A New Test of Throughput Invariance in Fitts' Law: Role of the Intercept and of Jensen's Inequality

Halla H. Olafsdottir, Yves Guiard, Olivier Rioul, Simon T. Perrault

▶ **To cite this version:**

Halla H. Olafsdottir, Yves Guiard, Olivier Rioul, Simon T. Perrault. A New Test of Throughput Invariance in Fitts' Law: Role of the Intercept and of Jensen's Inequality. Interfaces, the quaterly magazine of BCS interaction group, 2012, 93, 8 p. hal-00710498v2

HAL Id: hal-00710498
https://hal.science/hal-00710498v2

Submitted on 19 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A New Test of Throughput Invariance in Fitts' Law: Role of the Intercept and of Jensen's Inequality

Halla B. Olafsdottir
Institut Mines-Telecom
Telecom ParisTech
CNRS LTCI
halla@telecom-paristech.fr

Yves Guiard
CNRS LTCI
Institut Mines-Telecom
Telecom ParisTech
yves.guiard@telecom-paristech.fr

Olivier Rioul
Institut Mines-Telecom
Telecom ParisTech
CNRS LTCI
olivier.rioul@telecom-paristech.fr

Simon T. Perrault
Institut Mines-Telecom
Telecom ParisTech
CNRS LTCI
simon.perrault@telecom-paristech.fr

Fitts' law states that movement time varies linearly with the index of difficulty or, equivalently, that throughput (*TP*) is conserved across variations of the speed/accuracy strategy. Replicating a recent study by MacKenzie and Isokoski (2008), we tested the throughput invariance hypothesis with some fresh data and found the *TP* to be systematically affected by the strategy. This result, we suggest, pleads against the currently popular definition of the *TP* inherited from Fitts (1954), namely *TP* = *ID/MT*, which we recall is incompatible with the Shannon equation of Fitts' law. We also show that the statistical elaboration of the *TP* suffers from a problematic amount of uncontrolled variability due to the multiple inadvertent impact of Jensen's inequality.

Keywords: Throughput, Fitts' law, y-intercept, Jensen's inequality, statistics, aggregation order

1. INTRODUCTION

Humans have innumerable opportunities in everyday life to move their hand to some target location, for example to reach a light switch on a wall or to grasp some nearby object. In the specific context of human-computer interaction (HCI), the ubiquitous graphical user interface requires the people to express almost all their decisions by reaching and clicking target objects like icons, menu items or hypertext links. In all these cases users face a speed/accuracy dilemma—as everyone knows, the faster the reaching movement, the more likely the miss. This speed/accuracy trade-off is what Fitts' law is all about.

In the present paper, concerned with both the mathematical consistency and the empirical validity of Fitts' law modeling, we focus on a seldom-considered version of the law that takes the form of an invariance: if the equation is correct, a certain quantity, called the throughput, should be conserved across variations of the speed/accuracy balance. We will discuss, in light of some data, two difficulties that have hindered progress in the understanding of this conservation so far. One has to do with the controversial role of the equation's intercept and the other with the inadvertent influence of the order in which one computes the throughput and aggregates the data statistically.

2. FITTS' LAW

Fitts' law is a well known empirical regularity which predicts movement time *MT* as a function of target width *W* and target distance *D* [2,3]. HCI researchers generally use the Shannon equation [7,18]:

$$MT = a + b \cdot \log_2(D/W + 1) \qquad (1)$$

where *a* and *b* stand for adjustable constants and where the log term represents the task's index of difficulty (*ID*).

In fact there are many candidate mathematical models for Fitts' law (see Plamondon & Alimi [13], who list a dozen respectable equations), and not all models take the logarithmic form. In a famous contribution to the literature, Meyer et al. [12] have proposed to model *MT* as a power function of the ratio *D/W*, arguing that such a model encompasses the logarithmic model as a limiting case. This argument, however, has been recently challenged by Rioul and Guiard [16,17], who showed that mathematically Meyer et al.'s model is a quasi-logarithmic, not a genuine power model. Not only is Equation 1 of the logarithmic category, not only is it known to tightly fit most data sets, it is also of

© The Authors. Published by the British Informatics Society Ltd.

special importance in practice, being actually part of an ISO standard [6].

In this paper we re-examine the calculation of throughput (*TP*) from Equation 1 and draw attention to a previously unnoticed methodological difficulty that may hinder the empirical evaluation of the model.

3. LAW OF VARIATION VS. INVARIANCE

The Shannon model of Fitts' law is usually written as in Equation 1, which states a *law of variation* of the form $y = a + bx$. That is, y varies lawfully (linearly) with x. But the model may just as well be formulated as an *invariance*, as either

$$\frac{y-a}{x} = b \tag{2}$$

or

$$y - bx = a, \tag{3}$$

emphasizing that two quantities, a and b, are invariant across the variations of x.

It is noteworthy that, despite their mathematical equivalence, the law-of-variation formulation of Equation 1 and the invariance formulations of Equations 2-3 place the model in markedly different positions with regard to the risk of empirical falsification [14,15]. The Shannon equation (Equation 1) is in fact quite unlikely to be disproved by empirical data: at worst, one will obtain a disappointing fit, wondering whether one should continue to trust the model with an r^2 below .9, .8, or lower. But take the claim that $(y\text{-}a)/x$ must be independent of x, which has the form of a null hypothesis (H_0): if the data plead for the rejection of H_0, then one faces an empirical falsification of Equation 3—and, by implication, of Equation 1. Consistent with classic Popperian epistemology [14,15], this more challenging way of empirically testing the theory is commonplace in stronger domains of science like physics [11].

4. THE THROUGHPUT

The throughput (*TP*) of Fitts' law tasks is a standard of measurement widely used in the HCI community as a tool to quantify user performance with different input devices and different interaction techniques.

In classical Fitts' law experimentation participants are instructed to perform their movements as fast as possible given the *ID*, with a certain (ideally constant) level of accuracy. Although many factors such as mood, fatigue and alertness, may influence the *TP*, the Shannon model of Fitts' law says that the *TP* should be conserved within participant across variations of task difficulty (nominal *ID*) and movement accuracy (effective *ID*). Because the *TP* is a global index of performance which takes both speed and accuracy into account, its practical utility in the context of HCI research is very high.

MacKenzie and Isokoski [8] recently tested the robustness of the *TP* under three different instructional conditions: standard, speed emphasis, and accuracy emphasis. While, unsurprisingly, speed and accuracy of performance were both strongly affected by the change of instructions, the key outcome was the authors' failure to detect a significant effect of the instructional manipulation on the *TP*. MacKenzie and Isokoski argued that this result is evidence for the Shannon model of Fitts' law [7,18].

Our purpose below is two-fold. First we reanalyze the data of a recently published study [5] to test the null hypothesis of *TP* invariance across speed/accuracy variations. It occurred to us that because MacKenzie and Isokoski [8] varied instructions within a limited range, their test of the Shannon model of Fitts' law was somewhat lenient. Obviously, the issue being the demonstration that a certain experimental manipulation exerts no effect on a certain dependent measure, the larger the extent of the manipulation, the more persuasive the demonstration.

Thus, while our analysis below reproduces MacKenzie and Isokoski's lenient test on some fresh data, we will also report the results of a much tougher test in which the speed/accuracy strategy of our participants was made to vary over its whole spectrum, from maximum speed to maximum accuracy.

Our second purpose is to draw attention to a methodological difficulty that many authors may have incidentally noticed, without paying much attention to it, but that the results of present study forced us to consider seriously. The difficulty arises from the fact that the *order* in which one does the various operations required for the calculation of the *TP* affects the outcome to an appreciable extent. We will show that the problem is due to a mathematical result known as Jensen's inequality.

Twenty years have passed since MacKenzie [7] first proposed to replace Fitts' [2,3] original equation $MT = a + b\cdot\log_2(2D/W)$ with Equation 1. MacKenzie has convinced the HCI community that the Shannon formula is theoretically valid and empirically predictive, but there is still no agreement on the exact definition of the *TP*.

While Zhai [20] identified three candidate definitions in the literature, the basic dispute boils down to a simple mathematical dichotomy. What is not agreed upon is whether in the *TP* calculation

one should take into account the intercept *a* of the Shannon equation (Equation 1) and thus calculate the *TP* as

$$TP_{\mathrm{Z}} = \frac{1}{b} = \frac{ID}{MT - a} \tag{4}$$

or one should ignore the intercept and, in keeping with Fitts' initial suggestion [2], calculate the *TP* as

$$TP_{\mathrm{M}} = \frac{ID}{MT} \tag{5}$$

Equation 4 is a straightforward derivation of Equation 1. It gives the definition of the *TP* that Card et al. [1] used in their well-known pioneering study of Fitts' law in the context of HCI. More recently that definition was forcefully advocated by Zhai [20], hence the subscript Z.

As emphasized by Zhai, Equation 5 is inconsistent with Equation 1, whose intercept *a* it leaves aside. Nevertheless this definition of the *TP* has been inflexibly advocated by MacKenzie (hence the subscript M), based on the argument that this intercept should be zero [18]. In a recent study, Guiard and Olafsdottir [4] have argued that such an assumption regarding the value of Fitts' law intercept cannot be made because the *ID* runs on a non-ratio scale of measurement (*i.e.*, an equal-interval scale with no physical zero), meaning that the value of the intercept is arbitrary and uninterpretable. But the fact is, the TP_{M} has never ceased to be popular among HCI researchers and its credibility is now further strengthened by an ISO standard [6].

Recently Wobbrock et al. [19] warned against comparisons across the two categories of *TP*, which necessarily produce more or less discrepant estimates. But unfortunately there is room—even within one and the same approach, and to make this point below we will stick to MacKenzie's—for quite another sort of discrepancy in the calculation of the *TP*. Consider Equations 6 and 7, two concrete statistical implementations of the mathematical formula of Equation 5:

$$\overline{TP}_{\mathrm{M,CtA}} = \frac{1}{N}\sum_{i=1}^{N} TP_i \tag{6}$$

$$\overline{TP}_{\mathrm{M,AtC}} = \frac{\overline{ID}}{\overline{MT}} = \frac{\log_2\left(1 + \frac{\overline{A}}{4.133\cdot\sigma_A}\right)}{\overline{MT}} \tag{7}$$

The only difference lies in the order in which one performs the averaging and the computing: in Equation 6 one first computes a number of *TP* values and then averages them (the CtA order) while in Equation 7 one first averages the *ID*s and the *MT*s and then computes one value of *TP* (the AtC order). Both equations seem to be mathematically and statistically sound and researchers who have utilized both versions may have considered them equivalent. The *TP* description offered by the International ISO 9241-9 standard [6] hesitates between them. In our view there is reason to be concerned by this irresolution.

5. FIRST AGGREGATE THEN COMPUTE OR THE REVERSE ORDER: A JENSEN'S INEQUALITY ISSUE

To reiterate, the calculation of *TP* involves two sorts of operations. One is *averaging*, a statistical operation that compresses a set of numbers into a single summary value, typically a mean. The other is *computing* (e.g., calculating the quotient of a fraction), an arithmetic operation that also often combines several numbers into a single result. Unfortunately, the final *TP* value depends on the order in which the averaging and the computing are done, as shown in Figure 1 with a very simple numerical example.

Computing →

Averaging ↓

	ID	*MT*	TP_{Z} *ID*/(*MT*-*a*)	TP_{M} *ID*/*MT*
	1	0.225	8.0	4.444
	2	0.350	8.0	5.714
	3	0.475	8.0	6.316
	4	0.600	8.0	6.667
	5	0.725	8.0	6.897
	6	0.850	8.0	7.059
	7	0.975	8.0	7.179
	8	1.100	8.0	7.273
	9	1.225	8.0	7.347
	10	1.350	8.0	7.407
Mean	5.50	0.788	8.0	**6.630**
		6.984		$TP_{\mathrm{M,CtA}}$
		$TP_{\mathrm{M,AtC}}$		

Figure 1. *Two ways of calculating the* TP_{M} under the assumption that Fitts' law follows the Shannon formula, with *a* = 0.1s and *b* = 0.125s/bit. The third column TP_{Z} = *ID*/(*MT* – *a*) is constant by hypothesis.

Figure 1 displays a hypothetical set of ten *MT* values computed from Equation 1 whose coefficients have been set to arbitrary but plausible [20] values, *a* = 0.1s and *b* = 0.125s/bit. The figure may help to see that there are two ways to obtain a global value of TP_{M} from ten pairs of *ID* and *MT* values. One option is to start by computing TP_{M} in each row and to then average the ten values of TP_{M} at the bottom of the rightmost column—this is what we call the *Compute-then-Average* (CtA)

option. With the hypothetical data set of Figure 1 one obtains TP_M = 6.63bits/s.

The alternative option is to start by averaging the 10 *ID*s and the 10 *MT*s downward and to then compute the TP_M just once from the mean *ID* and the mean *MT*—this is what we call the *Average-then-Compute* (AtC) option. This option with the data of Figure 1 yields TP_M = 6.98bits/s, which is more (+5.3%) than 6.63 bits/s.

The problem one is encountering here is ***Jensen's inequality***, which states that for any convex function[1]

$$f(\overline{x}) \leq \overline{f(x)}, \tag{8}$$

while the opposite holds true if the function is concave.

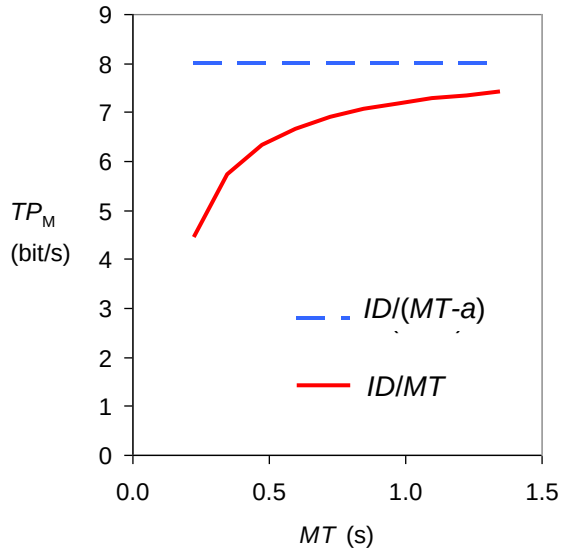

Figure 2. *The function $TP_M = f(MT)$ under the Shannon model.*

Figure 2, which plots the data of Figure 1, shows that the function TP_M / TP_Z = f (*MT*), or f (*MT*) = 1 – *a*/*MT*, is concave, thus

$$\overline{TP}_{M,AtC} \geq \overline{TP}_{M,CtA} \tag{9}$$

Let us return to real-world formulas like those of Equations 6 and 7. The impact of Jensen's inequality is complex and rather hard to guess for two reasons.

First, the TP_M formula involves not one, but five calculation steps. Since computing the TP_M involves a function of the form

$$\frac{\log\left(1+\frac{x}{\sqrt{y}}\right)}{t} \tag{10}$$

the computation of the TP_M requires five computation steps, represented from left to right in Figure 3.

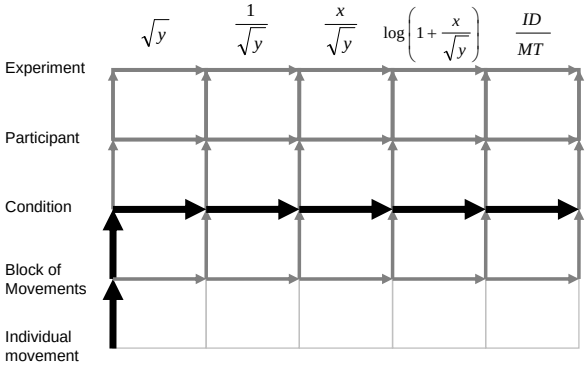

Figure 3. *The statistical-aggregation and calculation steps involved in the computation of a TP_M. One example path is highlighted, which delivers the TP_M value of one particular participant for one particular condition.*

Second, as illustrated also in Figure 3, a data set normally involves more than two levels of statistical aggregation. In fact the calculation of a TP_M may require up to four aggregation steps in a typical Fitts' law experiment (allowing for averaging over participants). Starting from the individual measure of *MT*—the atoms, so to speak—four successive aggregation steps, represented in the figure from bottom to top, can take place in a typical Fitts' law experiment:

a. the *movement-block* averages, each of which summarizes a number of individual measures;
b. the *condition* averages, each of which summarizes performance over a number of trial blocks;
c. the *participant* averages, each of which summarizes performance over a number of conditions; and
d. the *experiment* averages, each of which summarizes performance over a number of participants.

The important fact is that, contrary to the feeling that may arise from the simple comparison of Equations 6 and 7, there are many more than two ways of arranging the various computation and aggregation steps involved in the estimation of TP_M. The number of paths we are looking for is the number of possible ways of inserting 3 objects in 6 possible places

[1] A function is said to be convex (concave) if its graph lies above (resp. below) any tangent line.

$$\binom{6}{3} = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} = 20 \quad (11)$$

Thus, with five computation steps and four aggregation levels, there exist 20 possible paths, which all deliver different TP_M values.[2] Thus the Jensen inequality has many opportunities to operate, leading to a troublesome amount of uncontrolled variability in data processing.

The concrete example of the next section will show that this bias may be quite damaging. Depending on the CtA vs. AtC order, we found that our lenient test either succeeded or failed to replicate MacKenzie and Isokoski's result.

6. A RERUN OF MACKENZIE AND ISOKOSKI'S TEST

This section reports the results of a fresh test of the *TP*-invariance hypothesis based on a re-analysis of recently published data of ours [5]. Our test differs from MacKenzie and Isokoski's [8]—and by the same token from most standard Fitts' law tests—in three noteworthy respects.

First, we used discrete rather than reciprocal movements to obtain more reliable estimates of *MT*. As noticed by Fitts and Peterson [3], the discrete protocol allows more rigorous control over the variables of interest than is possible with the reciprocal protocol. In the reciprocal protocol movement time is the time it takes to carry out a movement *and* to evaluate the error inherited from the previous movement *and* to prepare the next movement. The discrete protocol, in contrast, measures the duration of a pure movement-execution process.

Second, the target was displayed as a one-pixel line, rather than as a band of width *W*. This feature does not mean that the experiment used a zero-width target, but rather that *W* was left unspecified, the one-pixel target serving to just indicate to participants what the amplitude of their movements should be on average. Accordingly, in our calculations the *ID* was computed from the ratio of mean movement amplitude (in fact always virtually equal to target distance *D*) to the standard deviation of the amplitude (rather than target width *W*). While the usual methodology uses *D* and *W* with a post-hoc adjustment for error because *W* provides a notoriously poor control over the actual spread of movement endpoints [18], our strategy is to forget once and for all about any tolerance specification and to simply consider actual spreads of movement endpoints.

Third and most importantly, our manipulation covered the complete range of speed/accuracy strategies, allowing a tougher and hence more informative empirical test of the Shannon model of Fitts' law.

6.1. Method [3]

Sixteen participants were presented with five sets of instructions, which formed an ordinal independent variable:

- max speed
- speed emphasis
- speed/accuracy balance
- accuracy emphasis
- max accuracy.

In the max-speed condition the only accuracy requirement was to terminate the movements on average in the vicinity of the target. At the opposite end of the instructions continuum, in the max-accuracy condition participants were to bring the cursor exactly to the target (zero pixel error), the only time constraint being to not waste any time. The three central levels of instructions, one unbiased (speed/accuracy balance) and two biased (speed emphasis and accuracy emphasis) were similar to those of MacKenzie and Isokoski.

The experiment used a computer screen and a Wacom™ tablet set to the absolute mode with a one-to-one mapping. The screen displayed two fixed vertical lines, 150 mm apart, indicating movement start and movement target, and a movable crosshair whose horizontal motion was controlled by the Wacom™ stylus. An L-shaped ruler was attached to the tablet to guide the stylus movement along the horizontal dimension, the shorter (vertical) leg of the L being aligned with the screen's start line, thus eliminating start point variability.

Each of the 16 participants ran five 15-movement blocks in each of the five instructional condition (25 blocks overall). In sum this experiment involved 15 movements x 25 blocks x 16 participants = 6,000 movements.

6.2. Data Analysis

We ran two within-participant one-way ANOVAs on TP_M. In one of them, aimed to replicate the lenient test of MacKenzie and Isokoski, the instructions factor was restricted to its three central levels, namely speed emphasis, speed/accuracy balance,

[2] This calculation takes account of the fact that the very first operation can only consist of an aggregation, because the standard deviation of movement amplitude is undefined below the level of the block of movements.

[3] For a detailed description of the method, see [5].

and accuracy emphasis. The other ANOVA considered all five levels, providing a much tougher test of the *TP* invariance hypothesis.

Individual-movement measures were *MT* (s) and amplitude (mm). For each or the 25 blocks we computed the three ingredients needed to calculate any *TP*, namely, median *MT* and the mean and standard deviation of amplitude.[4] We then computed the condition-level estimates of TP_M using both the CtA and the AtC order, ending up with two candidate dependent variables for the ANOVA test, $TP_{M,\ CtA}$ and $TP_{M,\ AtC}$. The figures below show averages computed over all 16 participants.

6.3. Results and Discussion

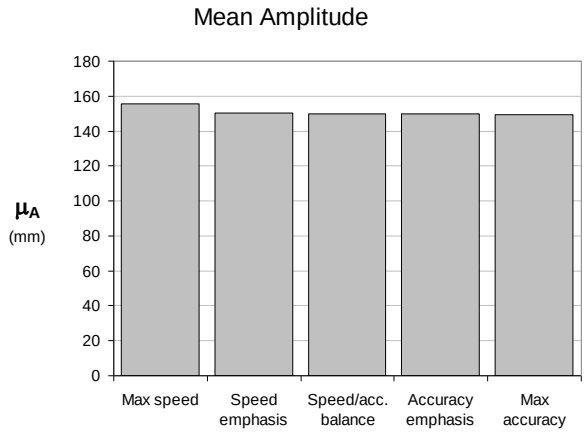

***Figure 4**. Mean amplitude vs. instructional condition.*

As shown in Figure 4, mean movement amplitude (μ_A) was very nearly a constant 150mm, as required. The participants were able to produce essentially unbiased aiming movements, the only exception being a 5.5mm overshoot error in the max-speed condition; although a statistically significant effect (t_{15}=4.50, p<.001) this is a remarkably small bias of +3.7%, which we shall not discuss here.

Rather than movement amplitude, what our instructional manipulation did influence were, unsurprisingly, the speed and accuracy of performance, two very strong effects just as they were in the MacKenzie and Isokoski [8] study. Figure 5 shows the gradual increase of median movement time (μ_T) from the max-speed condition (about 200ms) to the max-accuracy condition (more than a second), a considerable five-fold increase. Obviously a monotonic lengthening of movement time while movement amplitude remains a constant means a monotonic drop of average movement speed, an effect illustrated explicitly in Figure 6.

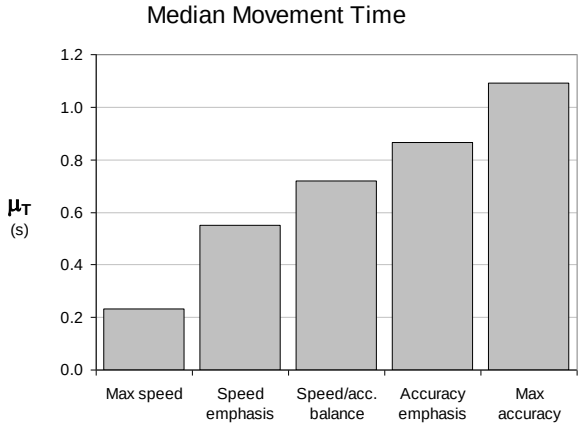

***Figure 5**. Median movement time as a function of instructional condition.*

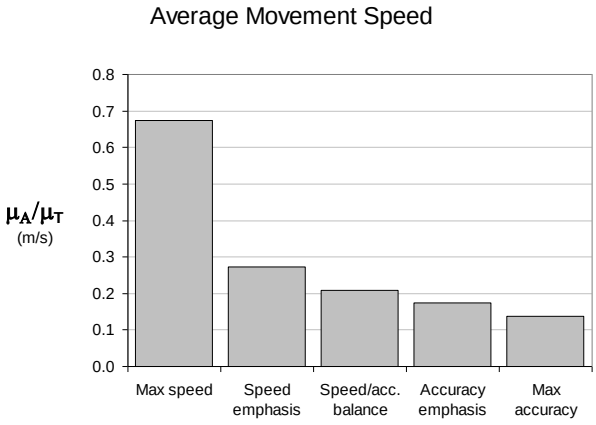

***Figure 6**. Average movement speed vs. instructional condition.*

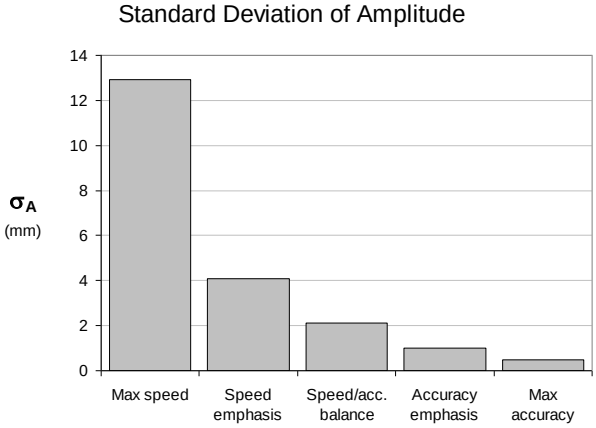

***Figure 7**. Endpoint spread vs. instructional condition.*

The other side of the evidence that our instructions were instrumental in modulating the participants'

[4] The distributions of movement time showing some positive skewness, we used the median, rather than the mean, for that dependent measure.

strategy is visible in Figure 7, which shows how the spread of movement endpoint, measured as the standard deviation of amplitude σ_A, declined gradually from the max-speed condition (with a standard deviation of amplitude σ_A of 13mm, or 8% of the mean) to the max-accuracy condition (0.5mm, or 0.3% of mean amplitude).

The crucial result of this experiment is shown in Figure 8, which plots $TP_{M,\,CtA}$ and $TP_{M,\,AtC}$, the two variants of the ISO estimate of *TP*, against the instructional factor.

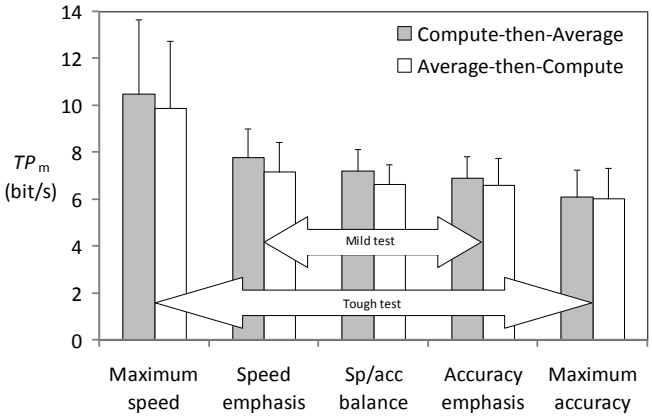

***Figure 8**. TP_M vs. instructional condition. Error bars show 95% confidence intervals based on between-participant standard deviations.*

Recall that according to MacKenzie and colleagues [7,8,18] the TP_M should *not* vary across variations of the speed/accuracy strategy. Tested over our complete set of instructions, the TP_M invariance hypothesis markedly failed. Whether computed with the CtA or the AtC order, the TP_M declined monotonically, from 10.5bits/s down to 6.1bits/s, as the instructions were shifted from the max-speed to the max-accuracy condition. This is a substantial effect, a 42% reduction of *TP*, and it is highly significant statistically (Table 1).

Turning to the lenient 3-level test, the outcome turned out to be equivocal. With the AtC order our lenient test replicated MacKenzie and Isokoski's non-rejection of H_0 ($p>.05$). With the CtA order, however, it did not ($p<.02$). This irresolution is a troublesome complication induced, we believe, by Jensen's inequality.

***Table 1**. Results of the tough and lenient ANOVAs conducted on the CtA and the AtC estimate of TP_M.*

	Tough 5 level			**Lenient 3 level**		
	F	***df***	***p***	***F***	***df***	***p***
CtA	**23.54**	**4**	**<.001**	**5.03**	**2**	**.013**
AtC	**19.69**	**4**	**<.001**	**1.86**	**2**	**.173**

In our view it is not necessarily the Shannon model of Equation 1 that should be questioned in light of the present data, but rather Equation 5, which implies the assumption that the intercept of Fitts' law is zero [20]—actually an untenable assumption given the non-ratio level of measurement on the continuum of μ_A/σ_A or *D/W* [4].

Would the test have been successful if the TP_Z had been used instead? We found with a simulation on our data set that the effect of instructions on *TP* would have been small and marginally significant, had the TP_Z been used instead of the TP_M. This result is doubtful, however, because a test of the invariance of $TP_Z = 1/b$ across variations of the speed/accuracy strategy requires the other coefficient, the intercept *a*, to be used in the calculation. This requiring that the Shannon equation be calculated beforehand, the test begs the question. Another sort of experimental test is needed to evaluate the invariance of the TP_Z.

7. CONCLUSION AND PERSPECTIVES

Routine TP_M measurement is an established norm of HCI, further strengthened since 2000 by an official ISO standard [6]. There is no question that standardization, which facilitates comparisons, is useful [18]. Twenty years of consensus about the Shannon model of Fitts' law have certainly been an asset for input research in HCI. However, failure to acknowledge Zhai's [20] demonstration that the standard method of measuring the *TP* (Equation 5) is inconsistent with the Shannon model (Equation 1) has been a handicap. Our data, which show that the TP_M not simply fails a tough invariance test but hardly passes a rather lenient test, support Zhai's [20] suggestion that researchers should return to the mathematically correct definition of the *TP* shown in Equation 4.

Another, no less important lesson to be learned from this study is that serious methodological work is needed to try to master the hidden variability that arises inadvertently in Fitts' law data due to Jensen's inequality. To our knowledge the impact of this methodological difficulty on data processing has not been yet correctly understood and we believe this general problem is worth a systematic investigation. There is reason to believe that this is a general methodological problem, with a scope extending far beyond the study of Fitts' law. Whether the solution rests on some mathematical or statistical principles or perhaps on some arbitrary conventions is an open question which we are currently investigating.

8. REFERENCES

1. Card, SK, English, WK, & Burr, BJ, (1978) Evaluation of mouse, rate-controlled joystick, step keys and text keys for text Selection on a CRT. *Ergonomics*, 21, 601 613.
2. Fitts, PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47, 381-391. Reprint: *Journal of Experimental Psychology: General*, 1992, 121, 262-269.
3. Fitts, PM, & Peterson, JR (1964) Information Capacity of Discrete Motor Responses. *Journal of Experimental Psychology*, 67,103-112.
4. Guiard Y., & Olafsdottir, HB (2011). On the measurement of movement difficulty in the standard approach to Fitts' law. *PLoS ONE* 6(10): e24389.
5. Guiard, Y, Olafsdottir, HB, & Perrault, ST (2011) Fitts' law as an explicit time/error trade-off. *Proc. CHI 2011.* New York: ACM, 1609-1628.
6. 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.
7. MacKenzie, IS (1992) Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7, 91-139
8. MacKenzie, IS, & Isokoski, P (2008) Fitts' throughput and the speed-accuracy tradeoff. Proceedings of *CHI'2008*. New York: Sheridan, 1633−1636.
9. MacKenzie, IS, Jusoh, S (2001) An evaluation of two input devices for remote pointing. Proceedings of the *Eighth IFIP Working Conference on Engineering for Human–Computer Interaction—EHCI 2001*. Springer, Heidelberg, Germany, 235–249.
10. MacKenzie, IS, Soukoreff, RW (2003) Card, English, & Burr (1978)—25 years later. Extended Abstracts of the ACM Conference on Human Factors in Computing Systems, *CHI 2003*, 760–761.
11. Meehl, PE (1967) Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34, 103-115.
12. Meyer, DE, Smith, JEK, Kornblum, S, Abrams, RA, & Wright, CE (1990) Speed-accuracy trade-offs in aimed movements: Toward a theory of rapid voluntary action. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 173-226). Hillsdale, NJ: Erlbaum.
13. Plamondon, R, & Alimi, AM (1997) Speed/accuracy trade-offs in target-directed movements. *The Behavioral and Brain Sciences*, *20*, 279–349.
14. Platt, JR (1964) Strong inference. *Science,* 146, N° 3642, 347-353.
15. Popper, KR (1959) *The Logic of Scientific Discovery*. New York: Basic Books.
16. Rioul, O & Guiard, Y (2011) "The power model of Fitts' law does not encompass the logarithmic model." *2011 Meeting of the European Mathematical Psychology Group* (EMPG 2011) Paris, France, August 29-31, 2011.
17. Rioul, O & Guiard, Y (in press) Power vs. logarithmic model of Fitts' law: A mathematical analysis. *Mathematics and Social Sciences*.
18. Soukoreff, RW, & MacKenzie, IS (2004) Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *International J. Human Computer Studies*, 61, 751-789.
19. Wobbrock, JO, Shinohara, K, & Jansen, A (2011) The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models. *Proc. CHI 2011.* New York: ACM, 1639-1648.
20. Zhai, S (2004) Characterizing computer input with Fitts' law parameters: The information and the non-information aspects of pointing. *International Journal of Human Computer Studies,* 61, 791-809.