Speed-accuracy tradeoff of aimed movement: A formal information-theoretic transmission scheme (FITTS)

Julien Gori, Olivier Rioul, Yves Guiard

▶ **To cite this version:**

Julien Gori, Olivier Rioul, Yves Guiard. Speed-accuracy tradeoff of aimed movement: A formal information-theoretic transmission scheme (FITTS). ACM Transactions on Computer-Human Interaction, 2018, 25 (5), pp.27:1-27:33. 10.1145/3231595 . hal-01690089

HAL Id: hal-01690089
https://hal.science/hal-01690089

Submitted on 22 Jan 2018

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.

Speed-Accuracy Tradeoff: A Formal Information-Theoretic Transmission Scheme (FITTS)

JULIEN GORI* and OLIVIER RIOUL, LTCI, Telecom ParisTech, Université Paris-Saclay
YVES GUIARD, LRI, Univ. Paris-Sud, CNRS, INRIA, Université Paris-Saclay and LTCI, Telecom ParisTech, Université Paris-Saclay

The rationale for Fitts' law is that the pointing task has the information-theoretic analogy of sending a signal over a noisy channel, therefore matching Shannon's capacity formula. Yet, the currently received analysis is incomplete and unsatisfactory: there is no explicit communication model for the aiming task, there is a confusion between central concepts of capacity (mathematical limit), throughput (average performance measure) and bandwidth (physical quantity) and between source and channel coding so that Shannon's Theorem 17 can be misinterpreted. We develop an information-theoretic model for pointing tasks where ID is the expression of both a source entropy and a channel capacity when misses are not allowed; then extend the model to include mistakes at rate ε and prove that ID should be adjusted to $(1-\varepsilon)$ID. Finally, we reflect on Shannon's channel coding theorem: Only minimum movement times, not performance averages, should be considered.

ACM Reference format:
Julien Gori, Olivier Rioul, and Yves Guiard. 0000. Speed-Accuracy Tradeoff: A Formal Information-Theoretic Transmission Scheme (FITTS). 1, 1, Article 00 (0000), 34 pages.
https://doi.org/0000001.0000001

1 FITTS' LAW: AN INTRODUCTION

The basic principles of the speed-accuracy tradeoff (e.g., that one can deliberately slow down one's movements to achieve a better precision) have been known for a long time by students of human motor control [56], but the best-known attempt to mathematically describe the tradeoff is due to Fitts [15]. Fitts' law, as we say today, predicts the movement time MT required to reach a target of width W located at distance D, through a parameter called the index of difficulty (ID) expressed in bits:

$$\mathrm{ID} = \log_2 \frac{2D}{W} \qquad \text{bit.} \tag{1}$$

The higher the value of the index, the more difficult the task, and the more time needed to reach the target. Thus, Fitts' law reads

$$\mathrm{MT} = a + b \cdot \mathrm{ID}, \tag{2}$$

*The corresponding author: julien.gori@telecom-paristech.fr

This is an author version, prior to reviewing.
0000. Manuscript submitted to ACM

where the intercept a and the slope $b > 0$ are constants to be empirically adjusted. The law has since been extended using an effective index of difficulty $\mathrm{ID_e}$ and compressed into a one-dimensional quantity, called the throughput, to which we will return.

Being successfully applicable to all sorts of conditions—e.g., with restricted visual feedback [57]) and with various types of participants (e.g., elders [3])—and in several environments—e.g., under water [28])—the law has proven to be impressively robust from the *empirical* point of view. Its *theoretical* foundation, however, has been challenged many times in many frameworks. Fitts originally used results from information theory [15] but other derivations have been put forth using feedback considerations [11, 12], [35], ballistic theory [26], control theory [6], and the theory of non-linear dynamical systems [5, 18].

This multiplicity of derivations, providing new ID's or entirely new formulations are discomforting and make Fitts' law much harder to extend. As Meyer *et al.* [36, p. 192] explained: "Although [Fitts'] empirical results were easy to replicate, the theoretical framework that he proposed to account for them was not well accepted [...]. Consequently, this triggered a search for other ways of explaining the logarithmic speed-accuracy tradeoff."

This paper is an attempt to articulate a thorough information-theoretic account of Fitts' law. While the information-theoretic framework will perhaps look archaic to some readers, we will suggest that, quite on the contrary, that framework is still alive and promising. The remainder of this introductory section will put into context and motivate the content of this paper through a short historical review in the remainder of this introduction section.

1.1 Fitts' Law and Shannon's Information Theory

In 1948, Claude Shannon published *A Mathematical Theory of Communication* [48], a paper that pioneered the modern analysis of digital communications. Fitts was inspired by Shannon's work, to which he explicitly referred [15, 17]. Shannon provided mathematically well-defined measures of important concepts such as the information contained in a message or the uncertainty about the possible occurrence of an event. He also described a generic paradigm for communications with a strict partitioning between the *source*, the *encoder*, the *channel*, the *decoder*, and the *destination*. Shannon was able to obtain operational results, such as the maximum achievable rate of transmission over a noisy Gaussian channel [48, Theorem 17]:

Shannon's Theorem 17. *The capacity of a channel of band*[1] B_W *perturbed by white thermal noise of power* N *when the average transmitter power is limited to* P *is given by:*

$$C = B_W \ \log_2\left(\frac{P+N}{N}\right) \qquad \mathrm{bit\,s^{-1}}.$$

As Meyer *et al.* [36, p. 189] explained: "To interpret his results concerning movement speed and accuracy, Fitts (1954) adapted some concepts from information theory, which was popular at the time (Shannon 1948)". In fact, Fitts [15] explicitly used the words *entropy* and *capacity*, and his interpretation of his finding rested on a direct analogy with Theorem 17.

Obviously Fitts was not the only experimental psychologist in the nineteen fifties to pick up concepts from Shannon's information theory. While Shannon developed his theory only to solve specific problems related to

[1] Band is an old-fashioned designation for bandwidth. We use the B_W notation for bandwidth rather than the original W, to avoid ambiguity with Fitts' law notations.

digital communications—in fact, Shannon preferred the name "communication theory"—the book [50] which reprinted Shannon's paper together with an expository introduction by Warren Weaver had an immense impact on many scientists at that time. Weaver advocated the use of information-theoretic concepts to any scientific field adressing broad communication issues, including linguistics, social sciences, and psychology.

One of the first successful applications of information theory to psychology is Hick's law [25], later extended by Hyman [27]. Hick's law states that the time it takes a person to select one item in a set varies linearly with the *entropy* of the set; in the simple case of equiprobable stimuli, the reaction time increases logarithmically with the number of possible choices.

Perhaps the most memorable application of information theory to psychology is due to G. Miller [37]. In this highly-cited and most-influential paper, Miller attributed the coincidence that absolute judgment and short-term memory share the same limits—the famous magic number seven plus or minus two—to the human *capacity* for processing information.

Most of the successes of the information-theoretic approach to psychology were summarized in 1959 in a book by Attneave [2] entitled *Applications of Information Theory to Psychology*.

1.2 Whatever Happened to Information Theory in Psychology

Today it is not uncommon to find information-theoretic approaches in statistics, probability, economics, biology, etc.; however it is less so in psychology. Information theory had become so popular in the nineteen fifties that many psychologists had perhaps become over-eager to use it: Many resulting applications were far fetched and unfruitful; Attneave [2] pointed to "pointless", or "downright bizarre" applications.

The use of information theory outside the sphere of communication engineering was challenged at the same time by Shannon and the information theory community. In a famous editorial, Shannon [49] himself wrote that information theory "has perhaps been ballooned to an importance beyond its actual accomplishments." He insisted that "the use of a few exciting words like *information, entropy, redundancy*, do not solve all our problems." Elias [13], an important figure of the information theory society, urged authors to stop writing papers using information theory outside of its intended scope. In retrospect, Attneave's survey of 1959 looks like a funeral tribute. Since the end of the sixties very few new articles in psychology have referred to information-theoretic principles[2].

In 1963, Crossman and Goodeve [11] proposed a different explanation for Fitts' law which did not rely on information-theoretic results. Their model, based on feedback considerations, assumed an aimed movement to be composed of a sequence of submovements each of fixed duration and covering a fixed fraction of the remaining distance. These authors essentially attributed the logarithmic nature of the law to a visual and/or kinesthetic iterative feedback mechanism. Although the model provided a nice rationale, it was faced with a number of limitations, mostly caused by its deterministic nature—in particular it failed to explain movement end-point variability and excluded the very possibility of target misses.

By the end of the eighties, Meyer *et al.* [35, 36] proposed a *stochastic* feedback mechanism for rapid aimed movements, thus eliminating the main flaw of the Crossman and Goodeve model. Meyer et al. proposed

[2]Another cause for the decline of the popularity of the information-theoretic approach in psychology was the discomforting discovery reported in 1961, by Bertelson [4] that Hick's law [25, 27] could be explained as a sequential effect independently of stimulus entropy. To understand this finding a more sophisticated understanding of information theory was in order, but then psychologists were more tempted by the cognitive approach [38].

what they called a power model of Fitts' law, rather than a logarithmic one. In fact, as shown by Rioul and Guiard [42, 43] mathematically the Meyer et al. model falls in the class of quasi-logarithmic models. The stochastic optimized submovement model of Meyer *et al.* [35] is now considered by many psychologists (e.g. [45]) as the leading explanatory theory of Fitts' law, illustrating the extent to which information theory has lost ground in modern experimental psychology.

In a suggestive title, *Whatever Happened to Information Theory in Psychology?*, Luce [30] explains that information theory is "no longer much of a factor" in psychology, essentially relegating information theory to the rank of a historical curiosity.

1.3 Fitts' Law, Shannon's Theory, and Human Computer Interaction

Fitts' law became popular in the human-computer interaction (HCI) community after a seminal study by Card *et al.* [7]. Unlike experimental psychologists, however, HCI researchers have apparently remained confident in the promise of the information-theoretic approach to Fitts' law thanks to Scott MacKenzie's sustained effort to develop a complete performance model of Fitts' law for HCI using the tools of information theory [32], including an improvement of Fitts' formula to make it more consistent with both Shannon's Theorem 17 and the available empirical data.

MacKenzie [32] later incorporated information-theoretic results such as the entropy of a Gaussian distribution to account for target misses in pointing. Importantly, the recent ISO standardization of the experimental methodology for the evaluation of pointing devices is explicitly based on information-theoretic principles [1, 51].

More recently, Soukoreff and MacKenzie [52] have proposed a "fundamental theorem of human performance" based on modified equations from information theory which the authors claim to explain the speed accuracy tradeoff.

Other HCI researchers working on Fitts' law have expressed the view that in this field the information-theoretic approach has been somewhat imprecise: for example, A. Newell wrote:

> *"Theories are approximate. Of course, we all know that technically they are approximate; the world can't be known with absolute certainty. But I mean more than that. Theories are also deliberately approximate. Usefulness is often traded against truth. Theories that are known to be wrong continue to be used, because they are the best available. Fitts' law is like that. How a theory is wrong is carried along as part of the theory itself."* [39, p.13]

One problem with approximate theories, however, is that one can always devise a slight variation of a model to obtain a better fit with some data, leading to a proliferation of variants. This is certainly the case with Fitts' law: for example Plamondon *et al.* [40] have listed a dozen formulations of the speed-accuracy tradeoff, most, but not all of which correspond to the logarithmic tradeoff function.

The three best-known logarithmic models based on an analogy with Shannon's capacity formula ($C \propto \log(\frac{P+N}{N})$) are:

$$\text{Fitts' index [15]} \qquad \text{ID} = \log_2 \frac{2D}{W}, \tag{3}$$

$$\text{Welford's index [54]} \qquad \text{ID} = \log_2 \left(\frac{1}{2} + \frac{D}{W} \right), \tag{4}$$

$$\text{and MacKenzie's index}^3 \text{ [31]} \qquad \text{ID} = \log_2 \left(1 + \frac{D}{W} \right). \tag{5}$$

MacKenzie's formulation has been almost unanimously accepted in HCI but many experimental psychologists still use Fitts' original formulation, and so it is a fact that no general consensus has been achieved regarding the exact formulation of the law. Natural questions that remain open are:

- why should D/W be analogous to P/N as defined in Shannon's Theorem 17?
- what is the bandwidth B_W of Shannon's Theorem 17 analogous to in Fitts' law ?
- since D and W are amplitudes while P and N in Shannon's Theorem 17 are powers, what happened to the squares?[4]
- which formulation for ID should we choose?

An important concern is that approximate theories may provide "local" results, but rarely do they propose a solid framework that allows a generalization of the law. Rephrasing Newell's quote: usefulness is not only traded for truth but also for generality. On second thoughts, that tradeoff is perhaps less well balanced than it seems.

1.4 Aim of the Present Study

Luckily, information theory does provide the solid theoretical framework we need. Among its appealing features let us mention that it makes it possible to "investigate all kinds of systems without needing to understand the machinery" [29]. There is little doubt that the modeling of so intricate a machinery as the human movement system may benefit from information theory. To continue Newell's quote:

> *"Grossly approximate theories are continuous launching pads for better attempts. Fitts' law is like that too."* [39, p.13]

Any attempt to achieve a sounder, more rigorous theory demands that the flaws of the current account be uncompromisingly acknowledged. We believe the information-theoretic treatment of Fitts' law that is currently received within HCI suffers from three fundamental weaknesses:

- there is no explicit *communication scheme* for the aiming task: no serious information analysis can dispense with such a scheme;
- Shannon's results on channel coding are misinterpreted: Theorem 17 concerns the *transmission*, not the generation of information;
- two concepts, the information-theoretic *capacity*, a mathematical limit, and the *throughput*, an average empirical measure, are usually confounded.

[3] The index of Equation 5 is usually known in HCI as the Shannon index, which suggests an exact match with Shannon's information theory. In this paper to take it for granted that the analogy with Shannon's Theorem 17 holds would amount to begging the question, and so we will refer to the "MacKenzie index", neutrally acknowledging the fact that it was first proposed by MacKenzie [31].

[4] The power of a random variable X is the average of X^2. If X has zero mean, power is identical to variance.

In Fitts' law research as well as in other fields, information theory has suffered the backlash from its popularity in the nineteen fifties—it has been literally a victim of its own success. Blatant abuses of Shannon's theory in a few scientific fields have led, possibly quite wrongly, to its global discredit in fields where its use was indeed promising—and still is.

Our goal in this paper is to show that a simple, yet rigorous communication model for human aimed movement is possible, and that this approach can provide useful results for HCI. The remainder of this paper is organized as follows.

We start in Section 2 by presenting the few fundamentals (known concepts and results) of information theory that will be needed throughout this paper. In Section 3 we then review previous information-theoretic approaches and in Section 4 we propose a simple model for errorless aiming as observable in task contexts where target misses are prohibited, or even technically impossible as was the case in Fitts' disc- and pin-transfer experiments [15]. From this model we will derive Fitts' law through the computation of the capacity of the so-called "uniform channel". in Section 5 we will then extend the model, computing the associated capacity so as to accommodate the occurrence of target misses: the index of difficulty will then become a simple function of the probability of the target miss. Finally, in Section 6 we will show that the very definition of the notion of capacity demands that Fitts' law be interpreted as a law of *extreme*, rather than average performance, a result whose implications for the statistical handling of experimental data are far reaching.

2 SOME KEY CONCEPTS OF INFORMATION THEORY

It is customary in HCI to use the terms of *capacity*, *throughput*, and *bandwidth*—three technical terms that receive different precise definitions in information theory—almost interchangeably when referring to the idea of information-transmission rate.

In the Fitts' law literature the term *capacity* is often used in a non-technical sense. This is the case for example in Fitts' own writings: in both [15] and [17] the word capacity is used three times (in the title as well as on the first and last pages of the paper) but it seems that in Fitts' mind the capacity was a general notion that neither required a formal definition nor afforded measurement.

Typically in HCI, the word *throughput* serves to denote the measured performance, but there has been a long controversy on the definition of that term. One option is to take the inverse of the slope of Fitts' law [7, 59], the other is to take the ratio $\frac{\mathrm{ID}}{\mathrm{MT}}$ [32, 51]. Both options conveniently compress the two parameters of Fitts' law into a single parameter but are not identical because of the existence of the intercept. Some definitions (e.g., in [1]) allow for the integration of the error rates [51]. Not only does the throughput appear to be an all-encompassing measure that lacks an information-theoretically justified definition, it is also confusing to have at one's disposal two incompatible definitions for it.

Finally, *bandwidth* is used either as a synonym of throughput [59] or as an equivalent to $\frac{1}{\mathrm{MT}}$ [34]. Likewise, the term information is loosely applied, often as a synonym for entropy (e.g., [15]), occasionally for mutual information (e.g., [10]).

This section recalls basic definitions and some fundamental results of information theory[5]. Although that material is very well known, it is absolutely essential for the understanding of this paper.

[5]Many important notions and proofs are omitted. The interested reader could look at [9, 41, 58] for more details and in-depth analyses.

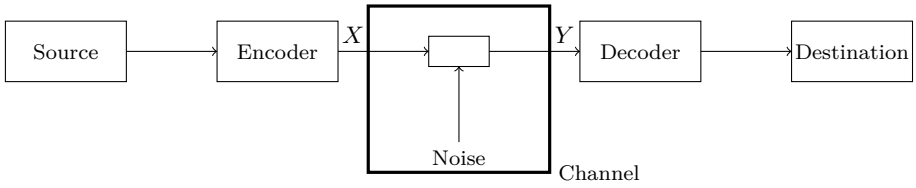

Fig. 1. Shannon's point-to-point communication paradigm

2.1 Shannon's Communication Model

Shannon [48] gave an accurate and generic description of a point-to-point transmission system (see Figure 1). His analysis of information transmission is based on this scheme, composed of five elements. To identify each of these elements in a pointing task is a necessary preliminary step that traditional Fitts' law research has skipped.

- The information *source* produces M, modeled as a random variable. One particular message m is the outcome of M, say $M = m$. The only aspect that matters is that we can assign a probability to each outcome, in line with Shannon's famous quote: "Semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages";
- The *encoder* adapts the message from the source to the channel, in at least two aspects: a physical adaptation in which the message is converted into a suitable signal for transmission (e.g., the variation of an electrical current); and a channel encoding in which certain operations are performed on the message to enhance transmission quality. One important feature is that the encoder performs *deterministic* operations;
- The *channel* is the medium that serves to transmit the signal from the *emitter* (source and encoder pair) to the *receiver* (decoder and destination pair). On its way from the emitter to the receiver, the signal may be corrupted by *noise*. If the input of the channel is X, and the output is Y, then the channel is completely described by the probability of Y conditional on X: $p(Y|X)$.
- The *decoder* also performs deterministic operations to get back to the message space while trying to correct the effect of transmission noise in such a way that the destination can understand the message.

Because of channel noise, a given message at the input of the channel may turn into an erroneous message at the output, so that we may not achieve a completely reliable communication. The revolutionary aspect of Shannon's work was to demonstrate that every channel possesses a non-negative parameter, called its *capacity*, below which every rate of information can be achieved reliably, that is, with an arbitrarily low error rate. In a sense one can trade off the rate of information by lowering the speed of transmission to obtain an accurate communication: this is where Shannon's paradigm almost naturally comes into play in the study of the speed-accuracy tradeoff.

The task of the electrical engineer is to find the encoding and decoding schemes that match the channel so as to ensure optimal transmission (maximizing the transmission rate while keeping a very low error rate). Students of the human motor system actually face a *reverse engineering* problem: all the key elements of

the motor system are in place, but they would like to determine, if possible, the properties of the motor system from what they are able to observe.

2.2 Shannon's Information Measures

Information in Shannon's sense is a measure of randomness. We now review the definitions of entropy and mutual information.

Definition 2.1 (Entropy of a discrete random variable X).

$$H(X) = -\sum_x p(x)\log_2 p(x) = -\mathbb{E}\log_2 p(X) \qquad \text{bit}$$

where X is drawn according to the probability distribution $p(x) = P\{X=x\}$, and where $\mathbb{E}(X) = \sum_x x \cdot p(x)$ (noted $\mathbb{E}X$ when no confusion is possible) denotes the mathematical expectation of the random variable X.

Entropy measures the uncertainty of the outcome of a random variable, and that uncertainty is a function of the probabilities assigned to the different values of the random variable: the higher the entropy of X, the more uncertain its outcome, the harder the prediction. Entropy measures "information" in the sense that the outcome of a random variable will increase the receiver's knowledge (or decreases the receiver's uncertainty).

In pointing studies, the entropy has been used to measure the "difficulty" of the task (e.g. [10, 15]) or the richness of the set of pointing possibilities [46]. In an equiprobable scenario where X is uniformly distributed, the entropy reduces to the logarithm of the number of choices [9, 25, 41].

Entropy is instrumental in proving source coding results: if the source of information produces n messages $X_1, \ldots, X_n$, the information rate

$$R = \frac{1}{n}H(X_1, ..., X_n) = \frac{1}{n}\mathbb{E}\log_2 p(X_1, ..., X_n) \qquad \text{bit symbol}^{-1}.$$

is the amount of information the source produces on average and represents the minimal bit rate at which it is possible to encode the source without distortion. However, since the messages are transmitted over a noisy channel, some information might be lost. *Mutual information*, or, synonymously, transmitted information is the measure we need to characterize the amount of information that is *effectively* transmitted through the channel.

Definition-Proposition 2.2 (Mutual information between random variables X and Y).

$$\begin{aligned} I(X;Y) &= \mathbb{E}\log_2\left(\frac{p(Y,X)}{p(X)p(Y)}\right) = \mathbb{E}\log_2\left(\frac{p(Y|X)}{p(Y)}\right) = \mathbb{E}\log_2\left(\frac{p(X|Y)}{p(X)}\right) \qquad \text{bit} \\ &= H(Y) - H(Y|X) = H(X) - H(X|Y) \end{aligned}$$

where X and Y are drawn according to the joint pdf $p(x, y)$.

Each of these diverse expressions is useful. Mutual information measures the difference between the receiver's uncertainty about the source before the transmission ($H(X)$) and after the transmission given the channel output ($H(X|Y)$). In an ideal (noise-free) transmission, we would have no residual uncertainty on X after receiving Y, so that $H(X|Y)$ would be zero and $I(X;Y) = H(X)$: that information would then be perfectly transmitted from the source to the destination.

2.3 Shannon's Capacity: Maximum Transmitted Information

Shannon's Theorem 17, which was explicitly considered by Fitts, is in fact a corollary to the more general channel coding theorem, which states that the maximum bit rate (capacity) of a so-called "memoryless" channel in a reliable communication scheme is the maximum mutual information.

Channel Coding Theorem. *The capacity of a memoryless channel:*

$$C = \max_{p(x)} I(X;Y) \qquad \text{bpcu},$$

expressed in bits per channel use (bpcu), is such that for any rate $R < C$ and any $\varepsilon > 0$, there exists a coding scheme with arbitrarily small probability of error $\mathbb{P}_e < \varepsilon$.

In other words, channel capacity C is computed as the maximum amount of mutual information $I(X;Y)$ conveyed in the channel. This maximum is usually taken over some cost constraint on $p(x)$ (that is, on the channel use)[6]. As long as rate R does not exceed capacity C, error probability $\mathbb{P}_e$ can be made as small as we like—this defines "reliable communication" as a mathematical limit.

2.4 Throughput and Bandwidth: A Matter of Units

It is customary in modern practice of communication theory to use as units bits per second ($\text{bit}\,\text{s}^{-1}$) or bits per channel use (bpcu) interchangeably when discussing information rates. This is because in almost all digital devices, any waveform can be sampled at a fixed time period T_s. In this case, time is in one-to-one correspondence with sample number: it is the sample number times T_s.

The *throughput* has no fixed definition in digital communications as its definition may vary depending on the application (wireless network communication, packet-based schemes, etc.). The basic idea is to measure an effective speed of data transmission, usually in bits per second.

In contrast, the bandwidth of a signal has a simple definition.

Definition 2.3. The bandwidth is the difference between the upper and lower frequencies in a continuous set of frequencies.

The bandwidth is measured in 'Hertz $\equiv \text{s}^{-1}$' and, therefore, is in no way equivalent to throughout or to capacity. The following sampling theorem[7] can be used to relate 'Hertz', '$\text{bit}\,\text{s}^{-1}$' and 'bpcu'.

Shannon-Nyquist Sampling Theorem. *If a function of time has a limited bandwidth B_W, it is completely determined by its values ("samples") taken at a series of discrete times regularly spaced $\frac{1}{2B_W}$ seconds apart*[8]*.*

By the sampling theorem, T seconds of a waveform of bandwidth B_W correspond to $2TB_W$ samples fed into the channel. Thus, to obtain units in '$\text{bit}\,\text{s}^{-1}$' from a quantity expressed in 'bpcu', one just has to

[6] This cost constraint is usually a power constraint on the transmitted signal, but as we shall see later other types of constraint can be useful.
[7] This theorem has many aliases, ranging from Shannon's sampling theorem to the Whittaker-Nyquist-Kotelnikov-Shannon theorem.
[8] In addition, it is possible to derive a practical procedure to reconstruct the waveform (function of time) from its samples.

multiply the quantity by $2B_W$. For additive white Gaussian noise under a power constraint, Shannon [48] calculated the channel capacity as

$$C = \max_{p(x)} I(X;Y) = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right) = \frac{1}{2}\log_2(1+\mathrm{SNR}) \qquad \text{bpcu}, \tag{6}$$

where $\mathrm{SNR} = \frac{P}{N}$ is the signal-to-noise power ratio. Multiplying by $2B_W$ gives Shannon's Theorem 17:

$$C = B_W \log_2\left(1+\frac{P}{N}\right) = B_W \log_2(1+\mathrm{SNR}) \qquad \text{bit}\,\text{s}^{-1}. \tag{7}$$

Similarly any transmission *rate* R in 'bpcu', when multiplied by twice the *bandwidth* $2B_W$, yields the expression of the *throughput* R in 'bit s^{-1}'.

2.5 Spectral Efficiency

The relation between throughput and bandwidth can also be clarified using yet another important quantity used in digital communications, *spectral efficiency*.

The actual transmission rate R in 'bpcu' (or the actual throughput R in bit s^{-1}) used in a communication system is virtually never equal to the capacity—the capacity is only a theoretical upper bound. However, as we have seen in the preceding subsection, both quantities, when expressed in 'bit s^{-1}', increase linearly with the bandwidth B_W. Since R in 'bpcu' is fixed by the practical coding scheme used in the system, for a fixed code, the only way to increase the throughput R in bit s^{-1} is to increase the bandwidth B_W of the system, which is probably the reason for the widespread conflation of bandwidth and throughput. The ratio between the two quantities is called spectral efficiency:

Definition 2.4. Spectral Efficiency

$$S_E = \frac{R}{B_W} \qquad \text{bit}\,\text{s}^{-1}\,\text{Hz}^{-1} (\equiv \text{bit})$$

where B_W is the available bandwidth and R the actual throughput of the communication scheme.

Thus in an ideal noise-free setup, spectral efficiency would be equal to the capacity in bpcu.

The following interpretation is quite useful: a communication that lasts T seconds, occupies a bandwidth B_W, and successfully transmits L bits will have spectral efficiency

$$S_E = \frac{L}{T \cdot B_W} \qquad \text{bit}$$

So in essence S_E is just the number of transmitted bits (load) divided by the resources (time window and bandwidth) used for transmission.

2.6 Errors vs. Erasures

Due to noise in the channel, transmission mistakes in the channel[9] may occur. These can be of two types: errors and erasures. In communication engineering, an error is said to have occurred when the received symbol differs from that originally sent. For example, the word BUTTER is received in the place of the sent

[9] It is important to distinguish channel mistakes from decoding mistakes. Channel mistakes will inevitably occur, yet they can be corrected. Shannon's channel coding theorem states that the decoding errors can be made arbitrarily rare, meaning we are able to correct nearly all channel errors.

word BATTER, the A having been accidentally replaced by an U. But suppose that the received word is B?TTER, with the question mark signaling a missing character: This is what is called an erasure.

One important difference between an error and an erasure is that the former conveys wrong information whereas the latter conveys no information but the error's position. In usual Fitts' law experiments the outcome of a pointing act can be either measured as an error, i.e., a distance from endpoint to target center, or categorized in an all-or-none way as a hit versus a miss.

The error vs. erasure distinction will be very useful below in Section 5, when we proceed to extend the error-less model to the more general model that allows for target misses.

3 PREVIOUS INFORMATION-THEORETIC DERIVATIONS OF FITTS' LAW

As pointed out in the introduction, Fitts' law has been derived in multiple ways, multiple times. In this section, we look at those derivations that make use of the information-theoretic concepts of entropy and capacity.

Throughout this paper, we will exclusively consider the case of a discrete (one-shot) aiming task because the so-called reciprocal task introduced by Fitts [15] allows a problematic overlap between processes involved in controlling successive movements—in particular, we have the drawback that the variability of the movement endpoint can be in part attributed to the variability of the starting point (See [17, 22] for a more detailed argumentation).

3.1 Difficulty as a Source Entropy: Aiming is Choosing

In an early book chapter that has attracted limited attention, Fitts [14] wondered whether the scope of Hick's law [25] could be broadened: "The selection of a particular response member is only one of the ways in which man can generate information. Another way is by selecting one of several directions or amplitudes of the movement of a designated body member" [14, p. 53].

Hick measured choice reaction time in response to one of several equally probable stimulus events and found that reaction time increased linearly with the logarithm of the number of possibilities. From Figure 2 taken from [14], we see how Fitts envisioned aiming as a *choice*: aiming towards a target of size W out of a distance D is made equivalent by Fitts to choosing one target out of $n = \frac{D}{W}$. Note that Figure 2 represents targets when direction is fixed; adding the choice of direction doubles the choice to $n = \frac{2D}{W}$. Fitts' formulation then becomes, in bits:

$$\mathrm{ID} = \log_2 n = \log_2\left(\frac{2D}{W}\right) \qquad \text{bit},$$

which is almost identical to Hicks' formula: Hick considered $\log_2(n+1)$ bits for n choices, because he considered as a possibility *not* to choose any of the targets.

Welford [54] derived his own index using the same "aiming is choosing" rationale, the difference consisting in the definition of the amplitude to be considered and the way in which the targets are laid out. Figure 3 illustrates layouts considered by Fitts vs. by Welford.

Using the same rationale, we can in fact derive MacKenzie's index of difficulty by taking the amplitude to be equal to D, and the first and last targets centered around the starting and stopping points, as illustrated in Figure 4. Assuming that the probability of hitting a target is only dependent on its geometry, the chance of

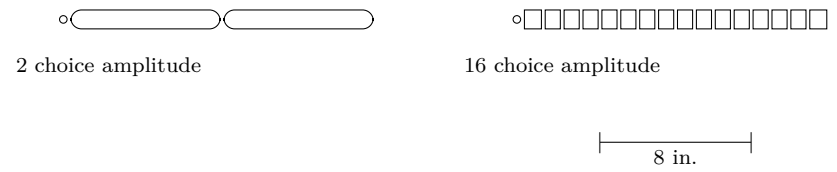

Fig. 2. Schematic drawing of two target patterns that would be used in an experiment to illustrate how the number of alternative movement amplitudes might influence decision time and movement time, reproduced from [14, Figure 2, p. 53]

.

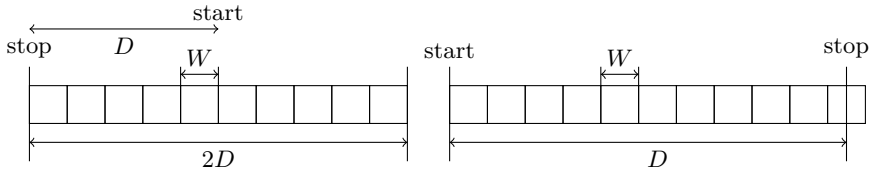

Fig. 3. Placing targets and identifying D and W in the context of the Fitts (left) and the Welford (right) formulation.

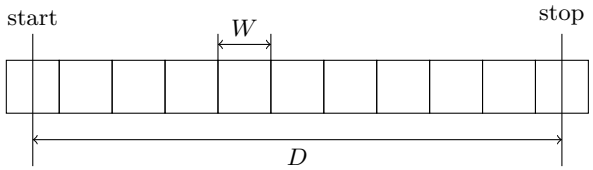

Fig. 4. Placing targets and identifying D and W in the context of the MacKenzie formulation

hitting a target of size W across a distance of $D+W$ is the ratio $p = W/(D+W)$. Provided that $(D+W)/W$ is a round number, the number of targets that fit inside $D+W$ is exactly $n = (D+W)/W$. Since distribution of the targets is uniform, the entropy H of this target distribution is simply

$$H = -\sum_{1}^{n} p \log_2 p = -\log_2 p = \log_2 n = \log_2\left(1 + \frac{D}{W}\right) \qquad \text{bit},$$

which yields an exact match with the MacKenzie ID.

The index of difficulty is here computed as a source entropy—there is no information transmission. The aiming task is simply identified to the creation of target identifiers using the "aiming is choosing" rationale. One may also argue that uniformly distributed (equiprobable) "targets" is a rather implausible hypothesis, but since the uniform distribution is the one that *maximizes* entropy [9, 41] it provides the least upper bound on the entropy for *any* target probability distribution. The resulting entropy H is thus the number of bits required to identify the target position without any prior knowledge whatsoever, and the index of difficulty arises as a measure of the uncertainty associated with the task of choosing one target. The more potential targets (the higher the ratio D/W), the more difficult the pointing task.

It is noteworthy that Fitts [14] explicitly used the term information "generation" rather than transmission and made movement time to depend upon the index of difficulty. This is consistent with his assumption that

the ID should serve to characterize target entropy—a *source* coding rate in Shannon's sense. It is somewhat surprising that to justify the same index in his famous article published one year later, Fitts [15] referred to Theorem 17—a *channel* coding rate in Shannon's sense, which is completely unrelated to source coding.

3.2 Difficulty as a Channel Capacity: An Analogy

The analogy with Theorem 17 was put forward first by Fitts [15], and later by MacKenzie [31]. In Shannon's capacity formula for the additive white Gaussian noise channel

$$C = B_W \log_2(1 + \mathrm{SNR}) \qquad \mathrm{bit\,s^{-1}},$$

MacKenzie [31] identified the bandwidth to the reciprocal of movement time $B_W = 1/\mathrm{MT}$, and $\log_2(1+\mathrm{SNR})$ to $\mathrm{ID} = \log_2(1 + D/W)$, so that $\mathrm{MT} = \mathrm{ID}/C$.

Fitts [15] followed the same steps, except that he identified $\log_2(1 + \mathrm{SNR})$ with $\log_2(2D/W)$ instead of $\log_2(1 + D/W)$. The addition of 1 to the term inside the log by MacKenzie was inspired by the visual shape of Shannon's capacity formula which can be expressed in two mathematically equivalent forms:

$$C = B_W \log_2\left(\frac{P+N}{N}\right) = B_W \log_2\left(1 + \frac{P}{N}\right) \qquad \mathrm{bit\,s^{-1}}. \tag{8}$$

MacKenzie [31] remarked that Fitts and Peterson's [17] formulation contained an "unnecessary deviation from Shannon's Theorem 17" (see also [33]) and that Fitts' index was actually based on an approximation of C for large SNR ($P \gg N$). Adding the one would give the true formula, because in Fitts' law it is not always true that $P \gg N$. Fitts and Peterson [17], however, considered the amplitude of the movement D to be equivalent to the signal *plus* noise power: $P + N$, and half the range of movement variability $W/2$ to be equivalent to noise power N; so in essence their formula also matches Equation 8. Therefore, as it turns out, MacKenzie's amendment boils down to a reformulation of the same idea as Fitts and Peterson's in which movement amplitude is made to correspond to the signal alone, instead of signal plus noise.

Recalling the exposition of Section 2, the analogy seems loose, whether with the Fitts or the MacKenzie version of the index:

(1) the SNR is a ratio of powers, while D/W is a ratio of amplitudes;
(2) there is no justification to identify B_W to $1/\mathrm{MT}$ beyond the fact that both have the same physical units $\mathrm{s^{-1}}$;
(3) the ID is in fact identified with *twice* the capacity $C = (1/2)\log_2\big((P+N)/N\big)$ in bpcu;
(4) most importantly, the channel, as well as the channel's input and output, are left undefined.

Overall, it is not clear how the proposed analogy may actually help tackle the problem of aiming.

3.3 Difficulty as an Entropy Difference

Definition-Proposition 2.2 makes it possible to calculate mutual information as the difference between two entropies. Crossman [10] was the first to use this result to compute what he called the "perceptual load" associated with an aiming task, arguing that "the perceptual load [...] is measured by the difference between initial and final entropy".

In keeping with Shannon's terminology, Crossman used $H(X)$ as the input entropy, and $H(Y)$ as the output entropy, but his formula[10] for information I:

$$I = H(Y) - H(X),$$

is questionable, since mutual information is in fact equal to

$$I(X;Y) = H(Y) - H(Y|X). \tag{9}$$

Now assume that the channel noise, represented as a random variable Z, is added to the channel's input X to yield the output $Y = X + Z$, where the noise Z is independent of X. This is known as an *additive* noise model. In this case we have

$$I(X;Y) = H(Y) - H(X+Z|X) = H(Y) - H(Z|X) = H(Y) - H(Z).$$

Thus information is obtained from the output (endpoint distribution) entropy by substracting the entropy of the *noise* Z, not the entropy of the input signal X like in [10].

Recently, a derivation in the same spirit was given by Hoffmann [26], who considered the difference in entropy between visually-controlled and ballistic movements for a distribution of movement endpoints. Hoffmannn's rationale, reminiscent of Woodworth's [56], was that visual control represents an extra process that must reduce the entropy of the endpoint distribution. Therefore, by taking the entropy difference between visually-controlled and ballistic movements, one should be left with the amount of information needed for the specific aiming process. Hoffman ended up with the following formula for mutual information:

$$I = \log_2(\sqrt{2\pi e}\sigma_b) - \log_2(\sqrt{2\pi e}\sigma_v),$$

where σ_v is proportional to W[11], and $\sigma_b = c + dD$, with the constants c and d to be evaluated empirically. This last relationship comes from ballistic movement theory, where movement time and movement variability are evaluated under a maximum torque condition [19].

The rationale of Hoffmann is the same as Crossman's, except that the start and endpoint entropies are evaluated differently—in both cases what is being evaluated is not information transmission. The problem is that the quantity considered has little to do with the idea of a capacity for transmitting information. In such derivations, Shannon's channel coding theorem has no light to cast.

There is another mismatch between Shannon's theory and ID in Soukoreff and MacKenzie's paper [53] where the entropy difference is considered between the (input) signal and the noise instead of between the output and the noise.

3.4 Soukoreff and MacKenzie's Fundamental Theorem of Human Performance

Soukoreff and MacKenzie [52] have proposed another account of the speed-accuracy tradeoff of rapid aimed movements based on modified information-theoretic inequalities. The main claim of the article is that the classical equation:

$$H(X|Y) = H(X) - I(X;Y) \geq H(X) - \max I(X;Y) = H(X) - C,$$

[10] Formula 3, page 5 in [10]

[11] This relationship was first used by Crossman [10], and supposes that the endpoints follow a Gaussian distribution (see Section 5).

should be accommodated to take into account the human's imperfect nature (Formula 8 in [52]) using a parameter $\alpha \geq 1$:

$$H(X|Y)_h = \alpha[H(X) - I(X;Y)] \geq \alpha[H(X) - \max I(X;Y)] = \alpha[H(X) - C] \tag{10}$$

These author's analysis, however, raises some doubts:

- The validity of the modified equation is evaluated in [52, Figure 5]. Although the maximum equivocation is 4 bit s^{-1}, throughput is extrapolated up to 12 bit s^{-1}. Also, all the points but three are clustered into an isotropic mass: removing the three data points corresponding to maximum equivocation, out of a total of 18 data points, would massively decrease the correlation. What further weakens the empirical analysis is that the data has been acquired by scanning Fitts' article [16], and that Fitts himself never actually described how he would estimate equivocation[12].
- Soukoreff *et al.* [52] treated the speed-accuracy tradeoff in general without tackling Fitts' law, perhaps the most important instance of a speed-accuracy tradeoff.
- Arguably twisting a fundamental information-theoretic formulation by introducing α is unsatisfactory—within an appropriate framework no twists should be needed.

Unlike Soukoreff and MacKenzie [52], we believe that Fitts' law results can in fact be accommodated within the standard information-theoretic approach. The goal of the model to be presented next is precisely to do that.

4 A CHANNEL CAPACITY FOR AN ERRORLESS MODEL OF FITTS' LAW

Many authors have adhered to the view that the human motor system can be modeled as a communication system composed of a source, a transmitter, a channel, a receiver, and a destination (see Section 2). Welford [54] discussed a single channel hypothesis with a structure for the chain of mechanisms involved in sensory-motor performance. More recently Zhai *et al.* [60] (page 106, Figure 2.1) proposed a model for stroke gestures, in which the human intention forms the source of the communication system. Figure 5 displays an adaptation of the stroke gesture's model to the case of pointing.

Message. The information source we consider is the user's intention, as in [60]. Following the "aiming is choosing" rationale, the participants' intention is that of choosing a target, i.e. locating its center. Thus, considering the (centered) partition of Figure 4, the message X takes value in the set $\{-\frac{D}{2}, -\frac{D}{2}+W, \ldots, \frac{D}{2}-W, \frac{D}{2}\}$. As seen in the previous Section, the entropy of the source then reduces to the MacKenzie index of difficulty: $H(X) = \mathrm{ID} = \log_2(1 + D/W)$. The description of the message shows a least-effort strategy for the description of the target: the smaller the targets, the higher the source entropy.

Channel. The message produced under the intention of the participant is encoded and sent through the noisy channel. The noise in the channel is presumably a reflection of the imperfection of neural and musculo-skelettal mechanisms, and should ultimately model movement end-point variability. If we want "target aiming" to become "target hitting", then the noise must have an absolute amplitude less than $\frac{W}{2}$, so that the constraint on the channel is an amplitude constraint, rather than the usual power constraint.

[12] Estimating information-related measures is far from trivial, and is actually a research question on its own.

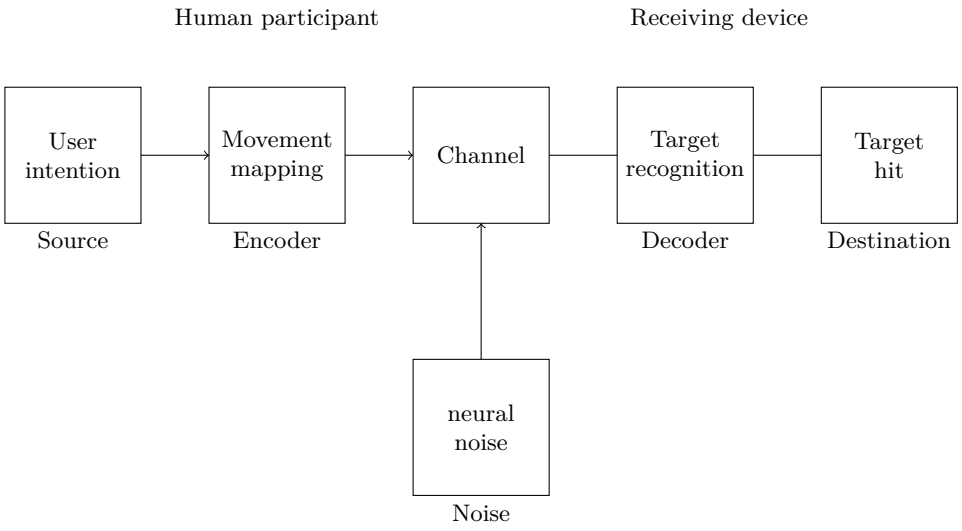

Fig. 5. The human motor system as a communication system

Destination. The receiver simply checks if the right target has been attained. It may be the participant herself, through a visual check (which suggests the possibility of some feedback). The right target may very well be always hit, which ensures errorless communication; this will be the case of our model.

In summary, our model for the aiming task is comprised of a source that represents the "aiming is choosing" paradigm and a limited-amplitude channel that allows the receiver to ensure that the target is never missed. A limited-amplitude channel is described next.

4.1 The Capacity of the Uniform Channel

A limited-amplitude channel was presented by Rioul and Magossi [44] to show that "Hartley's rule"[13] may yield Shannon's capacity theorem. The theorems and proofs of this subsection are directly inspired from this work.

Definition 4.1 (Uniform channel). The aiming task with target distance D and target width W is modeled as a channel with the following properties:

- discrete input: $X \in \{-\frac{D}{2}, -\frac{D}{2}+W, \ldots, \frac{D}{2}-W, \frac{D}{2}\}$
- uniformly distributed additive noise: $Z \in [-\frac{W}{2}, \frac{W}{2}]$
- output: $Y = X + Z$

The uniform channel's input is drawn uniformly in the set of messages relating to the center of the targets coming from the "aiming is choosing" rationale. The entropy of the input X is thus $H(X) = \mathrm{ID} = \log_2(1 + D/W)$. When the message enters the channel an independent noise taking values in $[-W/2, W/2]$ is added to it. Notice that relative to the previous subsection this definition adds the assumption that

[13] Hartley's rule is a formula which shares many similarities with the MacKenzie ID, particularly in the fact that it also involves the logarithm of a ratio of amplitudes, rather than a ratio of powers as in Shannon's capacity formula.

the noise is uniformly distributed (we will return to this assumption in Theorem 4.3). In line with the information-theoretic rationale, we now compute the capacity of this uniform channel:

Fig. 6. The uniform aiming channel under amplitude constraint

THEOREM 4.2. *The capacity C' of the uniform channel under the amplitude constraint $|X| < \frac{D}{2}$ is given by the following expression:*

$$C' = \log_2\left(1 + \frac{D}{W}\right) \qquad \text{bpcu.}$$

The proof can be found in the Appendix. Not only does the MacKenzie ID match the entropy of the target distribution, it also matches the capacity of the channel used in modeling the aiming task. An important result of the proof is that the capacity-achieving input distribution corresponds exactly to the uniform channel's input, meaning that no channel coding is required: sending messages from the source directly over the channel is optimal! What then distinguishes good from poor performances is bandwidth only, some participants having higher bandwidths than others. Theorem 4.2 also implies that:

$$C' = \max_{p(x)} H(X) - H(X|Y) = \max H(X), \tag{11}$$

meaning that in the optimal scheme, there is no information lost in the channel since $H(X|Y) = 0$. The choice of a uniform noise is motivated by the following bound:

THEOREM 4.3. *The capacity C_n of any limited-amplitude additive noise channel is lower bounded by the capacity of the uniform channel: $C_n \geq C'$.*

The proof is omitted here because it can be easily adapted from [44]. The argument used is that the uniform noise maximizes entropy under amplitude constraint, so that uniform noise is essentially a worst-case scenario. Any scheme, where noise is limited in amplitude to $[-W/2, W/2]$, whatever its distribution, will have higher capacity $C_n \geq C'$.

Since the MacKenzie's index involves a ratio of amplitudes D/W rather than a ratio of powers P/N, it is appropriate to compute it in terms of powers to further the analogy. The surprising result is that the ID is mathematically equivalent to Shannon's Capacity. This is expressed in the next theorem.

THEOREM 4.4. *Let $C = (1/2)\log_2(1 + P/N)$ denote the Shannon's capacity and $C' = \log_2(1 + D/W)$ denote the capacity for the uniform channel, then:*

$$C' = C \qquad \text{bpcu.}$$

The proof can be found in the Appendix. For this particular channel the index of difficulty and the Shannon capacity truly coincide, legitimizing the analogy with Shannon's Theorem 17!

4.2 A Remark on the Equivalence Between Indices

As we have shown above, $C = C'$ in bpcu is the amount of informational bits that can be sent per sample. We can define $\mathrm{ID} = \log_2(1 + D/W) = (1/2)\log_2(1 + \mathrm{SNR})$ but there are many other equivalent choices:

Proposition 4.5. *Any index of difficulty* ID *which is linearly related to* $\log_2(1 + D/W)$ *satisfies Fitts' law, in the sense that the relationship between* MT *and* ID *is linear.*

The proof is obvious: if $\mathrm{ID} = \alpha + \beta\log_2(1 + D/W)$ then $\mathrm{MT} = a + b\log_2(1 + D/W) = a' + b'\mathrm{ID}$ where $a' = a - b\alpha/\beta$ and $b' = b/\beta$. In fact the same argument shows that any two linearly related ID are equivalent:

Corollary 4.6. *Suppose that we have two* ID*'s such that* $\mathrm{ID}_1 = \alpha + \beta \cdot \mathrm{ID}_2$. *Then both will be equivalent in the sense of Fitts' law.*

Indeed, from Proposition 4.5, we will get $\mathrm{MT} = a_1 + b_1\mathrm{ID}_1 = a_1 + \alpha b_1 + \beta b_1\mathrm{ID}_2 = a_2 + b_2\mathrm{ID}_2$. Because both constants have to be measured from experimental data points, both indices are equivalent.

For example, Fitts' index [15] $\mathrm{ID} = \log_2(2D/W) = 1 + \log_2(D/W)$ is equivalent to Crossman's index [10] $\mathrm{ID} = \log_2(D/W)$. Also, the "mixed" Fitts-MacKenzie's expression $\mathrm{ID} = \log_2(1 + 2D/W)$ is equivalent to Welford's index since $\log_2(1 + 2D/W) = 1 + \log_2(1/2 + D/W)$.

As another illustration, Soukoreff *et al.* [53] proposed a novel formulation for ID:

$$I_D \text{ entropy} = m + \log_2(U) - \frac{1}{2}\log_2\left(\pi e\frac{W^2}{8}\right) + 1, \tag{12}$$

where U is the "size of the movement universe", i.e. the largest extent considered for movements. Grouping the logarithms together, we obtain

$$I_D \text{ entropy} = m + 1 + \log_2\left(\frac{2U}{W}\sqrt{\frac{2}{\pi e}}\right) = m + 1 + \log_2\left(2\sqrt{\frac{2}{\pi e}}\right) + \log_2\left(\frac{U}{W}\right). \tag{13}$$

Now considering the largest extent to be either D, $D + \frac{W}{2}$ or $D + W$, one recovers the respective indices of difficulty of Fitts, Welford and Mackenzie.

4.3 A Proper Analogy

Equipped with the above results, we are now able to formulate a proper analogy from Shannon's capacity formula rearranged in the following manner:

$$C = B_W \cdot \log_2(1 + \mathrm{SNR}) = 2B_W \cdot \frac{1}{2}\log_2(1 + \mathrm{SNR}) \qquad \text{bit}\,\text{s}^{-1}.$$

By theorems 4.2 and 4.4, we can now identify $(1/2)\log_2(1 + \mathrm{SNR})$ by MacKenzie's index of difficulty $\mathrm{ID} = \log_2(1 + D/W)$. Also, by virtue of the Shannon-Nyquist sampling theorem, $2B_W$ refers to the maximum number of samples that are sent per second which can be identified to $\frac{1}{\mathrm{MT}}$ as we effectively send one sample during MT seconds. We thus obtain Fitts' Formula [15]:

$$\mathrm{MT} = \frac{1}{C}\mathrm{ID},$$

but without intercept. Interestingly, Fitts did not refer to an intercept in his 1954 article. He introduced it later to make the model more flexible for experimental data. The interpretation of the intercept has been debated many times (e.g. [51, 59]). Although our formula seems to comfort those who believe the intercept

reflects the non-informational part of pointing [59], intercepts can also arise between two equivalent indices (see Corollary 4.6).

5 COMPUTING CAPACITY IN THE PRESENCE OF TARGET MISSES

In a Fitts' law experiments the outcome of a pointing act can be either measured as an error, i.e., a distance from end-point to target center, or categorized in an all-or-none way as a hit versus a miss. Information theory offers a useful distinction between transmission *errors* (the received symbol is wrong) and *erasures* (the received symbol is empty), see Section 2. This distinction seems to have escaped attention so far in HCI, where it has been a solid tradition, since MacKenzie [32], to measure movement endpoints from the center of the target and, assuming that the distributions of these measures is normal, to compute an effective index of difficulty $\mathrm{ID_e}$.

The goal of a Fitts' law experiment being to observe and study the speed-accuracy tradeoff, the choice of the metrics used to measure *speed* and *accuracy* is critical. While there has been unanimous agreement in the literature that movement times provide a satisfactory measure of speed, the measure of accuracy has been controversial from the outset [10]. Only recently was the adjustment for target misses standardized by ISO [1, 51], through the effective index of difficulty $\mathrm{ID_e}$. Unfortunately, as we will show below, the standardized method is not compliant with information theory.

Our uniform channel model predicts a null error rate, and is therefore sufficient as a description in a paradigm that does not allow mistakes, such as Fitts' pin and disc transfer experiments [15, Experiments 2 and 3]. However, in the majority of Fitts' law experiments target misses do occur, and so an extension of the model is needed.

There are three different ways of handling mistakes:

- *Ignoring the mistakes.* Fitts, who did not measure actual amplitudes, classified the movements in a dichotomous way as hits and misses. Although he did tabulate the (variable) error rates he obtained in his stylus-pointing experiments, he felt in a position to leave them aside because of the "small incidence" of target misses [15, p. 265].
- *Taking the error rate into account.* To our knowledge, Crossman [10] was the first to try to incorporate the error rate information into his ID measure, leveraging the standard Gaussian distribution model.
- *Taking the spread of endpoints into account.* This is the standardized way of measuring accuracy in Fitts' law [1, 51]. Recourse to the standard deviation as a measure of accuracy has the implication that the magnitude of the metrical error (the distance from target center) matters in the upcoming analysis: regardless of whether the outcome is a hit or a miss, the farther the endpoint from target center, the worse the performance. It also implies that there is equivalence between two movements hitting the target if and only if they end up at exactly the same distance from the center of the target (which, strictly speaking, never happens).

The ISO standard and Fitts' law literature in general treats pointing mistakes as *errors*, by referring to the standard deviation of the endpoints distribution—either by direct estimation or through a calculation from error rates. Thus in the *error* concept, the accuracy depends on the (continuous) distance between the movement endpoint and the target center.

This approach is not quite consistent with the all-or-none logic of Fitts' experimental paradigm: in an experiment that asks participants to *hit* the target 96 percent (or so) of the time, all movements that end up inside the W interval should be recognized as equivalent from the point of view of accuracy. The same equivalence is true in real-world interfaces: what matters is not precisely *where* the click takes place, but rather *whether or not* the click falls in the intended area. This corresponds to the information-theoretic concept of *erasures* described in Section 2.

Thus there is a conceptual mismatch between the standardized measurement of accuracy and the reality of the pointing task in both controlled experiments and real-world target acquisition tasks. But the established computation of $\mathrm{ID_e}$ suffers from other deficiencies.

5.1 Information-Theoretic Critique of $\mathrm{ID_e}$

The effective width $\mathrm{ID_e}$ is defined as $\log_2(1+\overline{D}/W_e)$, where $\overline{D}$ corresponds to the average covered distance, and W_e is the effective width (to be detailed just below). It is used as a replacement to ID in the movement time equation (Equation 2). The computation of effective width is detailed in [51]. Let σ denote the standard deviation of the end-point distribution, and ε the error rate, i.e., the proportion of target misses:

- If σ is available:
$$W_e = 4.133\sigma. \tag{14}$$
- Otherwise:
$$W_e = \begin{cases} W \cdot \frac{2.066}{z(1-\varepsilon/2)} & \text{if } \varepsilon > 0.0049\% \\ 0.5089 \cdot W & \text{otherwise.} \end{cases} \tag{15}$$

The received justification is as follows [32, Section 2]:

> "The entropy (H), or information, in a normal distribution is $H = \log_2\left((2\pi e)^{\frac{1}{2}}\sigma\right) = \log_2(4.133\sigma)$, where σ is the standard deviation in the unit of measurement. Splitting the constant 4.133 into a pair of z-scores for the unit-normal curve (i.e., $\sigma = 1$), we find that the area bounded by $z = \pm 2.066$ represents about 96 % of the total area of the distribution. In other words, a condition that target width is analogous to the information-theoretic concept of noise is that 96 % of the hits are within the target and 4 % of the hits miss the target [...]. When an error rate other than 4% is observed, target width should be adjusted to form the effective target width in keeping with the underlying theory."

This methodology raises three issues:

(1) The computation of W_e as 4.133σ as well as the computation leading to Equation 15 presumes a Gaussian distribution of endpoints [51]. This is somewhat unsafe as the validity of this hypothesis has been questioned empirically (e.g., [15] [55, Discussion]).

(2) To our knowledge Information Theory provides no justification to the relation $W_e = 4.133\sigma$. When Crossman [10] calculated the expression for W_e from the area under the standard normal curve, he took the 5% value as an arbitrary "permissible" error rate. MacKenzie [32] noticed that by changing the arbitrary rate from 5% to 3.88% (approximately 4%), the entropy of the rectangular distribution of width W_e would equal the entropy of the Gaussian distribution of standard deviation σ (see

Appendix), but this is no more than a coincidence: we can see no information-theoretic reason to equalize these two entropies.[14].

(3) The threshold of error rate placed at 0.0049% (Equation 15) is arbitrary. Even with a Gaussian distribution of endpoints, the one-to-one relationship between standard deviations and error rates is only true for strictly positive error rates. Indeed, when the error rate vanishes, so does the standard deviation, and so ID_e tends to infinity. To prevent this from happening, Soukoreff and MacKenzie [51] have recommended that below a certain error rate (0.0049%), ID_e should be kept constant. The justification of the threshold error rate of 0.0049% is that it "rounds to 0.00". As shown below, the existence of such a threshold and its value of 0.0049% is in fact adverse to the theory.

The standardized index of difficulty ID_e is thus questionable. It relies on the unsafe Gaussian hypothesis, two arbitrary constants, and one coincidence. Even more importantly, it has never been shown to be the correct expression of the capacity of a human-motor channel—the expected rationale behind Fitts' law if one chooses the information-theoretical framework.

We now propose a new effective index $ID(\varepsilon)$ that is compliant with Fitts' experimental design, does not rely on the Gaussian hypothesis and is justified theoretically as a channel capacity, through an extension to the model of Section 4.

5.2 A Compliant Index of Difficulty: $ID(\varepsilon)$

As noted above, treating target misses as transmission errors is not adapted to Fitts' paradigm—these events should rather be viewed as erasures. In fact, the design of the experiment entails a binary decision: there is a target and the movement either finishes inside (a hit) or outside (a miss). This is consistent with the instruction "try to hit the target" as opposed to "try to hit the center of the target". We now extend the model that does not allow or account for mistakes of the previous section with a channel which allows erasures.

Consider a channel that oscillates randomly between a good (G) state and a bad (B) state, with probability ε of being in state B and probability $1-\varepsilon$ of being in state G. When the channel is in its good state, it corresponds to the channel of capacity $\log_2\left(1+\frac{D}{W}\right)$ that we derived in Section 4, which we will refer to as the *Fitts channel*. However, when the channel is in its bad state it can only produce erasures—we call it an *erasure channel*. In Information Theory this configuration (Figure 7) is known as a *compound* channel [20].

Let us now evaluate the Shannon capacity of this compound channel. This will serve as a common ground to compare the performance of different participants operating at different accuracy levels (with different values of ε). The channel capacity corresponds to the maximum transmission rate that the participants would have achieved with an arbitrarily small error rate (refer to the Channel coding theorem of Section 2). We thus adjust the rate, to obtain the one that the participants would have had, had they never missed the target. Shannon's capacity of the compound channel of Figure 7 is given by the following theorem.

[14]Incidentally, these entropies can both be negative. Information Theory distinguishes the (discrete) entropy of a discrete random variable, which is non-negative and serves as a measure of information, and the (so-called *differential*) entropy of a continuous random variable such as a normal random variable, which is positive for large variances and negative for small variances and thus cannot be interpreted as a measure of information [9].

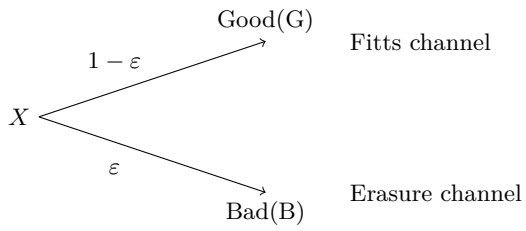

Fig. 7. Compound channel for an aiming task with target misses.

Theorem 5.1 (Compound Channel Capacity). *Consider a compound channel as in Figure 7, with probability ε of being in state B and probability $1-\varepsilon$ of being in state G. The capacity of such a channel is given by*

$$C = (1-\varepsilon)\log_2\left(1+\frac{D}{W}\right).$$

As expected, the obtained capacity is lower than the capacity $\log_2\left(1+\frac{D}{W}\right)$ that would have been achieved with 100% hitting success ($\varepsilon = 0$).

The formal information-theoretic proof is known [9] and summarized in the Appendix for completeness, but it is easy to sketch the reasoning: The participant is effectively time sharing both channels. With Fitts' channel, the transmitted information is $\log_2\left(1+\frac{D}{W}\right)$ bits and with the erasure channel the transmitted information is 0 bit, so that, on average, $C = (1-\varepsilon)\times\log_2\left(1+\frac{D}{W}\right)+\varepsilon\times 0$. In line with Fitts' parallel between capacity and ID, our new effective index is

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

where ε is no other than the traditional 'error rate' more cautiously designated here as the percentage of target misses.

5.3 Comparing the Two Indices

We now provide an analytical comparison of $\mathrm{ID_e}$ and $\mathrm{ID}(\varepsilon)$. The behavior of the standardized $\mathrm{ID_e}$ for vanishing error rates is problematic. The inverse Gauss error function[15] (see Appendix) $\mathrm{erf}^{-1}(1-\varepsilon)$ tends to $+\infty$ as ε vanishes, so that we should normally have

$$\lim_{\varepsilon\to 0}\mathrm{ID_e} = \infty.$$

Due to the 0.0049% bounding, however, instead we obtain

$$\begin{aligned}\lim_{\varepsilon\to 0}\mathrm{ID_e} &= \log_2\left(1+\frac{D}{0.5089W}\right)\simeq\log_2\left(1+\frac{2D}{W}\right)\\ &= 1+\log_2\left(\frac{1}{2}+\frac{D}{W}\right),\end{aligned}$$

[15]The inverse Gauss error function erf^{-1} has the following relation to the z-score: $z(x) = \sqrt{2}\,\mathrm{erf}^{-1}(2x-1)$

which is equivalent to the Welford index of difficulty [54], by direct application of corollary 4.6. The arbitrary choice to bound the index at the 0.0049% rate results in the index coincidentally tending to the Welford ID, not the MacKenzie ID. Thus there is no continuity[16] as epsilon approaches zero for ID_e.

In contrast, $ID(\varepsilon)$ *does* have the property of continuity towards zero since obviously $ID(0) = ID$.

Figure 8 shows the two indices $ID(\varepsilon)$, ID_e as well as the unbounded u-ID_e (for which the 0.0049% distinction is not made) for $D/W = 15$ as a function of ε in the interval $[0 - 1]$. The difference $ID_e - ID(\varepsilon)$ between ID_e and $ID(\varepsilon)$ is lowest around $\varepsilon = 0.1$. With higher values of ε, the difference increases but such high errors rates are not common. However, for very small values of ε, $ID(\varepsilon)$ can be up to 1 bit smaller than ID_e. Thus the difference between $ID(\varepsilon)$ and ID_e can be non-negligible for very careful participants or in conditions with a high emphasis on accuracy.

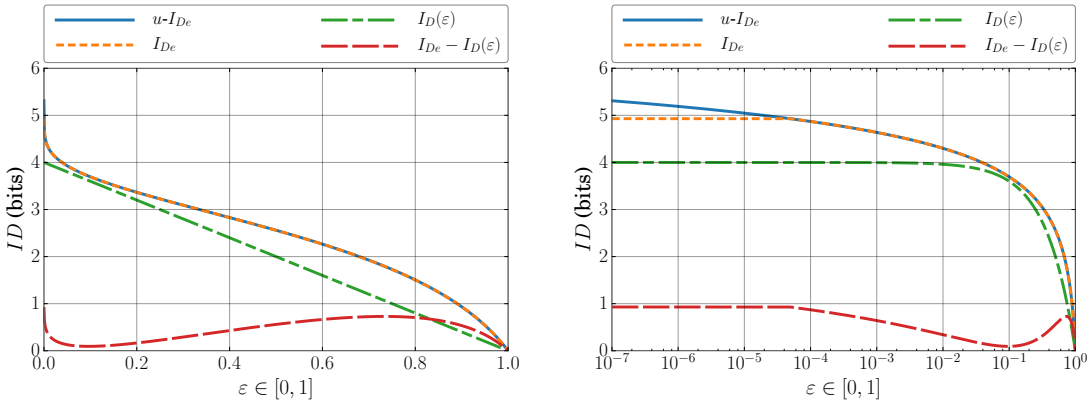

Fig. 8. Comparison of $ID(\varepsilon)$ and ID_e for erasure rate in $[0, 1]$, for $\frac{D}{W} = 15$. u-ID_e refers to ID_e where the 0.0049% distinction is not made. The scale is lin-lin left and lin-log right.

6 PERFORMANCE FRONTS FOR FITTS' LAW

Fitts' law has always been considered as a law of average performance. Although the notation does not make it explicit, MT, the dependent variable of Equation 2, typically denotes the *mean* of samples of movement time measures. Soukoreff and MacKenzie [51] state that " Each condition must be presented [...] many [...] times, so that the *central tendency* of each subject's performance [...] can be ascertained".[17]

Researchers have "agreed to disagree" on many issues of Fitts' law, e.g. on which formulation for the index of difficulty to use, on how to account for errors, and on the meaning of the intercept. However, almost all Fitts' law students have apparently agreed on recourse to *linear regression* to describe the relation between ID and MT. That technique provides both an estimate of parameters a (intercept) and b (slope) and a measure of goodness of fit, through the r-squared coefficient, in a very simple and rapid manner.

Likewise, the ISO standard's throughput must be computed as a mean of means. Its identification to a channel capacity seems, therefore, problematic since the channel capacity concept has nothing at all to

[16] Not in the sense of a mathematical continuity, but in the sense that there is switch from one index to an other.
[17] emphasis added.

do with average information transmission performance: only the *best* transmission schemes are capacity achieving.

In this Section, we build on recent work by Guiard and colleagues [23, 24], who challenged the common view that Fitts' law characterizes average movement time. These authors put forward the view that *only* the best movement times can serve to infer Fitts' law.

6.1 Fitts' Law as a Performance Limit

We see two reasons for which Fitts' law should be viewed as a performance limit rather than a law of average performance.

(1) Fitts' information-theoretic rationale for aiming considers the transmission of information about the target through a human motor channel, and as we have shown Fitts' law can be derived by computing the *capacity* of this channel, which is a theoretical upper bound—the maximum amount of information that can be transmitted reliably—and which is accordingly calculated as an extremum through the Channel Coding Theorem—the maximum of mutual information over all input distributions. Thus, only movements that maximize transmitted information should be relevant for the derivation of Fitts' law, i.e. those movements that for a fixed ID achieve the lowest MT, or conversely those that for a fixed MT achieve the highest ID;

(2) Guiard and Rioul [24] have shown that three paradigms, the time minimisation paradigm of Fitts [15], the spread minimisation paradigm of Schmidt *et al.* [47], and the dual minimisation paradigm of Guiard *et al.* [23] can receive a unified account provided that the participants are assumed to invest less than 100 percent of their resource in their performance.[18] Accordingly, only the best performing samples should be expected to describe the speed-accuracy tradeoff, and Guiard and Rioul could then merge the linear law of Schmidt *et al.* [47] and the logarithmic description of Fitts' law as different regions of the same speed-accuracy tradeoff function.

To understand the constraints on movement, one should consider the movements that are most constrained: one can only hope to model what can be modeled. In the real world, movements are weakly constrained, if not at all. One rarely tries to point as fast and as accurately as possible. Even in a controlled experiment, the participants' attention fluctuates. As we will now demonstrate, the front of performance is the most natural technique to reveal Fitts' law.

6.2 A Field Study Example

An example will help illustrate the front of performance approach. The data come from a pointing study run "in the wild" by Chapuis *et al.* [8]. While delivering very large data sets, field experiments (as opposed to controlled experiments) provide a beneficial magnification of the fact that not all resources are invested by participants for each movement.

For several months Chapuis *et al.* [8] unobtrusively logged cursor motion from several participants using their own hardware. The authors were able to identify offline the start and end of movements as well as the target information, for several hundreds of thousands of click-terminated movements. With this information,

[18] Not only does the less-than-total resource investment assumption match common sense, it matches the information-theoretic concept of capacity. The capacity is reached at the limit of a (perfectly) optimal coding scheme, channel bandwidth being exploited in full. Anything less will give lower transmitted information.

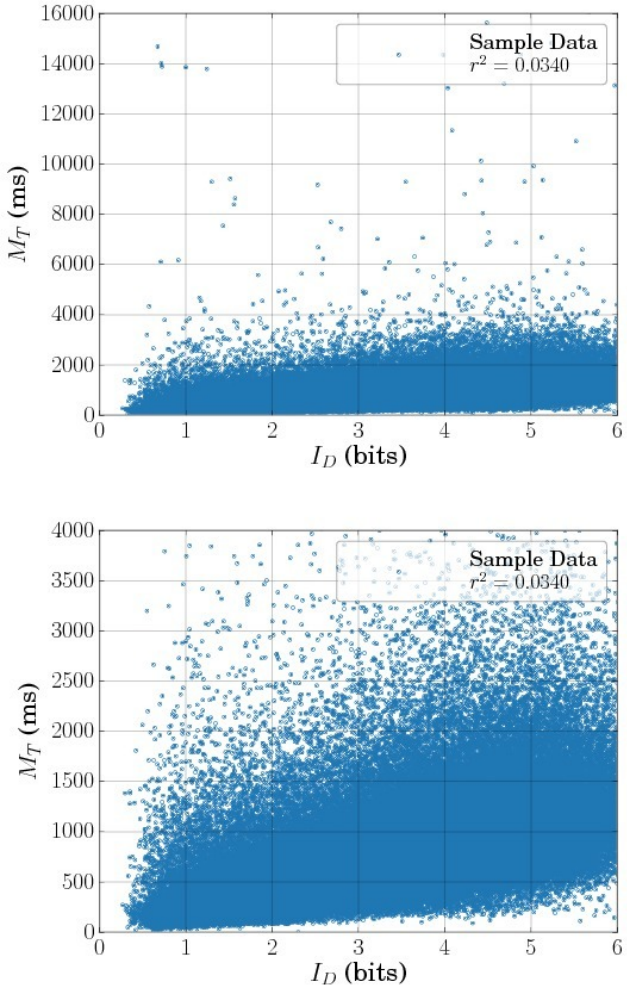

Fig. 9. Movement time as a function of task difficulty in one representative participant of Chapuis *et al.* [8]. Shown are over 90,000 individual movement measures. *Above*: MT up to 16 s. *Below*: MT up to 4 s. Cut-offs are here arbitrary but necessary as some movement times lasted minutes.

.

one can then represent the movements in a MT versus ID graph, as normally done in a controlled Fitts law study. To compute task difficulty in the 2D space of computer screens they followed the suggestion of MacKenzie and Buxton [34] and chose

$$\mathrm{ID} = \log_2\left(1 + \frac{D}{\min(H, W)}\right),$$

where H and W are the height and width of the target, respectively. Whenever an item was clicked , it was considered the target, meaning the rate of target misses was 0 percent and hence $\mathrm{ID}(\varepsilon) = \mathrm{ID}(0) = \mathrm{ID}$.

Figure 9 shows the data from one representative participant (P3) of Chapuis *et al.* [8]. The ID axis is truncated at 6 bits because beyond that level of difficulty the density of data points dropped dramatically. Obviously, the data obtained with no speeding instructions (and no experimenter to recall them) exhibits a huge amount of stochastic variability along both dimensions of the plot. While in the X dimension, most ID values fell in the range from 0.5 to 6 bits (presumably a reflection of the geometric composition of the graphical user interface), the variability along the Y dimension is extremely high. Judging by linear regression on this raw data, we find that movement time and the index of difficulty are essentially *uncorrelated* since the r-squared coefficient is very close to 0 ($r^2 = 0.034$). Thus, at first sight, this data fails to confirm Fitts' law, but it is important to realize that this first impression is quite false.

In the lower panel of Figure 9 which ignores all MT data above 4 s and thus zooms-in on the Y-axis towards the bottom of the plot, one can distinctly see that the bottom edge of the cloud of data points does not touch the X axis. Rather, in the downward direction, the density drops sharply: no matter the ID region considered, the distribution of performance measures has an unending *tail* above and what we call a *front* below, the latter being very steep in comparison with the former. The unending tail is understandable as "it is always possible to do worse" [24]. In contrast, the movement time cannot be reduced below a certain strictly positive critical value which accurately defines the front.

A closer look at the lower panel of Figure 9 reveals that the bottom edge of the scatter plot is approximately *linear*: this linear edge is what justifies Fitts' law. In other words, the empirical regularity in Fitts' law is, in essence, a front of performance, a lower bound that cannot be passed by human performance. Such a front of performance is observable in data from the field study of Chapuis *et al.* [8] because unsupervised everyday pointing does offer, albeit in a minority of cases, opportunities to perform with high levels of speed and accuracy. The difference between a field and a controlled experiment is thus one of degree, not of nature. Experimenters have recourse to pressurizing speed/accuracy instructions simply to get rid of endless, uninformative, tails in their distributions of MT measures.

Figure 10 shows the same plot with the Y axis zoomed-in further so that the range of MT measures approaches that commonly obtained in a typical controlled experiment. Even though the front edge is incomparably sharper than the tail edge, the zoomed-in view of Figure 10 reveals a number of presumable *fast* outliers. Many reasons may explain why a small proportion of data points "cross" the frontier, seemingly violating the theoretical lower bound. Some data points may just correspond to unreasonably fast but lucky movements, others to failures of the analysis software, which may have wrongly classified as target-directed movements which terminated with accidental clicks. Yet another possibility is that targets lying at the edge of the screen can be aimed at with a purely ballistic throw of the cursor which will remain on that edge. An empirical scatter plot will never exhibit a perfectly neat front of performance, and so an estimation procedure is still needed to actually estimate the front.

We devised a heuristic method to fit a straight line to the bottom of the edge of the scatter plot, robust enough to accommodate the imperfectness of the front. Figure 10 shows the resulting front fit, in red at the bottom edge of the plot. The obtained line is independent of *slow* outliers. In contrast, linear regression lines obtained with different threshold levels $[2s, 3s, ..., 9s, 10s]$ for outlier rejection (in white in Figure 10) show that they are highly dependent on the threshold level.

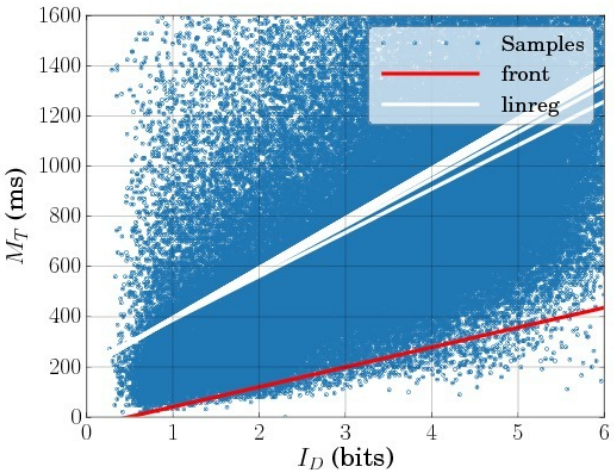

Fig. 10. Same data as in Figure 9, with the Y-axis cut at 1.6 s. Shown are linear fits from usual linear regression ('linreg' in white), using a number of different thresholds for the exclusion of outliers, as well as an estimation of the front of performance (in red).

Thus, an interesting characteristic of the front of performance approach is that it dispenses one with the difficult task of handling *slow* outliers, whose removal requires arbitrary choices. For example, some experimenters remove values k standard deviations away from the sample mean, k being typically chosen between 2 and 3. Some simply trim the data, by removing all samples above a certain limit, say $\mathrm{MT} > 2s$. One issue here is that the tolerance for outliers is variable across the ID scale. As illustrated Figure 10, the fit computed by linear regression highly depends on the arbitrary choice of tolerance. In contrast, the front of performance *by definition* will not depend on slow outliers at all, and in this sense it is far more robust.

Of course, the red line is quite different from the white lines in Figure 10: Characterizing Fitts' law by best rather than average performances is not a minor adjustment. Even though experimenters do their best to reduce the inherent variability of human aimed movement, a typical sample of measures exhibits quite large dispersion. The common practice of considering averages per block, rather than raw measures, reduces this dispersion artificially. This practice does not eliminate the fact that because of movement variability, the quantitative difference between average fit and best-performance fit is substantial.

7 CONCLUSION

Shannon's channel coding theorem expresses the best compromise between the rate of transmission and the probability of errors. As such it seems well fitted to the analysis of the speed-accuracy tradeoff. But as we have shown, existing information-theoretic derivations are compromised by the same long-standing misunderstanding: Channel capacity results were thought of as results on information generation, rather than on information transmission as they should be. Fitts originally derived ID as a source entropy by analogy with Hick's paradigm [14]. As we have shown, Fitts' original idea can be used to derive several known indices of difficulty (such as MacKenzie's) for different target layouts. One year later, he derived the

same ID by analogy with Shannon's capacity [15] but never precisely identified the channel or the noise. More recent attempts to use information bits were flawed because the estimate of transmitted information did not match the correct mathematical definition of mutual information. As as consequence, until now, the information-theoretic rationale for pointing tasks is judged "more metaphorical than mathematical" [59].

We have proposed a formal, detailed information-theoretic model for Fitts' pointing task in order to better understand Fitts' law in the light of Shannon's information theory. We showed that to ensure a correct execution of the aiming task, the amplitude of the channel noise should be limited to half the target's width. This rigorously defines a transmission channel for which its capacity turns out to be exactly equal to MacKenzie's ID, thus legitimizing its use. In addition, that index of difficulty truly coincides with the celebrated Shannon's capacity formula, which legitimizes the analogy with Shannon's Theorem 17.

In order to account for possible participant mistakes, we generalized our model to the case where targets can be missed. Target misses, as opposed to errors, correspond to the correct information-theoretic notion of erasure applied to Fitts' paradigm. The channel becomes a compound channel with erasures whose capacity is the modified index $(1-\varepsilon)$ID. We showed that this new index is not only more rigorous, it is also theoretically safer than the ISO index $\mathrm{ID_e}$ as it does not presuppose a Gaussian distribution of endpoints. It is also more convenient in practice since it allows the researcher to dispense with an arbitrary treatment of the 0 percent miss case and it is also much simpler to compute than the traditional $\mathrm{ID_e}$.

Finally, we argue that by its very definition, capacity is a law of extreme performance. This precludes any use of linear regression to estimate Fitts' law since regression is just an averaging method. Many experimenters claim to measure extreme performance, but end up reporting average performance only. Simply telling participants to do their best is not enough to ensure high-fidelity data.

We hope that this theoretical work and information-theoretic tools would eventually prove useful to the HCI researcher. In particular, we believe that the notion of front of performance is a promising tool and not just a theoretical curiosity, as long as efficient estimation methods can be found. In this direction, more work should be done to find a reliable method for fitting the front of performance (we have used a heuristic method).

APPENDIX

A.1 Proof of Theorem 4.2

The proof is based on the following lemma.

LEMMA A.1. *Consider an additive noise channel with input X, noise Z and output $Y = X+Z$. If the output is uniformly distributed in $[-(D+W)/2, (D+W)/2]$ and the noise is uniformly distributed in $[-W/2, W/2]$, then the input must be a uniform discrete random variable in the set $\{-D/2, -D/2+W, \ldots, D/2-W, D/2\}$.*

A rigorous proof can be found in [44]. A proof sketch is as follows. The probability density function p_Y of the sum of two independent random variables $Y = X + Z$ is a convolution product $p_X * p_Z$. If X is discrete uniform with $n = 1 + D/W$ values equally spaced by W, and Z is uniformly distributed in $[\frac{-W}{2}, \frac{W}{2}]$ (a rectangle of width W), then their convolution product if the juxtaposition of n rectangles W apart, which is a larger rectangle of width $D + W$.

Proof of Theorem 4.2. We can expand the mutual information $I(X;Y)$ as difference of differential entropies:

$$I(X;Y) = H(Y) - H(Y|X), \tag{16}$$
$$= H(Y) - H(X+Z|X), \tag{17}$$
$$= H(Y) - H(Z) \tag{18}$$
$$= H(Y) - \log_2(W), \tag{19}$$

where (16) is by definition of mutual information, (17) is by additivity of the channel, (18) is by independence of X and Z, and (19) is from the computation of differential entropy for a continuous uniform random variable. Maximizing $I(X;Y)$ is thus equivalent to maximizing $H(Y)$. Because $|X| \leq \frac{D}{2}$ and $|Z| \leq \frac{W}{2}$, we have that $|Y| = |X+Z| \leq |X| + |Z| \leq \frac{D+W}{2}$ by the triangular inequality. The maximum

$$C' = \max_{|X| \leq \frac{D}{2}} I(X,Y) = \max_{|Y| \leq \frac{D+W}{2}} H(Y) - \log_2 W.$$

will be attained when Y is uniformly distributed in $[-(D+W)/2, (D+W)/2]$ and from Lemma A.1, this is obtained when X is discrete uniform in the set: $\{-D/2, -D/2+W, \ldots, D/2-W, D/2\}$. It follows that

$$C' = \log_2(D+W) - \log_2(W) = \log_2\left(1 + \frac{D}{W}\right) \qquad \text{bpcu.}$$

as claimed. □

A.2 Proof of Theorem 4.4

Proof of Theorem 4.4. Let M be the cardinality of the set $\{-D/2, -D/2+W, \ldots, D/2-W, D/2\}$:

$$M = 1 + \frac{D}{W}.$$

The channel input's average power is:

$$P = \mathbb{E}(X^2) = \frac{1}{M}\sum_{k=0}^{M-1}\left(\frac{M-1}{2} - k\right)^2 W^2 = \frac{1}{M}\, 2W^2 \sum_{k=0}^{\frac{M-1}{2}} k^2 = \frac{M^2-1}{12} W^2$$

where we have used the well-known formula for the sum of consecutive squares. The noise power N of the uniformly distributed distribution in $[-W/2, W/2]$ is

$$N = \frac{W^2}{12}.$$

It follows that

$$C' = \log_2 M = \frac{1}{2}\log_2 M^2 = \frac{1}{2}\log_2(1 + M^2 - 1) = \frac{1}{2}\log_2\left(1 + \frac{P}{N}\right) = C.$$

as claimed. □

This coincidence can be explained quite easily when noticing that

$$1 + \text{SNR} = \frac{P+N}{N}$$

is the ratio between the power of the output Y over the power of the noise Z. In our case, both output and noise are uniformly distributed, the power is proportional to the square of the range of the distribution, so that

$$\frac{P+N}{N} = \left(\frac{D+W}{W}\right)^2,$$

Taking the logarithm gives $C' = C$.

A.3 Calculation of W_e

Consider the random variable for the endpoint location Y, such that $Y \sim \mathcal{N}(0, \sigma^2)$ and a target of width W. The event $|Y| > W/2$ defines an error. Width W and error rate $\varepsilon > 0$ are related through the following one-to-one relation

$$\varepsilon = 1 - 2\int_0^{\frac{W}{2}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt = 1 - \operatorname{erf}\left(\frac{W}{2\sqrt{2}\sigma}\right)$$

or

$$W = 2\sqrt{2}\sigma \operatorname{erf}^{-1}(1-\varepsilon),$$

where erf is the gaussian error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

These formulas are consistent with ISO's recommendations: taking $W = 4.133\sigma$, we find

$$\varepsilon = 1 - \operatorname{erf}\left(\frac{2.066}{\sqrt{2}}\right) = 3.88\%.$$

The multiplicative constant α such that $W_e = \alpha W$, where $W_e = 4.133\sigma$ is the width such that the error rate is 3.88%, is given by

$$\alpha = \frac{W_e}{W} = \frac{4.133\sigma}{2\sqrt{2}\sigma \operatorname{erf}^{-1}(1-\varepsilon)} = \frac{2.066}{\sqrt{2}\operatorname{erf}^{-1}(1-\varepsilon)}.$$

so that W_e is be given by the following formula:

$$W_e = \alpha W = \frac{2.066}{\sqrt{2}\operatorname{erf}^{-1}(1-\varepsilon)} W.$$

To compare this to the ISO recommendation, consider the z-score related to the area under a $\mathcal{N}(0,1)$ distribution by

$$\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = x \iff 1 - Q(z) = x \iff z = Q^{-1}(1-x)$$

where Q is Marcum's Q-function and Q^{-1} the inverse Q function. The inverse Q-function can be easily related to the inverse error function

$$Q^{-1}(y) = \sqrt{2}\operatorname{erf}^{-1}(1-2y),$$

and by replacing y by $1-x$, we find that

$$z = \sqrt{2}\,\mathrm{erf}^{-1}(2x-1).$$

Replacing x by $1-\frac{\varepsilon}{2}$ gives the final result:

$$W_e = W\frac{2.066}{z(1-\frac{\varepsilon}{2})} = \frac{2.066}{\sqrt{2}\,\mathrm{erf}^{-1}(1-\varepsilon)}W.$$

A.4 Proof of Theorem 5.1

Proof of Theorem 5.1. Since the only way to produce an erasure symbol is for the channel to be in state B, we have $I(X;Y) = I(X;(Y,E))$. This can be expanded as [9]

$$\begin{aligned} I(X;(Y,E)) &= \mathbb{P}(E=G)I(X;Y|E=G) + \mathbb{P}(E=B)I(X;Y|E=B) \\ &= (1-\varepsilon)\,I(X;Y|E=G) + \varepsilon I(X;Y|E=B). \end{aligned}$$

Here $I(X;Y|E=G)$ is the mutual information computed for the uniform channel, and $I(X;Y|E=B)=0$ because if the channel is in bad state, only an erasure can come out of the channel. Therefore the distribution that maximises the mutual information for the compound channel is the same than the one that maximises mutual information for the uniform channel. □

ACKNOWLEDGMENTS

The authors would like to thank Michel Beaudouin-Lafon and Olivier Chapuis for their fruitful discussions, the authors of [21] and [8] for having made available their data sets for us, and more generally all authors that give easy access to the data-sets they produce.

REFERENCES

[1] ISO 9241-9:2000. 2000. *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices.* Standard. International Organization for Standardization, Geneva, Switzerland.

[2] F. Attneave. 1959. *Applications of information theory to psychology: A summary of basic concepts, methods, and results.* H. Holt, New York.

[3] M. Bakaev. 2008. Fitts' Law for Older Adults: Considering a Factor of Age. In *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems (IHC '08)*. Sociedade Brasileira de Computação, Porto Alegre, Brazil, Brazil, 260–263. http://dl.acm.org/citation.cfm?id=1497470.1497502

[4] P. Bertelson. 1961. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology* 13, 2 (1961), 90–102.

[5] R. J. Bootsma, L. Fernandez, and D. Mottet. 2004. Behind Fitts' law: kinematic patterns in goal-directed movements. *International Journal of Human-Computer Studies* 61, 6 (2004), 811–821.

[6] D. J. Cannon and L. J. Leifer. 1990. Speed and accuracy for a telerobotic human/machine system: experiments with a target-threshold control theory model for Fitts' law. In *Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on*. IEEE, IEEE, , 677–679.

[7] S. K. Card, W. K. English, and B. J. Burr. 1978. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics* 21, 8 (1978), 601–613.

[8] O. Chapuis, R. Blanch, and M. Beaudouin-Lafon. 2007. *Fitts' Law in the Wild: A Field Study of Aimed Movements.* article 1480. "LRI. http://insitu.lri.fr/~chapuis/publications/RR1480.pdf 11 pages.

[9] T. M. Cover and J. A. Thomas. 2012. *Elements of information theory.* John Wiley & Sons, New Jersey.

[10] E. R. F. W. Crossman. 1957. *The speed and accuracy of simple hand movements.* Technical Report. .

[11] E. R. F. W. Crossman and P. J. Goodeve. 1963. Feedback control of hand-movement and Fitts' Law. (1963).

[12] E. R. F. W. Crossman and P. J. Goodeve. 1983. Feedback control of hand-movement and Fitts' law. *The Quarterly Journal of Experimental Psychology* 35, 2 (1983), 251–278.
[13] P. Elias. 1958. Two famous papers. *IRE Transactions on Information Theory* 4, 3 (1958), 99.
[14] P. M. Fitts. 1953. *The influence of response coding on performance in motor tasks, in Current Trends in Information Theory.* University of Pittsburgh Press, Pittsburgh, PA.
[15] P. M. Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.
[16] P. M. Fitts. 1966. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of experimental psychology* 71, 6 (1966), 849.
[17] P. M. Fitts and J. R. Peterson. 1964. Information capacity of discrete motor responses. *Journal of experimental psychology* 67, 2 (1964), 103.
[18] J. M. Flach, M. A. Guisinger, and A. B. Robison. 1996. Fitts's Law: Nonlinear Dynamics and Positive Entropy. *Ecological Psychology* 8, 4 (1996), 281–325.
[19] K. Gan and E. R. Hoffmann. 1988. Geometrical conditions for ballistic and visually controlled movements. *Ergonomics* 31, 5 (1988), 829–839.
[20] E. N. Gilbert. 1960. Capacity of a burst-noise channel. *Bell Sys. Tech. J.* 39 (1960), 1253–1265.
[21] K. Goldberg, S. Faridani, and R. Alterovitz. 2015. Two large open-access datasets for Fitts' law of human motion and a succinct derivation of the square-root variant. *IEEE Transactions on Human-Machine Systems* 45, 1 (2015), 62–73.
[22] Y. Guiard. 1997. Fitts' law in the discrete vs. cyclical paradigm. *Human Movement Science* 16, 1 (1997), 97–131.
[23] Y. Guiard, H. B. Olafsdottir, and S. T. Perrault. 2011. Fitt's law as an explicit time/error trade-off. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, 1619–1628.
[24] Y. Guiard and O. Rioul. 2015. A mathematical description of the speed/accuracy trade-off of aimed movement. In *Proceedings of the 2015 British HCI Conference*. ACM, New York, 91–100.
[25] W. E. Hick. 1952. On the rate of gain of information. *Quarterly Journal of Experimental Psychology* 4, 1 (1952), 11–26.
[26] E. R. Hoffmann. 2013. Which version/variation of Fitts' law? A critique of information-theory models. *Journal of motor behavior* 45, 3 (2013), 205–215.
[27] R. Hyman. 1953. Stimulus information as a determinant of reaction time. *Journal of experimental psychology* 45, 3 (1953), 188.
[28] R. Kerr. 1973. Movement time in an underwater environment. *Journal of Motor Behavior* 5, 3 (1973), 175–178.
[29] D. Laming. 2001. Statistical Information, Uncertainty, and Bayes' Theorem: Some Applications in Experimental Psychology. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Salem Benferhat and Philippe Besnard (Eds.). Lecture Notes in Computer Science, Vol. 2143. Springer, Berlin, 635–646. DOI:http://dx.doi.org/10.1007/3-540-44652-4_56
[30] R. D. Luce. 2003. Whatever Happened to Information Theory in Psychology? *Review of General Psychology* 7, 2 (2003), 183–188.
[31] I. S. MacKenzie. 1989. A note on the information-theoretic basis for Fitts' law. *Journal of motor behavior* 21, 3 (1989), 323–330.
[32] I. S. Mackenzie. 1992. *Fitts' law as a performance model in human-computer interaction.* Ph.D. Dissertation. University of Toronto.
[33] I. S. MacKenzie. 2013. A Note on the Validity of the Shannon Formulation for Fitts' Index of Difficulty. *Open Journal of Applied Sciences* 3, 6 (2013), 360–368. DOI:http://dx.doi.org/10.4236/ojapps.2013.36046
[34] I. S. MacKenzie and W. Buxton. 1992. Extending Fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, New York, 219–226.
[35] D. E. Meyer, R. A. Abrams, S. Kornblum, C. E. Wright, and J. E. Keith Smith. 1988. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological review* 95, 3 (1988), 340.
[36] D. E. Meyer, J. E. Keith-Smith, S. Kornblum, R. A. Abrams, and C. E. Wright. 1990. *Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action, in Jeannerod Attention and Performance.* Lawrence Erlbaum Associates, Inc, New Jersey.
[37] G. A. Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review* 63, 2 (1956), 81.
[38] U. Neisser. 1967. Cognitive psychology. (1967).
[39] A. Newell. 1994. *Unified theories of cognition.* Harvard University Press, Cambridge, MA, USA.
[40] R. Plamondon and A. M. Alimi. 1997. Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences* 20, 02 (1997), 279–303.

[41] O. Rioul. 2007. *Théorie de l'information et du codage*. Hermes-Lavoisier, London, UK.
[42] O. Rioul and Y. Guiard. 2012. Power vs. logarithmic model of Fitts' law: A mathematical analysis. *Math. Sci. hum. / Mathematics and Social Sciences* 50.3, 199 (2012), 85–96.
[43] O. Rioul and Y. Guiard. 2013. The power model of Fitts' law does not encompass the logarithmic model. *Electronic Notes in Discrete Mathematics* 42 (June 2013), 65–72.
[44] O. Rioul and J. C. Magossi. 2014. On Shannon's Formula and Hartley's Rule: Beyond the Mathematical Coincidence. *Entropy* 16, 9 (2014), 4892–4910. DOI:http://dx.doi.org/10.3390/e16094892
[45] D. A Rosenbaum. 2009. *Human motor control*. Academic press, Cambridge, MA, USA.
[46] Q. Roy, Y. Guiard, G. Bailly, E. Lecolinet, and O. Rioul. 2015. Glass+Skin: An Empirical Evaluation of the Added Value of Finger Identification to Basic Single-Touch Interaction on Touch Screens. In *INTERACT'15: IFIP International Conference on Human-Computer Interaction*. Springer, Bamberg, Germany, 55–71.
[47] R. A. Schmidt, H. Zelaznik, B. Hawkins, James S. F., and J. T. Quinn Jr. 1979. Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychological review* 86, 5 (1979), 415.
[48] C. E. Shannon. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27 (1948), 379–423, 623–656.
[49] C. E. Shannon. 1956. The bandwagon. *IRE Transactions on Information Theory* 2, 1 (1956), 3.
[50] C. E. Shannon and W. Weaver. 1949. *The Mathematical Theory of Communication*. University of Illinois Press, U.S.A.
[51] R. W. Soukoreff and I. S. MacKenzie. 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, 6 (2004), 751–789.
[52] R. W. Soukoreff and I. S. MacKenzie. 2009. An informatic rationale for the speed-accuracy trade-off. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. IEEE, , 2890–2896. DOI:http://dx.doi.org/10.1109/ICSMC.2009.5346580
[53] R. W. Soukoreff, J. Zhao, and X. Ren. 2011. The entropy of a rapid aimed movement: Fitts' index of difficulty versus Shannon's entropy. In *IFIP Conference on Human-Computer Interaction*. Springer, 222–239.
[54] A. T. Welford. 1960. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics* 3, 3 (1960), 189–230.
[55] A. T. Welford, A. H. Norris, and N. W. Shock. 1969. Speed and accuracy of movement and their changes with age. *Acta psychologica* 30 (1969), 3–15.
[56] R. S. Woodworth. 1899. Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements* 3, 3 (1899), i.
[57] J. Wu, J. Yang, and T. Honda. 2010. Fitts' law holds for pointing movements under conditions of restricted visual feedback. *Human Movement Science* 29, 6 (2010), 882 – 892. DOI:http://dx.doi.org/10.1016/j.humov.2010.03.009
[58] R. W. Yeung. 2008. *Information theory and network coding*. Springer Science & Business Media, Berlin, Germany.
[59] S. Zhai. 2004. Characterizing computer input with Fitts' law parameters—the information and non-information aspects of pointing. *International Journal of Human-Computer Studies* 61, 6 (2004), 791–809.
[60] S. Zhai, P. O. Kristensson, C. Appert, T. H. Andersen, and X. Cao. 2012. Foundational issues in touch-screen stroke gesture design-an integrative review. *Foundations and Trends in Human-Computer Interaction* 5, 2 (2012), 97–205.

B STATEMENT OF PREVIOUS RESEARCH

This paper has never been submitted anywhere. It reports original theoretical work by the authors. Some aspects of the work have already been published in the following papers:

- Gori, J., Rioul, O., & Guiard, Y. (2017). To Miss is Human: Information-Theoretic Rationale for Target Misses in Fitts' Law. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 260-264). ACM.
- Gori, J., Rioul, O., & Guiard, Y. (in press). One Fitts' Law, Two Metrics. Human-Computer Interaction – INTERACT 2017.
- Rioul, O., Magossi, J. C. (2014). On Shannon's Formula and Hartley's Rule: Beyond the Mathematical Coincidence. Entropy 16(9), 4892-4910.

Section 2 explains well-known results from information theory that are needed in the remainder of the paper but that HCI readers are unlikely to be familiar with.

Section 4 uses some proofs that were derived by Rioul and Magossi (2014). Some of these proofs have been added in the appendix of the submission for completeness.

Section 5 incorporates most of the content of Gori et al. (2017), although rephrased to match the context of this more general paper.

Section 6 incorporates and extends an example taken from Gori et al. (in press).

The goal of the paper is to provide a complete information-theoretic model for aiming in Fitts' law paradigm, providing new results, but also synthesizing and making sense of known results. This explains why some content has been re-used. Gori et al. (2017) provides a model and a formula for an index of difficulty; in Section 5 of the present submission, the same formula is shown to represent a generalized model for errorless pointing.