT_i} \right), \text{ where } N = |A| \times |W|, \quad (9)$$

$$TP_{inv} = 1/b. \quad (10)$$

Using either approach, the grand throughput for an entire experiment would be the mean throughput over all subjects.

## RESULTS

### Adjustment of Data

As is customary, trials flagged as spatial outliers were removed. A spatial outlier was defined from prior work [19,28] as an error whose movement was less than half the nominal distance  $A$ , or whose endpoint landed more than twice the target width  $W$  from the target center. In all, 9/7560 and 2/7560 test trials were removed from the 1-D and 2-D tasks, respectively.

### Error Rates

Error rates for Fitts' law studies should be around 4% [15]. Appropriately, over all  $A \times W$  conditions, the error rate for the 1-D task was 4.4% and for the 2-D task was 4.3%.

### Comparison of Fitts' Law Models ( $r$ , $a$ , $b$ )

Space precludes showing Fitts' law models for all 21 subjects, but average model fits (Pearson  $r$ ) and parameters ( $a$  and  $b$ ) for both the 1-D vertical ribbon task and the 2-D ring-of-circles task are shown in Table 1. For the 2-D task, the exact same pointing data was analyzed using univariate ( $SD_x$ ) and bivariate ( $SD_{x,y}$ ) endpoint deviation.

The average fit of the models for the 1-D task was  $r=.948$ .<sup>3</sup> The average fit for the 2-D task using univariate endpoint deviation ( $SD_x$ ) was  $r=.951$ . For the same 2-D data, using bivariate endpoint deviation ( $SD_{x,y}$ ) resulted in  $r=.962$ . Paired samples  $t$ -tests show that 2-D bivariate  $r$  values were significantly higher than both the 1-D and 2-D univariate  $r$  values ( $p<.05$ ), and that the univariate  $r$  values were not significantly different from each other.

The standard deviation of subjects' bivariate  $r$  values was lower than their univariate  $r$  values for the 2-D task (.016 vs. .021, respectively) given the exact same pointing data.

For model parameters, it is generally desirable to have intercepts  $a$  that are nonnegative and as close to zero as possible [26,30]. Intercepts in both 2-D models were not significantly different from those in the 1-D model, but were marginally different from each other ( $t(20)=2.04$ ,  $p=.05$ ). The 1-D, 2-D univariate, and 2-D bivariate models produced 5, 4, and 3 negative intercepts, respectively. Tests of slope parameters  $b$  showed that the 1-D and 2-D univariate slopes were not significantly different, but that both were different from the 2-D bivariate slopes ( $p<.01$ ).

### Throughput Analyses

As described above, the mean-of-means approach ( $TP_{avg}$ , Eq. 9) [26] computes the average bps over  $A \times W$  conditions, and therefore accounts for the contributions of both  $a$  and  $b$ . However, averages may be affected by the range of  $ID$ s used, and  $a$  and  $b$  are conflated such that averages may be the same whose slopes ( $b$ ) differ due to the influence of intercepts ( $a$ ). By contrast, the slope-inverse approach ( $TP_{inv}$ , Eq. 10) [30] defines throughput as  $1/b$ . However, this approach must report  $a$  separately, and has been shown to depend upon the range of  $ID$ s used [26].

### Univariate vs. Bivariate Throughputs ( $TP_{avg}$ )

Using the mean-of-means approach, the average throughput for the 1-D task was  $TP_{avg}=4.85$  bps ( $SD=0.41$ ). The average throughput for the 2-D task with univariate endpoint deviation was  $TP_{avg}=4.91$  bps ( $SD=0.43$ ). A paired samples  $t$ -test indicates that these throughputs were not significantly different, indicating that 1-D and 2-D univariate throughputs were comparable despite changes in

<sup>3</sup> As Fitts himself observed [8], it is typical and appropriate for models using corrected  $ID_e$  instead of nominal  $ID$  to exhibit lower correlation coefficients (cf. [26,31]).

<i>A</i> × <i>W</i> design	1-D Ribbon Task, Univariate $SD_x$					2-D Circle Task, Univariate $SD_x$							2-D Circle Task, Bivariate $SD_{x,y}$						
	<i>a</i>	<i>b</i>	<i>r</i>	$TP_{avg}$	$TP_{inv}$	<i>a</i>	<i>b</i>	<i>r</i>	$TP_{avg}$	$TP_{inv}$	$\Delta TP_{avg}$	$\Delta TP_{inv}$	<i>a</i>	<i>b</i>	<i>r</i>	$TP_{avg}$	$TP_{inv}$	$\Delta TP_{avg}$	$\Delta TP_{inv}$
Mean	98.36	187.56	.948	4.85	5.42	74.48	192.65	.951	4.91	5.28	0.21	0.56	72.90	207.28	.962	4.49	4.90	0.36	0.72
StdDev	66.55	24.67	.027	0.41	0.71	56.01	27.09	.021	0.43	0.68	0.16	0.44	52.33	27.63	.016	0.40	0.61	0.24	0.50

**Table 1.** Summary results over 21 subjects for Fitts' law regression coefficients (*a* and *b*), Pearson *r* model fits, and throughput calculations ( $TP_{avg}$  and  $TP_{inv}$ ) for 1-D pointing to vertical ribbons and for 2-D pointing to circles in a ring [13]. The exact same data is used for each 2-D analysis. The  $\Delta TP_{avg}$  and  $\Delta TP_{inv}$  columns show the absolute differences in throughputs from their respective 1-D throughputs. Because Fitts' law intercepts can be negative, means and standard deviations for those columns are of  $|a|$ , i.e., intercepts' distance from zero.

dimensionality. The average absolute difference between subjects' 2-D univariate throughput and their 1-D throughput was  $\Delta TP_{avg}=0.21$  bps ( $SD=0.16$ ).

The mean of 2-D bivariate throughputs was  $TP_{avg}=4.49$  bps ( $SD=0.40$ ), which was significantly lower than 1-D throughputs ( $t(20)=-6.89$ ,  $p<.0001$ ). It was also lower than the 2-D univariate throughput ( $t(20)=-33.43$ ,  $p<.0001$ ). The average absolute difference between subjects' 2-D bivariate throughputs and their 1-D throughputs was  $\Delta TP_{avg}=0.36$  bps ( $SD=0.24$ ). These differences were significantly greater than the 2-D univariate differences ( $t(20)=2.65$ ,  $p<.05$ ).

Thus, it seems that under the mean-of-means approach to calculating throughput, for 2-D task data, univariate endpoint deviation produces throughputs closer to those obtained in the classic 1-D task. However, 2-D bivariate models fit the same data better than 2-D univariate models.

#### Univariate vs. Bivariate Throughputs ( $TP_{inv}$ )

Using the slope-inverse approach, the average throughput for the 1-D task was  $TP_{inv}=5.42$  bps ( $SD=0.71$ ). The average throughput for the 2-D task with univariate endpoint deviation was  $TP_{inv}=5.28$  bps ( $SD=0.68$ ). These throughputs were not significantly different, indicating comparability across dimensionality as occurred for the mean-of-means approach. The average absolute difference between subjects' 2-D univariate throughput and their 1-D throughput was  $\Delta TP_{inv}=0.56$  bps ( $SD=0.44$ ).

In comparison, the average 2-D bivariate throughput was  $TP_{inv}=4.90$  bps ( $SD=0.61$ ), which, like for the mean-of-means approach, was significantly lower than both 1-D throughput ( $t(20)=-3.35$ ,  $p<.01$ ) and 2-D univariate throughput ( $t(20)=-11.11$ ,  $p<.0001$ ). The average absolute difference between subjects' 2-D bivariate throughputs and their 1-D throughputs was  $\Delta TP_{inv}=0.72$  bps ( $SD=0.50$ ). As with the mean-of-means approach, these differences were significantly greater than the 2-D univariate differences ( $t(20)=2.18$ ,  $p<.05$ ).

Thus, the findings from the slope-inverse approach to calculating throughput agree with those from the mean-of-means approach, but with different magnitudes.

#### Throughput Calculation Approaches ( $TP_{avg}$ vs. $TP_{inv}$ )

Thus far, we have compared univariate ( $SD_x$ ) and bivariate ( $SD_{x,y}$ ) endpoint deviation calculations for 2-D task data using (i) model fits and (ii) similarity to 1-D throughputs. While this was our main interest, we also reflect on the two throughput calculations,  $TP_{avg}$  and  $TP_{inv}$ . We do not advocate for one or the other here, but lend some empirical

evidence to them, as prior comparisons have turned largely on principled theoretical concerns [26,30].

First, we compare the throughputs  $TP_{avg}$  and  $TP_{inv}$  from Table 1. For 1-D throughputs, at 5.42 vs. 4.85 bps,  $TP_{inv}$  was significantly higher among subjects than  $TP_{avg}$  ( $t(20)=3.37$ ,  $p<.01$ ). The same was true of 2-D univariate throughputs, at 5.28 vs. 4.91 bps ( $t(20)=2.88$ ,  $p<.01$ ), and 2-D bivariate throughputs, at 4.90 vs. 4.49 bps ( $t(20)=3.92$ ,  $p<.001$ ). Clearly, the slope-inverse calculation yields higher throughputs than the mean-of-means calculation.

Besides producing higher throughputs,  $TP_{inv}$  also had higher variance among subjects than  $TP_{avg}$ . For the 1-D task, standard deviations were 0.71 vs. 0.41, respectively. For the 2-D task using univariate endpoint deviation, standard deviations were 0.68 vs. 0.43; using bivariate endpoint deviation, standard deviations were 0.61 vs. 0.40.

Second, while we learned that 2-D univariate endpoint deviation produces more agreement with 1-D throughputs than 2-D bivariate endpoint deviation, we have not considered whether this result is stronger for  $TP_{avg}$  or  $TP_{inv}$ . The average absolute difference between 1-D  $TP_{inv}$  and 2-D univariate  $TP_{inv}$  was 0.56 bps, while for  $TP_{avg}$  it was 0.21 bps, a significant difference ( $t(20)=3.47$ ,  $p<.01$ ). For the 2-D bivariate case,  $\Delta TP_{inv}$  was 0.72 bps, while  $\Delta TP_{avg}$  was 0.36 bps, again a significant difference ( $t(20)=2.90$ ,  $p<.01$ ). Thus, in both cases,  $TP_{avg}$  agrees better between 1-D and 2-D throughputs than does  $TP_{inv}$ .

#### Form × Scale (Guiard 2009)

Recently, Guiard [11] has argued that the typical procedure of separately manipulating *A* and *W* in Fitts' law experiments introduces potential confounds. He contends that the universally agreed upon variable of interest in Fitts' law, and the sole determiner of movement time (*MT*), is index of difficulty (*ID*), which is expressed using the equivocal ratio *A/W*. But typical experimental treatment of Fitts' law regards it as if it had three parameters—*A*, *W*, and *ID*—which introduces a confound when *A* or *W* has a disproportionate effect on *MT*. (As an example, Guiard points to Fitts' original disc-transfer experiment [7].) A remedy, Guiard argues, is to instead manipulate *Form* (*F*), which is *ID*, and *Scale* (*S*), which is either *A* or *W* but not both. Guiard's practical recommendation is to hold *A* constant and manipulate only *W* over the course of an experiment, thereby removing any potential confound.

In light of Guiard [11], it is conceivable that our throughput results could be affected by the traditional *A* × *W* design, and that an *F* × *S* design is worth exploring. We can retroactively

$F \times S$ design	1-D Ribbon Task, Univariate $SD_x$					2-D Circle Task, Univariate $SD_x$							2-D Circle Task, Bivariate $SD_{x,y}$						
	$a$	$b$	$r$	$TP_{avg}$	$TP_{inv}$	$a$	$b$	$r$	$TP_{avg}$	$TP_{inv}$	$\Delta TP_{avg}$	$\Delta TP_{inv}$	$a$	$b$	$r$	$TP_{avg}$	$TP_{inv}$	$\Delta TP_{avg}$	$\Delta TP_{inv}$
Mean	144.69	184.79	.956	4.87	5.66	84.28	198.50	.962	4.89	5.16	0.25	1.03	97.57	214.47	.969	4.49	4.78	0.43	1.22
StdDev	100.52	41.94	.053	0.41	1.19	66.10	32.86	.025	0.43	0.81	0.20	0.70	64.88	36.08	.021	0.40	0.74	0.24	0.74

**Table 2.** Summary results over 21 subjects for the  $Form \times Scale$  design [11] using only one level of  $A$  (384 pixels) and all six levels of  $W$ . The data used for each 2-D analysis is exactly the same. Column definitions are the same as for Table 1.

apply Guiard's suggestion by retaining, for example, our middle  $A$  value (384 pixels) and ignoring trials for which  $A$  was 256 or 512 pixels. This leaves 1  $A \times 6 W$  for six ( $ID_e, MT$ ) points per subject. Our range of nominal  $ID$ s becomes 2.00 – 5.61 bits. Our total test trials, excluding outliers, becomes 5039, about a third of our original amount. We re-ran our regressions and throughput analyses, this time keeping only trials for which  $A=384$  pixels. Summary outcomes for our new analyses are shown in Table 2.

#### Comparison of Fitts' Law Models ( $r, a, b$ )

Although the average Pearson  $r$  model fits improved slightly in the  $F \times S$  design over the  $A \times W$  design, none of these improvements were statistically significant. As for the  $A \times W$  design, the two univariate fits were not significantly different, and the 2-D bivariate fits were significantly higher than the 2-D univariate fits ( $t(20)=2.47, p<.05$ ). Unlike in the  $A \times W$  design, the 2-D bivariate fits were no longer better than the 1-D fits. As before, the standard deviation of model fits among subjects was low, with the 2-D bivariate model being lower than the 2-D univariate model (.021 vs. .025).

Although the average absolute intercept  $|a|$  increased in the  $F \times S$  design from the  $A \times W$  design for each of the three models, these increases were not statistically significant. The standard deviation among subjects of  $|a|$  increased as well, and there were more negative intercepts than previously. The 2-D univariate and bivariate intercepts were not significantly different from each other or from the 1-D intercepts.

Slope parameters  $b$  did not change significantly in the  $F \times S$  design from the  $A \times W$  design for each of the three models. As in the  $A \times W$  design, in the  $F \times S$  design the 1-D and 2-D univariate slopes were not significantly different from each other, but both were significantly different from the 2-D bivariate slopes ( $p<.01$ ).

#### Univariate vs. Bivariate Throughputs ( $TP_{avg}$ )

Using the mean-of-means approach to throughput calculation (Eq. 9), the  $F \times S$  design produced almost identical  $TP_{avg}$  values to those from the  $A \times W$  design. Within the  $F \times S$  design, as in the full  $A \times W$  design, 2-D univariate throughputs were not significantly different from 1-D throughputs. However, as in the  $A \times W$  design, 2-D bivariate throughputs were significantly lower than 1-D throughputs ( $t(20)=-5.64, p<.0001$ ). As before, they were also lower than 2-D univariate throughputs ( $t(20)=-23.19, p<.0001$ ). Finally, while the absolute differences in 2-D throughputs from 1-D throughputs ( $\Delta TP_{avg}$ ) were not significantly different from those in the  $A \times W$  design, as in the  $A \times W$  design, 2-D bivariate throughputs were

significantly further from 1-D throughputs than 2-D univariate throughputs ( $t(20)=2.59, p<.05$ ). Thus, despite switching to the  $F \times S$  design from the  $A \times W$  design, the significance conclusions hold the same for  $TP_{avg}$ .

#### Univariate vs. Bivariate Throughputs ( $TP_{inv}$ )

Using the slope-inverse approach to throughput calculation, the  $F \times S$  design produced similar  $TP_{inv}$  values to those from the  $A \times W$  design. Within the  $F \times S$  design, as in the  $A \times W$  design, 2-D univariate throughputs were still not significantly different from 1-D throughputs, but this time the result was only marginal ( $t(20)=-1.98, p=.06$ ). As in the  $A \times W$  design, 2-D bivariate throughputs were significantly lower than 1-D throughputs ( $t(20)=-3.59, p<.01$ ) and 2-D univariate throughputs ( $t(20)=-9.10, p<.0001$ ).

Until now, the findings from the  $F \times S$  design mirror those from the  $A \times W$  design for both  $TP_{avg}$  and  $TP_{inv}$ . However, the story changes when we examine  $\Delta TP_{inv}$ . These differences grew significantly in the  $F \times S$  design from the  $A \times W$  design creating, on average, a discrepancy of over 1 bps per subject for both univariate ( $t(20)=3.22, p<.01$ ) and bivariate ( $t(20)=3.47, p<.01$ ) endpoint deviation. Thus,  $TP_{inv}$  was affected far more by the reduction of  $A$  to one level than was  $TP_{avg}$ . This is important in light of Zhai's [30] contention that  $TP_{avg}$  is affected more than  $TP_{inv}$  by the range of  $ID$ s used, and Soukoreff and MacKenzie's counter to the contrary [26]. Finally, as in the  $A \times W$  design, 2-D bivariate throughputs were significantly further from 1-D throughputs than 2-D univariate throughputs ( $t(20)=2.33, p<.05$ ). Thus, in switching to the  $F \times S$  design from the  $A \times W$  design, the significance conclusions are the same for  $TP_{inv}$  but the discrepancy between 2-D and 1-D throughputs grew significantly to over 1 bps.

#### Throughput Calculation Approaches ( $TP_{avg}$ vs. $TP_{inv}$ )

In the full  $A \times W$  design, slope-inverse throughputs ( $TP_{inv}$ ) were significantly higher than mean-of-means throughputs ( $TP_{avg}$ ) for all three models ( $p<.05$ ). However, in the  $F \times S$  design, this relationship held for only 1-D throughputs ( $t(20)=2.89, p<.01$ ). In the 2-D univariate model, although average  $TP_{inv}$  was greater than  $TP_{avg}$ , the difference was only marginal ( $t(20)=1.70, p=.10$ ). In the 2-D bivariate model, the difference was also only marginal ( $t(20)=1.93, p=.07$ ). The marginal results are at least due in part to the greatly increased variance among subjects'  $TP_{inv}$  values in the  $F \times S$  design, which was much higher than for  $TP_{avg}$ .

In the  $A \times W$  design, 2-D  $TP_{avg}$  agreed more with 1-D  $TP_{avg}$  than did 2-D  $TP_{inv}$  with 1-D  $TP_{inv}$  for both the univariate and bivariate models. In the  $F \times S$  design, the same was true, with  $\Delta TP_{avg} < \Delta TP_{inv}$  for univariate models ( $t(20)=5.05, p<.0001$ ) and bivariate models ( $t(20)=4.83, p<.0001$ ).

The resiliency of Fitts' law to changes in experimental settings is often touted as part of the law's strength. Here we have seen that in light of Guiard's [11] concern over confounding the effects on  $MT$  of  $A$ ,  $W$ , and  $ID$ , switching to a  $F \times S$  design by holding  $A$  constant at 384 pixels and varying only  $W$  has largely produced the same significance conclusions with a third of the experimental effort.

## DISCUSSION

To facilitate our discussion, we first restate our key findings, each of which had statistically significant support. Findings A-F are from the full  $A \times W$  design, while finding G specifically addresses Guiard's  $F \times S$  design [11].

- A. *Fitts' law intercepts transcend dimensionality but slopes change with endpoint deviation.* Model intercepts  $a$  did not change significantly from 1-D to 2-D tasks, while slopes  $b$  changed significantly from univariate to bivariate endpoint deviation calculations.
- B. *Bivariate models fit 2-D pointing data best.* On the exact same 2-D pointing data, bivariate endpoint deviation gave better model fits, as judged by Pearson  $r$  coefficients, than univariate endpoint deviation.<sup>4</sup>
- C. *2-D univariate throughputs are closest to 1-D univariate throughputs.* Using univariate endpoint deviation for 2-D pointing data gave throughputs closer to 1-D throughputs than using bivariate endpoint deviation under both the mean-of-means throughput calculation and the slope-inverse calculation.
- D. *Bivariate models produce lower throughputs.* Using bivariate endpoint deviation produces lower throughputs than using univariate endpoint deviation for the same 2-D pointing data under both throughput calculation schemes. This should be no surprise, as in bivariate models, deviation is being accumulated that is ignored in univariate models. Using bivariate endpoint deviation increases  $W_e$ , decreasing  $ID_e$  for the same  $MT$  and thus lowering throughput.
- E. *Slope-inverse throughputs are highest and exhibit most variance.* In the  $A \times W$  design, slope-inverse throughputs were higher than mean-of-means throughputs in both dimensionalities and under both endpoint deviation schemes. They also exhibited greater variance among subjects.
- F. *Mean-of-means throughputs agree most across dimensionalities.* Mean-of-means throughputs calculated from 2-D pointing data are closer to 1-D throughputs calculated the same way than are slope-inverse throughputs. This finding holds for both univariate and bivariate 2-D endpoint deviations.
- G. *Guiard's [11] Form  $\times$  Scale design largely upholds the above findings.* Switching to Guiard's  $F \times S$  design

instead of the traditional  $A \times W$  design did not significantly change model fits or parameters, and findings A-D and F were categorically upheld. Finding E was upheld only for the 1-D case, although for the 2-D cases, the results were both marginal ( $.05 \leq p \leq .10$ ). Additionally, the  $F \times S$  design showed that the slope-inverse approach was more sensitive to the reduced  $ID$  range than the mean-of-means approach: 2-D throughputs became significantly further from 1-D throughputs, and variance among subjects increased considerably.

At this point, we are ready to ask: (i) how resilient is Fitts' law to changes in dimensionality?, and (ii) should univariate or bivariate endpoint deviation be used in 2-D studies? We take each of these questions in turn.

### Fitts' Law across Dimensionalities

Is Fitts' law invariant to dimensionality? On this, our findings suggest "yes" if univariate endpoint deviation is used, and "no" if bivariate endpoint deviation is used. Put another way, as long as a 2-D task is treated with 1-D calculations, the results will be comparable to 1-D tasks. Evidence lies in the stability across dimensionalities of model intercepts and slopes (finding A) and throughputs (finding C). However, if bivariate calculations are used, then slopes (but not intercepts), model fits, and throughputs all change significantly across dimensionalities (findings A-D). However, these results do not necessarily mean that univariate deviation is most appropriate for 2-D pointing data; other considerations besides agreement with 1-D pointing studies must be weighed carefully.

Which throughput calculation seems most stable across dimensionalities?  $TP_{avg}$  produced throughputs with greater agreement across 1-D and 2-D than did  $TP_{inv}$  (finding F).

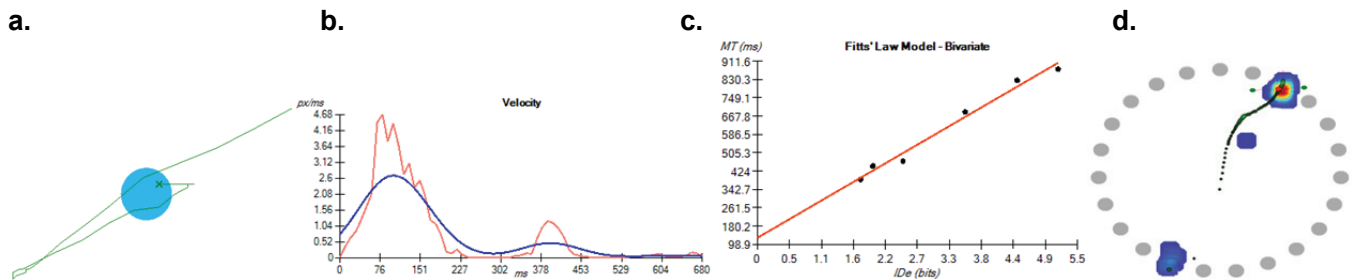
### Univariate or Bivariate Endpoint Deviation in 2-D?

Above, we saw that univariate endpoint deviation calculations for 2-D pointing data cause model parameters and throughputs to better agree with 1-D pointing results. In effect, by using univariate endpoint deviation, we are treating a 2-D task as if it were a 1-D task. But does this mean univariate endpoint deviation should be used, even in 2-D pointing studies?

Agreement across dimensionalities may be an important consideration. At the same time, there are both empirical and theoretical arguments that commend a bivariate approach. Empirically, bivariate endpoint deviation produces better model fits for 2-D pointing data than univariate deviation (finding B), and as noted in footnote 4, this cannot be attributed to an increase in model parameters. Better model fits using the same data and number of parameters are no small concern, for we usually want to prize models with the best possible explanatory power. Theoretically, there is satisfaction knowing that bivariate models do not simply ignore variation in a physically relevant dimension. Although Fitts' law is a model with only one dimension for target size, this does not mean that

<sup>4</sup>Note that the better model fits cannot be attributed to an additional model parameter. All linear regression models are still of the form  $MT = a + b \cdot ID_e$ . Endpoint deviation is a dimensionless scalar whether it is calculated from one coordinate or two.





**Figure 6.** The *FittsStudy* tool offers (a) zoomable path visualizations and path analysis measures [17], (b) graphical submovement profiles including velocity (shown), acceleration, and jerk over time, and (c) Fitts' law models (shown) and pointing error models [28] using both endpoint deviation schemes and throughput calculation approaches [26,30]. (d) An example project by students using *FittsStudy* shows an eye-tracking "heat map" overlaying *FittsStudy*'s ring-of-circles task.

targets or movements must be one-dimensional, just that size (nominal or effective) must be quantified with a dimensionless scalar reflecting *quantity* of tolerance ( $W$ ) or deviation ( $W_e$ ). Both  $SD_x$  and  $SD_{x,y}$  achieve this.

### Recommendations

In light of our findings, we recommend the following:

1. Consider using an  $F \times S$  design by choosing one level of  $A$  and varying only  $W$ . Doing so will save considerable effort and reduce the chances of factor confounds. Alternatively, use a full  $A \times W$  design but, afterwards, isolate a middle  $A$  value while retaining all  $W$  values and reanalyze the data as if it came from an  $F \times S$  design. If the conclusions change, report them and investigate a disproportionate effect of  $A$  or  $W$  on  $MT$ .
2. For 2-D pointing studies, use bivariate endpoint deviation ( $SD_{x,y}$ ), but examine results from univariate endpoint deviation ( $SD_x$ ) and report them if they significantly differ.
3. If a goal is to compare results between a 1-D task and a 2-D task, compute 2-D throughputs using the mean-of-means approach ( $TP_{avg}$ ) based on univariate endpoint deviation ( $SD_x$ ). Doing so will cause throughputs across dimensionalities to agree best.
4. Comparisons should generally not be drawn across throughput calculation approaches, *i.e.*, comparing throughputs from the mean-of-means approach ( $TP_{avg}$ ) to those from the slope-inverse approach ( $TP_{inv}$ ).
5. Given the prevalence of 2-D pointing in user interfaces, we encourage the use of the ISO 9241-9 ring-of-circles task in lieu of the vertical ribbon task.

### The *FittsStudy* Tool

To facilitate adoption of the foregoing recommendations, we built a downloadable software tool called *FittsStudy* (Figure 6). *FittsStudy* logs all pointing trials in a session in XML, from which it builds spreadsheets for statistical analysis. Although Fitts' law evaluation tools have been described in the literature [5,22,25], *FittsStudy* offers more features than previous tools, providing, among other things, trial browsing and zooming tools, submovement profile graphs, results from both endpoint deviation and throughput calculations, path analysis measures [17], a metronome for time-matching temporal-precision studies (*e.g.*, [23]), and

models of movement times and error rates [28,29]. Future work in *FittsStudy* includes offering not just the vertical ribbon and ring-of-circles tasks, but also random target fields (*e.g.*, [6]), and crossing and steering tasks [1]. *FittsStudy* can be downloaded for free from <http://depts.washington.edu/aimgroup/proj/fittsstudy/>.

### FUTURE WORK

This work has made progress in clarifying the effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on Fitts' law models and measures. However, it has relied on the two suggested ISO 9241-9 task arrangements [13], both of which use regularly-shaped targets. How endpoint deviation should be calculated for irregularly-shaped targets is a topic for future work, and will almost certainly involve probabilistic accounts of endpoint distributions [9,10]. Another item for future work is examining whether the findings comparing outcomes in 1-D and 2-D transfer to 3-D pointing environments. Yet another project could investigate whether different device types other than the mouse change the pattern of results seen here.

### CONCLUSION

The strength of our quantitative methods depends on deep understandings of our measures. In this paper, we have addressed Fitts' throughput and its relationship to task dimensionality, endpoint deviation, throughput calculation, and experiment design. Our findings indicate that Fitts' law is largely invariant to dimensionality (1-D vs. 2-D), but that this depends on using univariate endpoint deviation ( $SD_x$ ), even in 2-D, which ignores deviation in the orthogonal task dimension. Bivariate endpoint deviation ( $SD_{x,y}$ ), on the other hand, better models 2-D pointing data. We also offer support for using Guiard's [11]  $F \times S$  experiment design to save effort and avoid possible confounds. We built *FittsStudy* to facilitate exploration of these issues and ease the burden of these details on researchers and evaluators.

### ACKNOWLEDGEMENTS

The authors thank Ed Cutrell, I. Scott MacKenzie, T. Scott Saponas and Desney Tan. This work was supported in part by Intel Labs, Microsoft Research, and the National Science Foundation under grants IIS-0811063 and IIS-0952786.

## REFERENCES

1. Accot, J. and Zhai, S. (1997). Beyond Fitts' law: Models for trajectory-based HCI tasks. *Proc. CHI '97*. New York: ACM Press, 295-302.
2. Accot, J. and Zhai, S. (2003). Refining Fitts' law models for bivariate pointing. *Proc. CHI '03*. New York: ACM Press, 193-200.
3. Card, S.K., English, W.K. and Burr, B.J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics* 21 (8), 601-613.
4. Crossman, E.R.F.W. (1957). The speed and accuracy of simple hand movements. In *The Nature and Acquisition of Industrial Skills*, Crossman and Seymour (eds.). Report to the Joint Committee on Individual Efficiency in Industry.
5. Douglas, S.A., Kirkpatrick, A.E. and MacKenzie, I.S. (1999). Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. *Proc. CHI '99*. New York: ACM Press, 215-222.
6. Findlater, L., Jansen, A., Shinohara, K., Dixon, M., Kamb, P., Rakita, J. and Wobbrock, J.O. (2010). Enhanced area cursors: Reducing fine-pointing demands for people with motor impairments. *Proc. UIST '10*. New York: ACM Press, 153-162.
7. Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J. Experimental Psychology* 47 (6), 381-391.
8. Fitts, P.M. and Radford, B.K. (1966). Information capacity of discrete motor responses under different cognitive sets. *J. Experimental Psychology* 71 (4), 475-482.
9. Grossman, T. and Balakrishnan, R. (2005). A probabilistic approach to modeling two-dimensional pointing. *ACM Transactions on Computer-Human Interaction* 12 (3), 435-459.
10. Grossman, T., Kong, N. and Balakrishnan, R. (2007). Modeling pointing at targets of arbitrary shapes. *Proc. CHI '07*. New York: ACM Press, 463-472.
11. 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. *Proc. CHI '09*. New York: ACM Press, 1809-1818.
12. Hancock, P.A. and Newell, K.M. (1985). The movement speed-accuracy relationship in space-time. In *Motor Behavior: Programming, Control, and Acquisition*, Heuer, Kleinbeck and Schmidt (eds.). New York: Springer-Verlag, 153-188.
13. International Organization for Standardization. (2002). Ergonomic requirements for office work with visual display terminals (VDTs)—Requirements for non-keyboard input devices. *Ref. No. ISO 9241-9:2000(E)*.
14. Jagacinski, R.J. and Monk, D.L. (1985). Fitts' law in two dimensions with hand and head movements. *J. Motor Behavior* 17 (1), 77-95.
15. MacKenzie, I.S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction* 7 (1), 91-139.
16. MacKenzie, I.S. and Buxton, W. (1992). Extending Fitts' law to two-dimensional tasks. *Proc. CHI '92*. New York: ACM Press, 219-226.
17. MacKenzie, I.S., Kauppinen, T. and Silfverberg, M. (2001). Accuracy measures for evaluating computer pointing devices. *Proc. CHI '01*. New York: ACM Press, 9-16.
18. MacKenzie, I.S. and Soukoreff, R.W. (2003). Card, English, and Burr (1978)—25 years later. *Extended Abstracts CHI '03*. New York: ACM Press, 760-761.
19. MacKenzie, I.S. and Isokoski, P. (2008). Fitts' throughput and the speed-accuracy tradeoff. *Proc. CHI '08*. New York: ACM Press, 1633-1636.
20. Murata, A. (1996). Empirical evaluation of performance models of pointing accuracy and speed with a PC mouse. *Int'l J. Human-Computer Interaction* 8 (4), 457-469.
21. Murata, A. (1999). Extending effective target width in Fitts' law to a two-dimensional pointing task. *Int'l J. Human-Computer Interaction* 11 (2), 137-152.
22. Schedlbauer, M. (2007). An extensible platform for the interactive exploration of Fitts' law and related movement time models. *Extended Abstracts CHI '07*. New York: ACM Press, 2633-2638.
23. Schmidt, R.A., Zelaznik, H., Hawkins, B., Frank, J.S. and Quinn, J.T. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review* 86 (5), 415-451.
24. Sheikh, I.H. and Hoffmann, E.R. (1994). Effect of target shape on movement time in a Fitts task. *Ergonomics* 37 (9), 1533-1547.
25. Soukoreff, R.W. and MacKenzie, I.S. (1995). Generalized Fitts' law model builder. *Companion to CHI '95*. New York: ACM Press, 113-114.
26. 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'l J. Human-Computer Studies* 61 (6), 751-789.
27. Welford, A.T. (1968). *Fundamentals of Skill*. London, England: Methuen.
28. Wobbrock, J.O., Cutrell, E., Harada, S. and MacKenzie, I.S. (2008). An error model for pointing based on Fitts' law. *Proc. CHI '08*. New York: ACM Press, 1613-1622.
29. Wobbrock, J.O., Jansen, A. and Shinohara, K. (2011). Modeling and predicting pointing errors in two dimensions. *Proc. CHI '11*. New York: ACM Press. To appear.
30. Zhai, S. (2004). Characterizing computer input with Fitts' law parameters—the information and non-information aspects of pointing. *Int'l J. Human-Computer Studies* 61 (6), 791-809.
31. 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'l J. Human-Computer Studies* 61 (6), 823-856.