UC Irvine
UCI Open Access Publishing Fund

Title
Issues Related to HCI Application of Fitts's Law

Permalink
https://escholarship.org/uc/item/7vh7z6hk

Journal
Human-Computer Interaction, 28(6)

ISSN
1532-7051

Authors
Wright, Charles E
Lee, Francis

Publication Date
2013-11-02

DOI
10.1080/07370024.2013.803873

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

HUMAN–COMPUTER INTERACTION, 2013, Volume 28, pp. 548–578
Published with license by Taylor and Francis.
ISSN: 0737-0024 print / 1532-7051 online
DOI: 10.1080/07370024.2013.803873

Taylor & Francis
Taylor & Francis Group

Issues Related to HCI Application of Fitts's Law

Charles E. Wright[1] **and Francis Lee**[1]
[1]*University of California, Irvine*

Taking Fitts's law as a premise—that is, movement time is a linear function of an appropriate index of difficulty—we explore three issues related to the collection and reporting of these data from the perspective of application within human–computer interaction. The central question involved two design choices. Whether results obtained using blocked target conditions are representative of performance in situations in which, as is often the case, target conditions vary from movement to movement and how this difference depends on whether discrete or serial (continuous) movements are studied. Although varied target conditions led to longer movement times, the effect was additive, was surprisingly small, and did not depend on whether the movements were discrete or serial. This suggests that evaluating devices or designs using blocked data may be acceptable. With Zhai (2004) we argue against the practice of reporting throughput as a one-dimensional summary for published comparisons of devices or designs. Also questioned is whether analyses using an accuracy-adjusted index of difficulty are appropriate in all design applications.

1. INTRODUCTION

Fitts's law is a highly successful formulation that describes how the time to complete a movement depends on the distance to be covered and the spatial accuracy required. Although Fitts's law does not apply to all aimed movements (Wright & Meyer, 1983) and there has been interest in ways to escape the limitations it imposes in virtual environments (e.g., Balakrishan, 2004), the class of movements to which it

Charles E. Wright is a cognitive psychologist with interests in learning and control of skilled movements, analysis of movement trajectories, mechanisms for response selection, and visual attention; he is on the faculty of the Department of Cognitive Sciences at the University of California, Irvine. **Francis Lee** is a research assistant interested in cognitive modeling; he is expected to graduate with a BA in Cognitive Psychology from the University of California, Irvine in 2013.

© Charles E. Wright and Francis Lee

CONTENTS

1. INTRODUCTION
    1.1. Background
    1.2. Fitts's Law in Basic Research and Applied Settings
    1.3. Issue 1: Influence of Design Choices on the External Validity of Fitts's Law Studies
    1.4. Issue 2: Use of ID_e and the Accuracy Adjustment
    1.5. Issue 3: Summaries Based on *IP*, the Index of Performance
2. METHODS
    2.1. Participants
    2.2. Apparatus
    2.3. Design
    2.4. Procedures
3. RESULTS
    3.1. Latency
    3.2. Practice Effects
    3.3. Missed-Target Errors
    3.4. Dispersion of the Vertical Movement Endpoints
    3.5. Movement Time versus Index of Difficulty
    3.6. Target Utilization
    3.7. Effective Movement Distance
    3.8. Movement Time versus Effective Index of Difficulty
4. DISCUSSION
    4.1. Issue 1: Influence of Design Choices on the External Validity of Fitts's Law Studies
    4.2. Issue 2: Use of ID_e and the Accuracy Adjustment
    4.3. Source of the Speed-Accuracy Tradeoff for Serial vs. Discrete Movements
    4.4. Issue 3: Summaries Based on *IP*, the Index of Performance
5. SUMMARY

does apply is large and of immense practical significance. This broad applicability has stimulated interest in this formulation beyond its basic research origins. This interest has led both to a large increase in the number of papers appearing in traditionally applied outlets that use Fitts's law and to an international standard that specifies how Fitts's law results should be used to characterize and compare input devices (ISO, 2000).

The purpose of this article is neither to extend nor question Fitts's Law, which we take as a given. Instead we wish to explore three issues that have arisen as Fitts's law has been applied in human–computer interaction (HCI). The theme that unites these three issues is a concern that ideas and practices, which emerged from the basic research that provided the underpinnings of Fitts's law, have been adopted by applied researchers without sufficient scrutiny. Recent standardization efforts have made these issues more salient. Although standardization will almost certainly produce research results that are more consistent, standardization could also have negative effects if the research produced either is reported incompletely or is misleading when generalized to the situations that are of practical interest.

The central issue we explore is whether experimental design choices, which may make sense in a basic-research setting and are often used in applied studies, appropriately reflect the real-world situations to which the results are to be generalized. The two remaining issues emerged as we reflected on recommendations in the literature for how to analyze and report our data related to this central issue.

1.1. Background

Fitts's law holds that the time, T, to complete a speeded movement to a target is a linear function of an index of difficulty, ID, characterizing the movement:

$$T = a + b\,ID. \tag{1}$$

The index of difficulty depends on the target distance, D, from the starting point to the center of the target, and the target width, W. The definition of ID has evolved since the initial, admittedly ad hoc formulation proposed by Fitts (1954):

$$ID = \log_2\left(\frac{2D}{W}\right).$$

The version of the index of difficulty now typically used in HCI applications is (MacKenzie, 1992; Soukoreff & MacKenzie, 2004):

$$ID = \log_2\left(\frac{D}{W} + 1\right). \tag{2}$$

However, all of the proposed modifications of Equation 2 have the more general form $T = f(D/W)$; where $f(\,)$ is a simple (i.e., linear, logarithmic, or power) function of the dimensionless ratio D/W (Guiard & Beaudouin-Lafon, 2004).[1]

A second, important development related to Fitts's law has grown out of the observation that participants in these experiments often do not adjust their performance as much as might be expected when target width is changed. Specifically, changes in endpoint variability are typically much smaller than would be expected when target width is manipulated. To compensate for this, Welford (1968, pp. 147–148), based on earlier work of Crossman (1956), suggested that ID be replaced with an *effective* index of difficulty, ID_e, in which an estimate of the effective target width, W_e, replaces W, the nominal target width. W_e can be computed either from the observed

[1]In our opinion, the "best" formulation for the index of difficulty is a power function: $(D/W)^p$, where p is a fractional exponent ($0 < p \leq 1$). Kvålseth (1980) first made the case that this form generally fits movement time data better than one based on the logarithm, the shape of which approaches that of the power function as p gets close to zero. We prefer this form because it has been shown to generate fine-grained predictions about the resulting movement trajectories that have generally been confirmed (Meyer, Smith, Kornblum, Abrams, & Wright, 1990). However, for descriptive purposes, Equation 2 is fine, and, as it is widely used in this literature, we use it here.

incidence of errors (i.e., those movements ending outside of the target) or, preferably, from the variability of movement endpoints,

$$ID_e = \log_2\left(\frac{D}{W_e} + 1\right). \tag{3}$$

Although Fitts used an accuracy adjustment in his later work (Fitts & Radford, 1966), it has not always been adopted (MacKenzie, 1992).

Over the last half century, Fitts's law has been well studied and has proven to be highly successful. Data obtained using a large variety of input devices across a broad array of conditions are well fit by Equation 1 with R^2 values of .80 or higher (Plamondon & Alimi, 1997). The success of this formulation has not been universal, however. For example, it is well documented that this formulation breaks down as *ID* becomes small (Gan & Hoffmann, 1988) and when *W* is smaller than about 8 pixels (Chapuis & Dragicevic, 2011).

Fitts's papers contained elements of both basic and applied research. His formulation grew out an effort to understand human performance from the theoretical perspective of information theory. However, from a more practical perspective, he also proposed an index of performance (Fitts, 1954, Equation 2),

$$IP = \frac{ID}{MT}, \tag{4}$$

as a measure of *throughput* combining both speed and accuracy. This measure, which has the units bits of information per unit of time, was adapted from information theory, where it is used as a measure of the channel capacity. Fitts's expectation was that throughput would be a constant that could be used to characterize and compare operator performance with different devices and in different movement contexts.

1.2. Fitts's Law in Basic Research and Applied Settings

Although the specific form for the index of performance has been a subject of debate, Fitts's idea of using these or similar results to characterize movement situations and input devices has become increasingly influential in HCI (MacKenzie, 1992), especially after Card and his colleagues used the results from an application of Equation 1 (Card, English, & Burr, 1978) to justify commercialization of the mouse by Xerox. Of specific interest to HCI researchers, Fitts's law has been found to apply to pointing and dragging using a mouse, trackball, stylus, joystick, and touch screen. The results have been used both to assess and compare throughput and as part of larger models to predict performance in new user interfaces (e.g., Card, Moran, & Newell, 1983). This has led to the promulgation of an international standard, ISO9241-9, that provides guidelines for such evaluations (ISO, 2000). Other, more detailed recommendations for how these evaluations should be conducted have been proposed by Soukoreff and MacKenzie (2004). These attempts at standardization are important if they help reduce the confusion in the literature due to conflicting

results arising from methodological differences (MacKenzie, 1992). To evaluate this possibility for pointing movements made with the mouse, Soukoreff and MacKenzie compared nine studies that followed ISO9241-9 and 24 studies that did not. They found a dramatic increase in consistency of the results for the studies that followed ISO9241-9.

1.3. Issue 1: Influence of Design Choices on the External Validity of Fitts's Law Studies

Inferences about cause–effect relationships based on specific scientific studies are said to possess external validity if they may be generalized from the unique and idiosyncratic settings, procedures, and participants of those studies to other populations and conditions. This issue is often critical in design applications when published results are used to justify design decisions. Of course, the best way to settle external validity concerns is a replication using the settings, procedures, and participants of the intended application. However, for obvious practical reasons designers often prefer to generalize results from available, prior research when making design choices. Our central concern here is that the methodology of much of the research using Fitts's law, including studies adhering to the suggestions of ISO 9241-9 (ISO, 2002) and Soukoreff and MacKenzie (2004), may generalize poorly to the situations typically encountered in the HCI applications of that research: that is, the coefficients of Equation (1) derived from such research may deviate systematically from those obtained using procedures more like the situations encountered in the application environments.

One procedural aspect of concern here is the blocking of target conditions, where "target conditions" refers to combinations of D and W. Recommendation II of Soukoreff and MacKenzie (2004, p. 755) is in line with the practice followed in many studies based on Fitts's law. They suggest studying a variety of target conditions that include multiple levels of D and W chosen so that nominal ID values associated with the target conditions span a range between 2 and 8 bits. Each target condition should be presented enough times—they suggest between 15 and 25—that an accurate estimate of the central tendency can be ascertained for each participant using each target condition. Although there is no specific recommendation to this effect, either by Soukoreff and MacKenzie or in ISO 9241-9, a natural way to structure the repeated presentations for a particular target condition is to block them, that is, present a sequence of trials all having the same target condition. Blocking of the target conditions certainly is not necessary. Pastel (2011) is an example of a study that randomizes the target conditions. However, of the nine studies cited in Soukoreff and MacKenzie (2004) as examples that have followed ISO 9241-9, the six that we could obtain and that included enough detail to determine how the conditions were ordered, all blocked target conditions. It is also suggestive that some studies report not only blocking target conditions but discarding the first several trials within each block so that the data analyzed would better reflect optimum performance possible in that target condition.

Blocking target conditions and discarding initial trials within blocks may make perfect sense in the context of basic research where the goal typically is to estimate the best possible result that a participant can produce in a specific condition. However, it is far from clear that the results obtained using these procedures accurately describe performance in typical HCI applications. Making matters worse, we know of no studies that include or allow direct comparison of results obtained using blocked and fully varying target conditions.[2] The study reported here includes a factor—blocked/varied targets—to address that question.

In addition to manipulating the order of target conditions, this study also looks at the importance of movement preparation by including a second factor: discrete/serial. In the discrete movement task, the participant first freely moves the mouse to a specified starting point and then, after a signal, initiates a speeded movement to the target. After a pause, and possibly some feedback, this procedure is then repeated. In the serial movement task, after completing a movement, the participant immediately initiates a subsequent movement in the opposite direction; the process is repeated until the full sequence is done. In terms of opportunities for movement preparation, these procedures span a continuum that includes many typical HCI situations.

Unlike the blocked/varied factor, data have been reported that compare the discrete and serial movement conditions. In his first article, Fitts (1954) used a serial, stylus-tapping task (he described this condition as "continuous"). A decade later, Fitts and Peterson (1964) used the same apparatus in a discrete version of the task. Figure 2 of this second article compares the data from these two experiments. In this figure, *T* for the serial task is longer than that for the discrete task across the full range of *ID* values studied and gets larger as *ID* increases: For an *ID* of 2, the difference was roughly 100 ms and it increased to roughly 210 ms for an *ID* of 7. Fitts and Peterson expected these results both because *T* in their serial task included between-movement latencies that were excluded in their discrete task and, more fundamentally, because, in the discrete task, the participant starts each movement after having had time to program its parameters. However, the interpretation of their data is clouded by the presence of a substantial speed–accuracy trade-off: For the faster discrete task there were 10.5% target misses on average, almost 10 times as many as the 1.2% for the slower serial task.

Guiard (1997) provided a direct comparison of discrete versus serial movements using a linear positioning task. To make the conditions more similar, participants pushed a button on the manipulandum to signal the end of a movement and before

[2]This statement is qualified because there have been several studies (e.g., Megaw, 1975) that looked at limited variation of the target conditions in a serial, reciprocal, stylus-tapping task (like that studied by Fitts, 1954; see next). One experiment included conditions in which *D* was constant but the widths of the left and right targets were different. The analysis looked at *T* as a function of the width of that target and the target of the previous movement—as this was a serial, reciprocal task, the endpoint of the previous movement served as the starting point for the current movement. These were inversely related: *T* decreased substantially as the width of the previous target increased. Particularly striking was the observation that, when the previous target was larger than the current target, *T* was faster than it was in a condition in which the width of the two targets was identical. Megaw also studied successive movements that had different *D*s, but constant *W*. In this condition, *T* was about 35 ms slower and did not depend systematically on the target conditions of the previous movement.

initiating a discrete movement. Replicating the conclusion derived from Fitts's experiments, the slope of the function relating T to ID_e was larger in the serial condition (277 ms/bit) than in the discrete condition (205 ms/bit), and the direction of this difference was the same for all six participants. However, unlike the data from Fitts, Guiard found that, for all six participants, these functions crossed somewhere in the ID_e range between 2 and 6 and that this point of intersection was strongly correlated with a participant's overall movement time: the faster the participant, the higher the ID_e at the point of intersection.

The discrete/serial factor was included in our experiment for two reasons. First, there might be a main effect of this factor, which is important because, as just discussed, real-world tasks tend to fall between the extremes exemplified by the discrete and serial tasks in terms of the time available for movement preparation. Second, we suspected that the discrete/serial factor might interact with the blocked/varied factor: specifically, that any differences due to the blocked/varied factor might be larger with the serial task than with the discrete task. From an information-processing perspective and consistent with the expectation of Fitts and Peterson (1964), the implicit pressure to spend less time on movement planning in the serial task might impose a larger penalty when the target conditions are varied than when they remain constant. From the perspective of dynamical systems applied to the Fitts task (Guiard, 1993), the ability to recycle kinetic energy in the serial task appears to depend on the harmonicity of the repetitive movements. Guiard (1993, 1997) has shown that, with target conditions blocked, this advantage is reduced for more difficult movements. It seems plausible that this advantage might also be reduced when the target conditions are varied.

Much of the reported data for Fitts's law comes from the discrete task with blocked target conditions. However, many of the actual HCI situations for which designers might wish to draw inferences by generalizing these data provide less opportunity for movement preparation, as in the serial task, and/or involve target conditions that vary from movement to movement. Given this mismatch, it strikes us as important to have a clear sense for whether and, if so, how these two factors influence the coefficients of Equation 1.

1.4. Issue 2: Use of ID_e and the Accuracy Adjustment

Having collected the data to evaluate Issue 1, we encountered several issues about how to analyze and report it. As outlined in the Background section, the effective index of difficulty, ID_e (Equation 3), often is used when fitting Equation 1, replacing ID, the value computed from the nominal conditions. This substitution is recommended both by ISO 9241-9 (ISO, 2002, p. 30) and Soukoreff and MacKenzie (2004, pp. 755–757, Recommendation IV). Zhai, Kong, and Ren (2004) provided a thorough examination of the implications of this change. As they noted, the basis for this recommendation is largely pragmatic. For example, Fitts and Radford (1966, p. 479) say that they "feel" that using ID_e provides "a more precise estimate" even though the correlations with T were somewhat *lower*. The basis for this intuition appears to be the observation

that the spread of the endpoint distribution for movements to a target, defined by some combination of D and W, is not a fixed proportion of the nominal width of the target, W, as would ideally be the case. So, for example, when ID is smaller than about 3, movement endpoints are typically tightly clustered in the side of the target region closer to the starting point (Gan & Hoffmann, 1988). Because of this, T is larger than would be expected based on Equation 1. Zhai et al. (2004) formalized observations like these using an index of target utilization,

$$I_u = \log_2 \left(\frac{W_e}{W} \right). \tag{5}$$

Across a series of experiments, they found that I_u depends in a complex way on three factors: W, D, and operator intentions, which can be manipulated by instructions. Echoing Fitts and Radford (1966), Zhai et al. (2004) also found that replacing ID with ID_e, rather than improving fits obtained in several experiments, consistently produced a small decrement in R^2. Where they found an advantage of using ID_e was when they fit data obtained from several instructional conditions intended to induce participants to change their speed-accuracy trade-off. When the data from several conditions were fit simultaneously with a single line, the overall fit was substantially better using ID_e. In addition, separate fits for each instructional condition produced coefficients that were more similar when the predictor was ID_e instead of ID. These improvements occurred because using the effective rather than the nominal target width partially compensated for the changes induced by the instructional conditions.

These results suggest that using effective-target-width adjustment is a step that can be beneficial but should be used with caution, rather than automatically. In particular, we believe that basing design decisions on accuracy-adjusted data often may be a mistake. The problem is simple: In design situations, what is typically known are the nominal conditions not the effective conditions. Zhai et al. (2004, p. 826) and Chapuis and Dragicevic (2011, p. 13:23) also make this point. Predictions produced by inserting nominal ID values into Equation 1 with coefficients derived using ID_e will be biased, possibly badly; however, this procedure is precisely Recommendation VI of Soukoreff and MacKenzie (2004, p. 759). In addition, as Zhai et al. (2004, see Figure 17) concluded, an extension to Equation 1 that incorporates I_u in a principled way is not straightforward.

1.5. Issue 3: Summaries Based on IP, the Index of Performance

As noted in the Background section, Fitts (1954) initially proposed the measure of throughput, IP, in Equation 4 as a single-valued "index of performance." However, if Equation 1 is correct that T is a linear function of ID, then for IP to be even approximately constant across different values of ID, the constant term, a, in Equation 1 would have to be zero or at least relatively small. The parameter a is the solution to Equation 1 when $ID = 0$. Because that fact will play a critical role in the discussion that follows, we will refer to it henceforth as $\widehat{T}_{ID=0}$.

Because zero is outside the observable range of *ID* values,[3] the value $\widehat{T}_{ID=0}$ must be an extrapolation from the data. As with any extrapolation, estimates of $\widehat{T}_{ID=0}$ should be interpreted cautiously. Although it might seem intuitively plausible for *T* to approach zero as *ID* does, $\widehat{T}_{ID=0}$ values indistinguishable from zero have been observed only occasionally. Typically, $\widehat{T}_{ID=0}$ is found to be positive; however, in a few cases, including Fitts and Peterson (1964, p. 107) negative $\widehat{T}_{ID=0}$ values have been reported. According to one widely cited interpretation of $\widehat{T}_{ID=0}$, it should be positive reflecting the time required by fixed perceptual and/or motor processes (e.g., target selection) that, although required, are not influenced by movement difficulty (Welford, 1968). Whatever their source, as Zhai (2004) elegantly pointed out, the presence of a non-zero $\widehat{T}_{ID=0}$ implies that *IP* must vary, sometimes substantially, across target conditions with different *ID*.

Against this background, Soukoreff and MacKenzie (2004, Equation 10) recommended using what they call throughput (*TP*) obtained by a process of averaging accuracy adjusted *IP* values across a chosen set of target conditions as the preferred way of comparing different experimental conditions. Although there is much to recommend their carefully constructed approach for specific comparisons, we are concerned about its use when reporting results that are intended to be generic. The limitation of results reported using this approach arises when, for example, a designer wishes to apply the conclusions from a reported comparison to a current design choice. So long as the *ID* values used to make the reported comparison are close to those of the proposed application, then generalizing the reported results to the specific application may be justified. However, when the *ID* values differ across the two situations, basing a design choice on them will be problematic and a more nuanced description of the results underlying the original comparison, although more complicated, will probably be more useful. Also, as is demonstrated in the Discussion section, problems with *IP* constancy can arise for comparisons of conditions even within an experiment.

Perhaps because of similar concerns, Fitts subsequently proposed using the inverse of the slope coefficient,

$$IP = \frac{1}{b}, \tag{6}$$

as a measure of "relatively constant information capacity over a range of movement conditions" (Fitts & Radford, 1966, p. 476). Given the information theoretic approach that motivated Fitts's work, this definition makes perfect sense. Equation 6 also has the advantage over Equation 4 that, for any range of ID values over which

[3]Using Equation 2, the smallest plausible value of *ID* is either 0.585 or 1, depending on whether the task is discrete, in which case $D/W = 1/2$ and the target region extends back to the starting point, or the task is serial, in which case $D/W = 1$ and the edges of the target regions just touch.

Equation 1 holds, the expected value of this estimate will be constant.[4] As many have pointed out, the disadvantage of this approach is that focusing solely on the inverse slope coefficient and ignoring a non-zero $\widehat{T}_{ID\,=\,0}$ discards information that may be important. From an applied perspective, what typically matters is not a theoretical construct such as information capacity but rather the expected time required to complete pointing operations with different levels of difficulty and, as Equation 1 states, that time depends on both $\widehat{T}_{ID\,=\,0}$ and the slope. Thus, we agree with Zhai (2004) that both coefficients should always be reported when characterizing perceptual-motor systems.

Many reports using Fitts's law have included both the $\widehat{T}_{ID\,=\,0}$ and slope parameters. In fact, later in the article just cited (Fitts & Radford, 1966, p. 480), Fitts talks about using both to characterize the human motor system. Also, the first article to apply the Fitts's law approach to HCI, Card et al. (1978) also reported the full equation for each device. However, a disturbing number of subsequent articles in applied areas (see Zhai, 2004, p. 795, for one listing) have based their assessments on either only Equation 4 or Equation 6.

Even more troubling is that Annex B to ISO 9241-9, which describe procedures for testing the efficiency and effectiveness of input devices, states that the goal of testing should be to "provide a measure of throughput" (ISO, 2000, p. 28) and goes on to define throughput as ID_e/MT (p. 30). Soukoreff and MacKenzie (2004) provided useful elaborations and extensions of the procedural recommendations in ISO 9241-9. Their seven recommendations include detailed instructions for obtaining data and fitting it using the version of Equation 1 based on ID_e. These instructions implicitly acknowledge a role for both coefficients. However, their seventh recommendation is more in line with the position taken by ISO 9241-9. This recommendation is to be applied when the purpose of an analysis is to compare two or more conditions. Such comparisons are to be based on *TP*, their variant of the *IP* measure based on ID_e. They asserted that the advantage of this approach is that, "calculated this way, *TP* is a complete measure encompassing both the speed and accuracy of the movement performance" (p. 760).

Although we can understand the appeal of being able to characterize and compare different operators, conditions, or devices using a one-dimensional metric, the inconvenient truth of Equation 1 is that, for comparisons intended for generic use, this is not generally applicable. The one special case in which this approach works generally is when the difference between the conditions being compared is

[4] In support of their argument against this characterization, Soukoreff and MacKenzie (2004, p. 775) displayed an equation for $1/b$ that is related to an intermediate step in the standard derivation of the slope estimator for linear regression. They asserted, incorrectly, that the form of this equation supports their claim that estimates of the inverse slope are sensitive to the values of the independent variable, in this case of *ID,* associated with the observations that are included in the estimate. However, if the function relating *T* and *ID* is linear, as Equation 1 holds, then, unlike *IP*, the expected value of the inverse slope coefficient does not depend on the particular *ID* values observed. *ID* appears in the equation displayed by Soukoreff and MacKenzie to normalize the slope estimate. In effect this sets the scale, or units (e.g., bits per s), of the inverse slope coefficient. If the scale of the ID values were to change, for example if the logarithm were taken with a base of 10 rather than 2, this would simply change the units.

effectively confined to a difference in $\widehat{T}_{ID=0}$ values. When the slopes differ, even if the $\widehat{T}_{ID=0}$ values are zero, the size of the *T* difference between the conditions will depend on *ID*; when the functions cross, the sign of the difference will change. Any single-measure approach to comparing conditions ignores these differences.

2. METHODS

2.1. Participants

There were 13 participants (7 male); all had vision correctible to 20/20 or better and were right-handed. One female participant dropped out after the second session due to scheduling conflicts. Each of the remaining participants took part in three 1-hr sessions. Subjects were paid $10 per hour. The protocol for this experiment was approved by the UCI Institutional Review Board.

2.2. Apparatus

A PC running a program written in MATLAB was used to present stimuli and record responses. Stimuli were presented on a 17-in. CRT computer monitor running at a 60 Hz refresh rate with a resolution of 1280 × 1024 pixels. The screen was calibrated so that 1 pixel extended 0.25 mm in both the horizontal and vertical dimensions. Participants used a Logitech optical mouse (Model #M-98C) to make responses. All movement acceleration software was disabled so that a mouse movement of 1 mm produced a constant cursor movement of 5 mm (20 pixels). As is typical for mouse movements, the screen surface was oriented vertically in front of the participants at eye level and the mouse movements were made in a different plane, on the horizontal surface of the table at which the participants were sitting. The possible impact of this orientation difference was minimized because, as described next, only the left–right movements played an important role in this experiment. Participants adjusted the height of their chair to a comfortable height.

2.3. Design

There were three factors, all manipulated within subjects: the target condition factor (10 levels), specified *D*, *W*, and thus *ID* of a movement, the blocked/varied factor (two levels) specified how the target conditions were ordered, and the discrete/serial factor (two levels) specified whether successive movements were produced in a continuous series or discretely. As shown in Figure 1, the 10 levels of target condition included nine unique levels of *ID* constructed from six levels each of *D* and *W*.

The experiment was organized into blocks of 22 movements. The first two movements in each block were considered warm up and were not included in the analyses. Each hour-long session consisted of 40 blocks, organized into four sets

FIGURE 1. Nominal target conditions.

Target Distance, *D* (pixels)	Target Width, W (pixels)	Index of Difficulty, *ID* (bits)
50	18	1.92
100	20	2.58
200	28	3.03
200	20	3.46
100	8	3.75
400	32	3.75
800	30	4.79
600	18	5.10
400	8	5.67
600	8	6.25

of 10 blocks. One of the four combinations of blocked/varied and discrete/serial factors was used in each of these four 10-block sets. The order of these four task combinations was balanced across groups of four participants using a different digram-balanced, 4 × 4 Latin square for each of three groups of four participants.

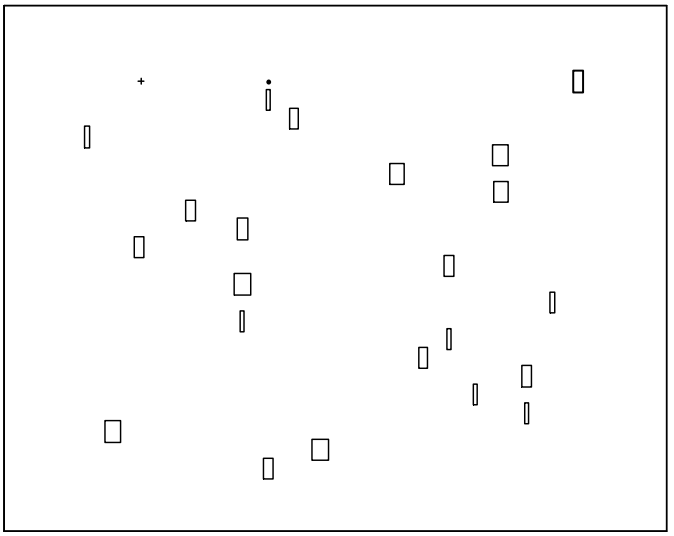
In the varied level of the blocked/varied factor, the target conditions occurred in a pseudo-random order generated with two constraints. First, each of the 10 target conditions had to occur twice in Movements 3 through 22 of each block; the target conditions for the first two movements were randomly selected without this constraint. Second, a target condition could not be selected for a particular movement if that would result in a target closer than 100 pixels to the left or right edge of the screen. Thus, across a set of 10 blocks, each target condition determined the target for 20 trials, ignoring the first two trials in each block. In the blocked level of the blocked/varied factor, a single target condition defined all of the movements in one block. Thus again, across a set of 10 blocks, each of the 10 target conditions determined the target for 20 test trials. The order of the 10 target conditions within a set was approximately balanced across participants using a digram-balanced, 10 × 10 Latin square.

2.4. Procedures

Before each block a displayed message stated whether the target conditions would be blocked or varied and whether the trials would be discrete or serial. At the end of each block, the display showed the mean movement time and number of targets missed. The mean movement time in this display excluded either the movement latency, for movements made in the discrete condition, or the dwell time, for movements made in the serial condition. (Latency and dwell time are described more fully next.) The experimenter compared these with previous values in similar conditions and verbally encouraged the participant to move quickly while minimizing errors.

Figure 2 shows a scale reproduction of an example stimulus display at the start of a block with varied target conditions. The small cross (5 pixels across) was the cursor

FIGURE 2. Reproduction of a possible stimulus display at the start of a varied target condition block.



that moved as the mouse moved. The dot (6 pixels in diameter) was the starting point; it was displayed as a filled red circle when the cursor was too far from the starting point to initiate a movement and a filled green circle when the cursor was within 3 pixels of the starting point. Once the cursor had remained in the start region for 0.5 s, a tone was presented. The tone indicated that the participant was free to start the movement at any time. The onset of the tone also began the timing of the latency period, that ended as soon as the mouse had been moved 5 pixels from its initial position. Having moved the mouse to bring the cursor within the target rectangle, the participant pushed a mouse button indicating that the movement was complete. If the cursor position when the button was pushed was outside of the target rectangle, an error was noted but no immediate feedback was given beyond that of the visible presence of the cursor outside the target.

The 22 rectangles in Figure 2 were the targets for the movements in one block. The vertical position of the targets reflects the order of each target in the upcoming sequence of movements. In this condition, the distance between targets, D, and their width, W, varied pseudo-randomly from target to target, as previously described. The vertical size and spacing of the targets were constant (35 pixels).

Although the display in Figure 2 looks cluttered, raising the concern that participants would have found it difficult to locate successive targets, after a little practice participants found these displays natural and easy to use. There were three reasons for this. First, after the initial movement, which was always to the right, movement direction alternated; so, for example, having completed a movement to the right, the participant would know to look back to the left to find the next target. Second, the next movement target was always visually quite salient because it was highlighted by being

displayed with a white outline on the gray background of the screen; the targets for all of the later movements were displayed with black outlines. (Note that in Figure 2 the highlighted box is indicated by a heavier line weight rather than luminance.) The third reason that successive targets were easy to locate was that the vertical center of the next target was always displayed at the same vertical position—as a corollary, the movements required were only horizontal. This was possible because, as soon as the movement to a target was completed, the box for that target disappeared from the screen, the remaining targets were moved up the screen an amount equal to the vertical spacing of the targets, and the next target, which was now at the same vertical position as the previous target was highlighted.

When the target conditions were blocked, all of the rectangles had the same W and, ignoring the alternating directions, were the same D apart. Thus the target rectangles appeared in two vertical columns. In all other respects, the displays and procedures were identical across these two levels of the blocked/varied factor.

In the serial level of the discrete/serial factor, the participant was free to start each successive movement as soon as the mouse button had been clicked to end the previous movement. Thus, a mouse button click both marked the end of one movement and began the timing for the latency period of the subsequent movement. The running program did, however, identify the time between the mouse click and the beginning of movement in the opposite direction as dwell time. In the discrete condition, as soon as the mouse button had been clicked to end a movement, the target display was modified as previously described, and a starting point circle was displayed at the center of what had just been the movement target. At this point the participant had to move the cursor to the starting point and wait for the go signal just as with the first movement in the block.

3. RESULTS

The movement time data are of primary interest in this experiment. Before presenting these data we summarize briefly several other aspects of the movements produced that could, depending on the results, suggest complications for the interpretation of the movement time data.

3.1. Latency

For discrete movements, the latency was the time from the GO signal until the start of the movement was detected. For all but the first movement in a serial movement block, there was no GO signal. For these blocks, what was recorded in lieu of the latency was the dwell time between the mouse button press, which ended the preceding movement, and the detection of the start of the subsequent movement in the opposite direction, in essence the dwell time. Given these procedural differences, the mean latency for the discrete movements, 525 ± 42 ms, is not directly

comparable to that for the serial movements, 72 ± 12 ms. (The notation, $X \pm Y$, gives a mean value, the X, followed by the half width of the 95% confidence interval for that value, the Y.) The blocked/varied factor had no effect on this measure, $t(11) = 0.589$. Also, consistent with the results reported by Munro, Plumb, Wilson, Williams, and Mon-Williams (2007) there was no effect of the *ID* of the upcoming movement, $t(11) = 1.202$, $p = .316$. This is important because had such an effect been observed it would call into question the procedure we used to segment movements into the latency and movement components.

3.2. Practice Effects

Each participant produced data in each condition on each of 3 days. Not surprisingly, performance improved with practice when it was assessed either as the mean *T* or as the slope relating *T* to *ID*. This improvement was larger and only statistically significant between Day 1 and Day 2. Based on this pattern, the data from Day 1 were excluded from the analyses that follow. Including the Day 1 data does not qualitatively change any of the results reported next, but it does reduce the precision of some of the comparisons.

3.3. Missed-Target Errors

A simple interpretation of the *T* versus *ID* relation is only possible if there are not large, systematic variations in the proportion of missed target errors across conditions. Overall, the error rate was quite low: 1.7% ± 1.2% of trials. However, this percentage did exhibit small, but significant, effects of both the discrete/serial factor and *W*. Errors were more frequent, $t(11) = 2.719$, $p = .020$, in the serial movement condition (3.0% ± 2.3%) than in the discrete movement condition (0.4% ± 0.3%).[5] There was not a difference between the varied and blocked conditions, $t(11) = 0.859$, nor was there an interaction of these factors, $t(11) = 0.081$. Consistent with the findings of Zhai et al. (2004) on target utilization, the percentage of errors also decreased with increasing *W* (slope = −0.13 ± 0.11), $t(11) = -2.604$, $p = .025$. This decrease was larger, $t(11) = 2.322$, $p = .027$, for serial movements (−0.20 ± 0.18) than for discrete movements (−0.05 ± 0.04). These results will play an important role in the interpretation of the effects of the discrete/serial factor on movement time.

3.4. Dispersion of the Vertical Movement Endpoints

The vertical size of the target boxes was intended to be large enough to constrain the movements minimally given that only horizontal movement was evaluated.

[5]These values are representative of all but one participant whose error rate for the continuous movements was 12.9% but only 1.2% for the discrete movements. However, excluding this subject did not change the qualitative description of these data. The overall error rate dropped to 1.2% ± 0.7% of trials. Errors still occurred more often, $t(10) = 3.368$, $p = .007$, for continuous movements (2.1% ± 1.3%) than for discrete movements (0.4% ± 0.3%).

Consistent with this expectation, the standard deviation of the vertical component of the endpoint positions was 3.6 ± 0.4 pixels, only slightly more than one tenth of the vertical size of the targets. Not surprisingly, there were small, but statistically reliable, increases in the vertical endpoint dispersion when the target conditions were varied and in the serial-movement condition.

3.5. Movement Time versus Index of Difficulty

Figure 3 displays data averaged across participants showing the relationship between T and ID for the four combinations of the blocked/varied and discrete/serial factors. Straight lines provide good fits to both the mean data (R^2 varied between .963 and .987 across the four conditions) and to the data for each participant (the median R^2 across participants for the four conditions varied between .827 and .899). Linear functions were fit separately to the data from each participant in each of the four running conditions. The resulting coefficients are summarized in Figure 4. Fitts's law is typically parameterized using the slope and a constant, $\widehat{T}_{ID\,=\,0}$. However, because this and the slope parameter must necessarily be highly correlated given the range of ID values, we prefer to report and focus on the slope and $\widehat{T}_{ID\,=\,4}$, which is essentially the average of T given that the mean of the ID values in this experiment was 4.03.

As shown in Figure 4, the average movement time, $\widehat{T}_{ID\,=\,4}$, was larger for discrete than for serial movements, $t(11) = 3.036$, $p = .011$. The effect of the blocked/varied factor was statistically unreliable, $t(11) = 1.692$, $p = .119$. However, these main effects were modified by their interaction, $t(11) = 2.832$, $p = .016$. One way to understand this interaction is that, as predicted in the Introduction, $\widehat{T}_{ID\,=\,4}$

FIGURE 3. ***T*** **averaged over participants as a function of** ***ID*** **for each of the four combinations of discrete/serial and blocked/varied.**

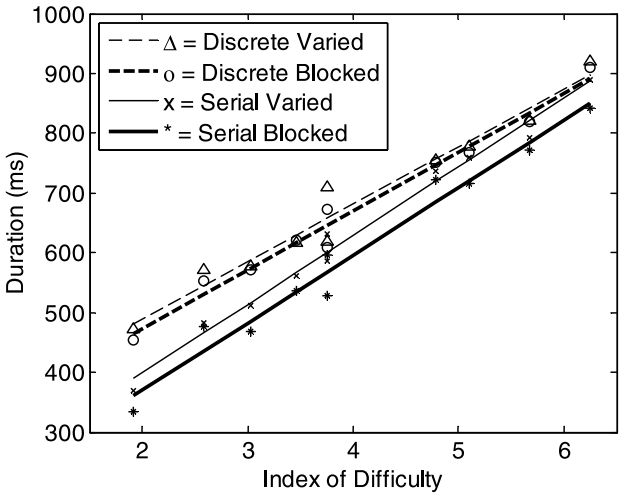


FIGURE 4. Movement time and the mean coefficients from the linear fit of *MT* versus *ID* averaged over participants for each of the four combinations of the blocked/varied and discrete/serial factors.

	Serial	Discrete	*M*	Varied − Blocked
$\widehat{T}_{ID\,=\,4}$, Mean Movement Time (ms)				
Varied	629 ± 42	681 ± 66	655 ± 50	22 ± 29
Blocked	596 ± 57	670 ± 84	633 ± 69	
M	612 ± 49	675 ± 74	644 ± 58	
Serial − Discrete	−63 ± 45			Interaction 11 ± 8
Slope of the Linear Fit: *T* versus *ID* (ms/bit)				
Varied	114 ± 12	96 ± 11	105 ± 10	−0.2 ± 6
Blocked	113 ± 12	99 ± 18	106 ± 14	
M	114 ± 12	97 ± 14	106 ± 12	
Serial − Discrete	17 ± 8			Interaction 2 ± 7
$\widehat{T}_{ID\,=\,0}$, *Y*-Intercept of the Linear Fit: *T* versus *ID* (ms)				
Varied	169 ± 44	297 ± 56	233 ± 38	23 ± 19
Blocked	144 ± 41	276 ± 61	210 ± 46	
M	157 ± 40	286 ± 56	221 ± 41	
Serial − Discrete	−130 ± 53			Interaction 2 ± 25

was larger than would have been expected given the main effects in the condition that provided the least opportunity for movement planning: that is, varied target conditions combined with the serial task.

Although the movements in the serial task took less time on average, the slope of the linear function relating *T* to *ID* was larger for the serial than for the discrete movements, $t(11) = 4.607$, $p = .001$. For the slopes there was neither a reliable difference between the varied and blocked conditions, $t(11) = 0.097$, nor an interaction, $t(11) = 0.676$. The upshot of this combination of effects due to the blocked/varied factor was that, although the *T* for serial movements was generally less than that of discrete movements, this difference was reduced as *ID* increased. Specifically, the movement time difference due to the discrete/serial factor completely disappeared at the highest IDs when the target conditions were varied (the lighter weight lines in Figure 3).

3.6. Target Utilization

The previous section examined how *T* depended on *ID*. However, as noted in the Introduction, if participants used different proportions of the target across levels of

either the blocked/varied or discrete/serial factors, this would suggest that a summary based on ID_e, the effective index of difficulty might provide a better summary of the results for at least some purposes. Figure 5 displays a summary of the relationship between the index of utilization, I_u, defined in Equation 5 as proposed by Zhai et al. (2004), and target width, W. I_u is equal to 0 when the effective target width, W_e, is the same as the actual target width. Because the collected data included the movement endpoints, W_e was calculated based on the standard deviation of the dispersion of those endpoints in the horizontal direction of movement, sd_x. Specifically, $W_e = 4.133\ sd_x$. The constant in this equation is specified by ISO 9241-9 (ISO, 2000, Annex B, p. 29) based on the expectation that participants have the goal to adjust the speed of their movements so that 4% end outside of the target. I_u values of −1 and +1 reflect W_e values that are, respectively, one half or twice W.

The straight lines relating I_u to $\log_2 W$, shown in Figure 5, capture a large proportion of the variance (R^2 across the four conditions varied between 0.86 and 0.93). The figure clearly suggests that there were two strong effects: I_u decreased linearly with the logarithm of target width and was larger for discrete movements than for serial movements. Consistent with this impression, there was a significant effect of the discrete/serial factor, $t(11) = 5.705$, $p < .001$: The mean of I_u, averaged across $\log_2 W$, was 0.01 ± 0.20 bits for the discrete condition and 0.45 ± 0.14 bits for the continuous condition. The magnitude of the slope relating I_u to $\log_2 W$ also depended on the discrete/serial factor, $t(11) = 2.621$, $p = .024$: For the discrete condition the slope was −0.39 ± 0.10 bits per pixel; for the continuous condition it was −0.26 ± 0.05 bits per pixel.

FIGURE 5. Index of target utilization, *Iu*, as a function of target width, *W*, for each of the four combinations of discrete/serial and blocked/varied.

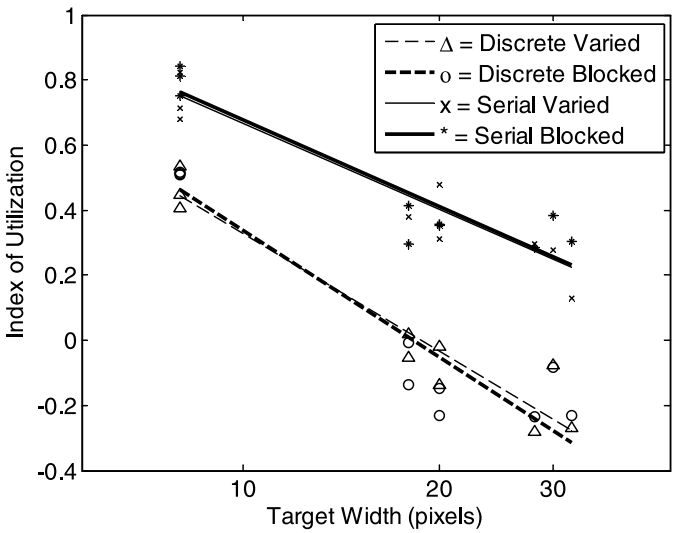
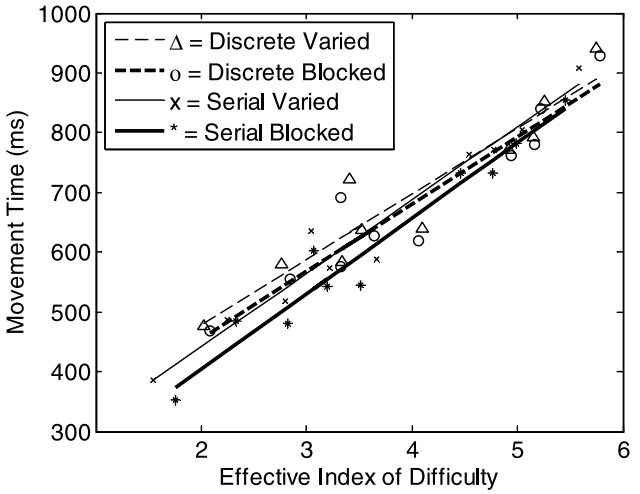


FIGURE 6. *T* averaged over participants as a function of ID_e for each of the four combinations of discrete/serial and blocked/varied.



3.7. Effective Movement Distance

Just as it could be dangerous to assume that participants would have responded to changes in W in a simple, consistent way, it is also possible that the movement distances produced might not have been simply related to the target distance, D. To assess this possibility, the actual average movement distance within each block was regressed against D and W simultaneously. This was done separately for the data from each participant for each of the four combinations of the discrete/serial and blocked/varied factors. An analysis of the resulting coefficients showed that the actual movement distances depended only on D—there was no discernible influence of W—that this relationship was proportional with a slope quite close to unity (0.9992 ± 0.00036), and that the slope did not differ across conditions.

3.8. Movement Time versus Effective Index of Difficulty

Figure 6 displays data, averaged across participants, showing the relationship between T and the effective index of difficulty, ID_e, for each of the four combinations of the blocked/varied and discrete/serial factors. ID_e here was computed as $\log_2(D/W_e)$, where $W_e = 4.133\ sd_x$. Figure 7 summarizes the coefficients, computed separately for each participant, of the linear relationship between T and ID_e.[6] Straight

[6]Because each participant produces different values of ID_e for the different ID levels and the four running conditions, the coefficients of the fitted lines in this figure differ slightly from the average of those obtained when this regression is done separately on the data from each participant. However, for these data, these differences are negligible.

FIGURE 7. Movement time and the mean coefficients from the linear fit of *MT* versus ID_e averaged over participants for each of the four combinations of the blocked/varied and discrete/serial factors.

	Serial	Discrete	*M*	Varied − Blocked
$\widehat{T}_{IDe=4}$, Mean Movement Time (ms)				
Varied	686 ± 34	698 ± 61	692 ± 47	24 ± 20
Blocked	655 ± 49	681 ± 74	668 ± 60	
M	671 ± 43	689 ± 66	680 ± 53	
Serial − Discrete	−19 ± 37			Interaction 7 ± 13
Slope of the Linear Fit: *T* versus ID_e (ms/bit)				
Varied	120 ± 11	108 ± 13	114 ± 10	−4 ± 5
Blocked	124 ± 11	111 ± 15	118 ± 12	
M	122 ± 10	110 ± 13	116 ± 11	
Serial − Discrete	13 ± 9			Interaction 0 ± 9
$\widehat{T}_{ID=0}$, Y-Intercept of the Linear Fit: *T* versus ID_e (ms)				
Varied	205 ± 33	265 ± 42	235 ± 27	38 ± 19
Blocked	158 ± 26	236 ± 54	197 ± 33	
M	181 ± 26	250 ± 44	216 ± 29	
Serial − Discrete	−69 ± 43			Interaction 9 ± 30

lines fit these data reasonably well, although not quite as well as the fits based on *ID* (the median R^2 across participants for the four conditions varied between .756 and .895). Still, compared with the fits based on *ID* in Figure 4, these fits generally had smaller confidence intervals suggesting that using W_e, rather than the nominal values, *W*, produced coefficient estimates that were more consistent across participants.

The most striking change in the analysis based on ID_e is that the difference of $\widehat{T}_{IDe=4}$ between discrete and serial movements was much smaller and no longer statistically reliable, $t(11) = 1.112$, $p = .290$. The difference of $\widehat{T}_{IDe=4}$ between the varied and blocked conditions was also slightly smaller in this analysis; however, because of the reduced variability across participants, this difference just reached the level of being statistically reliable, $t(11) = 2.581$, $p = .026$. The results in this analysis were also more straightforward because the interaction of these two effects was smaller and no longer statistically reliable, $t(11) = 1.227$, $p = .245$. As in the analysis based on *ID*, the slope of the linear function relating *T* to ID_2 was larger for the serial than for the discrete movements, $t(11) = 3.198$, $p = .008$. And again there was neither a reliable slope difference between the varied and blocked conditions, $t(11) = 1.585$, $p = .141$, nor an interaction, $t(11) = 0.102$.

4. DISCUSSION

4.1. Issue 1: Influence of Design Choices on the External Validity of Fitts's Law Studies

The primary purpose of this study was to determine whether and, if so, how the coefficients characterizing the Fitts's law relationship for the mouse depend on two experimental design choices that dictate task organization. The blocked/varied factor specified whether variations in the target conditions were blocked, as has often been true in published studies, or whether they varied from movement to movement, as is more typical in applications. Large differences due to this factor would undermine the use for applied design decisions of the large body of results obtained with blocked target conditions. The discrete/serial factor specified whether movements in a block were produced as separate, discrete movements or one movement series. In terms of opportunities for movement preparation, many target acquisition tasks of interest to HCI fall somewhere on this discrete–serial continuum. This factor has been shown previously to have effects on the Fitts's law slope. It might also be expected to interact with the blocked/varied factor. Theoretically, one might expect there to be differences across the four combinations of these two factors based on two intuitively plausible ideas: that immediate experience with a specific movement will improve performance when that movement is repeated and that performance will be better on a movement when there is plenty of time to plan it.

Although the pattern of results for both the blocked/varied and discrete/serial factors depended somewhat on whether the analysis was based on *ID* or ID_e, what is most striking is that the effects they produced were relatively small. The statistically significant effects are potentially important theoretically; however, the small size and nature of these effects diminishes their importance from an applied perspective.

The most consistent results involved the slope of Fitts's law, which estimates how movement time increases with difficulty. This slope did not change across the levels of the blocked/varied factor; however, it was larger in the serial version of the task than in the discrete version. In the analysis based on *ID*, this difference amounted to a 16% effect (the slope was 114 ms/bit in the serial task and 97 ms/bit in the discrete task); when the analysis was based on ID_e, the effect was only 11%, although the slope estimates were somewhat larger (122 ms/bit and 110 ms/bit, respectively). This difference is in the same direction as those reported by Fitts and Peterson (1964) and by Guiard (1997). However, the result here enlarges on those previously reported in that we found no evidence for a slope difference due to the blocked/varied factor or an interaction of these two factors. The direction of this effect makes sense if the discrete/serial factor is viewed from a planning perspective because it suggests that the movement time "cost" associated with reducing the opportunity to plan an upcoming movement is larger for more difficult movements. Because this is a difference of slopes, the practical impact of this effect is difficult to assess: For easy movements (e.g., $ID = 2$), the slowing in the serial task is only

about 30 ms; however, it increases to as much as 65 ms for more difficult movements (e.g., $ID = 6$). Keep in mind, however, that these estimates can be viewed as upper bounds, because, from the perspective of movement preparation, practical tasks fall somewhere in the middle of the continuum defined by extremes of this factor. Thus, the differences between these tasks will be larger than those between any practical task and either of these extremes.

The slope difference between levels of the discrete/serial factor can be understood as an interaction of that factor with *ID*. In addition to this interaction, main effects of both factors were reflected in the values of $\widehat{T}_{ID=4}$. However, the interpretation of these effects is difficult because they are different when the analysis is based on *ID* versus ID_e. This can be seen by comparing Figures 4 and 7.

Along with a larger slope, $\widehat{T}_{ID=4}$ for the serial task was 63 ms less (10%) in the analysis based on *ID*. In the analysis based on ID_e, this advantage shrank to just 19 ms, an estimate that was not reliably different from zero. This estimate of the overall advantage of the serial task shrank by two thirds because the spread of the movement endpoints for this task was substantially larger than that for the discrete task. This is shown by the larger I_u values in Figure 5. This increase is reflected in smaller values of ID_e for the serial task that, in turn move those data points to the left in Figure 6, producing a larger value $\widehat{T}_{ID_e=4}$ relative to the discrete task.

The interpretation of the blocked/varied factor is more straightforward, because in both analyses, the average movement time was larger when the target conditions varied from movement to movement compared to when they were blocked. Although the estimate of this effect was almost identical in the analyses based on *ID* or ID_e (22 ms vs. 24 ms), it was statistically reliable only in the analysis based on ID_e. This slowing with varied target conditions is not surprising. From a planning and control perspective, it seems quite plausible that there would be an advantage of repeating a movement. What may be surprising is that the cost associated with making a varied as opposed to a consistent series of movements is only about 3.5% of the average movement time and that there is no indication that this cost varies with movement difficulty.

That there is only a small, additive effect of the blocked/varied factor is an excellent outcome from an applied perspective because it suggests that this experimental design choice is of only secondary importance when generalizing results to a new situation. Also, because the effect is additive, compensating for it is straightforward if this is deemed necessary. To gain perspective on the size of this effect, consider that its impact is smaller than that of increasing *ID* by one fourth of a unit. Another way to view this difference is that across the 12 participants in this study, $\widehat{T}_{ID=4}$ ranged from 583 ms to 806 ms. Given the restricted population (college students) included in this study, the upper end of this range is undoubtedly very small compared with that of a less restricted population. Yet the effect of the blocked/varied factor is only 10% of the variation across participants observed here. Still, given that the small size of this effect was unexpected, it would be prudent to replicate it, both with the mouse and with other devices before taking it too seriously.

4.2. Issue 2: Use of ID_e and the Accuracy Adjustment

The results summarized in the previous section demonstrate that incorporating an accuracy adjustment, by substituting ID_e for ID when fitting results with Equation 1, led not only to quantitative differences in the estimated coefficients but also to qualitative changes in the overall pattern of results. Our results are consistent with those of Zhai et al. (2004) and those of others that they summarize in showing both that Fitts's law fits slightly less well when ID_e is substituted for ID and that the resulting coefficients are more similar across both conditions and participants.

The first of these results, the reduced R^2 fitting Equation 1, presumably reflects the fact that ID_e values were more variable than ID values. ID_e values should be more variable because they are estimated from the data rather than being manipulated by the experimenter. This increased variability reduces their predictive power. Countering this disadvantage is the suggestion that ID_e values improve on ID values by being less biased predictors of movement difficulty. The basis of this suggestion is the hypothesis that the planning and production of these movements is guided by a subjective transformation of the target width. Although this internal quantity, which we estimate using W_e, depends on W, it does so in a complex way that varies between participants, across movement conditions, and with speed-accuracy instructions (Zhai et al., 2004). Supporting this hypothesis are published demonstrations in which analyses based on ID_e appear to compensate, at least partially, for these differences (Zhai et al., 2004). It is this compensation that leads to the second result mentioned previously: that is, we found the coefficient estimates of the fitted functions were more similar across participants and conditions.

To see why using an accuracy-adjusted analysis based on ID_e can make Fitts's law results more similar across experimental conditions, consider the data for the index of utilization, I_u, displayed in Figure 5. These data illustrate two ways that the variation in W_e did not simply follow that of W. Keep in mind when looking at this figure that if W_e had been some fixed proportion of W, then the fitted lines in this figure would be flat. Instead, the fitted lines have negative slope because, as the W increased, W_e failed to increase proportionally. A second obvious difference in Figure 5 is that the two lines fit to the data for the serial task are clearly above those for the discrete task because, in the serial task, the dispersion of the endpoints was much broader than that in the discrete task. It is this difference that accounts for the qualitative discrepancy in the results summary based on the analysis based on ID (Figures 3 and 4) and that based on ID_e (Figures 6 and 7). Larger *effective* targets in the serial task would be expected to produce shorter movement times and more errors, and both of these differences were observed. The movement time analysis based on ID_e compensated for the differences in target utilization by displacing to the left all of the data points from the serial task. In the analysis based on ID, these serial-task data were aligned vertically with the data points from the discrete task that had the same nominal target conditions. Figure 6 shows that, after this displacement, the data points from the two tasks are almost collinear.

As we noted earlier, another advantage of analyses using ID_e is that may have more statistical power because they can compensate for between-subject variations in target utilization. These variations can be nontrivial: In this experiment, the mean I_u values for each participant ranged from –0.14 to 0.65. The results for the blocked/varied factor illustrate why this can be important. Because the values of I_u were essentially unchanged across the levels of blocked/varied (see Figure 5), a comparison of Figures 4 and 7 shows that the size of the effect associated with this factor was almost unchanged in the analyses based on *ID* and ID_e. However, because the analysis based on ID_e accounts for the between-subject differences in target utilization, the blocked/varied effect is statistically significant in the analysis based on ID_e, but this same difference fails to reach significance in the analysis based on *ID*.

These results illustrate both advantages and a disadvantage of including an adjustment for accuracy in applications of Fitts's law. Weighing these considerations, there appears to be a clear benefit of this approach for most applications. Not surprisingly then, this substitution is recommended both by ISO 9241-9 (ISO, 2002, p. 30) and Soukoreff and MacKenzie (2004, pp. 755–757, Recommendation IV).

Despite the persuasive argument for using ID_e, there is a serious issue with this approach that needs careful consideration. Because of the number of factors that influence I_u and the complexity of their influence (Zhai et al., 2004), values of ID_e can, at least for now, only be estimated from data. The crux of the problem, then, is that the power of ID_e as a predictor is only as good as the precision of its estimates. An extreme example where this consideration plays an important role is that of designers who have only available values of *ID*, based on a proposed design, and a set of Fitt's law coefficients from a plausibly similar movement situation. In such situations, the version of the coefficients used to make design choices should be based on *ID*, not ID_e; however, if, as is often now the case, a published study includes only coefficients computed with ID_e, then the data from such a study will not be useful in this way. Zhai et al. (2004, p. 826) and Chapuis and Dragicevic (2011, p. 13:23) also made this point.

Going beyond this obvious case, however, we believe that even when data are available to estimate ID_e, the decision to base an analysis or a comparison on these values probably should be considered carefully and not adopted automatically based on a blanket recommendation. Because W_e values are usually derived from standard deviations, more observations are required to obtain good estimates than are necessary for a summary based on means. Given that W_e and *T* data are typically derived from the same data set, this statistical reality can lead to data sets for which the large variability of the ID_e values more than offsets any advantage they might have through compensating for differences in target utilization.

4.3. Source of the Speed-Accuracy Tradeoff for Serial versus Discrete Movements

When *T* is regressed against *ID*, serial-task movements were found to be faster than discrete-task movements (see Figure 3) with the size of this advantage decreasing

for movements with large *ID*. As in the comparison of these tasks reported by Fitts and Peterson (1964), this counterintuitive advantage appears to reflect a speed–accuracy trade off. Evidence for the accuracy component of a trade off takes two forms: larger I_u scores (Figure 5) for serial movements, which reflect a larger spread in the distribution of endpoints, and a somewhat larger proportion of movements classified as errors because the mouse button was pressed with the cursor outside of the target. The suggestion that the serial task advantage may be due only to this speed–accuracy trade off gains credence based on the analysis using ID_e; when the differences in I_u are taken into account, the movement time advantage of serial movements largely disappears (Figure 6). Despite this explanation for the serial-task advantage, from a theoretical perspective, there are at least two reasons why this result is worth exploring further.

The idea that planning improves the performance of movements underlies most thinking about human motor control. From this perspective, the speed advantage of serial movements is surprising. The discrete task would appear to provide the best opportunity to create and carry out a movement plan, and yet, if anything, these movements were slower. The movement planning perspective can be reconciled with these results by assuming either that the necessary planning can be done quickly—it must be quite fast because, in the serial task, the mean time between the click to end one movement and the start of the next movement was 72 ms—or that the planning for the upcoming movement can be overlapped with the production of the previous movement. For either explanation, the observation that the speed advantage for serial-task movements can be understood as a speed–accuracy trade off is critical; it would be difficult to reconcile a true serial-task advantage with this perspective.

A second, theoretically interesting question about the serial task advantage concerns its locus within a movement: specifically, is there an identifiable part of a movement in the serial task that, because it takes less time, leads to movements that are less accurate? The trajectories of Fitts task movements are usually thought to consist of an initial, fast submovement, which ends close to or within the target, followed, if necessary, by one or more feedback-guided, corrective submovements (Meyer, Abrams, Kornblum, Wright, & Smith, 1988; Meyer, Smith, Kornblum, Abrams, & Wright, 1990). To look for a possible locus of the speed–accuracy trade off between discrete and serial movements we briefly summarize here two measures characterizing the movement trajectories: one that characterizes the initial submovement and a second that characterizes the corrective submovements.

A full decomposition of these movement trajectories into each of their component submovements would be complex and well beyond the scope of this article. However, because these submovements exhibit an invariance in velocity profile shape (Freund & Büdingen, 1978; Gordon & Ghez, 1987), the overall peak velocity of a movement, which is a quantity that can be determined reliably from movement trajectory data, provides one useful, rough characterization of the initial submovement in a movement trajectory. The mean peak velocity was 3243 $\pm$ 519 pixels/s. This value did not differ reliably across the levels of the discrete/serial factor, the blocked/varied factor, or their interaction. As would be expected, peak velocity did increase with *D*

(the slope of the linear fit was 9.1 ± 1.6 pixels/s per pixel); however, this slope did not depend on either of the experimental factors. These results suggest that the initial submovement was not the locus of the speed–accuracy trade off between the serial and discrete movements.

Looking at several summaries of the corrective submovement process, the one that best captures the serial task advantage is what might be called *click delay*: the time between when the cursor last enters the target region and when the mouse button is pushed to end the movement. Of interest, click delay was a large proportion of the overall movement time, *T*, and did not vary with *D* or *W*. However, click delay was substantially longer, $t(11) = 4.300$, $p = .001$, for discrete movements (372 ± 55 ms) than for serial movements (293 ± 33 ms).

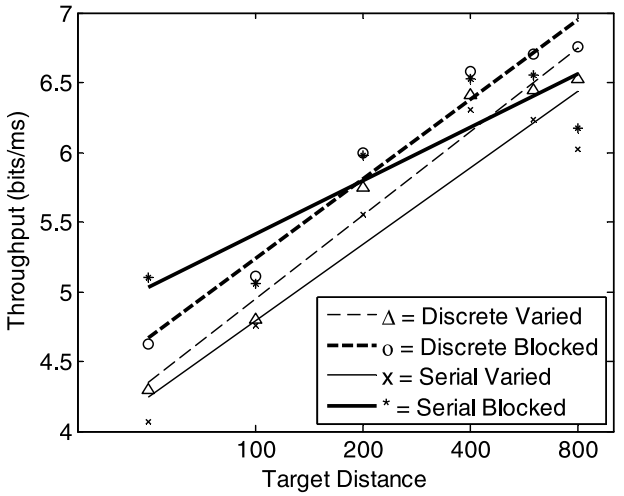
Taken together, these post hoc summaries are consistent with the interpretation that serial and discrete movements differed primarily in the way that they were terminated. The movements in both tasks began similarly; however, in the discrete task, participants were more careful to assess the position of the cursor before pressing the button to end the movement. By contrast, in the serial task, not only did participants end movements more aggressively, increasing both the dispersion of the movement endpoints and the incidence of actual errors, but, as the very short latencies suggest, they combined the button press ending one movement with the initiation of the subsequent movement.

4.4. Issue 3: Summaries Based on *IP*, the Index of Performance

As part of the recommendation VII, Soukoreff and MacKenzie (2004) exhorted, "If the purpose of this analysis is the comparison of two or more experiment conditions, then *throughput* (*TP*) is calculated" (p. 759). As defined in their Equation 10, *TP* is estimated for a participant as the ratio ID_e/T averaged across the target conditions in the experiment. Computing this quantity, we find an effect of the blocked/varied factor, $t(11) = 3.362$, $p = .006$, with higher *TP* when the target conditions are blocked (6.1 ± 0.5 bits/s) than when they are varied (5.8 ± 0.4 bits/s), but no effect of the discrete/serial factor or interaction between these factors. Because *TP* is based on the accuracy-adjusted ID_e values, it is perhaps not surprising, that this summary mirrors those based on ID_e summarized for $\widehat{T}_{ID_e = 4}$ in Figure 7. However, as stated in the Introduction, we have concerns about a general recommendation to use *TP*. Here we use our data to illustrate our concerns.

TP is an accuracy adjusted version of the throughput measure, *IP*, originally proposed by Fitts (1954). Fitts proposed this measure for two reasons: The concept of throughput made sense from his original information theoretic perspective, and he expected *IP* to be relatively constant across target conditions as defined by *ID*. That expectation proved to be incorrect in our data. Although accuracy-adjusted throughput was somewhat better in this respect, as shown in Figure 8, it was far from being constant across target conditions, at least for our data. Although there was only a little variation across levels of W, and thus accuracy-adjusted throughput was also

FIGURE 8. Throughput (ID_e/T) averaged over participants as a function of D for of each the four combinations of discrete/serial and blocked/varied.



only poorly correlated with *ID*, there was a strong relation between accuracy-adjusted throughput and *D* (across the four experimental conditions, R^2 ranged from .80 to .97). Thus, a comparison of two experimental conditions made using *TP* may depend on the particular target conditions included in the comparison.

The results in Figure 8 illustrate how the lack of invariance across target conditions of the accuracy adjusted measure of throughput both might or might not be a problem. Looking at Figure 8, it appears that this issue would be of little concern summarizing the effect of the blocked/varied factor: Although the effect may be getting smaller for longer target distances, throughput in the blocked condition was higher than that in the varied condition across the entire range of movement conditions studied.[7] Now consider the subset of the data in Figure 8 that might have resulted from an experiment that studied the effect of the discrete/serial factor using only blocked target conditions. If such an experiment had used only shorter target distances, *TP* would have been found to be higher for serial movements, but the opposite conclusion would have been reached if the focus had been only on longer target distances.

The point of this example is that there is an added danger when generalizing from results summarized using *TP*. The advantage of this measure, that it is a single-valued index of performance, also means that conclusions derived with it can only be safely generalized to situations with a similar set of target conditions. Of course, this issue is of little concern when Fitts's law data are collected for

[7] An interesting observation concerns the hypothesis by Fitts (1954) that throughput would be constant. In these data it appears that throughput does approach an asymptote between 6 and 7 bits/s for values of *D* above 400 pixels. Perhaps this reflects the maximum channel capacity of the motor system, and for shorter movements there is an additional bottleneck limiting throughput.

specific applications in user interface design or product development, because in these situations the target conditions studied can be chosen to match those of the intended application. However, such data are rarely published except, perhaps, in internal company documents. Although one could plausibly argue that ISO 9241-9 (ISO, 2002) is intended to provide guidance in only these applied situations, the recommendation from Soukoreff and MacKenzie (2004) to use *TP* does not seem to be similarly limited in its scope and certainly has been adopted by authors of a number of published studies. Our alternative suggestion is that published comparisons of experimental conditions always present at least the complete set of coefficients for Fitts's law. From this summary, designers can construct single-valued comparisons for an appropriately selected set of target conditions, using *TP* or some other measure. If accuracy adjustment makes sense for data being reported, then both adjusted and unadjusted coefficients should be reported, because, as we have previously noted, in many design situations good information on target utilization will not be available.

Separate from the question of whether *TP* should be used to report comparisons of conditions based on Fitts's law is whether *TP* is the best choice for such comparisons. The allure of *TP* for this role (and *IP* before it) is that it provides a single-valued index of performance. However, *TP* is not the only measure available for such comparisons, and, even in applications for which the experimental target conditions can be chosen appropriately, we are not convinced that it is necessarily the best such measure. As we have illustrated reporting our results, $\widehat{T}_{IDe\,=\,4}$ (and, without an accuracy adjustment, $\widehat{T}_{ID\,=\,4}$) is also a single-valued index of performance that, because it is an estimate at a difficulty level chosen to be close to the average of those used in this experiment, also provides an overall summary of performance that averages across the specific set of target conditions used. Given these similarities and considering that, as in this experiment, *TP* and $\widehat{T}_{IDe\,=\,4}$ often lead to qualitatively similar conclusions, we believe that the choice between $\widehat{T}_{IDe\,=\,4}$ and *TP* (or some other measure!) should depend on which measure is most natural in a particular application. *TP* estimates throughput, measured in bits per second. Throughput was of interest to Fitts because it was central to the theoretical perspective he proposed to interpret these results. Although throughput is a useful concept in many engineering situations, for many applications in interface design or product development, a comparison using units of time, such as $\widehat{T}_{IDe\,=\,4}$, is probably at least, if not more convenient.

5. SUMMARY

The data from mouse movements reported here show how the coefficients of Fitts's law are changed by two design choices. There were three effects. (a) Averaging across levels of difficulty movements took 3.5% longer to complete if the target conditions varied from movement to movement rather than being blocked. (b) Surprisingly,

discrete movements took almost 10% longer than serial movements; however, this additional time appears to have been spent verifying that the movements were ending in the target, and so the effect largely disappears using an analysis adjusting for endpoint accuracy. (c) The Fitts's law slope is higher for serial movements. These results have implications for theories about how movements are planned. However, for the more practical questions that motivated this article, the nature and size of these effects lead us to conclude that neither of these design decisions should present a serious impediment to the generalization of results summarized using Fitts's law. Thus, for example, one might well be justified to generalize results based on an experiment that used discrete movements with blocked targets when designing an application in which the movement "targets" vary and the movements are made serially.

As part of the analysis of these data, we have discussed two widely accepted recommendations concerning how our data should be analyzed and reported. One of these recommendations is to include an accuracy adjustment: that is, using ID_e instead of *ID* in Equation 1. There is much to be said for this approach because it has been shown to compensate for the differences in effective target width that are present across conditions and participants. Doing so allows more data to be described succinctly within the framework of Fitts's law. This approach also can increase the statistical power of comparisons. Both of these advantages were evident here. Despite these advantages, we would add two caveats to this recommendation. First, accuracy-adjusted analyses are only superior when the advantage of ID_e—its potential to reduce systematic error—is large enough to offset its inherent disadvantage—the increased variability of ID_e estimates. This is a trade off the researchers should assess before deciding to use an accuracy-adjusted analysis. Second, we disagree with Recommendation VI of Soukoreff and MacKenzie (2004) concerning how to proceed in design situations where ID_e values are not available. In this situation, we recommend that predictions be produced by inserting nominal *ID* values only into Equation 1 with coefficients estimated using *ID*. To make this possible, however, published reports of Fitts's law data should include, as we have here, the coefficients obtained both with and without the accuracy adjustment.

We also question the generality of Soukoreff and MacKenzie's (2004) Recommendation VII to use *TP*, an accuracy adjusted throughput measure, for comparing experimental conditions. We have two concerns. First, conclusions reached using *TP* may depend on the range of *ID* values sampled in an experiment. When those values closely match those of the to-be-generalized–to situation, there is no problem; however, if they do not, then the results cannot safely be assumed to apply. Thus, it would be a mistake to report only *TP* values in published results, as some authors, apparently following the recommendation of Soukoreff and MacKenzie have done; at a minimum both coefficients for Equation 1 need to be reported so that others can use a published summary to generate a comparison, perhaps using *TP*, for the specific *ID* values of interest. More broadly, however, *TP* is not the only possible single-valued index of performance that can be used to compare experimental conditions. Although, for many applications, throughput may provide a natural way to summarize

performance, for at least some other applications an alternative, such as average movement time, will be more useful. This is an issue that designers should consider rather than blindly following a blanket recommendation.

NOTES

Acknowledgments. We acknowledge the many fruitful discussions we have had about this research with our colleague Professor Charles Chubb and the many useful comments and suggestions we received on previous drafts of the manuscript from Yves Guiard and two anonymous reviewers.

Authors' Present Addresses. Charles E. Wright, Department of Cognitive Sciences, 2554 Social & Behavioral Sciences Gateway Building (SBSG), University of California, Irvine, CA 92697. E-mail: cewright@uci.edu. Francis Lee, 766 Vespucci Lane, Foster City, CA 94404. E-mail: francil@uci.edu.

HCI Editorial Record. Received September 21, 2011. Revision received September 4, 2012. Final manuscript received May 6, 2013. Accepted by Shumin Zhai.

REFERENCES

Balakrishnan, R. (2004). "Beating" Fitts' law: Virtual enhancements for pointing facilitation. *International Journal of Human-Computer Studies, 61*, 857–874.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics, 21*, 601–613. doi:10.1080/00140137808931762

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human–computer interaction*. Hillsdale, NJ: Erlbaum.

Chapuis, O., & Dragicevic, P. (2011). Effects of motor scale, visual scale, and quantization on small target acquisition difficulty. *ACM Transactions on Computer-Human Interaction, 18*, 13:1–32. doi:10.1145/1993060.1993063

Crossman, E. R. F. W. (1956). The measurement of perceptual load in manual operations (Unpublished doctoral dissertation). University of Birmingham, UK.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381–391. doi:10.1037/h0055392

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology, 67*, 103–112. doi:10.1037/h0045689

Fitts, P. M., & Radford, B. K. (1966). Information capacity of discrete motor responses under different cognitive sets. *Journal of Experimental Psychology, 71*, 475–482. doi:10.1037/h0022970

Freund, H. J., & Büdingen, H. J. (1978). The relationship between speed and amplitude of the fastest voluntary contractions of human arm muscles. *Experimental Brain Research, 31*, 1–12.

Gan, K.-C., & Hoffmann, E. R. (1988). Geometrical conditions for ballistic and visually controlled movements. *Ergonomics, 31*, 829–839.

Gordon, J., & Ghez, C. (1987). Trajectory control in targeted force pulses. II. Pulse height control. *Experimental Brain Research, 67*, 241–252.

Guiard, Y. (1993). On Fitts' and Hooke's laws: Simple harmonic movement in upper-limb cyclical aiming. *Acta Psychologica, 82*, 139–159.

Guiard, Y. (1997). Fitts' Law in the discrete vs. cyclical paradigm. *Human Movement Science, 16*, 97–131.

Guiard, Y., & Beaudouin-Lafon, M. (2004). Fitts' law 50 years later: Applications and contributions from human-computer interaction. *International Journal of Human-Computer Studies, 61*, 747–750. doi:10.1016/j.ijhcs.2004.09.003

ISO. (2000). ISO 9241-9 International standard: Ergonomic requirements for office work with visual display terminals (VDTs)—Part 9: Requirements for non-keyboard input devices: International Organization for Standardization.

Kvålseth, T. O. (1980). An alternative to Fitts' law. *Bulletin of the Psychonomic Society, 16*, 371–373.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction, 7*, 91–139.

Megaw, E. D. (1975). Fitts tapping revisited. *Journal of Human Movement Studies, 1*, 163–171.

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. K. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review, 95*, 340–370.

Meyer, D. E., Smith, J. E. K., Kornblum, S., Abrams, R. A., & Wright, C. E. (1990) Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 173–226). Hillsdale, NJ: Erlbaum.

Munro, H., Plumb, M. S., Wilson, A. D., Williams, J. H. G., & Mon-Williams, M. (2007). The effect of distance on reaction time in aiming movements. *Experimental Brain Research, 183*, 249–257. doi:10.1007/s00221-007-1040-y

Pastel, R. (2011). Positioning graphical objects on computer screens: A three-phase model. *Human Factors, 53*, 22–37. doi:10.1177/0018720810397353

Plamondon, R., & Alimi, A.M. (1997). Speed/accuracy trade-offs in target-directed movements. *Behavioural and Brain Sciences, 20*, 279–349.

Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts's law research in HCI. *International Journal of Human-Computer Studies, 61*, 751–789. doi:10.1016/j.ijhcs.2004.09.001

Welford, A. T. (1968). *Fundamentals of skill*. London, UK: Methuen.

Wright, C. E., & Meyer, D. E. (1983). Sources of the linear speed-accuracy tradeoff in aimed limb movements. *Quarterly Journal of Experimental Psychology, 35A,* 279–296.

Zhai, S. (2004). Characterizing computer input with Fitts' law parameters—The information and non-information aspects of pointing. *International Journal of Human-Computer Studies, 61*, 791–809. doi:10.1016/j.ijhcs.2004.09.006

Zhai, S., Kong, J., & Ren, X. (2004). Speed–accuracy tradeoff in Fitts' law tasks—on the equivalency of actual and nominal pointing precision. *International Journal of Human-Computer Studies, 61*, 823–856. doi:10.1016/j.ijhcs.2004.09.007