

An Informatic Rationale for the Speed-Accuracy Trade-Off

R. W. Soukoreff and I. S. MacKenzie
Department of Computer Science and Engineering
York University
Toronto, Ontario, Canada

Abstract—This paper argues that the speed-accuracy trade-off arises as a consequence of Shannon’s *Fundamental Theorem for a Channel with Noise*, when coupled with two additional postulates: that people are imperfect information processors, and that motivation is a necessary condition of the speed-accuracy trade-off.

Keywords—Speed-Accuracy Trade-Off, Information Theory

I. INTRODUCTION

The speed-accuracy trade-off arises spontaneously and universally across the full range of human activities – simply put, when we go faster we make more errors. Reported in academic papers dating back more than a century [5], mundane to the point of being proverbial (“Haste makes waste”) and steeped in common sense (we instinctively slow down to avoid errors), it is hard to imagine a more banal feature of human performance. And yet, ironically, the cause and underlying mechanism of the speed-accuracy trade-off remain a mystery that has persisted through the years, despite the work of some noteworthy researchers. For example, Fitts [1] proposed a random-walk model to explain the speed-accuracy trade-off in choice reaction time studies; Plamondon and Alimi [2] constructed a delta-log-normal impulse model of the human neuromuscular system that explains the speed-accuracy trade-off in rapid aimed movements; and Swensson [4] proposed an explanation of the speed-accuracy trade-off for visual discrimination tasks by combining the features of a random-walk with post-deadline guessing. But all of the explanations proffered so far are too narrow in scope to explain the speed-accuracy trade-off over the full range of human activities. Specifically, Fitts’ random-walk model of choice reaction time does not explain the empirical results pertaining to movement or visual discrimination. Likewise, the neuromuscular impulse model of Plamondon and Alimi does not apply to choice reaction time or visual discrimination tasks. Similarly, Swensson’s random walk with guessing does not explain the speed-accuracy trade-off as it pertains to movement. And so we have three explanations of the speed-accuracy trade-off that, together, raise an interesting philosophical question: is it reasonable to conclude that humans coincidentally embody (at least) three distinct mechanisms that give rise to three separate instances of the speed-accuracy trade-off, or, could these be three manifestations of a single underlying phenomenon?

If there is a single, unifying, explanation of the speed-accuracy trade-off, then it must affect the human machine at a very fundamental level in order to influence human

performance across the wide spectrum just listed, including vision (an example of the human *input* system), motion (the human *output* system), and choice reaction time (*internal* information processing). This paper presents an explanation of the speed-accuracy trade-off that is rooted in communication theory – specifically Shannon’s *Fundamental Theorem for a Channel with Noise* [3] – augmented with two additional hypotheses: that humans are imperfect information processors, and that motivation is a necessary condition for the speed-accuracy trade-off to exist. Choice reaction time tasks, rapid aimed movements, and visual discrimination tasks, are all activities that involve the movement of information, and consequently they are bound by the laws of communication theory. Our thesis is that the speed-accuracy trade-off arises as a predictable feature of communication within humans, and it is for this reason that the speed-accuracy trade-off influences the wide range of human activities reported.

II. COMMUNICATIONS THEORY

A. Source Entropy, Equivocation, and Throughput

Shannon’s ground breaking work *A Mathematical Theory of Communication* [3] deconstructs a communications system into three parts: the transmitter, the receiver, and the communications medium between them. The rate at which information is broadcasted by the transmitter is called the *source entropy*¹. The rate of information *successfully* received by the receiver, is called the *throughput*. (The quantifier ‘successfully’ is the defining characteristic of throughput; information that is transmitted but not accurately received, does not count towards throughput.) And the difference between source entropy and throughput is called *equivocation*. This quantity corresponds to information that has been lost in transit, corresponding in traditional communications theory to the corruption of the broadcast signal by noise, typically occurring as the signal traverses the communications medium. Mathematically, this relationship is,

$$E = I - R, \quad (1)$$

¹ Shannon defines a discrete information source as a producer of messages, where each message conveys a quantity of information called its *entropy*, usually measured in the units *bits*. Shannon then refers to the average entropy per message produced by an information source as “the entropy of the source” [3, page 13]. We find the term “entropy of the source” to be cumbersome, and so will refer to this quantity as the *source entropy*.

where, E denotes the equivocation, I is the source entropy, and R , the throughput. The units of these quantities can be either *bits per second*, or *bits per symbol*.

Rearranging (1) yields,²

$$R = I - E. \tag{2}$$

An additional minor detail concerning throughput is that it cannot be negative³. Thus, to be accurate, Shannon's definition of throughput should be written,

$$R = \begin{cases} I - E & \text{if } I - E > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

B. The Fundamental Theorem for a Channel with Noise

The primary consequence of Shannon's *Fundamental Theorem for a Discrete Channel with Noise*⁴ [3] is that there is a limit to the rate of information transmission achievable through a noisy communications channel. In this sense the fundamental theorem is a universal law – a constraint on the movement of information that cannot be surpassed by any means. Using (3) it is possible to define the capacity for a channel in the presence of noise. The channel capacity, C , is simply the maximum possible throughput,

$$\begin{aligned} C &= \text{Max}(R) \\ &= \text{Max}(I - E), \end{aligned} \tag{4}$$

where the maximum is taken over all possible information sources, communication protocols, and error correcting schemes. As such, C represents the highest throughput transmittable through the channel by any means. Consequently, we may write,

$$R \leq C. \tag{5}$$

² Equation 2 is taken directly from Shannon [3, pages 20 and 42], although the notation has been simplified. The source information I is written $H(x)$ by Shannon. The equivocation E is written $H_s(x)$ by Shannon, to indicate that it can be calculated as a conditional entropy (the mathematical implications of this do not concern us, the interested reader is referred to [3, pages 10-12, and 35-36]).

³ The reader familiar with information theory will know that continuous information sources can sometimes yield negative entropy values. However, because all rates and capacities are calculated as the difference between two entropies, rates and capacities can only be greater than or equal to zero, even in the continuous case. As Shannon writes: "The entropy of a continuous distribution can be negative. ... The rates and capacities will, however, always be nonnegative." [3, page 38].

⁴ Although we are using the discrete fundamental theorem here, there are no significant differences (other than the complexity of the mathematics) between the discrete case and the continuous version (viz., *The Capacity of a Continuous Channel*), the definitions of throughput and channel capacity are identical. See [3] and compare pages 20 – 23 with pages 41 – 44. As Shannon notes, "The entropies of continuous distributions have most ... of the properties of the discrete case." [3, page 35].

C. How does the Fundamental Theorem limit Throughput?

Equations (4) and (5) describe the effect of Shannon's fundamental theorem, but they do not describe the mechanism by which the capacity of a communications channel is limited by the fundamental theorem. In general, an information source is free to attempt to transmit information at any rate, including rates that exceed the channel capacity. However, above the channel capacity, noise corrupts the signal in such a way that the achievable throughput cannot exceed the channel capacity. Specifically, the fundamental theorem states that the lowest possible equivocation is is,⁵ [3, Theorem 11, page 22]

$$E_{min} = I - C. \tag{6}$$

Shannon's fundamental theorem stipulates the values of equivocation that are attainable during the transmission of information over a noisy channel, and this is what limits throughput to the channel capacity. The fundamental theorem is depicted graphically in Fig. 1.

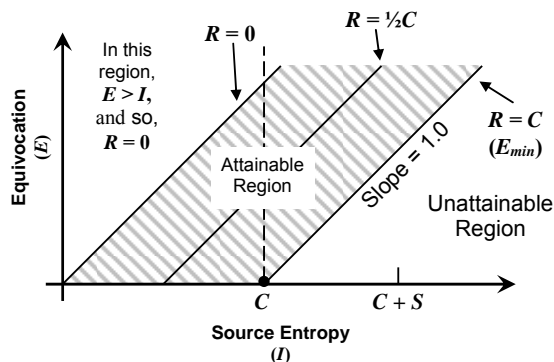


Figure 1. Depiction of Shannon's fundamental theorem for a discrete channel with noise

The hatched region denotes the ranges of source entropy and equivocation that are attainable. Throughout the hatched region there is a linear gradient of throughput ranging from 0 to C as indicated by the lines, $R = 0$, $R = \frac{1}{2}C$, and $R = C$. The units of the axes may be either *bits per second*, or *bits per symbol*. The point ' $C + S$ ' facilitates discussion in the text. (This diagram is based on Shannon [3, Figure 9, page 22].)

Consider the region of Fig. 1 to the left of the dashed vertical line extending upward from ' C '. In this region, the entropy of the information source I , is less than the channel capacity C , and so by (6) an equivocation of zero is possible (i.e., the hatched area extends all the way down to the x -axis). Intuitively, this means that the channel capacity is high enough, and the speed of information transmission low enough (i.e., below C), that perfect transmission is possible. Note that higher error rates are also possible in this region. One may choose a non-optimal communication protocol or even deliberately broadcast errors if one wishes to, but according to the fundamental theorem, an optimal protocol exists that can reduce the equivocation effectively to zero.

⁵ Again, note that negative values of entropy rates are not possible, so the minimum equivocation is zero if, $I < C$.

III. HUMAN PERFORMANCE

To the right of the $I = C$ dashed vertical line, the minimum possible equivocation rises above zero as stipulated by the fundamental theorem (6), reflecting the fact that throughput cannot exceed the capacity (and so, the hatched area only extends down as far as the $R = C$ line).

D. Exceeding the Capacity of a Channel

The relationship between the speed of transmission (i.e., source entropy rate, I) and throughput (R), at speeds that exceed the channel capacity (C), is demonstrated via the following calculations. Let us define $S > 0$ as the amount by which I exceeds C , (see the point ' $C + S$ ' in Fig. 1). Note that the value of S is arbitrary, and so the following calculations apply to all points where I is in excess of C , (i.e., everywhere to the right of the dashed vertical line in Fig. 1). Setting, $I = C + S$, and substituting this into Shannon's fundamental theorem (6), yields,

$$\begin{aligned} E_{min} &= I - C \\ &= (C + S) - C \\ &= S. \end{aligned}$$

Next, calculating throughput via (2), and remembering that E_{min} is the minimum allowable equivocation, we obtain,

$$\begin{aligned} R &= I - E_{min} \\ &\leq (C + S) - S \\ &\leq C. \end{aligned} \quad (7)$$

Thus at rates of transmission that are faster than the channel capacity, no matter how the information is encoded, nor what error-correcting codes are employed, the equivocation increases so that the throughput (the actual rate of useful information transmission) cannot exceed the channel capacity.

Treatises on information theory invariably omit discussion of communication at rates in excess of the channel capacity; because throughput cannot exceed channel capacity, there is no advantage to attempting greater-than-capacity transmission speeds. Consequently, information theorists focus on the region of Fig. 1 to the left of the $I = C$ dashed vertical line. But people are different than technology, they have the freedom to choose the speed at which they communicate, and they do attempt greater-than-capacity communication rates. We routinely drive too fast, talk in environments too noisy to be heard, practice to learn new skills, compete with one another physically and mentally, overload ourselves with too much to do, and leave things to the last minute. These are all practices that involve exceeding one's capacity to perform tasks. The application of information theory to human performance obliges us to consider the area to the right of the $I = C$ dashed line of Fig. 1, reflecting a fundamental difference between traditional communication theory and the human-centric communication theory developed here.

The preceding discussion comes more or less directly from Shannon's work [3]. The novel contribution made by this paper follows.

Two hypotheses concerning the application of information theory to human performance are now described. The first defines how humans differ from theoretically ideal information transmitters. The second characterizes the performance of people conscientiously performing an information transmission task. During these discussions, four claims are made that are used to develop the two hypotheses. These claims represent important concepts for which there exists strong supporting evidence. We begin with the first of these claims.

A. Claim 1: The Informatic Basis of Action

Humans have a capacity for processing information, and this capacity underlies many activities. This information theoretic basis implies that the properties of information theory underwrite many human activities.

This claim should not be surprising to students of Fitts' law or the Hick-Hyman law. The point made by the first claim is that there is an information theoretic basis for a wider array of human behaviors than merely the Fitts' and Hick-Hyman paradigms. Many activities would not be possible were humans unable to receive, transmit, or process information. Consequently, certain aspects of human performance fall within the sphere of communication theory.

B. Claim 2: Shannon's Fundamental Theorem Applies to Human Information Transmission

When operating below or at one's capacity, it is possible to achieve unequivocal (error-free) performance. Conversely, operating above one's capacity implies a strictly greater than zero equivocation (and error rate).

This claim is really a restatement of Shannon's fundamental theorem for a discrete channel with noise [3, Theorem 11, page 22] and thus is generally accepted as a universal law. Shannon's theorem places a limitation upon the movement of information that takes the form of a minimum equivocation arising when transmission rates exceed the capacity of a communication system. It is important to note that Shannon's equivocation is a minimum – one is free (and indeed likely) to commit more errors than are required by Shannon's theorem. Achieving Shannon's minimum equivocation is difficult, and in general, not always possible, because it requires perfectly optimal encoding of information. Consequently we expect humans to incur an equivocation that is greater than Shannon's theoretical minimum (the implication being that human performance is less than optimal).

C. The Fundamental Theorem of Human Performance

Humans are not optimal information processors, and in particular, the minimum equivocation achievable by a human is greater than the minimum required by Shannon's channel capacity theory.

Shannon [3, page 22], referring to Figure 9 in his paper (equivalent to Fig. 1 in this paper), states, "Any point above the ['Slope = 1.0' line of Fig. 1] in the shaded region can be attained and those below cannot. The points on the line cannot in general be attained..." Here, Shannon is making an important distinction between theory and reality. The theory is concisely stated by equations (6) and (5) above, derived in

detail in (7), and clearly depicted by Fig. 1. Undeniably, in theory, the ‘Attainable Region’ extends all the way down to the $R = C$ line. But in practice, it is very difficult to achieve performance that approaches the $R = C$ line. To do so requires ideal information encoding and the highest possible error mitigation, and this is not always practicable. And so a question arises: What mathematical function describes the lower-right edge of the ‘Attainable Region’ for *real-world* (and in our case, *human*) communication channels?

For reasons clarified later (see the ‘Supporting Evidence from the Literature’ section below), we posit the existence of a linear relationship between Shannon’s minimum equivocation and the minimum equivocation achievable by humans when operating above their capacity. This corresponds to the simplest function that remains above the $R = C$ line as required, and that passes through the point labeled C in Fig. 1 (a feat that every communications channel is capable of). Mathematically we write,

$$E_h = \alpha \cdot E_{min} \quad (8)$$

where E_h represents the minimum equivocation achievable by a human, and E_{min} represents Shannon’s minimum equivocation (defined by (6) above, $E_{min} = I - C$). The scalar, α , represents a fundamental constant of human performance, that is equal to the ratio of the human-achievable equivocation to the theoretically minimum equivocation (i.e., $\alpha = E_h / E_{min}$). Communication theory requires that $\alpha \geq 1$, because no process can achieve a lower equivocation than Shannon’s universal minimum (i.e., $E_h \geq E_{min}$). In practice, however, it is expected that α will always be greater than (as opposed to, equal to) one, simply because human information processing is not perfect.

The effect of the fundamental theorem of human performance becomes clear when Figs. 1 (above) and 2 (below) are compared. At information transmission speeds below the channel capacity (to the left of the $I = C$ dashed line), ideal information sources and humans may both achieve error-free performance (i.e., approach zero equivocation). However, above the channel capacity (to the right of the $I = C$ dashed line) the minimum equivocation achievable by humans (represented by the E_h line in Fig. 2) rises rapidly as the source entropy increases. This causes the achievable range of source entropy and equivocation to assume the triangular shape visible as the hatched area of Fig. 2. This contrasts with the range of performance characteristics available to theoretically optimal devices, which extends upwards and to the right indefinitely (observe the hatched area in Fig. 1).

D. Claim 3: Freedom of Choice Applies to Performance

Humans have control over the characteristics of performance (i.e., the rates of source entropy, and equivocation) with which they perform.

For example, when engaged in an information transmission task such as typing, one can find a comfortable speed where not too many errors are made (this speed is likely close to one’s capacity). But one can choose to type faster or slower than this, corresponding to changing one’s source entropy and thus shifting one’s place of performance horizontally in Fig. 2.

Also, one may choose to be careful or sloppy, corresponding to varying one’s equivocation, and so influencing one’s place of performance vertically in Fig. 2. However, the control humans have over their characteristics of performance is limited by information theory and the fundamental theorem of human performance, and so it is not possible to escape the hatched triangle of Fig. 2. And while the entire hatched triangle is theoretically achievable, not all locations within the triangle are equally preferable – this is discussed next.

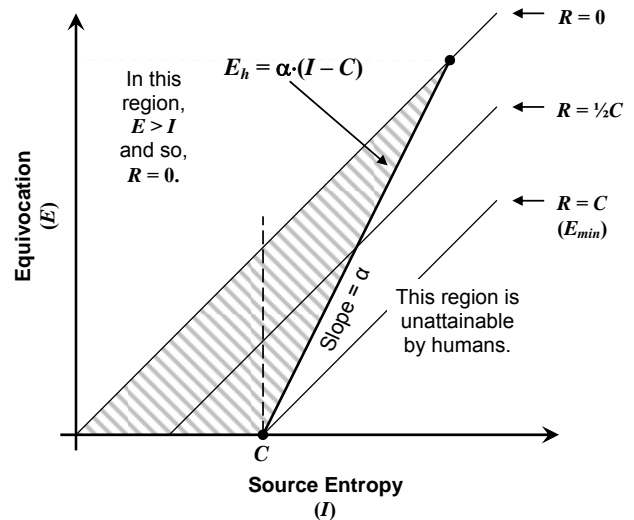


Figure 2. The Fundamental Theorem of Human Performance

Shannon’s fundamental theorem for a channel with noise [3, Theorem 11, page 22] implies that the area below the $R = C$ line is unattainable by any communications system (see Fig. 1). But for humans, the cost of exceeding the channel capacity is greater than Shannon’s minimum equivocation, and so the attainable region follows the E_h line instead. (Note that in this figure, a value of $\alpha \approx 2$ is depicted.)

E. Claim 4: Efficiency is Preferred

Although a wide range of values of source entropy and equivocation are achievable, humans generally prefer performance characteristics that are (at least) reasonably efficient.

Subjective comfort⁶ is not uniform across the range of available performance characteristics. Certain areas within the hatched triangle of Fig. 2 are preferable to others, and these predilections are predictable. Consider performance anywhere along the $R = 0$ line; to achieve zero throughput, equivocation must completely nullify source entropy. Few would choose to perform in this manner, expending effort and yet achieving nothing. Next, consider performance along the $R = \frac{1}{2}C$ line; there are several ways to achieve merely half of one’s capacity, for example, using a very slow source entropy (i.e., going at half speed), or by coupling a moderate source entropy with poor accuracy, or by exceeding one’s capacity and so expending considerable effort while enduring a devastating

⁶ Our choice of the rather vague term *comfort* here is deliberate. The intention is not to qualify or quantify satisfaction during information transmission tasks, but rather to predict, in general, the combinations of source entropy and equivocation that most will be drawn to as a result of subjective comfort.

equivocation. None of these alternatives seem appealing. Contrast this with performance near the point 'C' (on the x-axis of Fig. 2), where the highest throughput is achieved. Near 'C', source entropy is approximately equal to capacity, and equivocation is low. This combination is relatively comfortable, as the source entropy is not excessively fast (in fact it is slow enough that it is possible to achieve very few, or zero errors), and the low equivocation lessens the costly overhead of (and frustration associated with) error correction. To generalize, *the desirability* of a particular location in the information transmission space correlates positively with *throughput*.⁷

Due to the fundamental theorem of human performance, if source entropy and equivocation were measured and plotted for a group of people performing an information transmission task with their natural performance characteristics, then the performance characteristics must fall within a triangle similar to the hatched area of Fig. 2. But a positive correlation between throughput and subjective preferences implies that the distribution of the plotted performance points would not be uniformly distributed throughout the hatched triangle. Instead, the performance points would be concentrated near to the point 'C', with a decreasing number at performance points farther from 'C'. Additionally, it is unlikely that anyone would naturally choose to perform at less than half of their capacity – compare the stippled and hatched areas of Fig. 3.

So far we have considered the performance characteristics of people performing at the source entropy and equivocation point of their own choosing. Next we consider what happens when a person performs an information transmission task with either a speed or accuracy emphasis.

F. The Theory of Motivated Performance

If motivated to achieve a high level of performance, humans will increase their speed, decrease their error rate, or attempt a combination of the two. This has the effect of confining their performance characteristics to the E_h line (or possibly a small portion of the x-axis).

When motivated to perform a task quickly, people attempt to generate information rapidly, increasing their source entropy, and shifting their place of performance to the right in Fig. 3, impacting upon the E_h line. The range along the E_h line occupied by those with a speed-emphasis may be wide, as this range is only limited by an individual's tolerance of the frustration associated with the increased errors caused by exceeding their capacity. (Although, in accordance with Claim 4, we expect few will choose to go so fast as to halve their throughput.) Conversely, an accuracy emphasis causes people to slow down and be careful. In the context of Fig. 3, this has the effect of gravitating their performance point down to the x-axis, close to the point 'C'. But, because error-free performance is possible at any speed up to and including one's capacity, there is no advantage to slowing down less than C. Thus the range along the x-axis populated by those desiring

accuracy is likely not very wide. Because the E_h line goes through the point 'C', we conclude that, regardless of whether the goal is speed or accuracy, the effect of motivation is to shift performance characteristics onto the E_h line, either close to the point 'C' (if the goal is accuracy), or stretched-out along the E_h line (if the goal is speed).

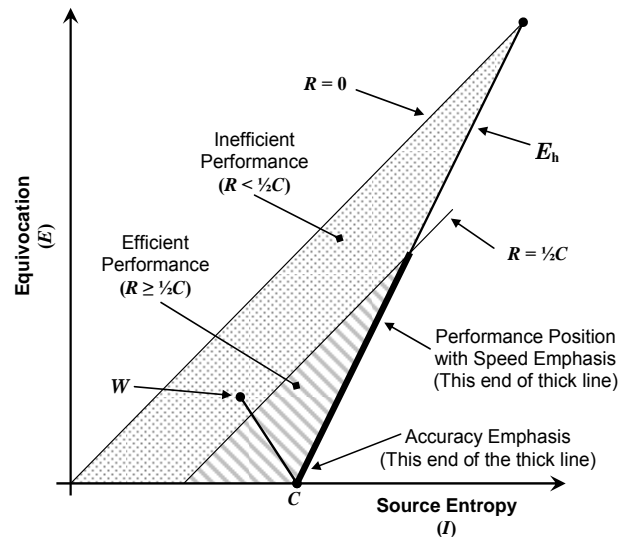


Figure 3. Effect of a speed or accuracy emphasis on performance

The triangle formed by the union of the stippled area and the hatched area indicates the range of source entropy and equivocation available to humans. However, we expect the performance of most people to fall into the hatched area where $R \geq \frac{1}{2}C$. Further, within the hatched area, an emphasis toward speed draws one's point of performance to the right, against the E_h line; an accuracy emphasis shifts the point of performance downward to the x-axis (but likely near to the point 'C', and thus still on, or very near to, the E_h line). The thickened line indicates the regions of the E_h line likely to be occupied by people performing with a speed or accuracy emphasis. The point W facilitates discussion in the text.

The theory of motivated performance draws a distinction between natural performance (constrained merely to the hatched triangle of Fig. 3), and motivated performance (constrained to the E_h line). Note that natural performance is underrepresented in the published literature, as most papers present studies of motivated subjects. The published literature reflects the fact that models of human performance are in fact models of motivated behaviour. Indeed, Fitts' law is often cited as a model of "rapid aimed movement", and studies of choice reaction time (viz., the Hick-Hyman law) and text entry usually instruct subjects to perform "as quickly and accurately as possible". This is because most researchers are interested in observing their subjects' maximum capacities – not some arbitrary speed below their capacity – and to ensure that the subjects perform to their maximum capability, researchers motivate their subjects.

G. Mathematical Implications

The theory of motivated performance states that the performance characteristics of motivated subjects fall along the E_h line. The defining characteristic of the E_h line is that it follows the minimum equivocation achievable by humans

⁷ Note that no claim is made here on the precise nature of the relationship between subjective comfort and source entropy, equivocation, or throughput – only that a positive correlation exists between relative desirability and throughput.

engaged in an information transmission task. The minimum human equivocation, E_h , is defined by the fundamental theorem of human performance via (8) as, $E_h = \alpha \cdot E_{min}$, where the minimum theoretical equivocation, E_{min} , is given by (6), $E_{min} = I - C$. Substituting (6) into (8) yields an expression describing the relationship between source entropy and equivocation along the E_h line,

$$\begin{aligned} E_h &= \alpha \cdot E_{min} \\ &= \alpha \cdot (I - C). \end{aligned} \quad (9)$$

Equation 9 defines the E_h line depicted in Figs. 2 and 3, having slope α , and extending upward and to the right of the point 'C'. Solving (9) for source entropy, I , and then substituting into (2) we obtain,

$$R_\alpha = E_h \cdot \left(\frac{1 - \alpha}{\alpha} \right) + C, \quad (10)$$

a useful expression describing throughput along the E_h line, as a function of equivocation. This expression for throughput (10) is depicted in Fig. 4.

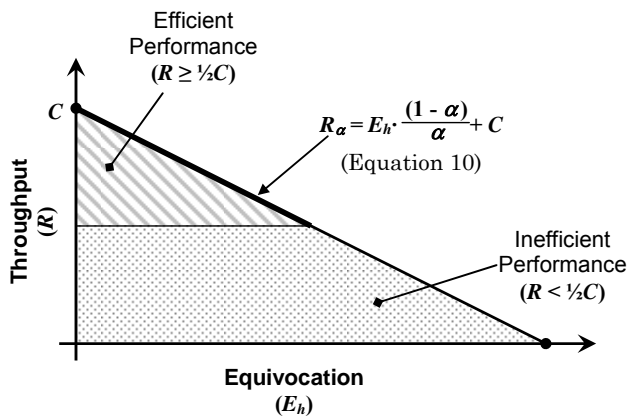


Figure 4. Throughput as a function of equivocation

This figure shows the relationship between throughput and equivocation along the E_h line. The thickened line indicates the portion of the E_h line occupied by those who are motivated to perform well (i.e., with good speed, high accuracy, or both). (Note that in this figure, a value of $\alpha \approx 2$ is depicted.)

H. Motivation and the Speed-Accuracy Trade-off

The mechanism by which motivation plays a causal role to the speed-accuracy trade-off follows. Motivation causes people to move their performance point onto the E_h line at the edge of the achievable performance space, and once at the edge, the freedom to individually affect speed separately from accuracy is lost. The nature of the speed-accuracy trade-off (i.e., that an increase in speed translates into an increase in equivocation) arises from the geometry of the E_h line. Mathematically speaking, the E_h line is *strictly increasing*, and so to accommodate a movement to the right in Fig. 3 (due to increased speed), the performance point must also rise (reflecting decreased accuracy) in order to remain within the achievable performance space, and in so doing, to satisfy the fundamental theorem of human performance.

In the absence of motivation, however, the speed-accuracy trade-off does not spontaneously arise. For example, consider the point W in Fig. 3. The point W represents an arbitrary performance point that anyone may achieve by deliberately performing slowly and carelessly. The point 'C' also represents a performance point that anybody may achieve, although with a higher speed, and yet a lower equivocation than the point W . Comparing the performance characteristics of these two points reveals that they contravene the speed-accuracy trade-off, whereby as speed increases, equivocation must increase. This example demonstrates that, without motivation forcing the performance point onto the E_h line, freewill may contravene the speed-accuracy trade-off.

I. The Role of α in the Speed-Accuracy Trade-off

It is worth noting that the argument given in the previous section does not require the equivocation ratio α , to be greater than one for the speed-accuracy trade-off to exist. Even if humans were ideal information sources (and consequently limited only by Shannon's fundamental theorem to the $R = C$ line, instead of to the E_h line), then the speed-accuracy trade-off would still be evident. Moving to the right along the $R = C$ line requires that equivocation increases as source entropy increases. The only difference would be the rate at which increased source entropy is traded for equivocation, (corresponding to whether, $\alpha = 1$, or, $\alpha > 1$).

IV. SUPPORTING EVIDENCE FROM THE LITERATURE

Due to the popularity of applying information theory in psychology, and in particular to reaction time studies, in the 1950s and 1960s, there are many published papers that report rates of information transmission (either throughput or estimates of channel capacity). The difficulty of finding supporting evidence lies in locating data that show the relationship between equivocation and source entropy or throughput (most publications present reaction *time* versus source entropy or throughput, and many papers overlook equivocation). Nonetheless, evidence supporting the two hypotheses presented above is provided in the paper by Fitts referenced earlier [1]. The purpose of Fitts' paper was to examine the means by which humans perform choice reaction time tasks. Ancillary to the experiment and analysis, Fitts included an "informatic analysis" where he noted that "a surprisingly consistent relation can be observed between response equivocation and information transmission rate" [1, page 856]. Although Fitts was unable to explain the linear relationship observed, the relation cited is that described above by (10), shown in Fig. 4.

Fitts did not include raw data in his paper, however he did include a figure [1, Figure 7, page 856] depicting data from all eighteen subjects, revealing the linear relationship between throughput and equivocation. Unfortunately, Fitts used the units *bits per second* for throughput, but *bits per response* for equivocation. This problem can be overcome, as Fitts included a figure showing the reaction time for each subject [1, Figure 3, page 852]. We were able to estimate Fitts' data by digitally scanning and enlarging these figures and measuring the location of the data points on the graphs. This yielded approximate values of throughput, equivocation, and reaction time, for all eighteen subjects. By dividing the equivocation

values (in *bits per response*) by the corresponding reaction times, equivocation values in *bits per second* were obtained for each of Fitts' subjects. These adjusted data appear in Fig. 5.

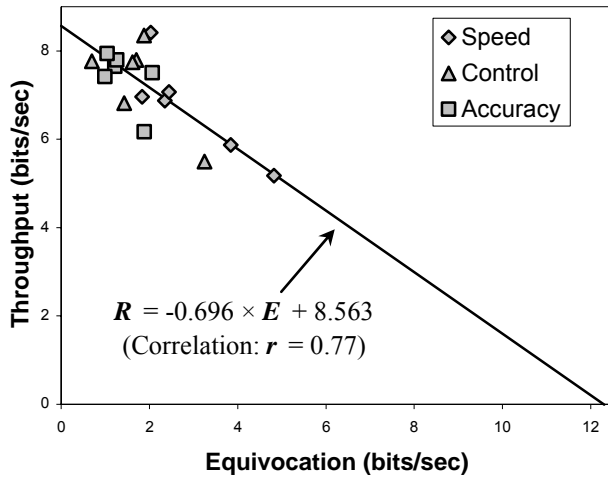


Figure 5. Throughput versus equivocation data adapted from Fitts [1]

This figure presents data from Figure 7 of Fitts [1] with a change of units. Equivocation was reported in *bits per response* in Fitts' paper, but here, the units are *bits per second*. Note the similarity between this figure and Fig. 4, above.

Linear regression of the units-adjusted data from Fitts' paper yields the linear relation,

$$R = -0.696 \times E + 8.563, \quad (11)$$

with correlation coefficient, $r = 0.77$. By applying the values from the regression equation (11), to the expressions for the slope and intercept from (10), we estimate that for Fitts' data,

$$C = 8.56 \text{ bits per second, and, } \alpha = 3.29.$$

Because Fig. 5 shows results for a group of people performing with performance characteristics falling along their own individual E_h lines, there is a fair amount of scatter among the data points (resulting in only a modest correlation coefficient value). But there is enough similarity between the individual E_h lines that Fitts was able to notice and report the linear relationship apparent in the data. Our fundamental theorem of human performance posits a linear relationship between source entropy and equivocation (implying a linear relationship between throughput and equivocation, see (10)), and requires that $\alpha \geq 1$. By comparing Fitts' adjusted data in Fig. 5 with Fig. 4, it is apparent that the predicted relationship between throughput and equivocation is confirmed. Also, note that the performance characteristics of the subjects fall in the upper half of Fig. 5, implying that people do seem to prefer efficiency (Claim 4, above).

V. CONCLUDING REMARKS

There are three facets of the speed-accuracy trade-off. The first is the inverse relationship between speed and accuracy that is observed over a wide range of human behaviors – as people speed up, they commit more errors, and as they slow down, they make fewer errors. The second facet is that the speed-accuracy trade-off is spontaneous and involuntary. And the third facet is that the speed-accuracy trade-off only manifests itself under motivated conditions (where the goal is either faster speed, higher accuracy, or both).

The theory of motivated performance explains why motivation is a necessary condition for the trade-off, and it reveals the spontaneous and involuntary nature of the trade-off; motivation limits the available range of source entropy and equivocation to the E_h line, and the shape of this line defines the nature of the trade-off. Although confined to the E_h line, motivated humans are free to occupy any place along the E_h line that they choose. One end of the E_h line corresponds to slow speed with high accuracy, the other to high speed with low accuracy. In the intervening range along the E_h line, the rate at which increased speed is traded for decreased accuracy is parameterized by the equivocation ratio, α ; see (9) and note that, essentially, it states that: $Error\ rate \propto \alpha \times Speed$, a mathematical statement synonymous with the speed-accuracy trade-off.

Mathematically, the distinction between the new theory described here and traditional information theory is the value of the equivocation ratio, α . We argue that when humans are engaged in an information transmission task, the resulting equivocation ratio is greater than one, and Fitts' 1966 data bares this out. However, if we set $\alpha = 1$, then the equations given here reduce to traditional information theory. That the distinction between the new and traditional models amounts to the value of a single parameter, is a strength of the new model, because, when either (9) or (10) are fit to empirical data, the value of the α parameter serves as a direct indicator of whether or not the new theory is sound and applicable.

REFERENCES

- [1] P. M. Fitts, "Cognitive aspects of information processing: III. Set for speed versus accuracy," *J. Exp. Psychology*, vol. 71, no. 6, pp. 849–857, 1955.
- [2] R. Plamondon and A. M. Alimi, "Speed/Accuracy trade-offs in target-directed movements," *Behavioral and Brain Sciences*, vol. 20, pp. 279–349, 1997.
- [3] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.
Note that page numbers appearing in references to Shannon 1948 refer to the 1998 republication, a freely downloadable .PDF file available from:
<http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html>
- [4] R. G. Swenson, "The elusive tradeoff: Speed vs accuracy in visual discrimination tasks," *Perception and Psychophysics*, vol. 12(1A), pp. 16–32, 1972.
- [5] R. S. Woodworth, "The accuracy of voluntary movement," *The Psychological Review*, vol. III, no. 3 (Monograph 13), pp. 1–114, July 1899.