Speed-accuracy trade-off in Fitts’ law tasks —
On the equivalency of actual and nominal pointing precision

Shumin Zhai
IBM Almaden Research Center, San Jose, California, USA

Jing Kong
Information Systems Engineering, Kochi University of Technology, Japan

Xiangshi Ren
Information Systems Engineering, Kochi University of Technology, Japan

ABSTRACT
Pointing tasks in human computer interaction obey certain speed-accuracy tradeoff rules. In general, the more accurate the task to be accomplished, the longer it takes, and vice versa. Fitts’ law models the speed-accuracy tradeoff effect in pointing as imposed by the task parameters, through Fitts’ index of difficulty ($I_d$) based on the ratio of the nominal movement distance and the size of the target. Operating with different speed or accuracy biases, performers may utilize more or less area than the target specifies, introducing another subjective layer of speed-accuracy tradeoff relative to the task specification. A conventional approach to overcome the impact of the subjective layer of speed-accuracy tradeoff is to use the a posteriori “effective” pointing precision $W_e$ in lieu of the nominal target width $W$. Such an approach has lacked a theoretical or empirical foundation. This study investigates the nature and the relationship of the two layers of speed-accuracy tradeoff by systematically controlling both $I_d$ and the index of target utilization $I_u$ in a set of four experiments. Their results show that the impacts of the two layers of speed-accuracy tradeoff are not fundamentally equivalent. The use of $W_e$ could indeed compensate for the difference in target utilization, but not completely. More logical Fitts’ law parameter estimates can be obtained by the $W_e$ adjustment, although its use also lowers the correlation between pointing time and the index of difficulty. The study also shows the complex interaction effect between $I_d$ and $I_u$, suggesting that a simple and complete model accommodating both layers of speed-accuracy tradeoff may not exist.

Key words and phrases: Pointing, input, speed-accuracy tradeoff, Fitts’ law, modeling, human performance.

1. Introduction
Acclaimed as one of the most successful human performance models (Newell, 1990), Fitts’ law has served as one of the few quantitative foundations for human-computer interaction research. In particular, it has been used as a theoretical framework for computer input device evaluation (e.g. Card et al., 1978; ISO, 2000; MacKenzie, 1992), a tool for optimizing new interfaces (e.g. Lewis et al., 1992; MacKenzie and Zhang, 1999; Zhai et al., 2002), as well as a logical basis for modeling more complex human-computer interaction (HCI) tasks (Accot and Zhai, 1997). Fitts’ law has also inspired alternative interaction techniques (e.g. Accot and Zhai, 2002; Kabbash and Buxton, 1995; Zhai et al., 1999) and gained new understandings, expansions, and applications in human-computer interaction research in recent years (e.g. Accot and Zhai, 2003; Guiard et al., 2001; McGuffin and Balakrishnan, 2002; Zhai et al., 2003).
Despite its impressive successes and critical importance, however, some of the fundamental issues in Fitts’ law, either as a general human performance model or as a tool for human-computer interaction research, are still not fully understood in the literature. Speed-accuracy tradeoff is one of them.

In essence, Fitts’ law is about revealing the rule of speed-accuracy tradeoff in human control performance. As in many human-performed tasks, the more precisely the task is to be accomplished, the slower it is. Conversely, the faster the task is completed, the less precisely the task tends to be performed. For pointing tasks, Fitts’ law precisely models how task precision affects pointing completion time:

\[ T = a + b \log_2 \left( \frac{D+W}{W} \right) \]  

(1)

where \( a \) and \( b \) are constants reflecting the efficiency of the pointing system, \( D \) is the pointing distance, \( W \) is the target width and \( T \) is the mean (expected) time of task completion. In other words, Fitts’ law formally models pointing speed (in the form of completion time \( T \)) as a function of relative task precision \( \frac{W}{D+W} \).

In Fitts’ law task precision is quantified by an index of difficulty (\( I_d \)), which has several forms in the literature. Here we use the so-called Shannon formulation since it is the most logical one at the boundary condition \( D = 0 \) (MacKenzie, 1989):

\[ I_d = \log_2 \left( \frac{D+W}{W} \right) \]  

(2)

Note that so far the precision parameters, \( D \) and \( W \), are \textit{a priori} task parameters. In this sense the original Fitts’ law is about the relationship of temporal (speed) performance and \textit{task} precision. Ideally the human performer uses all of the precision tolerance (\( W \)) that the task specifies, no more, no less. Statistically, this means the spread of the endpoints (hits) of Fitts’ aimed movements corresponds to the target width. For a Gaussian distribution, which the endpoints of a Fitts’ law task are expected to form, the spread of endpoints can be measured by its standard deviation \( \sigma \). Conventionally, the ideal case is that the central 96% of the endpoints (within \( \sqrt{2\pi}\sigma = 4.1325\sigma \)) corresponds to \( W \), i.e. \( \sigma = W/4.1325 \) (Figure 1).

---

\(^1\) In many scientific fields, accuracy and precision are distinguished: accuracy refers to the nearness of a measurement to true value (a constant), and precision is the closeness of repeated measurements of the same quantity (dispersion). By such a distinction, precision would be a better term here, since Fitts’ law is concerned with movement “resolution” (distance to width ratio). However, in the Fitts’ law literature the term accuracy is conventionally used. Furthermore, it has not been of interest to Fitts’ law research to study precise but inaccurate cases, namely, a group of trials repeatedly hitting a narrow band (precise) but off the center of the target (inaccurate). The dictionary definitions of the two terms are often interchangeable, and both are related to “exactness.” In the rest of paper we do not distinguish these two terms, due to these considerations.
In actuality, either when performing laboratory experiments or when selecting graphical user interface (GUI) widgets on a computer with a mouse, the human performer (or the computer user in the context HCI) may or may not comply with the task precision as specified by $W$ (Figure 1) – the width of the endpoints dispersion may depart from the target width $W$, causing either over or under utilization of the target area. In other words, the performer may introduce another layer of precision choice relative to the nominal task precision. The performer may be biased towards accuracy and use less area than the target gives, resulting in a more accurate than necessary but slower performance. Conversely, the performer may be biased towards speed and use more area than the target gives, resulting in a faster but more error prone performance. This second layer (or component) of speed-accuracy tradeoff is subjective and personal (hereafter referred as the subjective layer). In contrast, the bottom layer is objective (task specified) and nominal.

The existence of the subjective layer of accuracy complicates Fitts’ law. Facing such a complication, students of Fitts’ law have implicitly or explicitly taken two views. One is to continue to treat Fitts’ law as a task model as in Equation 1 based on $D$ and $W$, two nominal, a priori, and deterministic pointing task parameters. Alternatively, Fitts’ law can be re-written based on actual behavioral parameters:

$$T = a + b \log_2 \left( \frac{D_e + W_e}{W_e} \right)$$  \hspace{1cm} (3)

correspondingly,

$$I_{de} = \log_2 \left( \frac{D_e + W_e}{W_e} \right)$$  \hspace{1cm} (4)

where,

$$W_e = 4.133\sigma$$  \hspace{1cm} (5)
where $D_e$ and $W_e$ (or $\sigma$) are *a posteriori* and statistical measurement of the actual movements\(^2\). As practiced in the literature, they are called effective distance and effective target width, although the implication of the term effective is a biased one without proof. For distinction, we refer Equation 1 as the task form of Fitts’ law and Equation 3 as the behavior form of Fitts’ law. One of the goals of the current study is to compare and contrast the two forms.

There are both practical and theoretical implications to each form. We will discuss the theoretical implications shortly. For practical purposes, particularly for work in HCI, the desirability of task vs. behavior form of Fitts’ law isn’t clear cut. Often it is desirable to relate a user’s performance to the geometry of a graphical user interface. For example, for an interface designer it is important to know a user’s average selection time as a function of task parameters such as a GUI widget in certain size and location. For another example, in Fitts’ law based stylus keyboard optimization research researchers are concerned tapping time in relation to the geometrical variables – the relative location of keys (e.g. Lewis et al., 1992; MacKenzie and Zhang, 1999; Zhai et al., 2002). A task form of Fitts’ law is more desired in these cases. On the other hand, it is logically difficult to expect Fitts’ law in its task form to serve as a reliable tool in evaluating the performance of an input device characterized by $a$ and $b$ constants in Fitts’ law (cf. Zhai, this issue) if the performer’s actual pointing precision deviates too far from the specified task precision. In this case, it is reasonable to expect that Fitts’ law in its behavioral form, or at least some modification of the task form of Fitts’ law in consideration of the deviation, be more useful.

Instead of viewing the two forms as opposing models, a potentially more complete and more sophisticated approach is to represent the two layers, or two components, of speed-accuracy tradeoffs in one model. Such a model should relate time (or speed) to two independent factors, both concerning precision. The first is the task precision, as specified by target parameters $D$ and $W$. The second independent factor is the degree of target area utilization the performer chooses. This is relative to the specified task precision and subjectively introduced by the performer. The performer may choose to be less precise than the task specification and use more area than $W$ and hence gain a faster speed, or choose to be more precise than the task specification and use less area than $W$, causing a slower completion speed. Another goal of this study is to explore whether there exists a model of Fitts’ law that explicitly relate time to both layers of speed-accuracy tradeoff.

The degree of target area utilization has to do with the risk of missing the target the performer is willing to take. A more risky behavior tends to result in a wider endpoints distribution and cause over-utilization of the target area. A more risk-averse behavior tends to result in a narrower endpoints distribution and cause under-utilization of the target area.

2. **Theoretical and literature analysis**

The possible mismatch between the actual pointing precision and the nominal task specification has long been realized by researchers (e.g. Crossman and Goodeve, 1983/1963; Fitts and Radford, 1966; Welford, 1968), but the topic has lacked rigorous treatment in the literature, effectively being kept “under the carpet”. The most common method of compensating for this mismatch to date relies on the behavioral form of Fitts’ law. If the performer over- or under-utilized target width $W$, the *a posteriori* actual precision measured by $4.1325\sigma$ of the endpoints distribution, or $W_e$, would be used in lieu of nominal target width $W$. In the

\(^2\) Since the difference of $D$ vs. $D_e$ is comparatively smaller than that of $W$ vs. $W_e$, this study focuses on the latter.
Fitts’ law literature (e.g. MacKenzie, 1991, section 2.4), the justification for the use of $W_e$ is commonly traced to Welford (1968), which in turn attributes it to a report to the British Medical Research Council (MRC) by Crossman and Seymour in 1957. Based on Crossman and Goodeve (1983/1963), Crossman’s Ph.D. thesis (Crossman, 1956) appears to be the earliest formal source of this formulation, although Fitts and Radford (1966) suggested that the use of effective width could even be traced back long before Fitts’ law to Woodworth (1899). Crossman’s reasoning of using $W_e$ (Crossman, 1956), cited in Crossman and Goodeve (1983/1963), is as follows:

“I$_d$ could be viewed as the difference between two more fundamental quantities, $\log_2 W$ measuring the entropy of the endpoint distribution and $\log_2(2A)^2$, measuring the entropy of a hypothetical initial distribution of motion amplitudes. Since, as noted by Fitts, endpoints are observed to be normally distributed about the center of the target, the theoretically correct expression for endpoint entropy $H(o)$, is: $H(o) = \log_2 (2\pi e)^{1/2} \sigma_x$.”

Some researchers of Fitts’ law advocate $W_e$ since “This adjustment lies at the very heart of the information-theoretic metaphor that movement amplitude area analogous to ‘signals’ and end-point variability (viz. target width) is analogous to ‘noise’.” (e.g. MacKenzie, 1991, section 2.4). This information-theoretic foundation, however, isn’t very formal. First, the concepts of information in both Fitts’ initial work (Fitts, 1954) and Crossman’s reasoning, cited above, are more metaphorical than mathematical. Second, Crossman’s logic based on a two-component analysis of $I_d$ involving a separate entropy of “a hypothetical initial distribution of motion amplitudes” measured by $\log_2 (2A)$, is a difficult one. Third, the information-theoretic account of Fitts’ law, more specifically the analogy of signal transmission over a noisy channel (Shannon, 1948), is one among many competing theoretical explanations of Fitts’ law (e.g. Crossman and Goodeve, 1983/1963; Keele, 1968; Meyer et al., 1988). In the same article by Crossman and Goodeve cited above, they concluded, “It seems difficult to develop a truly information derivation of Fitts’ law” (p. 256) and went on to develop their control theory explanation of Fitts’ law. Paul Fitts’ own view on the theoretical basis of Fitts’ law was not rigid. In their discussion of different indices of difficulty, Fitts and Peterson clearly stated, “Since neither index has been derived formally from a theory, choice between them should rest on heuristic considerations” (Fitts and Peterson, 1964, p. 111).

For clarity of notations, we hereafter denote the index of difficulty based on the nominal target width, as defined by Equation 2, $I_{de}$, and use $I_d$ as a generic reference to index of difficulty.

In practice, the use of $I_{de}$ has been adopted by some, but not all, researchers. Conceptually the use of $I_{de}$ requires the assumption that the actual behavioral precision and the nominal task precision in fact have the same impact: a reduction or expansion of actual endpoint dispersion $W_e$ as a result of the performer’s subjective bias is equivalent to the same amount of reduction or expansion of the nominal target width $W$. A compelling theoretical question here is whether a posteriori measurement is indeed cognitively, or quantitatively, equivalent to a priori task specification.

Empirically, there is a scarcity of evidence to prove or disprove whether and how well substituting $I_{de}$ with $I_d$ could compensate for the influence of nominal and actual precision mismatch caused by the subjective layer of speed-accuracy tradeoff. It will be empirically informative to measure whether human performance adjusted with effective width is equivalent to the performance under a nominal width of the same size, had the performer complied with the exact target specification.

---

3 Due to its signal transmission analogy, the distance to target in Fitts’ law, denoted as $D$ in this paper, is also referred to as amplitude and denoted as $A$. Note also originally Fitts defined $I_d = \log_2 (2A/W)$, hence Crossman’s $\log_2(2A)$ here.
Aiming for the “standard” behavior of approximately 4% error rate, most Fitts’ law studies instructed the participants to perform “as fast as possible and as accurately as possible,” but do not systematically vary or control experimental participants’ actual precision relative to the nominal task precision; hence it is difficult to evaluate the influence of the subjective layer of speed-accuracy tradeoff and the correction effects using $I_{de}$. Fitts and Radford (1966) is a rare exception. They systematically manipulated three subjects’ operational bias towards accuracy (A), neutrality (N), and speed (S), by means of monetary award and penalty at 1 cent per point. For example, to bias the subjects toward a speed performance, the points for fast and correct, slow and correct, fast and wrong, and slow and wrong were +1, -1/2, -1/8 and -1, respectively.

Fitts’ thesis was that human information capacity in motor responses is relatively constant despite different experimental manipulations, so their paper did not focus on the effect of $I_{de}$ correction. Based on the data provided in (Fitts and Radford, 1966), however, we could construct Table 1 summarizing the average (over three participants) Fitts’ law parameters in two of their experiments, using both $I_{de}$ and $I_{dn}$.

<table>
<thead>
<tr>
<th>Regression coefficients</th>
<th>Bias</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$I_{de}$ based</td>
<td>$I_{de}$ based</td>
</tr>
<tr>
<td>a (ms)</td>
<td>A</td>
<td>-114</td>
<td>-148</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>-85</td>
<td>-103</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>-52</td>
<td>-75</td>
</tr>
<tr>
<td>b (ms/bit)</td>
<td>A</td>
<td>73</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>66</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>56</td>
<td>77</td>
</tr>
</tbody>
</table>

Table 1. Summary of (Fitts and Radford, 1966)

Table 1 shows that the experimental bias conditions yielded very different $b$ (and $a$) values using $I_{de}$. From condition A to condition S, $b$ varied 30% and 28.3% in Experiments 1 and 2, respectively. Risk-averse operation (A) tended to increase $b$ and risky operation (S) tended to decrease $b$. Using $I_{de}$, the $b$ values converged, but not completely. $b$ still varied 20.8% and 14.5% in Experiments 1 and 2. The same trend was true for $a$ – converging, but not completely. The use of $I_{de}$ did not appear to be able to completely compensate for performers’ biases. If it did, we would expect these parameters to converge to the ideal values.

However, Fitts’ and Radford’s study (1966) has several weaknesses as a foundation for understanding the two layers of speed-accuracy tradeoff in pointing. First, it was not targeted at this topic, so a more in-depth investigation based on their data is difficult. Second, the number of experimental participants was very small in their experiments. Third, the error rates (and hence the target area utilization levels) in the study were all very far from the expected 4%. For example, the error rates of the three bias conditions were 24%, 27% and 34%, respectively, in their Experiment 1. If the target utilization levels in Fitts and Radford (1966) were closer to the ideal and deviated to both the over and under utilization sides, the results could be different.

In a more recent systematic study of speed-accuracy tradeoff in aimed movements, Adam (1992) found that there was indeed a speed accuracy trade-off at the subjective level: given the same task constraints the performers either produced very accurate but slow movements or very fast but inaccurate movements, depending on the “objectives” of the performers in the accuracy-speed bias spectrum. Different from the current study, the focus of Adam’s study was how these objectives influence the kinematic features of reciprocal aiming movements. For example, the accuracy group of performers had lower peak speed than
the speed group. In terms of Fitts’ law modeling, Adam found that substituting nominal target width with effective width increased the percentage of time variance accounted by Fitts’ law. Fitts’ law modeling details, such as the performance parameters \( a \) and \( b \), were not reported in the study. The direction (under or over utilization of the target constraint) and amount of subjective bias were not quantified. The main conclusion of Adam’s study is that ambiguous task instructions, such as to move as quickly and accurately as possible, are vulnerable to different strategic interpretations varying from an emphasis on accuracy to an emphasis on speed (Adam, 1992).

The focus of this study is on the modeling aspects of speed-accuracy tradeoff in aimed movements. The first task of is to empirically evaluate the \( I_{de} \) based behavior form of Fitts’ law, when the target utilization levels are controlled to both sides of the ideal case. For clarity and convenience, we first formally define the notation of target width utilization. It is logical to assume that movement time is a function of both nominal task difficulty as quantified by \( I_{dn} \) and the level of target width (over) utilization by the performer, as quantified by \( I_u \). Without committing to a particular form of the function, we have:

\[
T = f(I_{de}, I_u)
\]

where the index of target width (over) utilization \( I_u \) is formally defined as:

\[
I_u = \log_2\left(\frac{W}{W'}\right)
\]

or

\[
I_u = \log_2\left(\frac{4.133\sigma}{W}\right)
\]

The logarithmic transformation here is merely for mathematical convenience and symmetry with \( I_d \). The absolute value of \( I_u \) indicates the degree to which the actual spread of the endpoints departs from the specified target width. A positive \( I_u \) means the performer over utilizes the target width and misses the target (error) with more than 4% probability. A zero \( I_u \) means the performer perfectly utilizes all variability specified by the task, no more, no less, and the error rate is exactly 4%. A negative \( I_u \) means that the performer under utilizes the target area, leaving a certain amount of safety margin, and misses the target with less than 4% probability.

In order to thoroughly study speed-accuracy tradeoff in pointing, in particular whether the impact of a non-zero \( I_u \) is compensated by replacing the task form of Fitts’ law \( T = f(I_{de}) \) with a behavior form of Fitts’ law \( T = f(I_{de}, I_u) \), empirical studies which systematically manipulate both index of difficulty \( I_{dn} \) and index of target utilization \( I_u \) have to be conducted.

### 3. Experiment 1

#### 3.1 Set-up and design

Twelve volunteers, of different gender (9 male and 3 female) and age (21 to 38 years, mean 26), participated in a target pointing experiment on a tablet computer (Fujitsu FMV Stylistic) with a screen size of 21 cm x 15.6 cm. Each pixel on the screen was 0.2055 mm wide. Similar to Fitts’ original experiment (Fitts, 1954), participants did reciprocal pointing on a pair of vertical strip targets with a stylus. The width \( W \) of the targets and the center-to-center distances \( D \) between the two strips were set at \( W = 12, 36, 72 \) pixels and \( D = 120, 360, 840 \) pixels. The order of the nine width and distance combinations was randomized. Twelve trials were presented in each \( W, D \) combination, with the first tap excluded in analysis. If tapped on the outside of the target, an auditory signal was played.

Each participant was instructed to repeat the experiment three times with different operating conditions biased toward accuracy or speed: accurate (A), neutral (N), and fast (F). They were instructed to tap the
targets “as accurately as possible” in Condition A, “as accurately as possible and as fast as possible” in Condition N, and “as fast as possible” in Condition F. The goal was to make the participants operate at different levels of target utilization.

The order of the A, N, F conditions was balanced by a Latin square pattern across the twelve participants.

### 3.2 Data processing and analysis

Occasionally, “accidental clicks” outside the general region of the target were registered, due to either the confusion of the participant, or instrument error. We used two simple and conservative rules to remove these “outliers” from further analysis to prevent their disproportional impact on modeling (Press et al., 1992). The first rule of removal was that the user hit the same target the trial started from or the user landed in the direction opposite to where the destination target was. This was determined by the distance between the hit point and the target center being greater than $D-W/2$. Twenty-five trials were removed by this rule. The second rule was that the distance of the endpoint to the target center was 8 times greater than the target size. Three additional trials were removed by the second rule. A total of 28 trials were removed by these two rules, constituting a small percentage of the total number of trials (3564).

#### 3.3 The index of target utilization $I_u$

The instructions for the operating strategy in the three experimental conditions had an obvious impact on participants’ target utilization and error rate. The average error rate in the A, N, F conditions was 3.2%, 10% and 19.4%, respectively. These rates overall were higher than what we hoped. Ideally the error rate in the N condition would be around 4% and A and F conditions be on the two opposite sides.

A wide range of target utilization levels was taken in the experiment by the participants (Figure 2). $I_u$ varied from -1 to 1.5 bit. Note that 1 bit of $I_u$ change means that the spread of the endpoints is twice or half of the specified target width. The operating conditions (A, N, F) changed the overall $I_u$ level. Furthermore, participants also shifted from target under utilization in low $I_{dn}$ to target over utilization in high $I_{dn}$ trials, regardless of the condition. Even under the N (neutral) condition, $I_u$ was not maintained at the ideal level (zero bit). This could be a deliberate choice of strategy shift by the human performers, or it could be fundamentally difficult to maintain the same level of target utilization facing targets with different $I_{dn}$. Guiard (2002) explains the same effect from a power constraint vs. precision constraint perspective, albeit in different terminologies. Note also that the amount of shift caused by $D$ and $W$ change were not the same (Table 2).

![Figure 2. The index of target utilization $I_u$ changes with instruction and $I_{dn}$](image-url)

<table>
<thead>
<tr>
<th>Instruction Bias</th>
<th>A</th>
<th>N</th>
<th>F</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4 The nominal $I_{dn}$ model

As a baseline for further analysis, we first applied the basic task form of Fitts’ law (Equation 1) to the data collected, using the task’s nominal index of difficulty $I_{dn}$. The result is shown in Figure 3.

![Figure 3. Linear regression $T$ vs. $I_{dn}$ in Experiment 1](image)

When analyzed separately, within each operating condition $T$ and $I_{dn}$ correlated strongly. 95% to 99% of variance in $T$ could be accounted for by the change of $I_{dn}$. There was a tendency that the more risky (faster-paced) the operating condition was, the stronger the correlation between $T$ and $I_{dn}$ was. The robustness of $I_{dn}$ prediction here is quite remarkable given the very different operating biases (different levels of target width utilization) in these conditions. However, the coefficients of the regression results (or $a$ and $b$ in Equation 1) varied from one condition to another. Table 3 summarizes the results. The statistical significance levels ($p$, based on F tests) are also listed in italic font under each parameter, and the overall regression significance is listed under the $r^2$ value of the regression.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>$a$</th>
<th>$b$</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>160.7</td>
<td>149.8</td>
<td>0.949</td>
</tr>
<tr>
<td>$p (F_{1,7})$</td>
<td>.013</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>N</td>
<td>122.6</td>
<td>118.1</td>
<td>0.981</td>
</tr>
<tr>
<td>$p (F_{1,7})$</td>
<td>.001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>F</td>
<td>123.1</td>
<td>92.4</td>
<td>0.994</td>
</tr>
<tr>
<td>$p (F_{1,7})$</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Mixed</td>
<td>135.4</td>
<td>120.1</td>
<td>0.696</td>
</tr>
<tr>
<td>$p (F_{1,23})$</td>
<td>.036</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Table 3. Summary of $T$ vs $I_{dn}$ regression in Experiment 1
Results in Table 3 clearly shows that Fitts’ law regression coefficients $a$ and $b$ using the $I_{de}$ model were influenced by the operating conditions (biases). $b$ varied 62% from Condition F to A in this experiment.

If we perform the same Fitts’ law regression based on the data from all conditions mixed, while still keeping the same unit of analysis $8 (I_{de} \text{ levels}) \times 3$ (operating conditions), we obtain a much weaker correlation ($r^2 = 0.696$); see the last row of Table 3, although the correlation is still statistically significant.

3.5 The effective $I_{de}$ model

We now test if the use of effective width, the behavior form of Fitts’ law, would be able to compensate for the difference of operating conditions (strategy biases). The linear regression results between mean trial completion time $T$ and the effective index of difficulty $I_{de}$ are shown in Figure 4 and Table 4.

Comparing Figure 3 vs. 4 and Table 3 vs. 4, the following observations can be made on the use of $I_{de}$.

First, the regression coefficients under different operating conditions were much closer to each other with $I_{de}$, showing the effect of compensation for the different levels of target utilization. The $a$ values from different conditions were all reduced and the $b$ values from different conditions became much closer to each other. From Conditions F to A, the $b$ value’s change was now 17.6% (in comparison to 62%).
Second, within each operating condition, $r^2$ between $T$ and $I_{de}$ decreased from the corresponding $r^2$ between $T$ and $I_{dn}$. Only 83% to 87% of the $T$ variance could be accounted for by $I_{de}$ (in comparison to 95% to 99% by $I_{dn}$). The same trend of $r^2$ reduction can be observed in the data of Fitts & Radford (1966). Why this was true will be discussed later. A counterexample to this trend was seen in Mackenzie’s recalculation of Fitts’ 1954 data (MacKenzie, 1991, section 2.5 Table 2), which found a slight increase in $r^2$ by using $I_{de}$ (from 0.983 to 0.99 and 0.98 to 0.988 respectively). Note that there the initial correlation was very high, and the change of correlation was small. Furthermore, since Fitts’ 1954 data did not have endpoints location recordings, the recalculation of $W_r$ was based on Z-score conversion from error rates – an imprecise or arbitrary estimation method when the error rate is low or zero.

Third, overall there was a shrinkage of the range of the independent variable from $I_{dn}$ to $I_{de}$. Within the same condition, particularly for the more risky condition F, there was a counter-clockwise rotation of the regression line.

Fourth, if we use data from all operating conditions together, the $r^2$ value between $T$ and $I_{de}$ increased from the corresponding $r^2$ value between $T$ and $I_{dn}$ regression (0.723 to 0.825), showing a stronger regularity of $T = f(I_{de})$ than $T = f(I_{dn})$ in modeling pointing time in the presence of a wide range of target utilization.

In summary, the use of $I_{de}$ demonstrated both benefits and drawbacks. It compensated operating biases (more converging $a$ and $b$ parameters), but not completely. Its robustness as a determinant of pointing time as measured by $r^2$ decreased within each operating condition but increased across conditions.

The utilization level in this experiment overall tilted to over utilization (positive $I_u$, see Figure 2). Even in Condition A, the overall error rate was 3.2%. It will be informative to also observe the more conservative under utilization side, as in the next experiment.

5. Experiment 2

The experiment had two operating conditions: A & F. In Condition F, the instruction was, “Move as fast as possible. It is okay if a few errors are made.” A gentle “ding” sound was played when an error was made. In Condition A, which in fact could be called the EA (extremely accurate) condition, participants were instructed to “Try to avoid any errors,” and a loud “ding” sound was played when a target was missed.

Eleven people of different gender (4 female and rest male) and age (20’s to 50’s), who had not been in the first experiment, participated in this experiment with both conditions. They were alternated between A to F and F to A order. An LCD display with stylus touch-sensitive surface (Wacom LCD Tablet Model PL-400) was used as the experiment apparatus. The rest of the experiment setup remained the same as Experiment 1.

Due to the quality of the tablet used in this experiment, many more erroneous trials were found. If the stylus struck the surface of the tablet too hard (and quickly bounced up), no click was registered. This meant the next click could be aimed at the “wrong” target. Using the same two rules outlined in Experiment 1, out of a total of 2178 trials in this experiment, 160 trials were removed by the first rule (tapped on the wrong side). Another 74 trials were removed by the second rule.

Under the instruction given in this experiment, participants indeed exhibited more conservative (risk-averse) behavior. Figure 5 shows the index of target utilization (cf. Figure 2). The $I_u$ values at almost all $I_{de}$ levels were negative (the average error rates were 0 and 0.5% respectively for A and F conditions, see Table 6), but participants were clearly more conservative in Condition A than in Condition F. We tried to bias the participants in this experiment to the direction opposite to Experiment 1 that was overall on the risky side. Note that the operating condition labels (A, F, etc) are relative within each experiment.

4 Note that two $W:D$ combinations, 12:120 and 36:360, resulted in the same $I_{de}$ but different $I_{de}$ due to the different amount of shift. The degrees of freedom in Table 3 and 4 were hence different.
Table 5 and Figure 6 show $I_{dn}$ and $I_{de}$ model regression results. Observations similar to those in Experiment 1 can be made here, but to a lesser degree: (1) As indicated by $r^2$ values $I_{dn}$ is a (slightly) more robust pointing time determinant than $I_{de}$ within each condition; (2) $I_{de}$ is more robust than $I_{dn}$ across conditions. (3) The range of $I_{de}$ shrunk from that of $I_{dn}$, but only from the low end of the index of difficulty this time. In fact for the A condition, the high end of $I_{de}$ was further extended to the right. (4) $I_{de}$ yielded more converging $a$ and $b$ parameters between the conditions than $I_{dn}$, hence it compensated for the different target utilization levels in different operating bias conditions (but not completely and to a lesser degree than in Experiment 1). Overall the changes caused by substituting $I_{dn}$ with $I_{de}$ are lesser in this experiment.

To verify the observations in these two experiments, we decided to conduct a more comprehensive experiment that covers a wide target utilization range on both sides.
5. Experiment 3

5.1 Set-up and design

Fifteen volunteers, 5 female and 10 male, aged 20 to 36 years old, participated in Experiment 3, in which the same experimental apparatus, software, and procedure as in Experiment 1 were used. The difference is that a greater range of target utilization is inducted: each participant was instructed to repeat the experiment five times with different operational strategies: extremely accurate (EA), accurate (A), neutral (N), fast (F) and extremely fast (EF). The following verbal instructions corresponding to each task were given by the experimenter to the participants: “Perform as accurately as possible and don’t worry about time or speed; try to avoid any error” in Condition EA; “as accurately as possible but keep some speed” in Condition A; “as accurately as possible and as fast as possible” in Condition N; “as fast as possible but keep some accuracy” in Condition F; and “as fast as possible and some errors are acceptable” in Condition EF.

5.2 Data processing and basic results

Similar to Experiments 1 and 2, 16 accidental trials were removed from the data pool out of a total of 7425 trials.

The error rates of the five conditions varied according to the verbal instructions, and the overall error rates in the EA, A, N, F, EF conditions were 0%, 1%, 4%, 9%, 22%, respectively, which were rather ideal because in condition N the error rate was at the standard 4% and the rest of the conditions were distributed symmetrically. As shown in Table 6, the overall $I_u$ levels of the first two experiments were either tilted to positive (Experiment 1) or negative (Experiment 2). This experiment is similar to a combination of the first two.

<table>
<thead>
<tr>
<th>Bias</th>
<th>$I_u$ max (bit)</th>
<th>$I_u$ min (bit)</th>
<th>$I_u$ max (bit)</th>
<th>$I_u$ min (bit)</th>
<th>$I_u$ max (bit)</th>
<th>$I_u$ min (bit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EA</td>
<td>3.2</td>
<td>-0.78</td>
<td>0</td>
<td>-0.33</td>
<td>1</td>
<td>-0.17</td>
</tr>
<tr>
<td>A</td>
<td>10</td>
<td>-0.88</td>
<td>4</td>
<td>0.23</td>
<td>0.83</td>
<td>-0.73</td>
</tr>
<tr>
<td>F</td>
<td>19.4</td>
<td>-0.44</td>
<td>9</td>
<td>2.04</td>
<td>-0.59</td>
<td></td>
</tr>
<tr>
<td>EF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. The $I_u$ range and error rate in each operating condition of the first three experiments; the highlighted cell are the most neutral conditions in each experiment.

The basic results of Experiment 3 with regard to the impacts of $I_{de}$ vs. $I_{di}$ as the determinant of mean trial completion time are summarized in Figure 7 and Table 6. The results in this more comprehensive experiment verified the trends observed in the first two experiments. First, $I_{de}$ was once again shown to be a remarkably robust determinant of the mean pointing time within each condition. In spite of the very different instructions and hence the very different levels of overall target utilization, the $r^2$ values of $T$ vs. $I_{de}$ linear regression were all above 0.9. There was a trend of increasing $r^2$ value from the more risk-averse conditions to the more risky conditions, consistent with the first two experiments. The $r^2$ values of $T$ vs. $I_{de}$ linear regression in each condition were uniformly lower than their corresponding $r^2$ values of $T$ vs. $I_{di}$ in the same condition (Figure 7). Within the same operating condition, $I_{de}$ was clearly a stronger determinant of mean pointing time than $I_{di}$.
Second, in contrast to the strength of $I_{dn}$ within each condition, $I_{de}$ is a stronger determinant than $I_{dn}$ when data from all conditions were merged in one regression. $I_{de}$ accounted for 78% of the variance of mean trial completion time caused by both different levels of index of difficulty and the very different five operating strategies (See Figure 8). In comparison $I_{dn}$ could account for only 46%. In this sense $I_{de}$ clearly has the ability to convert some impact of the different levels of target utilization to index of difficulty.

Third, there was an overall rightward shift of $I_{de}$ values from their corresponding $I_{dn}$ values at the low (left) end of index of difficulty, but on the high end, the shift depended on the operating condition. For Condition
EF and F, the high end $I_{de}$ points moved towards left, for Condition EA and A the high end $I_{de}$ points actually further extended to the right, for condition N there was little change.

Fourth, and perhaps most strikingly, the regression lines of $T$ vs. $I_{de}$ were much closer across different conditions than those of $T$ vs. $I_{dn}$, particularly for the more risky (faster) conditions (N, F, EF) (See Figure 7). The regression coefficients $a$ and $b$ hence were more converging between the conditions with $I_{de}$ than with $I_{dn}$. Across the five conditions (from EA to FF), the $a$ values were between 139 ms and 209 ms for the $I_{dn}$ model and between 7.5 ms and 76 ms for the $I_{de}$ model. The $b$ values were between 83 ms/bit and 214 ms/bit (158% difference) for the $I_{dn}$ model and between 143 ms/bit and 203 ms/bit (42% difference) for the $I_{de}$ model. This also supports that $I_{de}$ could at least partially overcome the different levels of target utilization due to operating biases and produce more stable estimates of Fitts’ law coefficients.

![Figure 8. T vs. $I_{dn}$ and T vs. $I_{de}$ of all conditions combined](image)

5.3 The interplay of $I_u$, $I_{dn}$, and $W_0$

We now examine the distribution of $I_u$, which is related to the third point above on the range and location of $I_{de}$ shift. Figure 9 shows $I_u$ across different $I_{de}$ values in each of the five conditions in Experiment 3. As a result of different operating conditions, target utilization levels as measured by $I_u$ shifted up or down over a wide range (-1.33 to +2.04 bit) as a result of the instruction bias. Furthermore, $I_u$ also changed with $I_{dn}$. While overall $I_u$ was correlated with $I_{dn}$ (the higher the $I_{dn}$, the higher the level of over utilization the participants tended to make), the degree of such a dependency changed with the operating condition. The more risky (faster-paced) the overall strategy was, the stronger the dependency was. In Condition EF the dependency was the strongest, with $r^2 = 0.831$. In Condition EA, in contrast, participants were quite consistently risk-averse, keeping $I_u$ well below 0 across all $I_{dn}$ values ($r^2 = 0.14$).
As we can see in Figure 9, even within the same operating condition and at the same $I_{dn}$ level, $I_u$ could still be very different. This leads us to examine the influence on $I_u$ separately by $W$ and $D$ (Figure 10, see also Table 8). $W$ was shown to have much stronger influence on $I_u$ than $D$, as indicated by the slopes and $r^2$ values. This gives rise to an explanation of the fact that $I_{de}$ was a weaker determinant than $I_{dn}$ of completion time $T$ within each condition. Since $T$ and $I_{dn}$ form a very strong correlation within each condition, any adjustment of $I_u$ will only weaken the strength of the correlation, unless all $I_u$’s were changed with the same proportion. The $I_{de}$ adjustment, however, depends on $W_e$ (relative to $W$), which is influenced more by $W$ than by $D$. This means the Fitts’ law regression points at the same or similar $I_{dn}$, determined by $(D+W)/W$ ratio, may shift laterally to a different extent in $I_{de}$ adjustment. Figure 11 shows the relative shifts in one experimental condition (N) of the regression points when $I_{dn}$ (diamonds) is changed to $I_{de}$ (squares).

Figure 9. The index of target utilization $I_u$ changes with instruction and $I_{dn}$

Figure 10. The index of target utilization $I_u$ as a function of $W$ (left), $D$ (right), and instruction condition

Figure 11. $T$ vs. $I_{dn}$ points (diamonds) shift different amount to $T$ vs. $I_{de}$ points (squares), causing lower correlation

<table>
<thead>
<tr>
<th>Condition</th>
<th>$W$</th>
<th>$D$</th>
<th>12</th>
<th>36</th>
<th>72</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.58</td>
<td>4.59</td>
<td>8.57</td>
</tr>
<tr>
<td>EA</td>
<td>120</td>
<td>1.58</td>
<td>4.59</td>
<td>8.57</td>
<td></td>
</tr>
<tr>
<td></td>
<td>360</td>
<td>1.52</td>
<td>4.27</td>
<td>9.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>840</td>
<td>1.87</td>
<td>4.31</td>
<td>6.92</td>
<td></td>
</tr>
</tbody>
</table>
Table 8 gives the standard deviation of the endpoints in Experiment 3 under different $D$ and $W$ values. When accuracy was emphasized, standard deviation was mostly decided by $W$. When accuracy was less emphasized, $D$ began to exert impact on the standard deviation.

### 5.4 Preliminary conclusions of $I_{dn}$, $I_{de}$ and the two layers of speed accuracy tradeoff

These three experiments systematically examined speed-accuracy tradeoffs in Fitts’ pointing tasks. They showed that both the nominal task precision and performer’s bias in over or under utilizing the given task precision tolerate change pointing completion time. The nominal index of difficulty $I_{dn}$ is a remarkably robust predictor of completion time within each operating condition: the mean task completion time could be well accounted for by $I_{dn}$ within each operating condition, even though the conditions were at very different overall levels of target utilization as quantified by $I_u$. The problem with the $T = f(I_{dn})$ model, however, is that the regression coefficients ($a$ and $b$ in Fitts’ law) change with the overall level of $I_u$.

The experiments also showed that performers could rarely completely match the nominal task precision specification. $I_u$ is not only affected by the performers’ overall bias towards speed (over utilization of target) or accuracy (under utilization), it also interacts with $I_{de}$. Higher $I_{de}$ tended to cause over utilization. Even if the performers overall operated at the standard error rate (4%), as in Condition N in Experiment 3, $I_u$ still changed over 1 bit from low $I_{de}$ to high $I_{de}$ trials.

Overcoming some of the limitations of the $T = f(I_{dn})$ model, the $T = f(I_{de})$ model could compensate for some of the difference of $I_u$, particularly when the average $I_u$ was around 0 or negative. As a result of using $I_{de}$, the $a$ and $b$ estimates were much less affected by the operation conditions, and the $r^2$ value between $T$ and $I_{de}$ was much higher when both $I_{de}$ and $I_u$ varied widely (data mixed from all conditions) than the $r^2$ value between $T$ and $I_{dn}$ (cf. Figure 7).

However, the correction effect of the $I_{de}$ model is at the expense of a weakened $T$ vs. $I_u$ relationship. Within each operating bias condition, $T$ vs. $I_{de}$ consistently yielded lower $r^2$ than $T$ vs. $I_{dn}$. This was partly due to the fact that the amount of shift from $I_{dn}$ to $I_{de}$ was influenced more by $W$ than by $D$, so $I_{dn}$ to $I_{de}$ shifts were not always the same. Given the strong regularity of $T$ vs. $I_{dn}$, any uneven change from $I_{dn}$ would only result in a weaker regularity.

In sum, the behavioral model $T = f(I_{de})$ offers a compromise. It could absorb some of the mismatch between the nominal task specification and the performer’s actual pointing precision, at the expense of a strong $T$ vs.
The results of \( T vs. I_u \) regression in terms of coefficients \( a \) and \( b \) were more stable than those of \( T vs. I_{in} \) across operating biases, but they still did not completely converge. It appears that \( W_e \) as a \textit{posteriori} adjustment does not fully account for the time performance difference caused by the second and subjective layer of speed-accuracy tradeoff — the performance’s incompliance with the task specification (a none-zero or varying \( I_u \)) resulting in an overall faster or slower speed.

### 6. Experiment 4

To observe more directly the exact extent of \( I_u \) impact on time performance, we conducted yet another experiment that systematically controlled effective width \( W_e \) and nominal target width \( W \) to a similar amount in two experimental conditions. If \( W_e \) could compensate \( I_u \) variance, we would expect the \( T vs. W_e \) relationship in the presence of varying \( I_u \) to be identical or similar to the \( T vs. W \) relationship when the performers obediently complied with the target size specification \( W \) with no or little \( I_u \) variance.

#### 6.1 Set-up and experimental design

Ten volunteers, eight males and two females (averaging 24.2 years old), participated in this experiment. Some of them had participated in Experiment 1 or Experiment 2. The experiment was conducted on the same apparatus as in Experiments 1 and 3 with a similar experimental procedure. It consisted of two parts (or schemes): part A (“the target width incompliant scheme”) and part B (“the target width compliant scheme”). In both parts, the participants performed reciprocal target tapping with a fixed distance \( D \) of 400 pixels. In part A (“target incompetent”), \( W \) was fixed at 20 pixels. Participants performed under the five sets of strategy instructions as in Experiment 3 (EA, A, N, F, EF). Under each instruction set, they performed 14 trials. There was a total of 700 trials collected (= 5 instructions \( \times \) 14 trials \( \times \) 10 participants). No accidental trials were observed. The goal of part A was to produce a set of time measurements under the same nominal target width, but very different effective target width \( W_e \) due to different levels of target width utilization. Based on the experience of Experiment 3, \( I_u \) was about -1, -0.5, 0, 0.5 and 1 bits in EA, A, N, F and EF bias conditions, respectively, which map 20 pixels (\( W \)) to five different \( W_e \) values to approximately 10, 14, 20, 28 and 40 pixels.

In part B (“target compliant”), \( W \) was set at 10, 14, 20, 28 and 40 pixels, corresponding to the expected \( W_e \) values in part A. The goal of Part B was to produce a set of time measurements when participants obediently complied with the given target widths to an (almost) ideal extent: the \( |I_u| \) value should be less than 0.1 bit (i.e. \( W_e \) matches \( W \) within 7% margin). To achieve that, we used a target width enforcement method inspired by and refined from the verbal feedback method of Guiard and colleagues (Zhai et al., 2003). During the experiment (after the first 5 trials in each block), if the running \( I_u \) value was greater than 0.1 (i.e. \( W_e > 1.072W \)), which meant that the participant took too much risk, a sign appeared in the middle of the two target strips to remind the performer to slow down. In contrast, if \( I_u \) was less than -0.1 (i.e. \( W_e < 0.933W \)), a sign of different color appeared to remind the participant to speed up. If no sign was displayed, it meant the participant’s current endpoints dispersion corresponded to \( W \) within a 7% margin so the participant could keep his or her current pace.

The method of measuring the running \( I_u (W_e) \) value was as follows: Before the participant performed the 15th trial in a \( W \) condition, the program calculated the standard deviation of the endpoints distribution based on all of the past trials (from 1 to 14). From the 15th trial the program calculated the standard deviation of the endpoints, based on the most recent 14 trials (i.e. a 14-trial moving window was used). The experiment program stopped the current \( W \) condition and began the next one once a block of a 14 trials whose \( |I_u| \) value was less than 0.1 was captured (i.e. their \( W_e \) matched \( W \) by a less than 7% margin). These 14 trials were used in later analysis. The program would have also aborted the current \( W \) condition if the participant had performed 30 trials without reaching a 14 trial block that met the requirement. In the actual experiment, none of the participants needed to use up the maximum 30 trials. We analyzed the endpoints of the last 14 trials and confirmed that they were normally distributed. The standard deviation calculation was...
based on the distance of the endpoints to the center of their distribution, not the center of the target (Isokoski and Raisamo, 2002). The order of the $W$ condition was randomized in our experiment.

Therefore, through part A, we could observe the relationship between mean trial completion time and the effective target widths under different levels of target utilization as a result of the participants’ disrespect of the nominal target width to varying directions and extent. Through part B, we could observe the relationship between mean trial completion time and the nominal target width when the participants obediently and effectively complied with the accuracy tolerance specified by the target. Comparing the target width compliant scheme and the target width incompliant scheme we could directly examine whether and how well $W_e$ reconciles the two layers of speed-accuracy trade-off.

6.2 Results

Table 9 shows the results of Experiment 4. Figure 12 (left) shows the logarithmic regression results from the two experiment schemes (Part A and B). Figure 12 (right) show the Fitts’ law regression of Part A and Part B, using $I_{de}$ and $I_{dn}$ respectively. The two sets of scatter plots (and regression lines and curves) show that the relationship between completion time and $W_e$ in the target incompliant scheme (part A) and the relationship between completion time and $W$ in the target compliant scheme (part B) were consistent in direction, but different in extent. In the near-zero range of $I_u$ ($<0.5$ bits, or $W/2^{0.5} < W_e < W/2^{-0.5}$), the difference was relatively small. When $I_u$ was beyond such a range, this difference increased rapidly. The greater $|I_u|$ was, the greater such a difference was.
The results of this experiment shed more light on the effect that \( I_{dc} \) only partially compensates for the time variance caused by different operating biases (different target utilization levels): while the \( T \) vs. \( W_e \) relationship in the presence of \( I_u \) variance was similar to \( T \) vs. \( W \) relationship when the performers obediently complied with the target width specification (with no or little \( I_u \) variance), they did not exactly match in extent. The impact of the \( W_e \) adjustment lagged behind the impact of \( W \) changes in the same amount.

7. Further Analyses

The foregoing analyses clearly indicate that the use of \( W_e \) is an imperfect and insufficient adjustment of \( W \) to overcome \( I_u \) variances in Fitts’ law tasks, although the direction of adjustment was empirically correct. Is it possible to find an adjustment that more fully compensates the impact of \( I_u \) variance? Given that \( W_e \) under-corrects when the endpoints dispersion deviated significantly from the nominal width \( W \) (shown in Figure 12), a more exaggerated effective width could possibly account for more of the remaining difference in completion time. This suggests the following modified effective width is worth investigating:

\[
W_m = W \left( \frac{W_e}{W} \right)^{\alpha}
\]

(9)

i.e.

\[
W_m = W \left( \frac{4.133\sigma}{W} \right)^{\alpha}
\]

(10)

Where \( \sigma \) is the standard deviation of the endpoints distribution.

\( \alpha \) should be greater than 1 in order to cause \( W_m \) to nonlinearly exaggerate the impact of deviation from \( W \): When \( W_e \) is close to \( W \), \( W_m \) has a similar value to \( W_e \); When \( W_e >> W \) or \( W_e << W \), \( W_m \) is much greater or much smaller than \( W_e \). The greater the \( \alpha \) value is, the more pronounced difference there is between \( W_m \) and \( W_e \). When \( \alpha \) is 1, \( W_m \) reduces to \( W_e \).

Accordingly,

\[
I_{dm} = \log_2 \left( \frac{D + W_m}{W_m} \right)
\]

(11)

and

\[
T = a + bI_{dm}
\]

(12)

Based on the empirical data from Experiment 4, 1.5 is a good estimate of \( \alpha \). Figure 13 (left) shows the mean trial completion time as a function of \( W_m \) with \( \alpha = 1.5 \) (the target incompliant condition of Experiment 4). The relationship between time and \( W_m \) matches almost exactly the relationship between time and nominal \( W \) when participants closely obeyed the target width (the target compliant condition of Experiment 4). Figure 13 (right) shows the corresponding Fitts’ law regressions using \( I_{dc} \) and \( I_{dm} \).
Of course, the value of $\alpha$ based on one small experiment (Experiment 4) is not likely to be an accurate estimate for other data sets, although it is plausible that a better compensation may be achieved by $W_m$ with certain $\alpha$ value. A potential difficulty that $W_m$ faces, however, is the limitations of $W_e$ adjustment that we see in the first three experiments: while the $W_e$ adjustment reduced the discrepancy of Fitts’ law coefficients measured under different operating biases, it also weakened the strong regularity within each operating condition as modeled by $T = f(I_{dn})$. Since $W_m$ is a more exaggerated version of $W_e$, the correlation of $T$ vs. $I_{dm}$ may be reduced further from that of $T$ vs. $I_{dn}$ within each operating condition. Another obvious weakness of the notion of $W_m$ is that since it is based on $W_e$, $W_e$ and another parameter $\alpha$, its direct definition is lacking or to be discovered in the future.

To test the possibly stronger compensation power of $W_m$ and its likely drawbacks, we reanalyzed the first three experiments, substituting $I_{de}$ with $I_{dn}$ (with $\alpha = 1.5$). The results are summarized in Figures 14 to 16 and Table 10.

\[
\text{A: } y = 187.6x - 106.6 \\
R^2 = 0.869
\]

\[
\text{F: } y = 137.1x + 12.6 \\
R^2 = 0.890
\]

**Figure 15.** \( T \) vs. \( I_{dm} \) regression of Experiment 2 (Compare with Figure 6)

\[
\text{EF: } y = 142.5x + 64.5 \\
R^2 = 0.452
\]

\[
\text{EA: } y = 197.9x - 0.0769 \\
R^2 = 0.801
\]

\[
\text{A: } y = 180.3x + 41.6 \\
R^2 = 0.638
\]

\[
\text{N: } y = 154.0x + 41.2 \\
R^2 = 0.770
\]

\[
\text{F: } y = 153.1x + 15.6 \\
R^2 = 0.809
\]

**Figure 16.** \( T \) vs. \( I_{dm} \) regression of Experiment 3 (Compare with Figure 7)

<table>
<thead>
<tr>
<th>Expt</th>
<th>( I_d )</th>
<th>( I_{dn} )</th>
<th>( I_{de} )</th>
<th>( I_{dm} )</th>
<th>( I_{dn} )</th>
<th>( I_{de} )</th>
<th>( I_{dm} )</th>
<th>( \Delta b/b_{min} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expt 1</td>
<td>180.2, 154, 0.73</td>
<td>80.2, 154, 0.73</td>
<td>123, 118, 0.98</td>
<td>123, 92.4, 0.99</td>
<td>135, 120, 0.70</td>
<td>49%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expt 2</td>
<td>111, 173, 0.93</td>
<td>111, 173, 0.93</td>
<td>111, 119, 0.92</td>
<td>132, 161, 0.83</td>
<td>132, 161, 0.83</td>
<td>16%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expt 3</td>
<td>201, 215, 0.90</td>
<td>201, 215, 0.90</td>
<td>171, 131, 0.96</td>
<td>141, 113, 0.99</td>
<td>138, 83.8, 0.99</td>
<td>45%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Expt 1 | 123, 118, 0.98 | 123, 92.4, 0.99 | 135, 120, 0.70 | 49% |
| Expt 2 | 111, 119, 0.92 | 132, 161, 0.83 | 132, 161, 0.83 | 16% |
| Expt 3 | 171, 131, 0.96 | 141, 113, 0.99 | 138, 83.8, 0.99 | 45% |

| Expt 1 | 126.2, 105, 0.87 | 126.2, 105, 0.87 | 126.2, 105, 0.87 | 126.2, 105, 0.87 | 126.2, 105, 0.87 | 126.2, 105, 0.87 | 126.2, 105, 0.87 | 126.2, 105, 0.87 | 126.2, 105, 0.87 | 49% |
| Expt 2 | 111, 119, 0.92 | 111, 119, 0.92 | 111, 119, 0.92 | 111, 119, 0.92 | 111, 119, 0.92 | 111, 119, 0.92 | 111, 119, 0.92 | 111, 119, 0.92 | 111, 119, 0.92 | 49% |
| Expt 3 | 171, 131, 0.96 | 171, 131, 0.96 | 171, 131, 0.96 | 171, 131, 0.96 | 171, 131, 0.96 | 171, 131, 0.96 | 171, 131, 0.96 | 171, 131, 0.96 | 171, 131, 0.96 | 49% |
Table 10. Comparison of $I_d$, $I_{dn}$, and $I_{dm}$ in Experiment 1, 2 and 3; highlighted is the most neutral condition in each experiment (cf. Table 6).

Based on the degree of convergence of the regression coefficients (see Column 10 of Table 10), $I_{dm}$ compensates for the different levels of $I_d$ slightly better (Experiment 1, Experiment 2) or better (Experiment 3) than $I_{dn}$. Taking the $b$ value in Experiment 3 (the most comprehensive and balanced experiment) as an example, under the EA, A, N, F, EF operating conditions $b$ was 215, 186, 131, 113, 83 ms/bit with $I_{dn}$; and 204, 187, 149, 143, 147 ms/bit with $I_{dm}$. With $I_{dm}$, $b$ was 198, 180, 154, 153, 143 ms/bit. If we use the estimates under Condition N as the best possible estimate, the $b$ values can be written as percentage changes from its best estimate, as shown in Table 11. $I_{dm}$ appears to offer better correction of operating biases than $I_{dn}$ (See also Figure 16 in comparison to Figure 7).

<table>
<thead>
<tr>
<th></th>
<th>EA (%)</th>
<th>A (%)</th>
<th>N (base)</th>
<th>F (%)</th>
<th>EF(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_{dn}$</td>
<td>64.1%</td>
<td>42%</td>
<td>131</td>
<td>-13.7%</td>
<td>-36%</td>
</tr>
<tr>
<td>$I_{dm}$</td>
<td>36.9%</td>
<td>25.5%</td>
<td>149</td>
<td>-4%</td>
<td>-1.3%</td>
</tr>
<tr>
<td>$I_{dn}$</td>
<td>28.6%</td>
<td>16.9%</td>
<td>154</td>
<td>-0.6%</td>
<td>-7.1%</td>
</tr>
</tbody>
</table>

Table 11. $b$ estimate as percentage variation from the neutral condition (N) in Experiment 3

However, the limitations of $I_{dm}$ were also apparent. First, it still could not completely compensate for the speed accuracy tradeoff caused by different levels of target utilization. The coefficients measured under different operating biases still did not fully converge. Second, since $W_m$ is a nonlinear amplification (or reduction) of $W_e$, it weakened the regularity found in $T$ vs. $I_d$ even more than $W_e$, as indicated by the further decreased $r^2$ values within each condition (Table 10). When mixing all biased conditions, the $r^2$ of $T$ vs. $I_{dm}$ was similar to the $r^2$ of $T$ vs. $I_{dn}$, weaker in Experiment 1, but stronger in Experiments 2 and 3.

One could argue that stronger results with $I_{dm}$ could be achieved with different $\alpha$ values. As Table 12 shows, this was indeed true, but there was not a $\alpha$ value that is optimal for all experiments with different ranges and sets of operating biases.

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$ =1.2</th>
<th>$\alpha$ =1.3</th>
<th>$\alpha$ =1.5</th>
<th>$\alpha$ =1.8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>$r^2$</td>
<td>a</td>
</tr>
<tr>
<td>EXP. 1</td>
<td>13.9</td>
<td>162.1</td>
<td>0.808</td>
<td>19.0</td>
</tr>
<tr>
<td>EXP. 2</td>
<td>-44.6</td>
<td>168.4</td>
<td>0.835</td>
<td>-57.1</td>
</tr>
<tr>
<td>EXP. 3</td>
<td>-100.7</td>
<td>207.3</td>
<td>0.806</td>
<td>-98.2</td>
</tr>
</tbody>
</table>

Table 12. Regression results with mixed conditions at different $\alpha$ values

Mathematically, the $\alpha$ term in Equations 9 —12 can be separated from $I_{dn}$. Considering that $D$ is typically greater or much greater than $W$, and when $W_d/W$ is not too distant from 1, the following equation approximates Equation 12:

$$T = a + b \log_2 \left( \frac{D + W}{W} \right) - b \alpha \log_2 \left( \frac{W_e}{W} \right)$$

or

$$T = a + b I_d - b \alpha I_n$$
where \( c = -b \alpha \).

Multiple regression results \((a, b, \text{ and } c)\) based on the first three experiments, however, tended to be highly dependent on which experimental conditions were included in the regression, suggesting the complex interactive nature of the \( I_{dn} \) and \( I_u \) effect.

Figure 17 shows the mean of the completion time \( T \) as a function of \( I_{dn} \) and \( I_u \) based on data from Experiment 3 (all conditions mixed). As we can see, while \( T \) increased with \( I_{dn} \) and decreased with \( I_u \), in general, they did not form a strictly monotonic function, suggesting complex interaction effects between \( I_{dn} \) and \( I_u \). This means that it is difficult, if not impossible, to establish a model that captures both layers of speed-accuracy tradeoffs in a complete and yet simple (linear) manner.

8. General Discussion and Conclusions

We have systematically explored the two layers of speed-accuracy trade-off in Fitts’ aimed movement tasks. Fitts’ law in its original form reveals the speed-accuracy tradeoff relationship between pointing completion time and task precision based on the nominal, objective, geometrical parameters of the target \((D \text{ and } W)\). However, in actuality speed-accuracy relationship in pointing also contains another subjective layer which depends on how obediently the performer complies with the specified target width and what bias the performer takes (toward either accuracy or speed). This performer introduced accuracy layer causes a discrepancy between the nominal task precision and the actual behavior precision. We defined an index of target utilization, \( I_u = \log_2(4.133\sigma/W) \), to quantify the degree of this mismatch. As is implicitly realized in the Fitts’ law literature, and as this study has explicitly and systematically shown, Fitts’ law tasks tend to involve both layers of speed-accuracy tradeoffs. Our study shows \( I_u \) is never constant in an experiment, even within the same instruction set, such as “as accurately as possible and as fast as possible,” except when an enforcement method is applied, as in Experiment 4. The overall \( I_u \) level can be influenced by the experimental instruction in the laboratory, or by performers’ preference and task strategy in real world tasks.
This study clearly demonstrated that varied \( I_u \) values influence Fitts’ law regression modeling, resulting in different \( a \) \( b \) coefficients which in the context of human-computer interactions are often used to characterize an input system’s efficiency. The classic approach to correct the influence of varying \( I_u \) is the so-called “effective target width” method. This approach takes the performer’s actual behavioral parameter, \( W_e \), rather than the nominal target width \( W \), as the basis of index of difficulty calculation. Such an approach has not had a strong theoretical or empirical foundation in the literature.

In a set of four experiments, we deliberately manipulated \( I_u \) over a wide range or controlled to specific levels through instructions and feedback control, which enabled us to investigate the two layers of speed-accuracy tradeoff systematically. Our investigation has led to the following conclusions.

First, the task form of Fitts’ law \( T = f(I_u) \) is a very strong model. The nominal \( I_{dn} \) is an impressive determinant of mean movement time, accounting for up to 99% time variance within an experimental condition. Revisions of index difficulty to a behavior form, either through \( W_e \), or its more aggressive version \( W_{ms} \), consistently weaken the regularity within an particular operating bias condition. On the other hand the resulting coefficients of \( T vs. I_{dn} \) can be easily swung by \( I_u \) levels. In the context of HCI, this poses a serious challenge to assessing the quality of various input systems (Zhai, 2004).

Second, \( I_{de} \) partially incorporates the second, subjective accuracy layer into Fitts' law model by adopting an actual and behavior parameter \( W_e \). The first three experiments showed that adopting \( W_e \) reduced the discrepancy of \( a \) and \( b \) estimates between different experimental conditions. Experiment 4 shows that, although not completely congruent, the impact of the \textit{a posteriori} effective target width was similar to that of \textit{a priori} nominal target width with which the performer obediently complied. The compensation effect of \( W_e \) was also shown by the higher \( r^2 \) value of \( T vs. I_{de} \) regression across different operating biases (mixed data from all conditions) than the \( r^2 \) value of \( T vs. I_{dn} \) regression. This study, to our knowledge, provided the first systematic empirical foundation for the use of \( W_e \). However, the compensation effect of \( W_e \) is gained at the cost of weakened regularity within each experimental condition.

Third, \( I_{dn} \), the more aggressive version of \( I_{de} \), takes a step further than \( I_{de} \) in both \( I_{de} \!’s \) pros and cons: it fully compensates for the \( I_u \) impact but further weakens regularity within each operating strategy condition.

Fourth, in the absence of the subjective layer of speed-accuracy tradeoff (the performer completely complies with task specification, keeping \( I_u \) at or near zero), both \( I_{de} \) and \( I_{dn} \) revert to the nominal \( I_{dn} \). In that sense, no harm can be done by adopting \( I_{de} \) and \( I_{dn} \).

Fifth, the level of target utilization in Fitts’ law tasks, as measured by the index of target utilization \( I_u = log(W_e/W) \), can be influenced by three factors: the overall operating bias (in the lab by instruction, in reality by user’s strategic choice); nominal target \( W \); and target distance \( D \). However, \( W \) and \( D \) have very different degrees of impact on \( I_u \) with the former being far greater than the latter. This means that the change from \( I_{dn} \) points (determined by \( D \) and \( W \)) to \( I_{de} \) points (determined by \( D \) and \( W_e \)) is not uniform.

Given that \( T \) and \( I_{dn} \) forms a very strong correlation within each strategy, the use of \( I_{de} \) or \( I_{dn} \) would only weaken this correlation. A compromise has to be made between the goodness of fit within each condition and better estimates of \( a \) and \( b \). Fundamentally, the use of \( I_{de} \) or \( I_{dn} \) is to incorporate a different layer of speed-accurate tradeoff (the subjective layer) into a very strong time and task accuracy relationship as modeled by \( T = f(I_u) \).

Sixth and finally, the two layers of speed accuracy tradeoff in pointing interact in a complex manner. They do not cause a simple additive effect, hence the difficulty of using \( T = f(I_{de}) \) or \( T = f(I_{dn}) \) and any other potential relationship as a strong general model.

In summary, this systematic investigation reveals the nature and the relationship of speed and accuracy in pointing. The implication of this study depends on the specific purpose of use. Theoretically we now know that the two layers of speed-accuracy tradeoffs, the objective and task layer and the subjective and behavior...
The impact of actual pointing precision $W_a$ and the impact of nominal pointing precision $W$ are not equivalent, but are numerically similar, particularly when $W_a$ is not too distant from $W$. The findings in this work also have more general implications to human motor control theory. One of the central issues in early motor control literature was the closed-loop vs. open-loop nature of human motor skills (Stelmach, 1976). This study shows that even for a low level tapping task, both external visual feedback and internal bias settings contribute to the control process. In a manual stabilization task, Pew (1966) showed that humans not only react instantaneously to the position of a controlled object, but can also adjust higher level control parameters to achieve stabilization. This experiment shows that in pointing tasks performers could adjust their overall bias towards speed or accuracy and integrate such an internal high level setting with the external low level visual feedback to manage a pointing process.

Practically, the findings in this study suggest that in order to accurately measure Fitts’ law parameters, $I_u$ should be kept as close to zero as possible and its variance should be kept as low as possible. One possible method of controlling $I_u$ in Fitts’ law studies is to use endpoints standard deviation-based feedback, as we did in Experiment 4. When $I_u$ is highly varied or when it is not near zero, however, this study provides an empirical foundation for the application of $W_a$ or its more aggressive and more complete version, $W_{ma}$, to adjust for $I_u$ changes. These adjustments consistently yield more logical Fitts’ law parameter estimates, $a$ and $b$, although one should also be aware of the limitations and side effects of $W_a$ or $W_{ma}$, including reduced correlation between pointing time and index of difficulty within each operating strategy and their incomplete compensation for the subjective layer of speed-accuracy tradeoff.

**Acknowledgements**

The work was partially supported by High-Tech Research Center Development Program (No.7512002002) in Japan. We thank Masafumi Ogasawara of Kochi University of Technology, Huaiyu Zhu and Barton Smith of the IBM Almaden Research Center for their assistance, support, and discussions of this work. We also thank the anonymous reviewers and in particular Yves Guiard for their extensive critique of an early version of this paper. We adopted many of their suggestions in revising this work. Their contributions are gratefully acknowledged.

**References**


