

The Entropy of a Rapid Aimed Movement: Fitts' Index of Difficulty versus Shannon's Entropy

R. William Soukoreff[†], Jian Zhao[†], Xiangshi Ren[‡]

[†] Department of Computer Science, University of Toronto, Canada

[‡] The School of Information, Kochi University of Technology, Japan
will@dgp.toronto.edu, jianzhao@dgp.toronto.edu, ren.xiangshi@kochi-tech.ac.jp

Abstract. A thought experiment is proposed that reveals a difference between Fitts' index of difficulty and Shannon's entropy, in the quantification of the information content of a series of rapid aimed movements. This implies that the contemporary *Shannon formulation* of the index of difficulty is similar to, but not identical to, entropy. Preliminary work is reported toward developing a model that resolves the problem. Starting from first principles (information theory), a formulation for the entropy of a Fitts' law style rapid aimed movement is derived, that is similar in form to the traditional formulation. Empirical data from Fitts' 1954 paper are analysed, demonstrating that the new model fits empirical data as well as the current standard approach. The novel formulation is promising because it accurately describes human movement data, while also being derived from first principles (using information theory), thus providing insight into the underlying cause of Fitts' law.

Keywords: Fitts' law, Human Performance Modelling, Entropy

1 Introduction

Fitts' law [3] describes the relationship between movement time, distance, and target width, for people engaged in rapid aimed movements. See Figure 1.

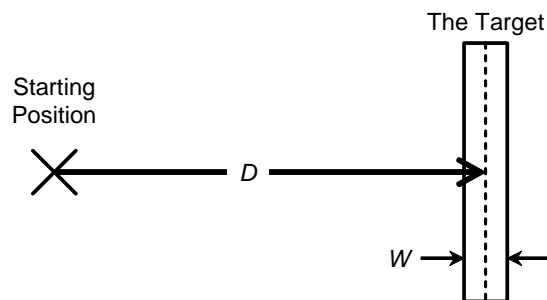


Fig. 1. The Fitts' law movement paradigm.

The *index of difficulty* of a given movement is defined,

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

where D represents the distance from the origin of the movement to the centre of the target, and W denotes the width of the target, both typically measured in millimetres or pixels. Equation 1 is the *Shannon formulation* of the index of difficulty, a formulation that is theoretically and empirically preferable to Fitts' original [8, 9].

Intuitively the index of difficulty, ID , represents the difficulty of completing the movement task defined by D and W , as ID increases for longer distances or narrower targets. By convention the units of the index of difficulty are taken to be *bits*. This convention diverges from the usual definition of bits as defined by information theory [15, 16], as the weighted average of the base-2 logarithm of the probabilities of the set of outcomes of a stochastic process – a point we will return to shortly.

The crux of Fitts' law is that there is a linear relationship between index of difficulty and *movement time*, the average time required to complete the movement task, typically measured in seconds or milliseconds. Commonly in the academic literature this linear relationship is written,

$$MT = a + b \times ID \quad (2)$$

where a and b are empirically determined constants, although in Fitts' original publication [3] the intercept term, a , was zero.

Fitts' index of difficulty is taken (if not formally then at least in name) to be a measure of information (viz., *bits* in the sense of entropy); a consequence of this is that the reciprocal of the slope coefficient b in Equation 2 may be interpreted as *throughput*, the rate of transmission of information, having the units *bits per second*,

$$\text{Throughput} = \frac{1}{b} = \frac{ID}{MT - a}. \quad (3)$$

Note that the ISO standard on pointing device evaluation [5] explicitly defines throughput as the ratio of the effective index of difficulty to movement time (essentially, Equation 3 with $a = 0$). (The technical distinction between $1/b$, and the ISO definition of throughput is not relevant here, under normal circumstances they are very similar, and in either case, ID is taken to be a measure of information entropy.)

Fitts' law typically yields very good fits to empirical data (with R^2 values above 0.9), and has been demonstrated to hold over a wide variety of movement conditions and circumstances (see [12], pages 281-282). In the field of human-computer interaction, Fitts' law is used to model the performance of a number of popular pointing devices including the mouse, stylus, and dactylic touch-screen – the primary applications being the prediction of movement times, and the comparative evaluation of pointing devices. When Fitts' law is used to make predictions of movement time, the meaning of Fitts' index of difficulty is of little importance, because due to the robust linear relationship between ID and MT , Fitts' law provides accurate predictions regardless of the precise meaning of the slope parameter b . However, the meaning of Fitts' index of difficulty is central to the evaluation of pointing devices, because in this case, the Fitts' throughput is used as a comparative dependent measure. This is

not merely a point of theoretical interest – it is critical to the development, improvement, evaluation, and study of pointing devices that the dependent measure used to compare them is sound. It is impossible to determine the tallest man in a room solely by comparing their weights (although we might get close); similarly, it is impossible to be certain that one pointing device is superior to another if we do not know what units of measurement have been used to compare them. Without Fitts’ law’s throughput we would have no way to compare the over-all performance of two pointing devices, yet if Fitts’ law is to be used as a performance metric, then we must know for certain that the Fitts’ law throughput means what we think it means.

In the field of information theory, throughput refers to the number of useful bits of information conveyed in a period of time, or more precisely, the entropy per unit of time. Shannon [15, 16] defines entropy as a quantity of information, measured in the units, *bits*. A coin-toss (resulting in a “heads” or “tails” with equal likelihood) produces 1 bit of entropy. But this definition of entropy and throughput is very different from the terms *index of difficulty* and *throughput* as used in the context of Fitts’ law. The questions we will look at next are: What is the precise relationship between the index of difficulty and entropy? Is it valid to interpret *ID* as a measure of information? And by extension, is it valid to interpret Equation 3 as throughput?

2 The Information Theoretic Basis for Fitts’ Law

From the very beginning Fitts attributed his law to information theory. The hypothesis Fitts stated in his seminal 1954 paper was: “If the amplitude and tolerance limits of a task are controlled by the experimenter, and the subject is instructed to work at his maximum rate, then the average time per response will be directly proportional to the minimum average amount of *information per response*¹ demanded by the particular conditions of amplitude and tolerance” [3, page 383]. Additionally, because in his original experiments the results of using a 1 oz and 1 lb stylus were similar, Fitts concluded that the relation he discovered was due to information processing instead of a physical mechanism, “The finding that relatively small differences in performance result from the change in stylus weight, and the validity of predictions of performance rate from the index of task difficulty lend support to the basic thesis of this paper, that it is the degree of control required over the organization of a response, i.e., *the amount of information required to specify a response*, which is the major factor limiting the rate of motor performance” [3, page 390].

Crossman had a significant impact upon our understanding and treatment of Fitts’ law through his invention of *the adjustment for accuracy* (a post-hoc analysis technique frequently employed in Fitts’ law studies). Crossman states that entropy underpins Fitts’ law: “the perceptual load in the current control of hand-movement is measured by the *difference between the initial and final entropy* of the extent of the movement” [2, page 74].

MacKenzie’s contribution to Fitts’ law was the Shannon formulation of the index of difficulty (Equation 1), and he states, “in executing a movement task, *ID* is the

¹ Italics have been added to the quotations appearing in this section for emphasis.

number of bits of information transmitted, and $[1/b]$ is the rate of transmission” [9, page 98].

In addition to the statements by the eminent contributors to Fitts’ law quoted above, similar statements appear in textbooks and review papers. For example when Card, Moran & Newell discuss the comparison of pointing devices, they conclude that “the match of the Fitts’ law slope of the mouse to the $[b] \approx .100$ sec/bit constant observed in other hand movement and manual control studies means that *positioning time is apparently limited by central information-processing capacities* of the eye-hand guidance system” [1, page 247]. Also Schmidt & Lee, although employing Fitts’ original formulation of the index of difficulty,

$$ID_{Fitts} = \log_2 \left(\frac{2D}{W} \right) \quad (4)$$

unequivocally characterise it as a quantity of information: “Recall the discussion of Hick’s law of choice reaction time ... the equation for that relationship also had a Log_2 term. The $\text{Log}_2(N)$, in which N was the number of equally likely stimulus-response alternatives, was a measure of the amount of information (in bits) required to resolve the uncertainty about N alternatives. The Log_2 term in Fitts’ law can be seen in a similar way: $2D/W$ is related to the number of possible movements, and the $\text{Log}_2(2D/W)$ is the *information required (in bits)* to resolve the uncertainty among them” [13, page 212]. Lastly, in a recent review paper Seow writes: “This index of performance $[1/b]$, or commonly called throughput (TP) by some HCI researchers ... is measured in bits per unit time and is homologous to the rate of gain of information in Hick’s (1952) paradigm and analogous to the channel capacity in Shannon and Weaver’s (1949) theory” [14, page 332].

These quotations support the perspective that the index of difficulty, ID , is a measure of information, and that the reciprocal of the slope, $1/b$, is throughput; however, this perspective is not universal. Both Kvålseth [6] and Lai et al. [7] provide analyses of empirical data indicating that the original formulation of the index of difficulty (Equation 4) differs from entropy. However, the divergence reported by Kvålseth could arise from his participants overemphasising speed at the expense of accuracy, in the absence of the adjustment for accuracy. Lai et al. analysed only the entropy of the spread of movement endpoints, as opposed to the entropy of the entire movement tasks, which undermines their argument.

The conclusion that one may draw from the literature is that the index of difficulty is possibly (even perhaps, likely) a measure of information, but it may not be. Next we shall consider whether our doubts regarding the entropic nature of the index of difficulty can be resolved by examining the derivation and foundation of the index of difficulty.

2.1 Shannon’s Theorem 17

Fitts cited Shannon’s theorem 17 as the basis of his law [3, page 382]. Shannon’s theorem 17 [16, page 43] describes the information capacity of a band limited continuous channel under an average power limitation in the presence of noise, finding that the channel capacity is equal to the difference between the entropy of the

signal and the noise, or,

$$C = H(\text{signal}) - H(\text{noise}) = B \cdot \log_2 \frac{S+N}{N} \quad (5)$$

where $H(\cdot)$ represents the entropy of the given quantity, B represents the bandwidth of the communications channel, and S and N represent the average power of the signal and noise respectively. Fitts' original formulation of the index of difficulty (Equation 4) has been largely supplanted in the recent literature by the Shannon formulation (Equation 1), originally proposed by MacKenzie [8, 9]. Comparing Equations 1 and 5, the analogy is clear to see. Fitts took movement distance to be analogous to signal power, and target width to noise power; so the theoretical foundation that Fitts' law rests upon is an analogy to information theory. This is not a robust foundation – movement distance and target width are simply not equivalent to average power levels, which are the product of root-mean-squared voltage and current levels. These equations look similar to one another, but similarity is not enough to convince us one way or the other. Fitts' law agrees with human performance data very well under a wide variety of circumstances, suggesting that the analogy to information theory is a fitting one, but in what subtle way may it differ? Over the years several researchers have proposed alternate derivations (or formulations) of Fitts' law, usually rooted in the characteristics of movement biophysics, motor control theory, or feedback loops (see [11, 12]), however currently no universally accepted explanation of Fitts' law exists (see [19] page 761, and footnote 17).

Without a robust theoretical basis underpinning Fitts' law, not only are we unable to answer the index of difficulty versus entropy question, but we also lack an explanation for why the Fitts' law phenomena occurs – why is there a linear relationship between movement time and the index of difficulty?

3 Index of Difficulty Versus Entropy – A Thought Experiment

An inconsistency between Fitts' index of difficulty and Shannon's entropy can be demonstrated via a thought-experiment that compares how these two quantities measure information transmission during an imaginary data entry task. The idea is to compare the information transmitted during a data entry single-finger typing task, as predicted by the index of difficulty of the finger movements versus the entropy of the typed text. See Figure 2.

The idea that Fitts' law could be used to predict data entry times, and consequently could be used to improve keyboard designs, originated with Card, Moran & Newell [1, pages 51-57], but was extended and popularised by others (for a review see [10]). The idea is depicted in Figure 3 – Fitts' law can be used to model the time to move between keys on a keypad or keyboard. By applying Fitts' law to the entirety of the keyboard (i.e., from every key to every key), and coupling this with a statistical model of the pairs of letters in a language, a model of single-finger text entry can be constructed.

Empirically, the model described above applies best to highly-trained “expert” data entry behaviour, where typing-time consists solely of movement time, no time being

needed for searching for the next key, cognitive processing, fatigue, etc. However, in our thought-experiment, the expertise of the hypothetical data entry clerk has no effect upon the outcome because the results of this thought-experiment concern quantifying the amount of information transmitted *per keystroke*, not *per unit of time*. Thus the implications are independent of the speed of typing during the imaginary data entry task.

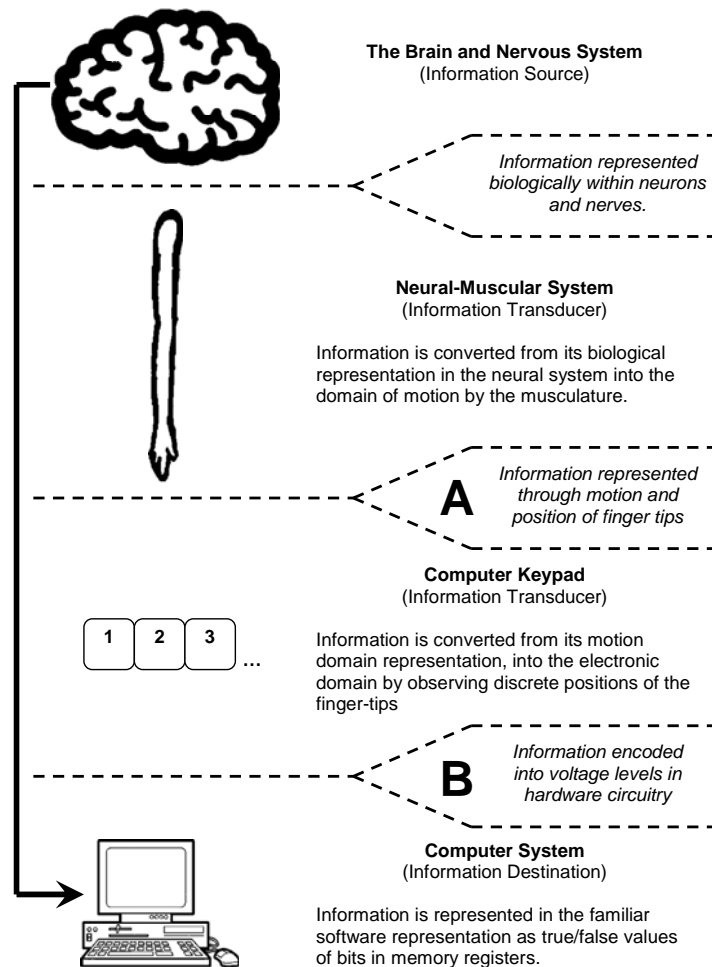


Fig. 2. The task of typing digital numeric characters is represented as a communications network containing an information source and destination with two intervening information transducers. Two specific locations within the network are identified. The point *A* corresponds to the interface between the *motion* transmission medium and the transmission medium of electronics within the keypad – the information content of the data stream at point *A* can be measured with the index of difficulty. The point *B* denotes the place within the computer where the information is represented as a series of characters – and entropy can be measured using Shannon’s definition of the entropy of discrete data.

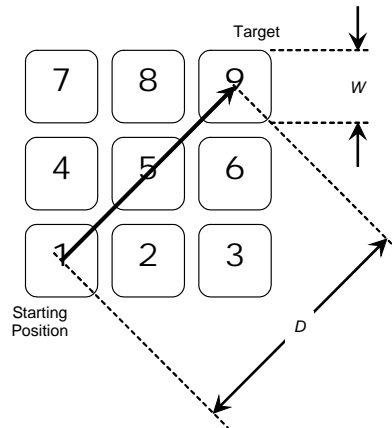


Fig. 3. Single-finger typing is depicted as a Fitts movement task. The distance between the keys, and the size of the keys², can be found from the geometry of the keypad. Application of Equation 1 yields the index of difficulty that corresponds to any key-to-key movement made by the finger across the keypad, representing the quantity of information needed to accomplish the keystroke according to Fitts' law. By averaging over all such movements the mean index of difficulty per character of data can be found, and if the index of difficulty is indeed a measure of entropy, then this quantity should equal the average entropy per character.

The apparatus for our thought-experiment is the *two-key binary keyboard* – an imaginary keyboard that contains only two keys, labelled '0' and '1'. A degenerate keyboard such as this would not facilitate as easy text entry as a full-size Qwerty keyboard, and yet it would be universal – with a suitable decimal, ASCII, or unicode-to-binary conversion table (and some patience), any text or numeric data could be entered. A movement model has been constructed for the two-key binary keyboard, see Figure 4.

Figure 4 presents the index of difficulty corresponding to all four possible key-to-key transition-movements (i.e., moving from a previous key to the subsequent key), that could occur during data entry on the two-key binary keyboard: 0→0, 0→1, 1→0, and 1→1. (Note that "0→0" and "1→1" denote key repetition of the '0' and '1' key respectively. The concern that the reader may have with regard to using Fitts' law to model key repeats will be dispelled in section 3.2.)

By coupling this movement model with a statistical language model it is possible to determine the average number of bits per character predicted by the Fitts' law movement model. Imagine that the "language" to be entered on the two-key binary keyboard consists of a random sequence of binary digits ('0' and '1'), where both digits have the same likelihood of occurrence,

$$p(0) = p(1) = \frac{1}{2}, \quad (6)$$

² Key width varies with the angle of approach to rectangular keys, this is not a factor in the 1-dimensional thought experiment described in the text, so we disregard this detail here.

and occur independently of one other,

$$p(0|0) = p(0|1) = p(1|0) = p(1|1) = \frac{1}{2}.$$

In this case the likelihood of each of the four key-to-key transitions is $\frac{1}{4}$, because,

$$p(0 \rightarrow 0) = p(0) \cdot p(0|0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, \quad (7)$$

$$p(0 \rightarrow 1) = p(0) \cdot p(1|0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$

$$p(1 \rightarrow 0) = p(1) \cdot p(0|1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, \text{ and}$$

$$p(1 \rightarrow 1) = p(1) \cdot p(1|1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

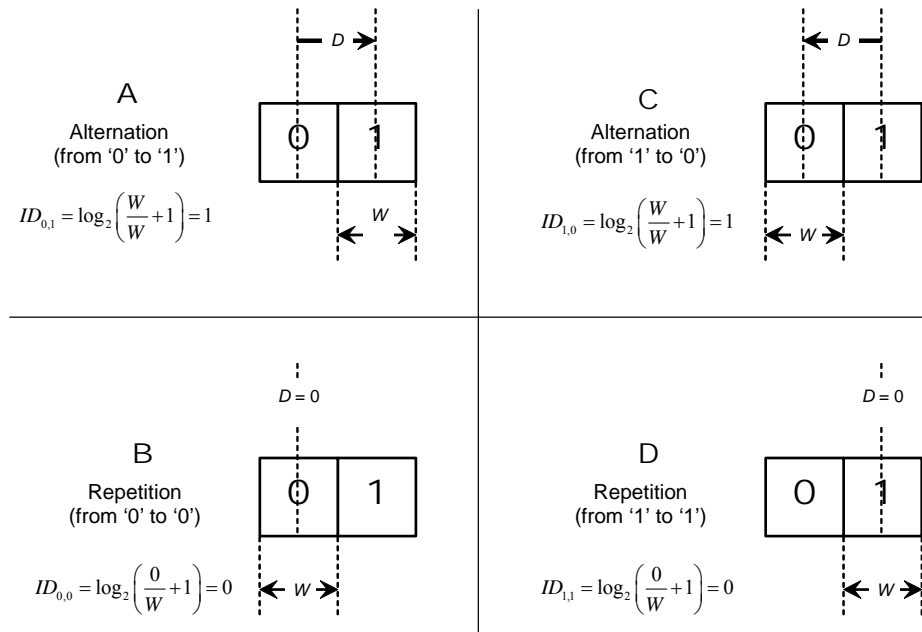


Fig. 4. The movement model for the two-key binary keyboard consists of all of the possible movements that may occur during data entry. Four possible key-to-key transitions are shown, 0→0, 0→1, 1→0, and 1→1.

3.1 Analysis of the Thought Experiment

The average number of Fitts' law bits per key-to-key transition, is simply the average of the index of difficulties weighted by the likelihood of each key-to-key transition. So, calculating a weighted average of the index of difficulty values from Figure 4 using the transition probabilities from Equation 7 we obtain,

$$\begin{aligned}
 \sum_{i,j \in \{0,1\}} p(i \rightarrow j) \cdot ID_{i,j} &= p(0 \rightarrow 0) \cdot ID_{0,0} + p(0 \rightarrow 1) \cdot ID_{0,1} \\
 &\quad + p(1 \rightarrow 0) \cdot ID_{1,0} + p(1 \rightarrow 1) \cdot ID_{1,1} \\
 &= \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 0 \\
 &= \frac{1}{2} \text{ bit per key-to-key transition.}
 \end{aligned} \tag{8}$$

According to Fitts' law, given the language model consisting of random binary digits, the average number of bits per key-to-key transition on the two-key binary keyboard is 0.5 bits. And because there is a one-to-one correspondence between keystrokes and key-to-key transitions (each keystroke being preceded by a Fitts key-to-key transition)³, the act of entering a random sequence of binary digits results in the transmission of 0.5 bits of "Fitts' information" per keystroke. However, this quantity does not match the number of bits of information transmitted as defined by Shannon [16] Theorem 2. Shannon defines entropy (information content arising from the outcome of a stochastic process) as,

$$H(\{p_i\}) = -\sum p_i \cdot \log_2(p_i) \tag{9}$$

Substituting the probabilities defined in Equation 6 into Equation 9 yields,

$$\begin{aligned}
 H(\{\text{binary digit}\}) &= -\sum p_i \cdot \log_2(p_i) \\
 &= -p(0) \cdot \log_2 p(0) - p(1) \cdot \log_2 p(1) \\
 &= -\frac{1}{2} \cdot \log_2(\frac{1}{2}) - \frac{1}{2} \cdot \log_2(\frac{1}{2}) \\
 &= 1 \text{ bit per binary digit.}
 \end{aligned} \tag{10}$$

Comparing the results in Equations 8 and 10 it is clear that Fitts' law and Shannon's entropy assign different quantities of information to the same task. This disparity is very troubling in a theoretical sense, because it implies that a data entry clerk can perform a series of physical motions averaging only 0.5 bits of information each (at point **A** in Figure 2), yet somehow when these motions are interpreted by the two-key binary keyboard, one entire bit of information is communicated per movement (at point **B** in Figure 2). Because there is only one communications channel between the information source and the destination, and hence no other means for the missing $\frac{1}{2}$ bit of information to be transmitted to the keyboard, there appears to be spontaneous generation of information, contradicting a fundamental tenet of information theory (specifically, Theorem 7 on page 15 of Shannon [16], which states that a finite-state transducer cannot increase the entropy of a signal). It

³ We assume that the clerk positions their finger over the keyboard prior to commencing the data entry task, so that even the first keystroke is preceded by a key-to-key transition.

should not be possible for there to be less information at point *A* than at point *B* of Figure 2.

In this analysis Shannon’s quantity of entropy was calculated via the application of a very straightforward equation (Equation 10), and the result matches our intuition – we would expect each digit in a sequence of (equiprobable and independent) binary digits to contain one bit of information – and so it seems logical to conclude that the problem lies in the Fitts’ law side of the analysis. The index of difficulty does not account for all of the information transmitted via motion.

3.2 Key Repetitions Are Not the Problem

Looking at Figures 4B and 4D, it seems that the cause of the Fitts-Shannon disparity may be the key repetition keystrokes. Key repetition is a degenerate form of “movement” where the distance moved is zero, and the corresponding Fitts’ law index of difficulty for key repeats (defined by Equation 1) is also zero. In the case of the two-key binary keyboard, key repeats account for half of all key-to-key transitions, and consequently in our thought-experiment, key repetition plays a significant role in reducing the average index of difficulty. Further there is a theoretical consideration, it could be argued that Fitts’ law applies only to rapid aimed *movements* – it is not entirely clear whether it can or should be used to model in-place tapping (this argument has been made before, see [18]).

We can investigate the key repetition question in two ways. (1) By enlarging the keyboard used in our thought-experiment we can reduce the frequency of key repeats. For example, in the case of a three-key keyboard, key repetition represents only 33% of the keystrokes, and for a four-key keyboard only 25%. And in general, an $n \times 1$ key keyboard would have n^2 possible unique key-to-key transitions, of which only n would be key repeats, and so the effect upon our model of key repetition can be reduced to insignificance by simply choosing a large enough value for n .⁴ (2) Alternatively, we can repeat the same thought-experiment calculations with a modified language model that omits key repeats (i.e., the sequence of digits to be entered by the hypothetical data entry clerk is perfectly random and equiprobable, except that it contains no key repeats). Figure 5 depicts the results of analysing theoretical keyboards with up to $n = 30$ keys, in a similar manner to the thought-experiment described previously; the keys of the keyboards were presumed to be in an $n \times 1$ physical arrangement, and in each case the alphabet size was presumed to be n . This calculation is presented twice – once including key repeats (see the *with key repeats* curve), and once where key repeats were excluded (*without key repeats* curve).

Mathematically, the formulation for the *with key repeats* curve presented in Figure 5 is a parametric combination of two formulae that both take n (the number of keys on the keyboard) as a parameter. The ordinate (y -coordinate) side of the parametric pairing is the Fitts’ index of difficulty, obtained by generalising the average index of difficulty calculation from Equations 1 and 8,

⁴ Note that the language model must correspond to the size of the keyboard in use, so the data to be entered should be random ternary digits when $n = 3$, quaternary digits when $n = 4$, and in general, n -ary for larger values of n .

$$\overline{ID} = \frac{1}{n^2} \cdot \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \log_2(|i-j|+1), \quad (11)$$

which corresponds to the weighted average of a series of index of difficulties (Equation 1), with target width $W=1$, and movement distances, $D=|i-j|$, that correspond to the distance between keys i and j . The abscissa (x -coordinate) side of the parametric pairing, the Shannon side of the relationship, was generalized from Equation 9, presuming an alphabet of n independent and equally probable symbols, where $p = 1/n$,

$$\begin{aligned} H(n) &= -\sum p_i \cdot \log_2(p_i) \\ &= \sum_{i=1}^n \frac{1}{n} \cdot \log_2(n) \\ &= \log_2(n). \end{aligned} \quad (12)$$

Note the implicit assumption in the definition of entropy given in Equation 12, that the data entry activity is performed without errors. As already described, our interest is in the rate of information per keystroke, not per unit of time, and so the data entry clerk in the thought experiment is free to type as slowly as they wish, and consequently may reduce the error rate to a negligible level. This formulation of entropy is valid given the assumption of negligible errors.

The formulation of the *without key repeats* curve presented in Figure 5 is also a parametric combination of two formulae, derived by removing the repeat conditions from Equations 11 and 12. The ordinate was calculated via,

$$\overline{ID} = \frac{1}{n^2 - n} \cdot \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \log_2(|i-j|+1), \quad (13)$$

while the abscissa was,

$$H(n) = \log_2(n-1). \quad (14)$$

The results depicted in Figure 5 demonstrate that the divergence between the Fitts' index of difficulty and Shannon's entropy are systemic, and not the result of key repeats. The trend of the divergence between the Fitts and Shannon models is clear – as the size of the keyboard increases the models do not converge (neither curve, *with key repeats* nor *without*, converge toward the reference line), a non-vanishing difference between these curves and the reference line is apparent. As we would expect, the *with key repeats* and *without key repeats* curves converge towards one another as n increases, as the effect of key repeats diminishes with increasing n . The data points toward the right of Figure 5 on the *with key repeats* curve represent the situation where the keyboard contains 30 keys, the alphabet contains 30 unique characters, and key repetition accounts for merely 3% of the keystrokes. At this point, the difference between the average entropy (4.91 bits) and the average index of difficulty (3.07 bits) is almost 2 bits, representing a significant disparity between the index of difficulty and entropy.

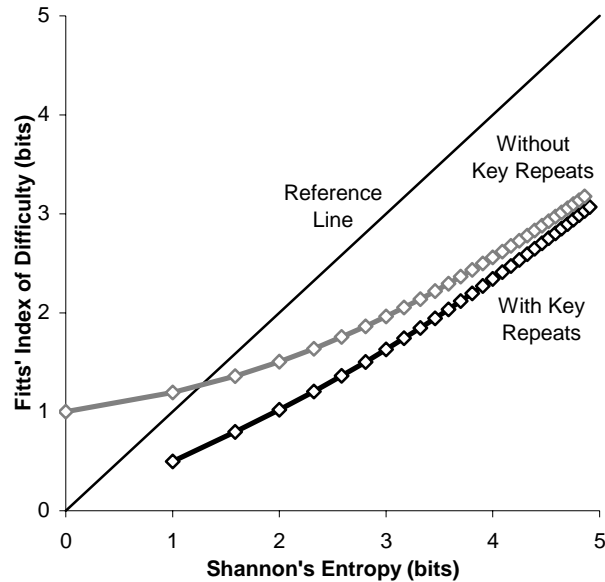


Fig. 5. Thought-experiment results for keyboards with 2 to 30 keys, with and without key repeats. The *Reference Line* (through the origin, with a slope of one) marks the position where the data points should lay if there was equality between the Fitts and Shannon models.

3.3 The Validity of the Comparison within the Thought Experiment

Could the disparity revealed by the thought experiment be an indication that the comparison between the index of difficulty and entropy proposed by the thought experiment is invalid – like “comparing apples and oranges”? No. The thought experiment constructs a series of Fitts movements for which it is possible to know the precise real entropy. This means that the thought experiment provides a means to quantify from outside of the body (entropy), the minimum information internal to the body responsible for the movement of the fingertip (Fitts’ law). Or in other words, the thought experiment allows us to compare the quantity of information *encoded within* the movement, with the information necessary to *prescribe* the motion.

As the data path depicted in Figure 2 is the sole means of communication from the information source to the destination, the information must traverse the point *A*, wherein the data stream is entirely encoded in the movements of the operator’s fingertip. The fingertip movements are interpreted by the keypad as it converts the movement signals into digits, and because the information is conveyed via the movement of the fingertip, it is the motions themselves that must contain an adequate quantity of entropy for this communication to be possible. So, the movements of the fingertip can be equivalently described as either a series of digits or as a series of distances to be moved, but because the intended outcome is identical regardless of which encoding of the messages is used, why should the entropy differ?

It is possible to imagine a situation wherein the information necessary to prescribe a movement is greater than the information encoded within the movement (for example, if entropy was lost due to corruption with noise), but this is not what the thought experiment reveals. Figure 5 shows that the information encoded within the fingertip movements is greater than the quantity of information thought necessary to prescribe those motions as described by the index of difficulty – which is troublesome. We are forced to conclude that there is an inaccuracy in the formulation of index of difficulty – it underestimates the quantity of information shown to be encoded within the fingertip movements.

4 The Entropy of a Rapid Aimed Movement

Having concluded that the index of difficulty as defined by Equation 1 is not an accurate measure of entropy, the question that presents itself is, what is the entropy of a rapid aimed movement? The purpose of this section is to address this question using only principles of information theory, with the smallest number of additional assumptions possible. This is not to say that a larger set of assumptions would not have the potential to yield a superior model, but our point here is that a well performing model can be created from an uncomplicated basis, and a small set of assumptions aids in interpreting the model and explaining the existence of Fitts' law.

The model presented here combines three things, an estimation of the entropy pertaining to the distance moved, an adjustment corresponding to the natural variability and inaccuracy of human performance (i.e., errors), and an additional term that accounts for the fact that each movement is a part of an ensemble of movements.

$$ID_{Entropy}(D, W) = H(D) - H(W) + H(Ensemble). \quad (15)$$

Note that, so long as the probability distribution of the distances moved (viz., the signal) is statistically independent of the probability distribution of the errors (viz., the noise), then these quantities may be combined through simple subtraction as we have done here (see Shannon [16], page 43, Theorem 16).

4.1 The Entropy Pertaining to the Movement Distance

Entropy provides a measure of the uncertainty within the set of possible outcomes of an information-generating process. So to calculate entropy, we must first identify the set of possible outcomes, and their probabilities. The ultimate goal of a rapid aimed movement is to enter a specific area of space, and so the movement endpoints seem to be the natural quantity of interest that reflects the result of the movement task. In general, the probability distribution of the movement endpoints is unknown.⁵ An important detail that arises here is that it is not the software designer's,

⁵ Note that the distribution of the scatter of movement endpoints is known to be approximately normally distributed, for example [3, 20]. But this is not the same thing as the distribution of the endpoints, which also depends upon the magnitude of the movement.

experimenter's, nor all knowing being's understanding of the probability distribution that is important, it is the understanding possessed by the user, the experiment participant, or (in the case of the thought experiment) the data entry clerk, that affects their output (see Hyman [4], page 189).

We will consider two possibilities of what someone performing a pointing task may know regarding the probability distribution of movement distances. Either they know the maximum extent of motion, $x \in [0, U]$, (where U is a specific positive distance denoting the size of the movement 'universe'), or nothing $x \in [0, \infty)$. We may estimate the probability distributions corresponding to these assumptions using *principle of maximum entropy*, which states that the probability distribution that best represents a given phenomenon is the distribution yielding the largest entropy, subject to what is known about the phenomenon. The respective entropies are presented in Table 1 (derivations are provided by Shannon [16], pages 35-37). Note that for the case where the maximum distance is not known in advance, $x \in [0, \infty)$, then we must include an additional assumption, which is that the expected distance (i.e., the first moment or mean, of the distribution) is equal to the required movement distance U , which seems reasonable as we would expect the average movement to be centred upon the target, with equal likelihoods of exceeding or undershooting that distance.

Table 1. Maximum entropy distributions for the two cases, where the distribution of maximum movement distances (U) is known by the mover, or not. (Note that conversion from *nats* (natural information units) to *bits* was accomplished by division by $\log_e 2$, see [16, pages 1-2].)

A Priori knowledge	$x \in [0, U]$	$x \in [0, \infty)$
Maximum Entropy Distribution	Uniform Distribution	Exponential Distribution
Entropy (Bits)	$H_{Uniform}(U) = \log_2(U)$ (16)	$H_{Exp}(U) = \frac{1}{\log_e 2} + \log_2(U)$ (17)

4.2 Noise, Errors, and the Distribution of Movement Endpoints

The distribution of the scatter of movement endpoints is known to be approximately normal [3, 20]. But there is another argument for modelling the scatter of movement endpoints with a normal distribution that can be used without having to explicitly assume that the shape of the distribution is normal. If we presume that errors take the form of deviations in both directions (closer and farther) from the ideal movement distance, then the principle of maximum entropy would have us use the distribution with the maximum entropy defined on the range $(-\infty, \infty)$, which is the normal distribution. The entropy of the normal distribution is, (from [16], page 37)

$$H_{Gaussian}(\sigma) = \frac{1}{2} \cdot \log_2(2\pi e \cdot \sigma^2). \quad (18)$$

The relationship between the target width, W , and the standard deviation of the movement endpoints, σ , will affect the error rate. If we assume that the target width represents ± 2 standard deviations as Crossman ([2], pages 75-76) suggests it does, then we will match the 4% error rate standard commonly in use today (see [19], page 764). Inserting, $\sigma = W/4$, into Equation 18 yields,

$$H_{Errors}(W) = \frac{1}{2} \cdot \log_2 \left(\pi e \cdot \frac{W^2}{8} \right). \quad (19)$$

4.3 The Entropy of the Ensemble of Movements

The entropy of an individual outcome depends upon the statistic properties of the set of possible outcomes. For example, consider the entropy of the letter ‘A’? If this letter is a part of an ensemble of Unicode UTF-8 characters, then 8 bits of information are needed to represent it. The ‘A’ could have been any of the other 256 characters of that ensemble, and (see Equation 12), $\log_2 256 = 8$. Alternatively, if the ‘A’ is part of an ensemble of alphabetic text (26 letters + space) then $\log_2 27 = 4.8$ bits are needed to represent it. Lastly, if the ‘A’ is part of an ensemble of letters in English text, then it represents between 1 and 2 bits of entropy (see [17]).

Similarly, the entropy of a rapid aimed movement depends upon the statistical properties of the other movements that one may be required to perform. The theoretical implications of this idea, and the practical aspects of accommodating this concept, both deserve more space than we have room for in this paper. Here, we will focus on one particular facet of pointing – Fitts’ law does not accommodate the direction of movement, and yet the thought experiment data entry task does. For example, for the $n = 10$ case (where the keyboard contains 10 keys, 0 through 9), if one’s finger has just finished hitting the ‘5’ key, and we know the distance to the next key is ± 2 , there is no way to know which key should be pressed next, ‘3’, or ‘7’. Ambiguity arises because there are two equally likely movements in the ensemble that must be chosen between. We will work around this issue by adding 1 bit of entropy to account for the directional information (*left* or *right*) that would be otherwise lacking from our model.

4.4 Putting It All Together - The Entropy of a Rapid Aimed Movement

Our model is found by inserting Equations 16 or 17, 19 and the constant 1 into Equation 15. Because the difference between Equations 16 and 17 is a single constant we combine these two cases by introducing a parameter, m ,

$$ID_{Entropy}(U, W) = m + \log_2(U) - \frac{1}{2} \cdot \log_2 \left(\pi e \cdot \frac{W^2}{8} \right) + 1. \quad (20)$$

where m has a value of 0 if the mover knows the distribution of distances to be moved, or $1/\log_e 2$, (which equals 1.443), if the distribution of distances is unknown.

5 Evaluating the Entropy Formulation of the Index of Difficulty

Reanalysing the thought experiment paradigm using the new entropy-based index of difficulty $ID_{Entropy}$,⁶ reveals that over the range of difficulty values examined (1-5 bits), and particularly for larger values (of n and ID), it fits the real (Shannon) entropy better than the traditional index of difficulty. The two variations of the movement distance entropy calculation, sandwich the Shannon entropy between two bounds, with the exponential distance entropy above, and the uniform below, see Figure 7. This suggests that the assumptions and approach are on the right track, but future work remains to determine the role of the m parameter in Equation 20 (is it valid to allow m to take values between the two extremes identified?), and to explore the possibility of using other (non-maximum entropy) distributions.

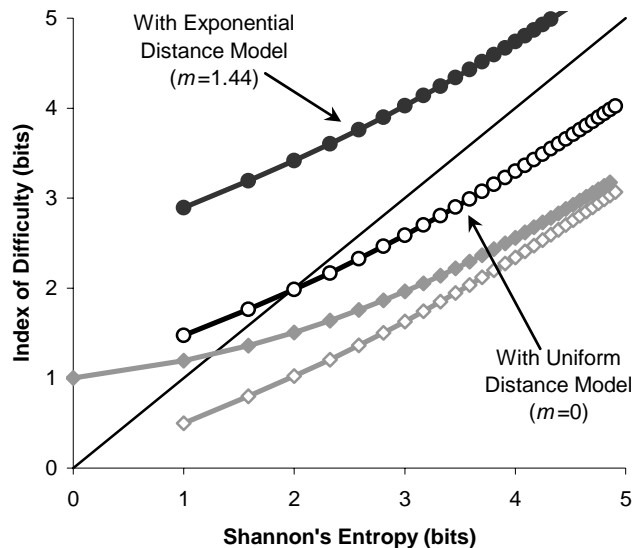


Fig. 6. This figure shows the results of applying the new index of difficulty formulation to the thought experiment results for keyboards with 2 to 30 keys. Two variations of the new model are depicted in black as indicated in the figure. The ordinate of the curve is the new model defined by Equation 20, the abscissa is the entropy found by Equation 12. The line through the origin denotes the position where the model should lay if there was equality between them. The two conditions evaluated previously, *with key repeats* and *without*, appear as the grey curves with empty diamond data points and filled diamonds, respectively.

Figure 7 presents the empirical data from Fitts' 1954 paper [3], analysed with the new and traditional formulations for index of difficulty.

⁶ The concept, in Equation 20, of the size of the movement universe, is different from the usual Fitts' law distance. It was taken to be the distance between the farthest extents of the *to* and *from* keys of the keyboard (i.e., $U=D+1$) representing the largest valid key-to-key movement.

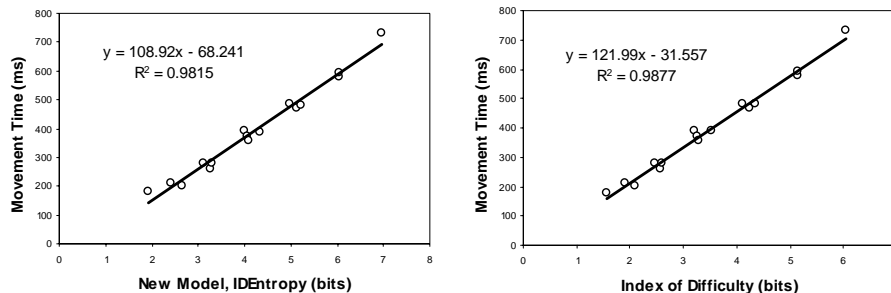


Fig. 7. These figures present an analysis of the empirical data provided by Fitts in his 1954 paper [3], for the 1oz stylus movement task, analysed two ways. On the left we present an analysis of Fitts' data using the entropy formulation of the index of difficulty, Equation 20, assuming the uniform distance entropy variant ($m = 0$). (Note that the value of the m parameter does not affect the fit of the model to the data, it only changes the intercept.) On the right we present an analysis using the standard formulation of the index of difficulty, Equation 1. In both cases the adjustment for accuracy was applied. Note that the R^2 values are similar.

6 Conclusions

The primary contribution made by this paper is the thought experiment that demonstrates the inconsistency between entropy and the Shannon formulation of the index of difficulty. The difference between the index of difficulty and entropy (as depicted in Figure 5) is systemic but not catastrophic, which likely explains why this disparity has not been noticed before, and why many researchers have been able to successfully use Fitts' law to comparatively evaluate the performance of pointing devices in the past. And no doubt the Shannon formulation of the index of difficulty will continue to be used in this fashion in the future.

The problem that has been illuminated here will primarily be of concern when the absolute quantification of information is desired, either to compare pointing devices with other devices for which the true entropy of interaction is known, or to measure human performance in absolute terms (i.e., "How many bits/s can a human output?").

This work also supports the continued use of the adjustment for accuracy in Fitts' law analyses, because performance cannot be measured in terms of entropy without accounting for the equivocation (the bits lost due to noise or errors). For the index of difficulty to be analogous to entropy, errors will have to be similarly accounted for.

Additionally, this paper makes progress toward developing a new entropy-based index of difficulty, that (1) is theoretically rigorous, (2) attempts to resolve the problem demonstrated by the thought experiment, and (3) provides the customary good fit to empirical data, as well as the traditional formulation of Fitts' law does. A small set of assumptions rooted in information theory were used in constructing the entropy ID (which with some mathematical rearrangement resembles the traditional formulation), thus implying that what gives rise to Fitts' law is indeed a limitation of human information processing, as opposed to mechanical or biological factors.

Refining the model, exploring the effect that the ensemble has upon the entropy of an individual movement task, and deciphering the parameter m , remains future work.

Acknowledgements. The authors are grateful to the reviewers for their thorough, copious, and very helpful criticism and comments during the review process. This study has been partially supported by Grant-in-Aid for Scientific Research (No. 23300048) in Japan.

References

- [1] Card S. K., Moran T. P. & Newell A. (1983) *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum.
- [2] Crossman E. R. F. W. (1956) *The Measurement of Perceptual Load in Manual Operations*, PhD Thesis, University of Birmingham, UK.
- [3] Fitts P. M. (1954) The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381-391.
- [4] Hyman R. (1953) Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, 45, 188-196.
- [5] ISO (2002) *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices (ISO 9241-9)*. Reference Number: ISO 9241-9:2000(E). International Organization for Standardization.
- [6] Kvålseth T. O. (1979) Note on information capacity of discrete motor responses. *Perceptual and Motor Skills*, 49, 291-296.
- [7] Lai S. C., Mayer-Kress G., Sosnoff J. J., Newell K. M. (2005) Information entropy analysis of discrete aiming movements. *Acta Psychologica*, 119, 283-304.
- [8] MacKenzie I. S. (1989) A note on the information-theoretic basis for Fitts' law. *Journal of Motor Behavior*, 21, 323-330.
- [9] MacKenzie I. S. (1992) Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7, 91-139.
- [10] MacKenzie I. S. & Soukoreff R. W. (2002) Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction*, 17, 147-198.
- [11] Meyer D. E., Smith J. E. K., Kornblum S., Abrams R. A. & Wright C.E. (1990) Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In *Attention and Performance XIII*, M. Jeannerod (Ed.). Hillsdale, New Jersey: Lawrence Erlbaum, 173-226.
- [12] Plamondon R. & Alimi A. M. (1997) Speed/accuracy trade-offs in target-directed movements. *Behavioural and Brain Sciences*, 20, 279-349.
- [13] Schmidt R. A. & Lee T. D. (2005) *Motor Control & Learning: A behavioral Emphasis*, 4th ed., Champaign, IL: Human Kinetics.
- [14] Seow S. C. (2005) Information theoretic models of HCI: A comparison of the Hick-Hyman law and Fitts' law. *Human-Computer Interaction*, 20, 315-352. Lawrence Erlbaum Assoc.
- [15] Shannon C. E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379-423 and 623-656. Republished and easier to obtain as [16], below.
- [16] Shannon C. E. (1998) *A mathematical theory of communication*. Freely available online from: <http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html>, Bell Labs.
- [17] Shannon C.E. (1951) Prediction and entropy of printed English. *Bell System Technical J.*, 30, 50-64.
- [18] Soukoreff R. W. & MacKenzie I. S. (2002) Using Fitts' Law to Model Key Repeat Time in Text Entry Models. Poster presented at *Graphics Interface – GI 2002*.
- [19] Soukoreff R. W. & MacKenzie I. S. (2004) Towards a standard for pointing device evaluation, Perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61, 751-789.
- [20] Welford A. T. (1968) *Fundamentals of Skill*. Methuen, London.