Pointing Beyond Fitts: Strategy, Movement Time Distributions, Local-Global Tradeoffs, and Generative Models

JULIEN GORI, Sorbonne Université, CNRS, Inserm, Institut des Systèmes Intelligents et de Robotique, ISIR, France

Traditional Fitts' law captures average trends in pointing tasks but ignores variability, strategy effects, and complex dependencies between movement time (MT) and effective index of difficulty (ID_e). In this work, we adopt a probabilistic perspective to model the full distribution of MTs and ID_e, and how this relationship depends on user strategy. We analyze data from four empirical studies using three different experimental protocols that manipulate task constraints and user strategy across different input modalities, examining how speed and accuracy are traded off locally (within conditions) and globally (across conditions). We find strong global dependencies but minimal local dependencies, suggesting movements within a condition are largely functionally equivalent. To capture behavioral distributions, we introduce a probabilistic ID_e model, a bivariate Gaussian model linking ID_e and mean movement time with strategy, as well as three models based on copulas and Exponentially Modified Gaussian (EMG) noise distributions that express the complete joint (ID_e, MT) distribution, all of which are parametrized by user strategy. These three models are compared on their ability to faithfully reproduce summary statistics of the pointing datasets. This work offers a richer, distribution-and-strategy-based perspective on pointing performance. Code and parameters are provided for reproducibility and future research.

ACM Reference Format:
Julien Gori. 2024. Pointing Beyond Fitts: Strategy, Movement Time Distributions, Local-Global Tradeoffs, and Generative Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 46 pages. https://doi.org/XXXXXXXXXXXX

1 INTRODUCTION

Pointing tasks, such as selecting an icon with a mouse, are among the most fundamental interactions in human-computer interaction (HCI) [7, 59]. The dominant quantitative model of pointing, Fitts' law [21, 30, 59], predicts the mean movement time (MT) required to reach a target of width W at distance D:

$$\mathrm{MT} = a + b \log_2(1 + \tfrac{D}{W}), \tag{1}$$

$$\equiv a + b\,\mathrm{ID}. \tag{2}$$

Here, $\mathrm{ID} = \log_2(1 + D/W)$ is the index of difficulty, and the parameters a and b are empirically estimated. Fitts' law captures the speed-accuracy trade-off (SAT) in motor control: smaller targets, which require more precision, take longer to acquire. It has been validated across devices, modalities, and populations, and remains a cornerstone of input evaluation and interface design [7, 59].

Yet, Fitts' law provides only a partial account of the SAT. It predicts average performance but ignores trial-to-trial variability. In reality, MTs are not fixed values but they are spread across a distribution, and some attempts are unexpectedly fast and precise while others are slow and inaccurate. Fitts' law also leaves no room for user strategy, where participants may favor speed or accuracy depending on instructions or personal preference [38, 71]. As we will

Manuscript submitted to ACM

show, variability in MT and user strategy strongly influence observed responses; ignoring them oversimplifies pointing behavior and limits the predictive power of models. To address these limitations, we adopt a *probabilistic* perspective that models the full distribution of outcomes and explicitly incorporates user strategy. Rather than assuming that going faster always means being less accurate, or that smaller targets always lead to lower MTs, we examine how user strategies and task properties influence the relationship between the observed MTs and accuracy.

We examine the dependence between movement time and accuracy at two levels of analysis. At the *local* level, we study the relationship between MT and accuracy within individual task conditions (given D, W, and strategy). At the *global* level, we pool data across conditions. This distinction is important, as we find that the SAT is often weak or absent when examined locally, but emerges clearly at the global level. Understanding this contrast helps explain when trade-offs appear in experimental data and how they should be modeled.

To model this dependence, we rely on the exponentially modified Gaussian (EMG) distribution, which is parametric model that was previously validated to model MT distributions [24, 27, 43, 73], but without accounting for strategy. The EMG model describes MT conditional on accuracy with a linear mean and a quadratic variance, thus imposing a specific dependency structure. While this captures important aspects of the SAT, it imposes the form of dependence. Thus, to allow for more flexibility, we also consider models based on *copulas*, which describe the joint distribution of MT and accuracy without assuming a fixed pattern of dependence. Together, these approaches let us compare descriptions of the SAT across experimental conditions.

To ground these perspectives, we analyze pointing data collected under three experimental protocols that differ in how accuracy is defined: (1) the classic *Fitts* protocol, where accuracy is set by target width W; (2) a *pure strategy* protocol, where accuracy is determined by participants in line with experimenter instructions; and (3) a combined *Fitts-with-strategy* protocol that merges both task and strategy constraints. Together, these protocols reveal how distributions and strategy interact to shape pointing performance.

Contributions. This paper makes the following contributions:

- A theoretical analysis that shows why Pearson correlation is not adapted to evaluate the fitness of models that predict the joint distribution between MT and accuracy.
- New probabilistic models of pointing linking movement time to accuracy, building on EMG, copulas and a bivariate Gaussian model that capture the dependency between MT and accuracy in different ways.
- Empirical findings across three protocols, that show that the SAT is often weak locally but emerges globally, and that strategy systematically shifts these patterns.
- Open resources: reusable model parameters and code for generating realistic pointing data, supporting replication and extension by the HCI community. The code used for this paper is available at the following repository.

By moving beyond deterministic, strategy-agnostic that describe average movement times, this work provides a richer account of the SAT in pointing and introduces tools that better reflect the variability and strategies observed in real interaction.

2 RELATED WORK

Pointing is an instance of a goal-directed movement, which has been extensively described in the motor control literature (see *e.g.*, [61, Section 2] or [18] for an introduction). The SAT in pointing has been known at least since Woodworth [67], and is classically captured by Fitts' law for goal directed movements, and Schmidt's law [57] for fast,

open loop movements. In HCI, pointing is one of the fundamental elementary interactions of users with GUI; the topic has, and continues to attract attention.

2.1 Fitts' law extensions

Many studies have extended Fitts' law, and namely its applicability to novel paradigms such as 2D and 3D pointing, by proposing a novel index of difficulty. For instance, Accot and Zhai [2] review several indices for bivariate pointing, with additional insights provided by [33], while Murata and Iwase [49] extend the framework to trivariate pointing. These models however do not provide MT distributions and overlook the impact of user strategy on performance. Another line of research focuses on modeling the spatial distribution of movement endpoints. Notable examples include the *sum-of-Gaussians* models [6, 69, 72]. While these models capture endpoint distributions, their treatment of movement time remains based on averages. Additionally, no sum-of-Gaussians model incorporates parameters that directly account for user strategy to our knowledge.

2.2 Pointing with different strategies

Most people will regularly over or under utilize the target width W [59], creating, according to Zhai et al. an additional subjective layer of speed-accuracy tradeoff [71]. This strategic choice may be affected by context, for example people will favor accuracy in when safety matters critically [37]. Various works have thus investigated the effect of strategy on pointing performance. This includes work on the invariance of throughput for different user strategies [4, 46, 54] and the comparison of evaluations of different interaction techniques [68], but has also lead researchers to define variations on Fitts' protocols, further described Appendix A. Here we focus on existing models that account explicitly for strategy.

2.2.1 The effective index of difficulty ID_e. Because participants do not always follow the accuracy constraint as enforced by W, Crossman [11] suggested a correction that was popularized in HCI by Mackenzie [30, 59], called the effective index of difficulty[1] ID_e, which is a function of the actual measured standard deviation of endpoints σ

$$\mathrm{MT} = a + b\ \log_2\left(1 + \frac{D}{4.133\sigma}\right) \equiv a + b \log_2\left(1 + \frac{D}{W_e}\right) = a + b\,\mathrm{ID}_e. \tag{3}$$

ID_e, as a function of σ, is a direct measure of accuracy, and may only be computed a posteriori on observed data.

2.2.2 The index of target utilization. Zhai *et al.* [71] investigated the effect of strategy on pointing by introducing an index of target utilization

$$I_U = \log \frac{W_e}{W}. \tag{4}$$

I_u measures the effect of strategy, because, as shown in [30, Appendix 3], when the spatial distribution of endpoints is Gaussian[2] $W_e = \frac{2.066}{\sqrt{2}\mathrm{erf}^{-1}(1-\varepsilon)} W$, where ε is the rate at which participants miss the target. Thus, I_u is a non-linear transformation of the miss rate ε. The key finding from Zhai et al. [71] for our purposes is that the post-hoc correction W_e of Equation 3 does not fully eliminate the influence of participant strategy: even when using ID_e, the parameters of Fitts' law vary with strategic choices. Moreover, they showed that D and W have effects on movement time beyond their ratio D/W. These results imply that ID_e should not be treated as a pure one-dimensional proxy for task difficulty, but rather as a quantity shaped jointly by ID, D, W, and participant strategy. This observation motivates our modeling of its distribution, see section 4.

[1] The index generalizes to multiple dimensions, but not without controversies, see *e.g.*, [25, 45, 66].
[2] which is a reasonable assumption that holds well in practice [28, 59]

The main outcome of their study that is relevant to the present work is that the post-hoc correction W_e of Equation 3 can not fully counterbalance strategic choices by the participants *i.e.*, the parameters of Fitts' model depend on participant strategy even when using ID_e, and that there are effects of D and W on movement time other than through the ratio D/W. Together, this motivates modeling ID_e as a function of ID, D, W and strategy.

2.2.3 The WHo model. To our knowledge, Guiard and Rioul's WHo model [38], which follows from the ressource allocation model described [37], is the only parametric model that describes pointing data with an explicit parameter for strategy. However, the WHo model does not describe a distribution of movement times. Instead it characterizes the "best average" samples, which is operationalized by fitting not on all data points but only on the convex hull of block averages: In their work, the WHo model is fit on only 10 block averages out of 400, but Guiard and Rioul never described what the distribution of movement times could look like within their WHo model.

2.3 Pointing models that describe the distribution of MT

2.3.1 Fitts' law, regression, and the implicit noise model . The expression for Fitts' law Equation 3 masks an underlying statistical model that assumes additive, centered noise[3]:

$$\mathrm{MT} = a + b\mathrm{ID}_e + \text{noise (additive)}, \quad \mathbb{E}[\text{noise}] = 0 \text{ (centered)}. \tag{5}$$

Pearson's r is commonly used as a goodness-of-fit measure. However, as shown by Drewes [17] and Gori *et al.* [32], the Pearson correlation between MT and ID_e, $r(\mathrm{MT}, \mathrm{ID}_e)$, often yields moderate to low values in practice on pointing datasets, with some cases showing correlations close to zero [30].

In fact, to mitigate variability in MT and increase the goodness-of-fit measure, researchers typically compute the sample average of MT, denoted in the remainder of the work as $\overline{\mathrm{MT}}$[4], before calculating r. Consequently, the model actually considered by many researchers is:

$$\overline{\mathrm{MT}} = a + b\mathrm{ID}_e + \text{noise (additive)}, \quad \mathbb{E}[\text{noise}] = 0 \text{ (centered)}, \tag{6}$$

with the goodness-of-fit measure being $r(\overline{\mathrm{MT}}, \mathrm{ID}_e)$. Fitts' law is thus essentially a Gaussian model for average pointing time $\overline{\mathrm{MT}}$.

2.3.2 Asymmetric MT distributions and the Exponentially Modified Gaussian (EMG) noise model . Rather than averaging out the noise as in Equation 6 and describing only $\overline{\mathrm{MT}}$, recent works model the entire noise distribution instead, and show that the conditional distribution of MT for a given ID or ID_e level is distributed according to an asymmetric, positive (right) skewed [9, 24, 27, 41, 43, 53, 73] distribution. While several distributions have been suggested, including gamma, Gumbel and lognormal, we follow Gori and colleagues [24, 27, 43], and Zhao *et al.* [73] who considered an Exponentially Modified Gaussian (EMG) noise model: it is positive, is assymetric, has parameters that are easy to interpret and has shown good performance, including when modeling steering data [64].

Gori [24] further showed that a *quadratic variance* model, where the standard deviation of the MT distribution increases linearly with its mean, provided a good fit to empirical data, which was later confirmed by Zhao et al. [73] who similarly found that quadratic variance models provided the best fit among the models they tested.

[3]Because most researchers applying Fitts' law rely on ordinary least squares (OLS), they also implicitly assume the noise is uncorrelated and homoscedastic (has equal variance), as OLS is only optimal under these conditions.
[4]See [31] for further discussion on considering the sample average.

The probability density function of the EMG model with linear increasing standard deviation reads

$$f_{\mathrm{MT}}(mt) = \frac{1}{2\lambda\mathbf{x}} \exp\left(\frac{1}{2\lambda\mathbf{x}}(2\beta\mathbf{x} + \frac{s^2}{\lambda\mathbf{x}} - 2mt)\right) \operatorname{erfc}\left(\frac{\beta\mathbf{x} + \frac{s^2}{\lambda\mathbf{x}} - mt}{\sqrt{2}s}\right), \tag{7}$$

where erfc is the complimentary error function, $\mathbf{x} = [1, \mathrm{ID}_e]^T$, $\lambda = [\lambda_0, \lambda_1]$ and $\beta = [a, b]$. Although this density is analytically quite complex, it has a simple conditional mean, where $\overline{\mathrm{MT}}$ increases linearly with ID_e

$$\mathbb{E}[\mathrm{MT}|\mathrm{ID}_e] = \beta\mathbf{x} + \lambda\mathbf{x} = a + \lambda_0 + (b + \lambda_1)\mathrm{ID}_e, \tag{8}$$

and a quadratic conditional variance, where the variance of MT increases quadratically with ID_e

$$\mathrm{Var}(\mathrm{MT}|\mathrm{ID}_e) = s^2 + (\lambda\mathbf{x})^2 = s^2 + \lambda_0^2 + 2\lambda_0\lambda_1\mathrm{ID}_e + \lambda_1^2{\mathrm{ID}_e}^2. \tag{9}$$

Throughout this work, we fitted the EMG distributions with maximum likelihood estimation.

2.4 Models that simulate entire trajectories

2.4.1 Optimal feedback control models. Several optimal feedback control (OFC) models for pointing have been applied in HCI [20, 42], following from the seminal work by Todorov and Jordan [60]. These predict complete movement trajectories as well as their spatial variability. Unfortunately, these models are not well suited to predict movement time and its associated variability, since OFC doesn't have an intrinsic mechanism to stop controlling, and thus it requires defining a stoping condition (see *e.g.*, [56, Section 3.4]). [5] Solutions to this problem can be considered, such as integrating an explicit cost for time, as in [5], but to our knowledge there is currently no OFC model that predicts realistic MT distributions. Reinforcement learning approaches, which can be viewed as methods to solve optimal control problems when the dynamics of the controled system are not assumed, share the same limitation: although they can learn policies that lead to realistic movement trajectories and can produce data compatible with Fitts' law [10, 40], they similarly require stopping criteria or cost of time [20], and to our knowledge, have never been shown to generate realistic distributions of movement times.

2.4.2 Simulation models. Optimal control models typically have closed-form solutions or can be computed numerically with sufficient speed to enable real-time implementation. In contrast, more complex approaches, to which we refer as simulation models, have also been proposed. For instance, Do et al. [15] present an intermittent control model for pointing that incorporates various sources of variability, including motor and perceptual noise. However, this model does not report movement time (MT) variability, and the role of user strategy in its framework remains unclear. Additionally, the model required six days of training. Similarly, Moon et al. [47] introduce a movement model trained using proximal policy optimization, a deep reinforcement learning technique, but the study does not report on MT variability and required approximately 40 hours of training. Thus, while simulation models can in principle describe MT variability, to our knowledge, no such results have been reported. Moreover, while it should be possible, it remains unclear at this point how these models could be adapted to account for user strategy. Finally, their lengthy training times and the efforts required for their implementation undermine their utility as simple, parametric models, which are the focus of this work.

[5] There are two possibilities: either a fixed horizon is used, which requires specifying in advance the movement duration, or an infinite or receding horizon is used. In that case, movements are perpetual, except if an arbitrary stopping conditions is formulated.

3 PRELIMINARIES

3.1 Theoretical analysis of the $r^2(\overline{\mathbf{MT}}, \mathbf{ID}_e)$ correlation

The goodness of fit of Fitts' law is usually justified via Pearson correlation $r^2(\overline{\text{MT}}, \text{ID}_e)$, which is close to 1 in many experiments where accuracy is manipulated by W [59]. It turns out that many MT distributions can achieve $r^2(\overline{\text{MT}}, \text{ID}_e)$ close to 1, which makes it a weak criterion for assessing model fit.

Proposition 3.1. *Any conditional distribution $p(MT|ID_e)$ will reach $r^2(\overline{MT}, ID_e) = 1$ if it has a conditional expectation $\mathbb{E}[MT|ID_e]$ that is linear in ID_e.*

The proof is given in Appendix B, and relies on the fact that Pearson correlation is a function of the covariance, which measures linear deviations from the mean. Several distributions that have been previously used for modeling pointing times satisfy this condition

- The EMG model with quadratic variance [24, 27, 43, 73] (Equation 8) has linear conditional mean.
- A gamma [41] distribution with shape k and scale θ has mean value $k\theta$. Having k or θ increase linearly with ID_e would fulfill the condition.
- A generalized extreme value [9] distribution $\text{GEV}(\mu, \sigma, s)$ has a mean value that increases linearly with μ. Having μ as a linear function of ID_e would thus also fulfil the condition.

There are many more distributions possible, since one can force a conditional linear mean into any continuous distribution whose mean can be set freely. Two important takeaways follow: First, models of movement time must have a conditional mean linear in ID_e to satisfy Fitts' law. Second it motivates the need to turn to measures of dependences that are more sophisticated than Pearson correlation.

In practice, $\overline{\text{MT}}$ may refer to different aggregated quantities, depending on how data are collapsed across conditions, participants, and repetitions [54]. For example: (i) averaging across all (D,W) pairs [59], (ii) averaging over ID levels (noting that multiple (D,W) pairs can map to the same ID [32]), or (iii) aggregating across participants [1, 46]. When blocks are repeated, aggregation may again be performed over repetitions [36, 41]. Each additional aggregation step tends to inflate $r^2(\overline{\text{MT}}, \text{ID}_e)$ by the same mechanism outlined in the proof of the Proposition.

3.2 Measuring the dependence between two random variables with bivariate copulas

In the previous section, we argued that Pearson correlation r^2 is a limited tool for evaluating Fitts' law models because it only reflects linear associations between mean MT and ID_e, and ignores other aspects of the joint distribution. To move beyond this limitation, we need methods that capture dependence structures more generally.

Copulas are mathematical functions that describe the dependence structure between two random variables, independent of their individual distributions. Sklar's theorem [52] guarantees that any bivariate joint distribution can be decomposed into two parts: the marginals, which describe the behavior of each variable separately, and a copula, which describes how the two variables are linked. Thus copulas "couple" [52, Chapter 1] two marginals, and capture the dependence between two variables even when the actual scales or distributions of the variables have been factored out, going well beyond what linear correlation can reveal.

To illustrate the usefulness of copulas, consider a model for generating synthetic human height and weight data. Both height and weight are likely to follow a Gaussian distribution, and with access to a basic anthropomorphic dataset, we can estimate their parameters. A simple, yet flawed, approach to generating realistic human data would be to treat height and weight as independent variables: first sampling height, then independently sampling weight. This approach,

however, ignores the natural correlation between height and weight: taller individuals tend to weigh more than shorter individuals, and this relationship needs to be captured for a realistic model. Copulas allow us to model the dependence structure directly; different copulas specify different forms of dependencies between weight and height.

Figure 1 illustrates how the Gaussian copula shapes a joint distribution: In (d), the two marginals (a) and (b) are independently combined. In (e), the two marginals are coupled with a Gaussian copula (c). For this illustration, we have used realistic marginal distributions for MT (a) and ID (b). Note that in both jointplots (d) and (e), the estimated marginals are identical, but notice how the joint distribution differ. In particular, the distribution obtained with the Gaussian copula has an increase in variance with ID levels that is characteristic of pointing data [24].

Copulas are parametric and describe how the two variables are associated, for example imposing a prescribed (rank) correlation, everywhere, or only at their tails. Different types of copulas introduce different types of dependencies; for example a Gumbel copula introduces so-called upper tail dependence [52], where if one variable takes on an "extreme" high value, the other is also likely to take on an "extreme" high value. These copulas are used often for variables that "reinforce" themselves *e.g.*, when modeling extreme financial events [14]. Identifying the best fitting copula fully describes the dependence structure between two variables, which include linear/non linear and tail dependencies. This can be done with standard inference methods such as maximum likelihood estimation; several software libraries exist to perform these inferences off the shelf. The way we identify copulas in this work is further described subsection C.1. All copulas used in this work can be visualized for two sets of marginal distributionsFigure 24.

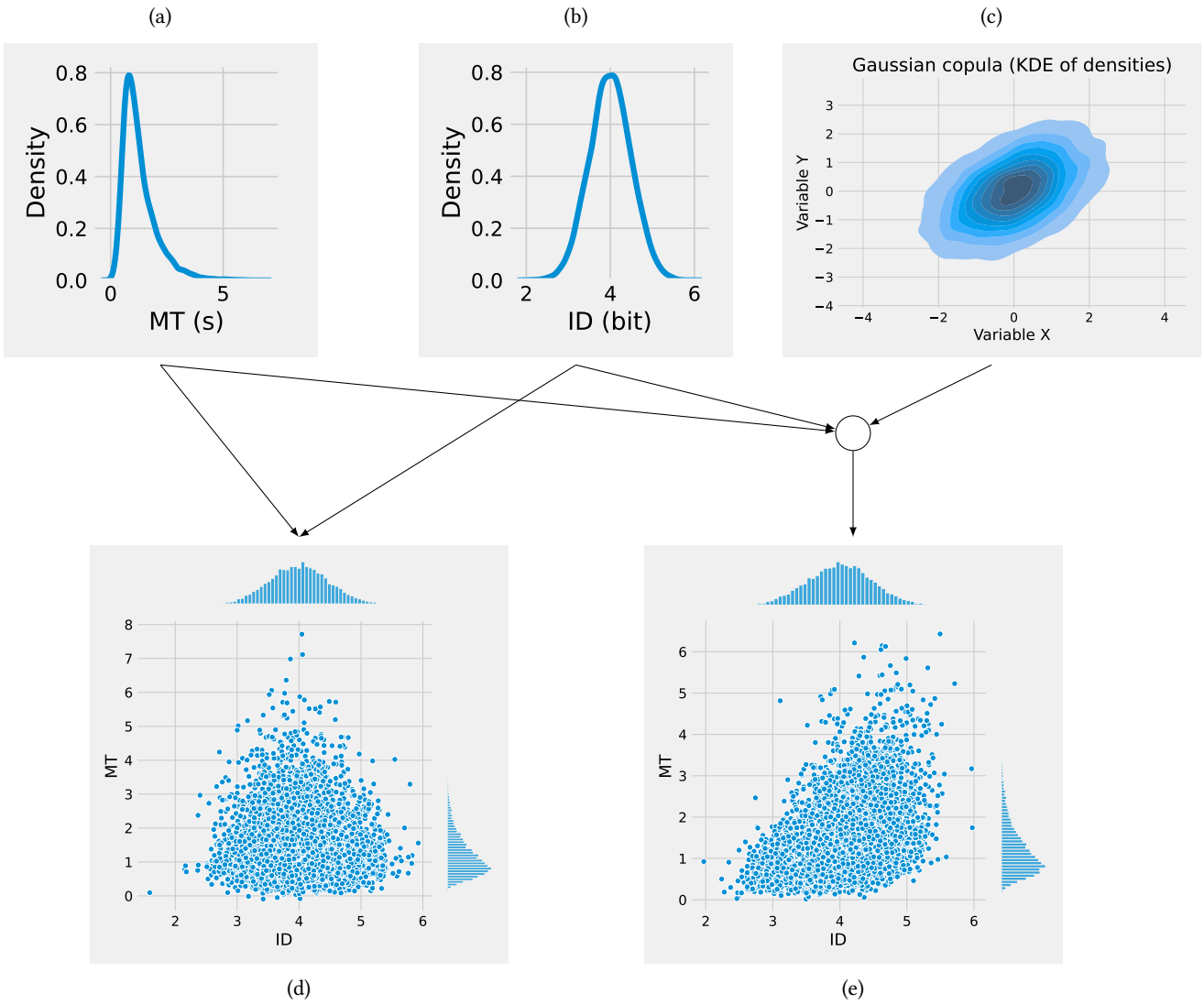

Fig. 1. How copulas couple two marginals. (a) represents the marginal distribution for MT, here an EMG model. (b) represents the marginal distribution for ID, here a Gaussian model. One can sample independently from MT and ID to obtain a joint (MT, ID) distribution as in (d), however, it does not reflect the fact that low ID movements are usually shorter. (c) is a Gaussian copula, which introduces a dependence where both low and high values have more chance of occurring together. This creates a joint distribution (e) where low (resp. high) values of ID are associated with low (resp. high) values of MT, as characteristic of empirical pointing data.

4 OVERVIEW OF THE EMPIRICAL ANALYSES OF LOCAL AND GLOBAL SPEED-ACCURACY DEPENDENCIES

In this work, we model the joint distribution of movement time (MT) and effective index of difficulty ID_e at two levels of analysis, by re-analyzing data from three previously conducted experiments. At the *global* level, we consider the

distribution aggregated across all task conditions:

$$p(\mathrm{MT}, \mathrm{ID}_e). \tag{10}$$

At the *local* level, we condition on task parameters D and W, and strategy s:

$$p(\mathrm{MT}, \mathrm{ID}_e | s, \mathrm{D}, \mathrm{W}). \tag{11}$$

The local distribution captures how MT and ID_e relate within a single experimental condition, for given values of D, W, and s[6]. The global distribution pools data across conditions and strategies, capturing overall patterns in the dataset. These two perspectives lead to different conclusions: within individual conditions, MT and ID_e often appear weakly dependent. When conditions are aggregated however a global trade-off consistent with the SAT emerges. Figure 2 illustrates this effect: locally, dependencies are weak, but globally, high values of MT align with high values of ID_e.

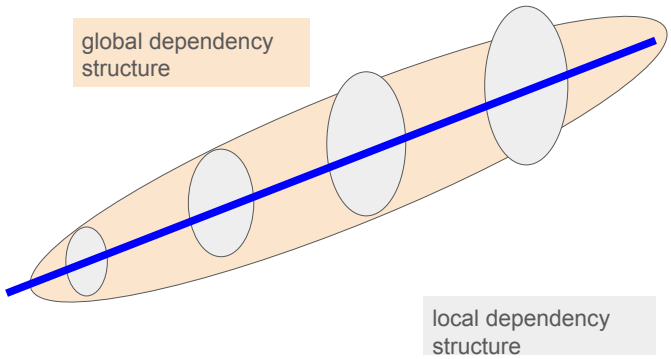

Fig. 2. The difference between the local and global dependency structure. The blue ellipses, which represent 95% confidence ellipses for data associated with one block, display an uncorrelated structure between MT and ID_e, since they are oriented vertically. The yellow ellipsis, which represents the 95% confidence ellipsis for the aggregated data shows that MT is correlated with ID_e because it is slanted.

To estimate these joint distributions, we consider two complementary approaches.

(1) **Conditional EMG approach.** We use the chain rule of probability:

$$p(\mathrm{MT}, \mathrm{ID}_e | \mathrm{D}, \mathrm{W}, s) = p(\mathrm{MT} | \mathrm{ID}_e, \mathrm{D}, \mathrm{W}, s)\ p(\mathrm{ID}_e | \mathrm{D}, \mathrm{W}, s). \tag{12}$$

We then make the simplifying assumption that MT depends on task parameters only through ID_e:

$$p(\mathrm{MT} | \mathrm{ID}_e, \mathrm{D}, \mathrm{W}, s) = p(\mathrm{MT} | \mathrm{ID}_e). \tag{13}$$

This reduction avoids considering too many variables and uses the fact that ID_e already summarizes task geometry and strategy *i.e.*, the remaining effect of s, D and W on MT are likely second order. The conditional

[6]Depending on the protocol, not all of D, W and s may be available. For example, in a standard Fitts protocol, we will look at $p(\mathrm{MT}, \mathrm{ID}_e | \mathrm{D}, \mathrm{W})$.

distribution of MT given ID_e can be modeled with the exponentially modified Gaussian (EMG) described in subsubsection 2.3.2, which assumes a linear conditional mean and quadratic conditional variance. The remaining term, $p(\mathrm{ID}_e \mid D, W, s)$, is then modeled, enabling the model to be used not just descriptively, but *predictively*. Because the EMG has already been validated in prior work, our goal here is not to re-compare candidate models for $p(\mathrm{MT}|\mathrm{ID}_e)$. Instead, we focus on showing how the distinctive features that make the EMG model successful (asymmetry and quadratic variance) contribute to better fits, while placing greater emphasis on modeling $p(\mathrm{ID}_e|\mathrm{D}, \mathrm{W}, s)$ which has received less attention and is central to understanding SAT.

(2) **Copula-based approach.** The EMG approach fixes the dependency structure between MT and ID_e through its quadratic variance model. To allow more general forms of dependence, and thus study the SAT from a perspective that is agnostic to any form of dependence, we instead model the joint distribution using copulas, which capture the dependency between two variables independently of their marginals. This approach removes the restriction imposed by the EMG variance model and provides a more flexible description of MT and ID_e.

Learning the global dependency $p(\mathrm{MT}, \mathrm{ID}_e)$ proceeds analogously by dropping the conditioning on D, W, and s.

We apply these two modeling approaches to three datasets collected under progressively richer experimental protocols, moving from a classic Fitts' law task and a strategy-based task to a combined task.

Study 1: Fitts protocol. In the first study (section 5), we re-analyze data a the standard Fitts protocol (subsection A.1), where accuracy is controlled by target width W. Because this dataset has already been used to validate the EMG distribution previously [24, 73], we focus here on copulas. Several candidate copula models were considered to describe the global and local dependencies; the rotated Gumbel copula was selected as the best fit for the global distribution, capturing strong correlation in the lower tail. For the local distributions, a low-correlation Gaussian copula provided the best fit. A linear model relating ID_e to ID is proposed to form the conditional distribution $p(\mathrm{ID}_e|\mathrm{ID}, \mathrm{D}, \mathrm{W})$. This study shows that the speed-accuracy tradeoff emerges globally, not locally.

Study 2: Pure strategy protocol. In the second study (section 6), we analyze data from a pure strategy protocol (subsection A.3), where accuracy is determined only by the participant's strategy s. As a first step, we examine the joint distribution of $\overline{\mathrm{MT}}$ and ID_e, and propose a bivariate Gaussian model parametrized by s. This simplified, aggregated view, highlights systematic effects of strategy on speed and accuracy without the additional variability present in single trials. This is important since unlike task parameters D and W, the influence of strategy alone on the speed-accuracy trade-off is less well modeled. Thus, this preliminary analysis provides a new, simple model and provides a broad trend and some intuition for the more detailed analysis that follows. Next, we show that the EMG model with quadratic variance also fits this dataset, extending its applicability. Finally, we find that the Gaussian and t-copulas provide the best fit for $p(\mathrm{MT}, \mathrm{ID}_e|s)$ and $p(\mathrm{MT}, \mathrm{ID}_e)$. This study shows that user strategy directly shapes the dependency structure, highlighting the importance of modeling strategy explicitly.

Study 3: Fitts-with-strategy protocol. In the third study (section 7), we combine both approaches to manipulate accuracy, by analyzing data from a Fitts-with-strategy protocol (subsection A.4), where accuracy is influenced by both target width W and strategic instructions s. We model the local distribution $p(\mathrm{MT}, \mathrm{ID}_e|s, \mathrm{D}, \mathrm{W})$, and confirm that the EMG with quadratic variance and asymmetric distribution provides good fits. As in the previous study, Gaussian and t-copulas yield the best joint models. We also extend the linear model relating ID_e to ID by including a term for strategy. This extension makes explicit how strategy interacts with task difficulty to shape user accuracy, as captured through ID_e.

5 FIRST STUDY: DEPENDENCE BETWEEN ID_e AND MT IN FITTS' PROTOCOL

5.1 Dataset presentation

We used an existing dataset by Jude et al. [41] (JGP). The JGP dataset was generated using a 2D multi-directional tapping task [59] using the FittsStudy software [66], with gestural interaction as input method. The experiment was replicated six times over three days, twice a day, for 15 participants with factors W (64, 96, 128 (px)) and D (256, 512, 1024, 1408 (px)) fully crossed, which makes 10 unique ID levels ranging from 1.52 to 4.58 bit. At 20 trials per block, the entire dataset consists of 21600 movements. We noticed a bug in the dataset we downloaded, where the 2 last replications are identical to the two first, so we discarded the 2 last replications from the dataset, leaving us with 4 replications (14400 trials remaining). The JGP dataset is pre-processed; ID_e and MT are already calculated.

5.2 Fitting EMG on MT given ID_e values

We fitted the EMG model for each participant and for each replication. In line with existing works on this dataset [24, 73], the EMG fit was largely superior to the Gaussian fit obtained via linear regression; We namely found AIC differences routinely above 300; typically a difference of 10 is already considered good support. A pair-plot of the fitted parameters can be found in the supplementary materials.

5.3 Fitting copulas on (MT, ID_e)

5.3.1 Local dependency structure. To test whether the SAT appears within individual task conditions, we examined the local dependency between MT and ID_e through $p(\mathrm{MT}, \mathrm{ID}_e|\mathrm{D}, \mathrm{W})$. If the SAT held locally, one would expect systematic dependence between the two variables when D and W are fixed.

We compared candidate copula models using the model evidence ratio $\mathcal{R}$ as explained in subsection C.2. This ratio expresses how likely one model is relative to the best-fitting model: for example, $\mathcal{R} = 0.1$ means the model is about 10 times less likely to be the true model than the current best fitting. This allows us to compare copulas on a common scale.

Pooled over participants and across D×W pairs, the independent copula was most often selected as the best-fitting model (see Figure 3). This indicates that within a given condition, longer movement times were not reliably associated with higher effective difficulty. In other words, the local dependency between MT and ID_e was weak or absent.

An exception occurred for the smallest target condition ($W = 64$), where both Spearman's ρ and Kendall's τ correlations were slightly negative. This suggests that in this extreme case, faster movements could even be associated with greater accuracy, the opposite of the expected SAT. While weak, this paradoxical result highlights that the large trial-to-trial variability can obscure or even invert the trade-off locally.

These findings suggest that the SAT is not strongly expressed at the local level, where performance appears largely independent once task parameters are fixed.

5.3.2 Global dependency structure. We then turned to the global dependency $p(\mathrm{MT}, \mathrm{ID}_e)$, pooling data across task conditions. In contrast to the local case, clear evidence of dependence emerged.

The rotated Gumbel copula consistently provided the best fit across participants (Figure 4). This model captures strong lower-tail dependence, meaning that low MT values align closely with low ID_e values. This means that when tasks are easier, movements are faster and much more consistent; as tasks become more difficult, variability in MT increases and the relationship between ID_e and MT weakens. The rotated Galambos copula was the second-best model, this can likely be explained by the fact that for some parameter values it is extremely similar to the rotated Gumbel [23],

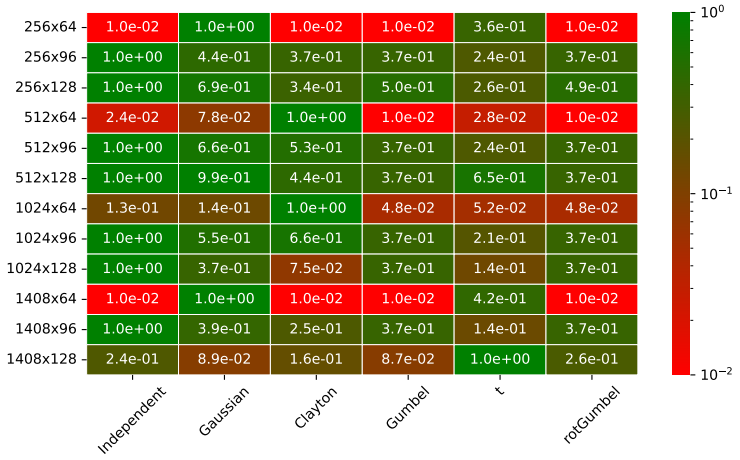

Fig. 3. Model evidence ratios ($\mathcal{R}$) for pooled participant data across D×W conditions. Except for the smallest target condition ($W = 64$), the independent copula provided the best fit, indicating no consistent local dependency between MT and ID_e. $\mathcal{R} = 0.1$ means the model is about 10 times less likely than the best-fitting copula.

Together, these results show a contrast: while the SAT is weak at the local level, it emerges clearly in the global distribution once conditions are aggregated.

5.4 Model between ID_e and D, W, ID

To test whether W and D influence ID_e beyond their joint contribution through ID, we fitted four linear mixed-effects models (LMM) with random intercepts for participants (see Appendix E for an explanation of LMM and how they were fitted and compared): a full model with ID, W, D, and all interactions, two simpler models with only ID+W or ID+D, and a baseline with ID alone. A summary of the results are displayed Table 1, while the full fits are available in the supplementary materials.

The full model is difficult to interpret due to collinearity between ID, W, and D, and model comparison shows that it was not supported relative to simpler alternatives, so it was excluded. The simpler models revealed a clear effect of D: both its main effect and its interaction with ID were significant ($\alpha = 0.05$), while W showed no significant contribution. Log-likelihoods for the D and W models were essentially equivalent. The ID-only model was rejected based on significantly lower log-likelihood. The two simple model fits are shown in Figure 5, which illustrates that D significantly shifts the ID_e-ID relationship upwards. This is concistent with general results in psychophysics where longer distances lead to overall larger variability. This result diverges from Zhai et al. [71], who reported W effects[7]. Two factors may explain the discrepancy. First, our W values (64–128 px) were larger than theirs (12–72 px), and W–ID interactions likely manifest more strongly at smaller target sizes.[8] Second, fully disentangling W from ID requires a fully crossed shape-scale design [32, 35], which was not available in our dataset. Overall, the additional effects of D and especially W beyond ID, while they can be of theoretical interest under specific experimental designs, are likely negligible in most practical contexts.

[7]Figure 5 suggests an effect of W on the slope of the ID_e-ID relation, but this effect did not turn out significant

[8]However, comparisons must be taken cautiously, since effective motor control depends on display density and input device characteristics [8].

Fig. 4. Model evidence ratios $\mathcal{R}$ for global copula fits across participants. The rotated Gumbel copula consistently outperformed other candidates.

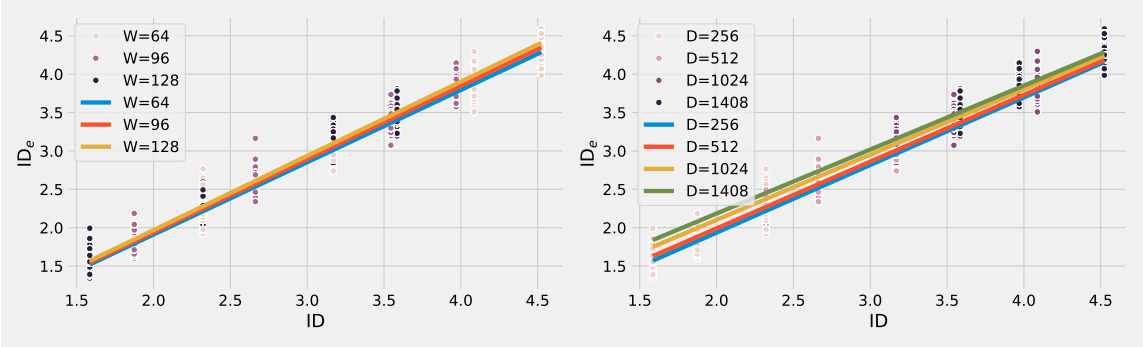

Fig. 5. The relationship between ID_e and ID depending on W values (left) and D values (left). The filled circles represent the observed ID_e values for prescribed ID levels for varying W or D values, while the lines represent the linear model fits for varying W or D values.

Table 1. Mixed effect model fit results for ??, ?? and ??. Each line (except the last, which displays the log-likelihood associated with each model) displays the estimated coefficient associated with the variable in the left column as well its associated p-value (in bold if p<0.05).

	ID ∗ W ∗ D + (1|Participant)	ID ∗ W + (1|Participant)	ID ∗ D + (1|Participant)	ID + (1|Participant)
Intercept	0.262 (p=0.402)	0.016 (p=0.833)	0.1 **(p=0.016)**	0.079 **(p<0.001)**
ID	0.846 **(p<0.001)**	0.918 **(p<0.001)**	0.890 **(p<0.001)**	0.937 **(p<0.001)**
w	-0.992 (p = 0.589)	0.162 (p=0.827)		
ID:w	-0.664 (p=0.763)	0.379 (p=0.102)		
D	0.593 (p=0.707)		0.296 **(p<0.001)**	
ID:D	-0.042 (p=0.837)		-0.040 **(p=0.025)**	
w:D	5.022 (p=0.389)			
ID:w:D	-1.378 (p=0.429)			
log-likelihood	401	401	401	380

5.5 Interpretation of the first study

This study examined dependencies between ID_e and MT under Fitts' protocol, where accuracy is imposed by target width W. At the local level, copula fits showed little to no correlation between ID_e and MT within a given condition, with the smallest target size even yielding a slight negative correlation. This suggests that, when accuracy is specified a priori, movements within a fixed (D,W) pair are essentially equivalent with respect to the SAT: variability in execution reflects motor noise more than systematic planning trade-offs. The analysis at the global level revealed a clear dependence, showing that the SAT only emerges across task conditions. The dependence was best captured by the rotated Gumbel copula, which emphasizes lower-tail dependence. This implies a strong correlation between low values of ID_e and MT *i.e.*, a small spread of the MT data for low values of ID_e, which is consistent with the quadratic variance model used in the EMG model. Finally, we analyzed the relationship between ID_e and ID. Possible D and W effects were weak, and not always significant, but likely ill-estimated due to collinearity issues. Thus, to adress collinearity issues, we fit the linear models with ridge regression with cross-validation[9]. Although detailed results are not reported here, the estimated coefficient for ID was 0.94 *i.e.*, closer to 1 than the estimates from the presented linear models. This suggests that the presented effects of D and W are inflated by collinearity. Thus, this further supports that ID is the dominant predictor for ID_e, and that additional effects of D and W are negligible.

6 SECOND STUDY: DEPENDENCE BETWEEN ID_e AND MT IN THE PURE STRATEGY PROTOCOL

In the pure strategy protocol (see subsection A.3), accuracy is not manipulated by W but by the experimenter instructing the user to conform to strategies favoring speed or accuracy.

6.1 The GO dataset

We investigate the effect of participant strategy on movement time and ID_e, using the Guiard-Olafsdottir (GO) dataset [36]. Participants had to aim at a line (1D task), with five different strategies, from speed emphasis allowing a large spread of endpoints, to precision emphasis where the goal is to target the 1-pixel-wide line. For comparison, the instruction in a

[9]Ridge regression is a technique for fitting linear models when collinearity exists among predictors, as it produces more stable and reliable estimates of effects [34]. The method introduces a regularization term controlled by a hyperparameter known as the ridge parameter. When the ridge parameter is set to 0, ridge regression reduces to the traditional ordinary least squares approach. As the ridge parameter increases, the method progressively shrinks the variance of the coefficient estimates, improving stability and reducing sensitivity to multicollinearity. The optimal value of the ridge parameter is determined through cross-validation.

standard Fitts experiment is usually between the *speed* and *balanced* condition, depending on the experimenter. The 16 participants performed 5 strategies, each replicated 5 times. We thus have 25 blocks per participant, each of about 15 to 20 movements each. We used a pre-processed version of the dataset used in [28]. One participant had partially corrupt data; we removed this participant entirely for safety. We also removed 79 outlier points, given the recording procedure was sometimes unreliable. The resulting dataset is composed of $15 \times 5 \times 5 = 375$ different experimental blocks for a total of 5858 movements.

6.2 A model between $\overline{\text{MT}}$ and ID_e parametrized by strategy

Correlation between ID_e and $\overline{MT}$ and conformity with Fitts' law. A summary of the GO dataset, shown in Figure 6 confirms that data from the pure strategy protocol conforms with Fitts' law Equation 3. Participants covered a wide range of ID_e (1 to 9 bits), depending on the strategy. While aggregating data per strategy yields a very high correlation between $\overline{\text{MT}}$ and ID_e ($R^2 = 0.97$), block-level averages show more variability ($R^2 = 0.58$)[10]. To capture how strategy affects behavior without overly masking this variability, we first model $\overline{\text{MT}}$ conditional on strategy. This provides a simple model of strategic effects, useful for simulation and prediction, and helps to build intuition for the subsequent trial-by-trial analysis.

A bivariate Gaussian model for $\overline{MT}$ and ID_e for each strategy. To characterize how each strategy affects MT and accuracy, we fit a bivariate Gaussian to the observed ($\overline{\text{MT}}$, ID_e) data for each of the five strategies in the GO dataset:

$$\begin{pmatrix} \text{ID}_e \\ \overline{\text{MT}} \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_i \\ \mu_t \end{bmatrix}, \begin{bmatrix} \sigma_i^2 & r\sigma_i\sigma_t \\ r\sigma_i\sigma_t & \sigma_t^2 \end{bmatrix} \right) \tag{14}$$

where μ_i, σ_i^2 are the mean and variance of ID_e, μ_t and σ_t^2 those of $\overline{\text{MT}}$, and r the Pearson correlation. We chose the bivariate Gaussian model because it is simple, easy to interpret, it makes the dependence between $\overline{\text{MT}}$ and ID_e completely specified via a single parameter (Pearson's r). The conditional mean of MT given ID_e is linear:

$$\mathbb{E}[\overline{\text{MT}}|\text{ID}_e] = \mu_t + r\frac{\sigma_i}{\sigma_t}(\text{ID}_e - \mu_i). \tag{15}$$

which, in line with Proposition 3.1, ensures that, at the level of strategy, the model is consistent with Fitts' law while still capturing variability in both MT and ID_e.

Gaussian fits for each strategy in the GO dataset. Each ellipse shows the 95% prediction region of a bivariate Gaussian fitted to (ID_e,$\overline{\text{MT}}$). The separation of ellipses reflects overall differences between strategies, while their orientation shows the local correlation between ID_e and $\overline{\text{MT}}$[11]. Strategies produce higher means and larger spreads with consistent positive correlation, whereas the *speed emphasis* condition shows little to no correlation.

Strategy Dependent Bivariate Gaussian model for $\overline{MT}$ and ID_e. We combined the five strategy-specific Gaussian fits into a single model by assigning each strategy a numerical score on [-1,1] where -1 stands for complete speed emphasis, 0 for balanced, and +1 for complete precision emphasis. This simple, symmetric coding allows us to examine how the Gaussian parameters vary with strategy, without relying on post-hoc measures such as ID_e.

[10] It is usual to report very high values of R^2 in Fitts law evaluations, but this is for a large part due to the practice of aggregating and averaging data, which masks the actual data's variability [32].

[11] The covariance matrix of a Gaussian bivariate is the rightmost term of Equation 14 where ρ is the correlation between the two components. The angle of the prediction ellipsis is given by $\frac{1}{2}\arctan\left(\frac{2\rho\sigma_i\sigma_t}{\sigma_i^2-\sigma_t^2}\right)$.

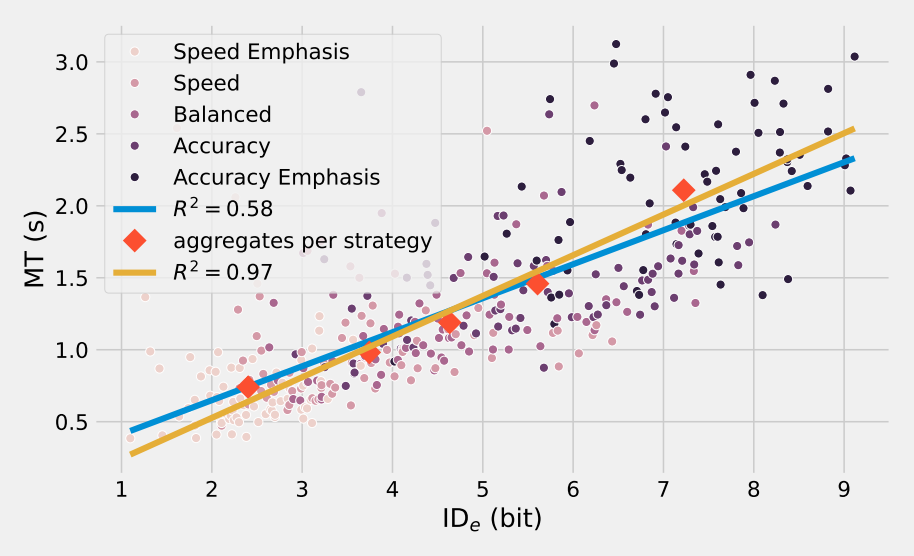

Fig. 6. Fitts' law evaluated

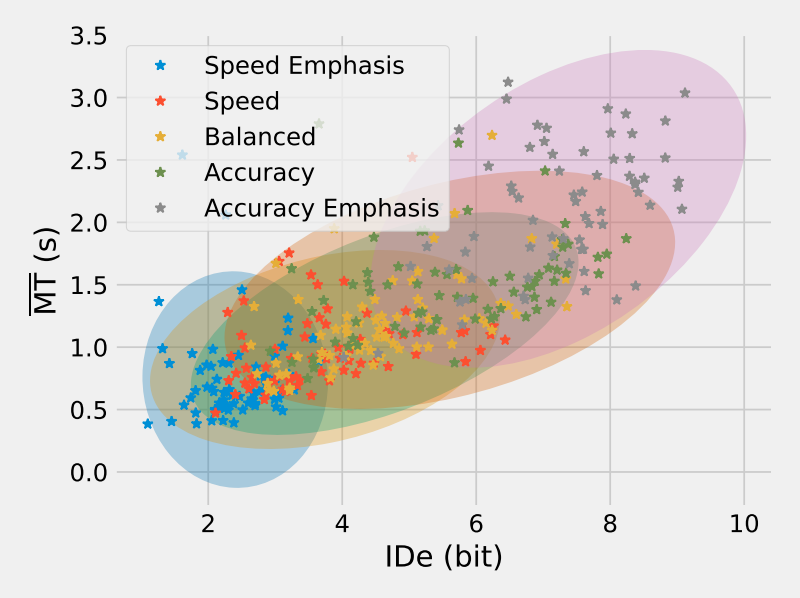

Fig. 7. Gaussian fits for each strategy in the GO dataset. Each ellipse shows the 95% prediction region of a bivariate Gaussian fitted to (ID_e,$\overline{MT}$). The separation of ellipses reflects overall differences between strategies, while their orientation shows the local correlation between ID_e and $\overline{MT}$. Strategies oriented toward precision produce higher means and larger spreads with consistent positive correlation, whereas the *speed emphasis* condition shows little to no correlation.

We compared linear models of each Gaussian parameter against constant (no-effect) models using AIC (Table 2). It results in three distinct cases:

(1) Both μ_i and μ_t increase linearly with strategy, with strong statistical support. This captures the intuitive shift from faster but less precise movements under speed emphasis to slower, more precise movements under precision emphasis.
(2) The spreads (σ_i, σ_t) show weaker trends. For σ_i, the constant model was favored overall, but the speed condition was a clear outlier. For σ_t evidence slightly favored a linear increase.
(3) The correlation r is best described as constant across strategies, except for the speed emphasis condition where it collapsed toward zero. In other words, under speed emphasis, accuracy decouples from mean movement time.

Table 2. Model comparison for μ_i, μ_t, σ_i, σ_t, ρ. The table gives the AIC for each model fit; lower is better, differences of about 10 indicate substantial support for the lower AIC model. The AIC in bold signals this was identified as the best model.

parameter of Equation 14	strategy model AIC	constant model AIC
μ_i	**1**	21
μ_t	**-4**	9
σ_i	1	**3**
σ_t	**-17**	-10
ρ	-3	**1**

Crucially, the linear dependence of μ_i and μ_t on strategy implies that their relationship is itself linear, reproducing a Fitts-like law across strategies:

$$\mu_i = \alpha' + \beta' s \longrightarrow s = \frac{\mu_i - \alpha'}{\beta'} \tag{16}$$

$$\mu_t = \alpha + \beta s = \alpha + \frac{\beta}{\beta'}(\mu_i - \alpha') = \alpha - \frac{\beta}{\beta'}\alpha' + \frac{\beta}{\beta'}\mu_i, \tag{17}$$

Thus, the Gaussian model is compatible both with local Fitts' law (within a given strategy, via Equation 15) and with global Fitts' law (across strategies, via the linear trend in means). The model therefore provides a bridge between individual strategies and Fitts' law.

6.3 EMG fitting

To complement the Gaussian bivariate analysis, we next examined whether the EMG model, originally fitted on data from Fitts' protocol, also accounts for the distributions observed in the pure strategy protocol. We compared two model comparisons:

- First we compared a Gaussian model with constant variance against one with quadratic variance, to evaluate the effect of the variance model.
- Second, we compared the Gaussian model with quadratic variance to the EMG model with quadratic variance, to evaluate the impact of the shape of the distribution: symmetric (Gaussian) versus asymmetric with long tails (EMG).

Figure 9 summarizes the model comparisons via AIC differences. Two main findings emerge. First, allowing variance to scale with difficulty is strongly favored over constant variance, confirming that heteroscedasticity (non constant

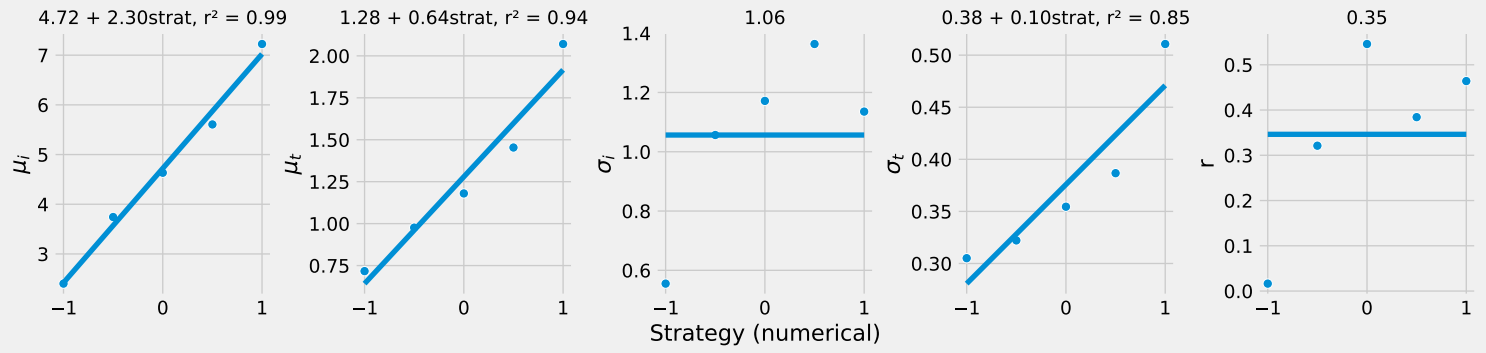

Fig. 8. The parameters of the mean and covariance of the bivariate Gaussian model defined in Equation 14 plotted against the five numerical strategies. The associated linear fits and r^2 are displayed above each plot when applicable.

variance) is a general feature of movement times. Second, the EMG model consistently outperforms the symmetric Gaussian with quadratic variance, showing that skewness and long tails are not specific to the Fitts' protocol but arise under strategy manipulations as well.

Finally, a pairplot of EMG parameter estimates (see Supplementary Materials) reveals essentially the same correlation structure as in the JGP dataset, with only a single exception. This further supports the generality of the EMG description across protocols.

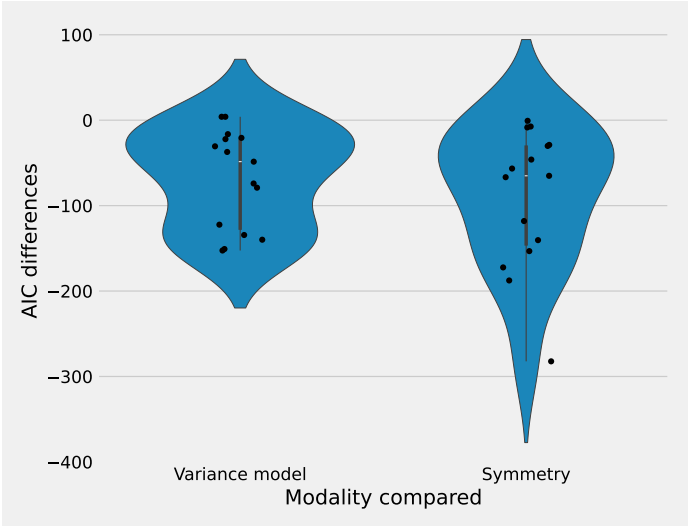

Fig. 9. Model comparisons for the pure strategy protocol. Each violin shows the distribution of AIC differences across participants, with individual values overlaid (black dots). Left: allowing variance to increase with difficulty (quadratic variance) is strongly favored over constant variance. Right: the EMG model is favored over the symmetric Gaussian (both with quadratic variance), indicating that skewness and long tails are characteristic of movement times under strategy manipulations.

6.4 Fitting copulas

To complement the Gaussian bivariate analysis, we examined the dependence between ID_e and MT using copulas. As in subsection 5.3, we compared several candidate copulas based on their model evidence ratio $\mathcal{R}$, fitting them separately for each participant and strategy ($5 \times 15 = 75$ fits).

The results displayed in Figure 10 show that the Gaussian copula provided the best local fit, except for the balanced strategy, where the Galambos copula was slightly favored. Although the Galambos family is typically associated with strong upper-tail dependence, the estimated parameter values here ($\theta \approx 0.5$[12]) make it similar to the independent copula, as can be seen Figure 24.

For the Gaussian copula, the fitted correlation parameters were small but positive (Table 3), in close agreement with the corresponding Pearson correlations obtained in the ($\overline{MT}$, ID_e) bivariate model; in particular the correlation for Gaussian copula was close to null for the speed emphasis conditions, as before.

We also pooled strategies to assess the *global* dependence structure at the participant level. As shown in Figure 11, the Gaussian and t-copulas provided the best fits for 10 of 15 participants, with an average correlation of $\rho \approx 0.71$. The t-copula degrees of freedom ν varied widely across participants, suggesting some heterogeneous tail behavior. For 4 of the remaining participants, the Galambos copula was selected, again with θ values close to independence.

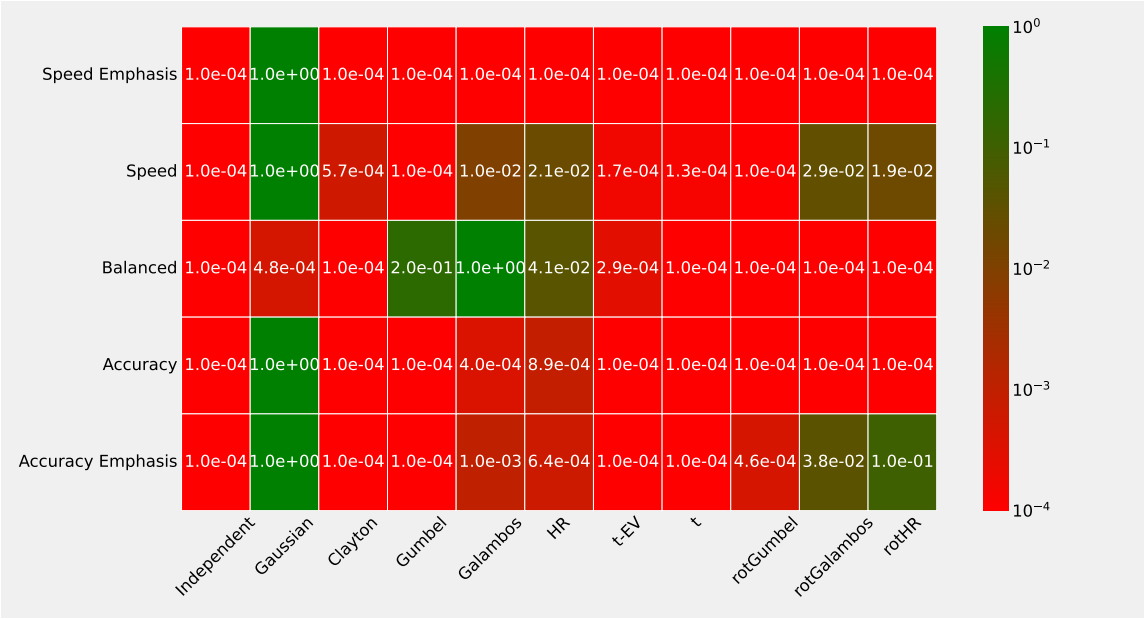

Fig. 10. Model evidence ratio $\mathcal{R}$ of the tested copulas against the best copula, for different strategies.

6.5 Interpretation of the second study

This second study examined the dependency between ID_e and MT within the pure strategy protocol, where accuracy is specified by instructions rather than target width. As in the first study, we found strong *global* dependence between

[12] A violin plot of the estimated Galambos copula parameters is provided in the Supplementary Materials.

Table 3. Mean ρ values of the Gaussian copula and associated t-statistics and p-values for the one way Student's t-test.

Strategy	T-Statistic	P-Value	Mean	Count
1 (speed emph.)	0.298	0.770	0.032	15
2 (speed)	1.67	0.118	0.134	15
3 (balanced)	6.91	**<0.001**	0.321	15
4 (accuracy)	1.30	0.214	0.117	15
5 (accuracy emph.)	3.26	**0.006**	0.253	15

Fig. 11. Model evidence ratio $\mathcal{R}$ of the tested copulas against the best copula, for different participants.

ID_e and MT when data is pooled across conditions (here, strategy), while within conditions, *local* dependence remains weak.

We first confirmed that the data from the pure strategy protocol are compatible with Fitts' law. When aggregated by strategy, ID_e and $\overline{MT}$ were almost perfectly correlated ($\rho = 0.97$). To capture variation without erasing it completely through averaging and the effect of strategy, we introduced a bivariate Gaussian model parameterized by strategy. This model is both simple and interpretable: it reproduces Fitts' law globally, and shows positive within-strategy correlations between ID_e and $\overline{MT}$, except in the speed-emphasis condition. The latter exception aligns with the idea of ballistic, feedforward movements [11, 22], where accuracy no longer relates to movement time.

Concerning distributional modeling of MT, as expected, the EMG model outperformed simpler Gaussian alternatives, thanks to its ability to capture both long-tailed distributions and variance increasing with difficulty. This replicates the findings from the Fitts protocol, suggesting that the properties making EMG effective are general characteristics of movement, and not protocol-specific.

Finally, we assessed dependence using copulas. At the local level, low correlation Gaussian copulas provided the best fit in nearly all conditions, indicating a weak but symmetric form of dependence between MT and ID_e. The balanced strategy was a slight exception, where the Galambos copula gave a marginally better fit, though with parameter values close to independence. At the global level, pooling across strategies revealed much stronger dependencies: Gaussian and t-copulas dominated[13], with an average correlation of $\rho \approx 0.71$. This shows that the speed-accuracy trade-off becomes most apparent once strategic variation is aggregated, consistent with the global-local contrast observed in the first study.

7 THIRD STUDY: DEPENDENCE BETWEEN ID_e AND MT IN THE FULL PROTOCOL

The third study builds directly on the first two by combining their respective manipulations of accuracy. Whereas the first study controlled accuracy through target width W and the second study controlled accuracy through explicit strategy instructions, the YKORM dataset [68] used a *Fitts-with-strategy* protocol in which both W and strategy were varied simultaneously (see subsection A.4). This allows us to test whether the patterns observed separately in the Fitts and pure strategy protocols also hold when both factors operate together.

7.1 The YKORM dataset

The original study's objective was to compare two existing interaction techniques (bubble cursor and bayesian touch criterion) against regular pointing, across strategies. The study used an ISO circle-of-circles tasks with distractors. The experiment fully crossed 4 factors: D (400, 770)px, W (8, 24, 70)px, strategy (fast, balanced, accurate) and a parameter for the distractor density (3 levels), which provides an ideal design to examine how task geometry and strategy together shape the joint distribution of MT and ID_e. For comparison with the two previous studies, we only used regular (mouse-based) pointing data, and aggregated data across distractor density levels. Twelve participants completed 23 trials per condition, which makes 12 participants × 3 W × 2 D × 3 strategy × 69 movements = 216 sets × 69 movements = 14904 trials. The first 3 trial of each set were discarded (648) with some outliers (244), leaving us with 14012 data points.

7.2 Conformity with Fitts' law

We first verified (see Figure 12) that the YKORM data is consistent with Fitts' law, as in the previous studies. To do so, we fit Fitts' effective law using a mixed-effects linear model with fixed effects for ID_e, strategy, and their interaction, and random intercepts for participants

$$\text{MT} \sim \text{ID}_e * \text{strategy} + (1|\text{Participant}). \tag{18}$$

We find that one unit of strategy augments the slope by approximately 73 ms/bit. Since one unit of strategy separates the speed and accuracy conditions, this amounts to a change of about 35 ms/bit between each consecutive condition, consistent with the study from Zhai et al. [71] who found differences of about 30ms/bit in a similar study.

7.3 Model between ID_e and $\overline{\text{MT}}$

As in Study 2, we modeled the joint distribution of ID_e and $\overline{\text{MT}}$ using the bivariate Gaussian formulation (Equation 14). Here, however, the YKORM dataset allows us to disentangle the effects of both strategy and ID, since each condition

[13]We consider the pair Gaussian / t-copula because the t-copula with parameters ρ and ν tends to the Gaussian copula with parameter ρ for large values of ν, just like the t-distribution is equivalent to the Gaussian distribution with many degrees of freedom.

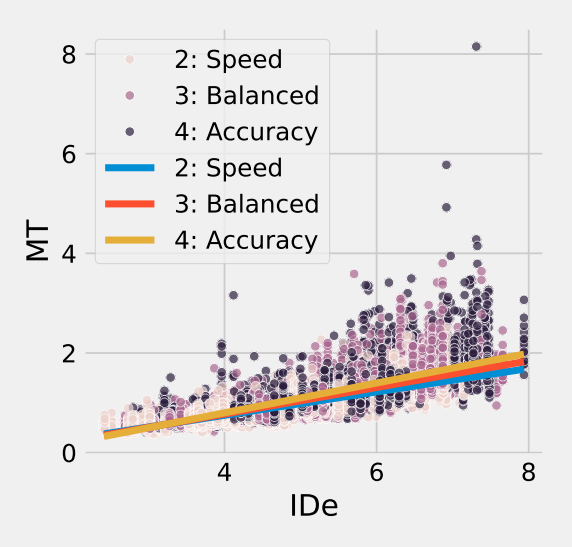

Fig. 12. Data of the YKORM dataset and adjusted Fitts' law obtained by computing linear regression on $\overline{MT}$ with ID_e and strategy as predictor. The fitted model is $\overline{MT}$ = -0.24 + 0.266·ID_e -0.244·s + 0.073·s · ID_e.

combines one of three strategies with one of six ID levels. We therefore fit the Gaussian separately for each of the 18 strategy × ID combinations; the resulting 95% ellipses are shown in Figure 13.

Compared with the pure strategy dataset (section 6), the spreads of the Gaussians are smaller. This is expected: the Fitts-with-strategy protocol constrains effective width more tightly, while strategy only modulates how closely participants adhere to the target constraints.

To examine how Gaussian parameters vary with ID and strategy, we compared nested models of decreasing complexity:

$$X \sim \text{ID} + \text{strategy (full model)} \tag{19}$$

$$X \sim \text{ID (ID model)} \tag{20}$$

$$X \sim \text{strategy (strategy model)} \tag{21}$$

$$X \sim 1 \text{ (constant model)} \tag{22}$$

where X is one of μ_i, μ_t, σ_i, σ_t, ρ as defined Equation 14.

The model comparisons results, based on AIC, are given in Table 4. Two findings merge:

- For the means, the full model with ID and strategy was preferred, (ΔAIC > 10). This confirms that both factors shift the joint distribution.
- For the variance and correlation terms, simpler models were more appropriate. σ_t was best modeled by the ID model, while σ_i and ρ were best modeled by a constant model.

Figure 14 illustrates these fits, with parameter estimates and r^2 values reported in the panel titles.

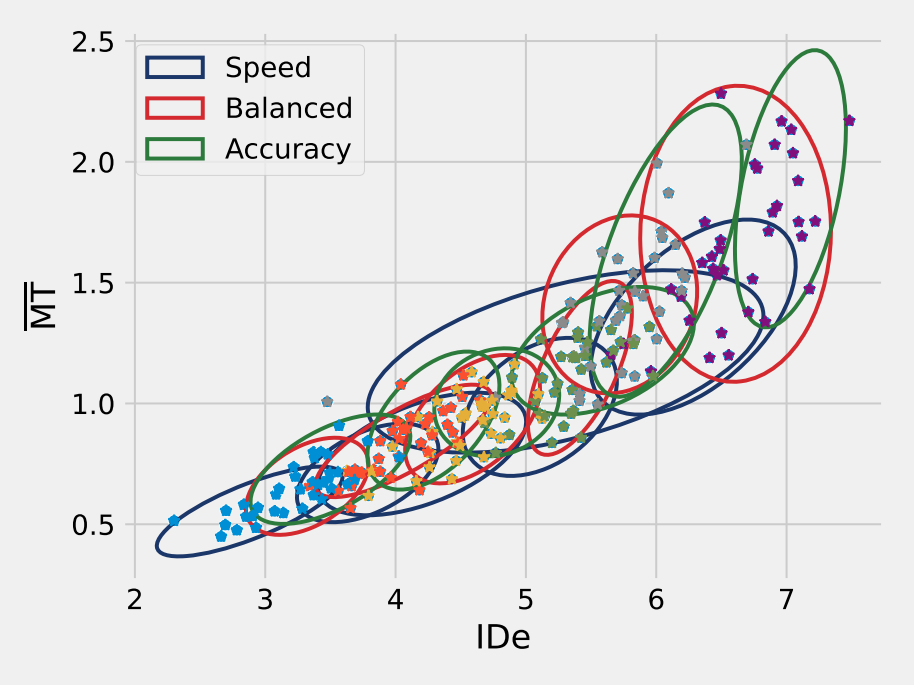

Fig. 13. 95% prediction ellipses of the bivariate Gaussian model fitted to the YKORM dataset. Each model is fit to ($\overline{\text{MT}}$,ID_e) data for a given strategy and ID level. The different colors of the points indicate different ID levels.

Table 4. Model comparison for μ_i, μ_t, σ_i, σ_t, ρ. The table gives the AIC for each model fit; lower is better, differences of 10 are meaningful. The AIC in bold signals this was identified as the best model.

Equation 14 parameters	Full model AIC	ID model AIC	strategy model AIC	constant model AIC
μ_i	**-30**	7	59	58
μ_t	**-34**	-14	17	17
σ_i	-40	-38	-40	**-39**
σ_t	-107	**-106**	-90	-91
ρ	-6	-7.5	-4	**-5.7**

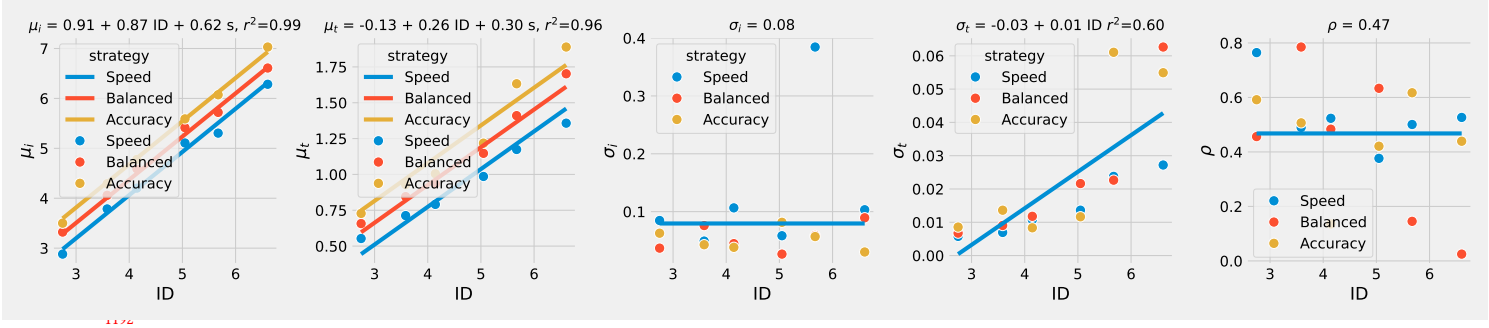

Fig. 14. Model of the bivariate Gaussian model fitted to ID and strategy levels. The model parameters are visible in the title, as well as the goodness of fit when applicable (r^2).

7.4 Model between ID_e, W, D ID, and strategy

In the first study, we examined how ID_e varied with W, D, and ID in the Fitts protocol. There, ID was by far the strongest predictor, while main or interaction effects of W and D were weak. With the YKORM dataset, we can extend this analysis by including strategy as an additional factor. This allows us to ask whether the presence of explicit strategy instructions alters the way ID_e relates to task geometry.

To keep the analysis tractable, we compared a set of simplified mixed-effect models (all with random intercepts for participants). The baseline model included ID and strategy as additive fixed effects. We then tested whether adding W or D (and their interactions with ID or strategy) significantly improved model fit:

$$ID_e \sim \text{ID} + \text{strategy} + (1|\text{Participant}) \tag{23}$$

$$ID_e \sim \text{W} * \text{ID} + \text{W} * \text{strategy} + (1|\text{Participant}) \tag{24}$$

$$ID_e \sim \text{D} * \text{ID} + \text{D} * \text{strategy} + (1|\text{Participant}) \tag{25}$$

$$ID_e \sim \text{strategy} + \text{ID} * \text{W} + \text{ID} * \text{D} + (1|\text{Participant}) \tag{26}$$

Model comparison results are given in Table 5. As expected, both ID and strategy had strong and highly significant effects on ID_e. The more complex models, which added main or interaction effects of W or D, provided no meaningful improvement in fit: the small improvement in log-likelihood was insufficient compared with the increase in parameter number ($\Delta\text{AIC} < 2$).

The baseline model is plotted in Figure 15. It shows the additive effect of strategy: accuracy-oriented strategies shift ID_e upward, while speed-oriented strategies shift it downward. The slopes with respect to ID are nearly parallel across strategies, with a small but nonsignificant interaction ($p = 0.101$).

In short, the results mirror those from the JGP dataset: ID_e is primarily driven by ID, with W and D offering little explanatory power. The new contribution here is that strategy exerts a systematic vertical shift on ID_e.

Table 5. Estimation and comparison of the 4 models for ID_e

	Equation 26	Equation 24	Equation 25	Equation 23
Intercept	1.607 **(p <0.001)**	1.034 **(p <0.001)**	0.970 **(p <0.001)**	1.003 **(p <0.001)**
strategy_num	0.604 **(p <0.001)**	0.666 **(p <0.001)**	0.646 **(p <0.001)**	0.604 **(p = 0.001)**
ID	0.716 **(p <0.001)**	0.851 **(p <0.001)**	0.845 **(p <0.001)**	0.872 **(p <0.001)**
W	-0.863 (p=0.314)	-0.783 **(p = 0.003)**		
ID:W	0.172 (p=0.722)	0.260 **(p <0.001)**		
D	-0.184 (p=0.911)		0.155 (p=0.592)	
ID:D	0.108 (p= 0.627)		0.023 (p=0.707)	
strategy_num:W		-0.181 (p=0.150)		
strategy_num:D			-0.072 (p=0.686)	
log-likelihood	-236	-239	-241	-246

Comparison to the first study and synthesis. Taken together, the JGP and YKORM datasets point to a consistent conclusion: ID_e is primarily explained by ID, with only weak and inconsistent evidence for additional effects of W or D. While Zhai et al. [71] reported stronger contributions, our re-analyses suggest these are at best second-order. What is robust, however, is the influence of strategy: across conditions, strategy systematically shifts ID_e upward or downward.

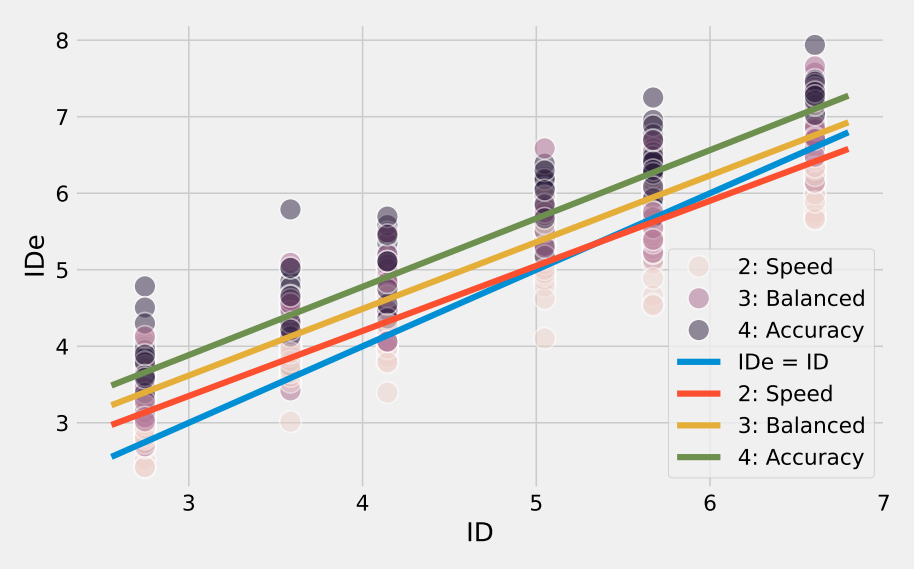

Fig. 15. The simple additive model of Equation 23 displayed on top of the YKORM data. The identity line ($ID_e = ID$) is also plotted.

A simple and well-supported model is therefore

$$ID_e = \alpha + \beta \cdot ID + \gamma \cdot \text{strategy}, \tag{27}$$

with $\beta \approx 0.9$ across datasets, positive intercepts α, and a strategy term γ that moves the effective ID line up or down. Interestingly, because the intercept is positive and $\beta < 1$, the ID_e line will cross the $ID_e = ID$ line. Thus, combined with the fact high IDs are overwhelmingly obtained with small values of W [32, 35], this model explains both the over- and under-utilization of targets noted by Zhai et al., while also accounting for strategic modulation.

7.5 EMG

We next assessed whether the EMG model, with its quadratic variance and asymmetric long tails, also provides a good account of MT distributions in the YKORM dataset. As before, we compared constant-variance vs. quadratic-variance models, and symmetric Gaussian vs. asymmetric EMG models, fitting per participant and per strategy while pooling over D and W.

The AIC differences (Figure 16) again show strong support for both quadratic variance and asymmetry. This mirrors the findings from the GOP dataset, reinforcing the conclusion that the EMG model captures general distributional properties of movement times, robust across protocols and user strategies.

7.6 Fitting copulas

We next assessed whether the joint dependency between ID_e and MT in the YKORM dataset is consistent with the results of the previous studies. Copulas were fitted per participant and per condition, with model evidence ratio $\mathcal{R}$ used for comparison.

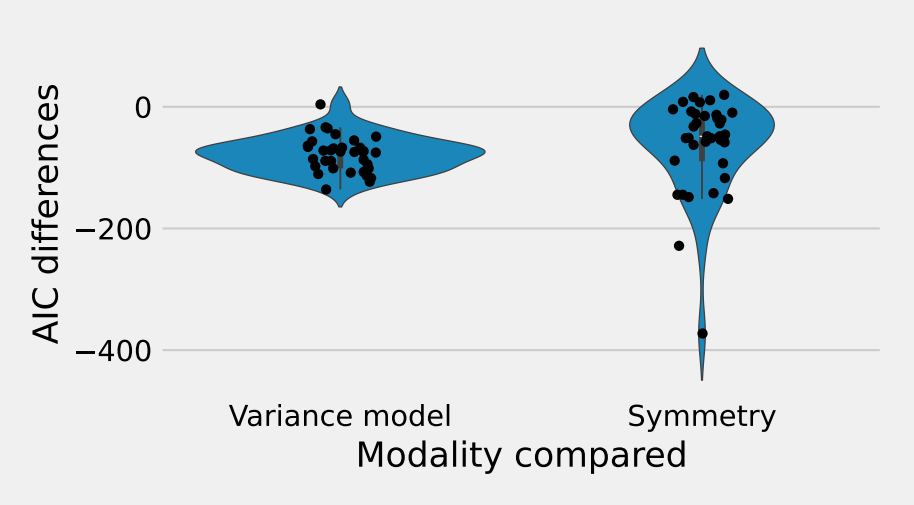

Fig. 16. Violin plots for the AIC differences when comparing variance and symmetry models. Refer to the caption of Figure 9 for more details.

Effect of (D,W) condition. Pooling across strategies, we obtained 72 datasets. The results (Figure 17) show that Gaussian copulas again provide the best overall fit, with a an exception for the smallest width target (W = 8), where rotated Galambos or HR copulas could score higher. The average Gaussian copula correlation was $\rho = 0.41$, indicating a moderate *local* dependency. Unlike in the previous datasets, this larger ρ likely reflects the fact data was pooled over strategies. No significant effects of D or W on ρ were found.[14]

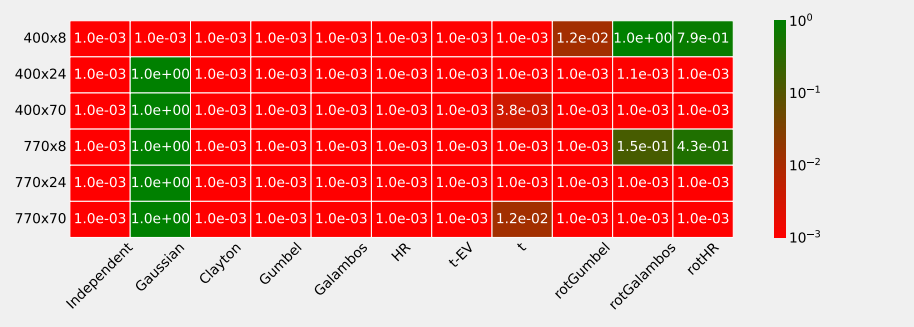

Fig. 17. Model evidence ratio $\mathcal{R}$ of the tested copulas against the best copula for different (D, W) pairs of the YKORM dataset.

Effect of strategy. When pooling across (D,W) conditions (36 datasets), Gaussian and t-copulas largely dominated (Figure 18). Copula parameter estimates for the t-copula (Figure 19) showed no systematic dependence on strategy. This complements the earlier results: strategy shifts the marginal distributions of MT and ID_e, but does not alter the dependency structure between them.

[14] A barplot of the ρ values and the fit details can be found in the supplementary materials.

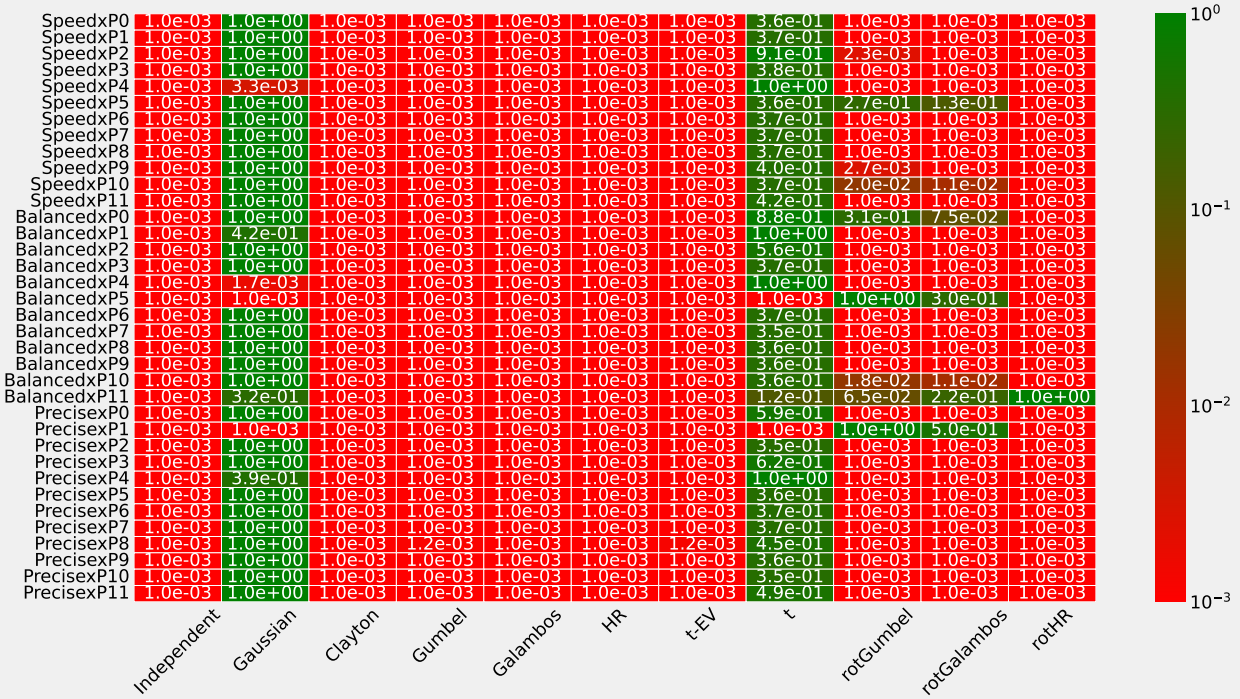

Fig. 18. Model evidence ratio $\mathcal{R}$ of the tested copulas against the best copula, for different strategies and participants of the YKORM dataset.

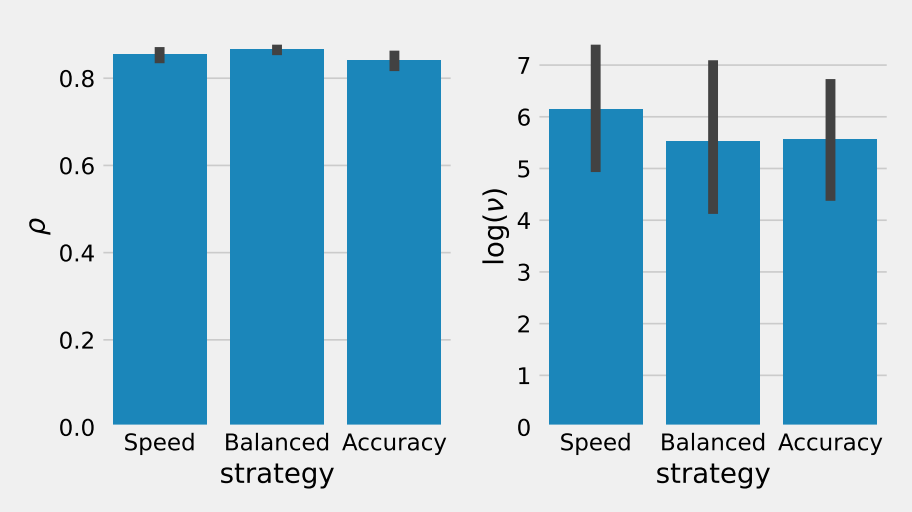

Fig. 19. Parameters of the t-copula for different strategies.

8 SYNTHESIS OF THE THREE STUDIES: COPULA FITS AND ID_e MODEL

8.1 Copula fits

Across the three studies, copulas consistently revealed a weak local dependence between ID_e and MT. This suggests that, at the trial level, faster movements can also be more precise. Strategy introduces slightly stronger dependence, which we interpret as participants adjusting performance relative to the previous block, adding variability in the actual performed strategy.

In contrast, the global dependence across conditions is strong and robust, in line with the classical speed-accuracy tradeoff. The first study favored the rotated Gumbel copula (mean $\theta \simeq 2$), while the second and third studies favored Gaussian/t-copulas (average $\rho \approx 0.8$). Importantly, these copulas yield visually similar joint distributions (see Figure 24, where the top right panel and right panel of the third row are hardly dinstinguishable), so the discrepancy does not undermine the main finding: ID_e and MT are tightly coupled at the global level.

Could the difference between Gumbel and Gaussian fits be an artifact of model identifiability [26, 65]? To test this, we simulated datasets with realistic marginals and sample sizes comparable to our studies (Appendix F). The analysis shows that the rotated Gumbel should be identifiable given our data, ruling out a spurious explanation. We therefore attribute the difference to the experimental protocols and input modalities: gesture input in study 1, physical arm movements in study 2, and mouse pointing in study 3. We return to this interpretation in subsection 8.4.

8.2 Model between ID_e and ID

Across studies 1 and 3, we found no consistent evidence that D or W influence ID_e beyond their ratio in ID. This is not surprising: in standard full-factorial designs, D and W are strongly correlated with ID [32], which makes disentangling their effects statistically fragile. A shape × scale design would be needed to test these effects more decisively [35].

Importantly, the simple linear model proposed in Equation 27 already captures the main phenomenon of interest described by Zhai et al. of target over- and under-utilization. We therefore advocate this model as a parsimonious and practically useful account. Expressed probabilistically with a standard Gaussian noise formulation, it reads:

$$\mathrm{ID}_e \sim \mathcal{N}\left(\alpha + \beta\,\mathrm{ID} + \gamma\,\mathrm{strategy}, \sigma^2\right). \tag{28}$$

We assessed the variability of ID_e by estimating $\sigma \approx 0.3$ for the YKORM dataset and $\sigma \approx 0.15$ for the JGP dataset. We return to the interpretation of these values in subsection 8.4.

8.3 EMG model

Across all three studies, the EMG distribution consistently provided a better account of movement time than Gaussian alternatives. More specifically, models with quadratic variance and asymmetry were strongly supported, highlighting that the long right tails and variance growth with difficulty are robust features of pointing data. While we did not rule out that other asymmetric distributions might provide an even closer fit, the EMG model strikes a good balance between parsimony and descriptive power, and its validity generalizes across protocols, tasks, and input modalities.

8.4 Revisiting results with the PDD dataset

To further validate and generalize our findings, we analyzed the Pointing Dynamics Dataset (PDD) [48]. This dataset is particularly informative because it uses a shape × scale design (D = 212, 353 mm; ID = 2, 4, 6, 8), which allows to

Table 6. Mixed effect model fit results for ??, ?? and ??. Each line (except the last, which displays the log-likelihood associated with each model) displays the estimated coefficient associated with the variable in the left column as well its associated p-value (in bold if p<0.05).

	Full model (??)	W model (??)	D model (??)	only ID (??)
Intercept	0.954 **(p<0.001)**	0.709 **(p<0.001)**	1.085 (**p<0.001**)	0.792 **(p<0.001)**
ID	0.864 **(p<0.001)**	0.958 (**p<0.001**)	0.844 (**p<0.001**)	0.934 **(p<0.001)**
w	0.104 **(p<0.001)**	0.038 **(p<0.001)**		
ID:w	-0.068 **(p<0.001)**	-0.015 **(p<0.001)**		
D	0.145 **(p<0.001)**		-0.029 (**p<0.001**)	
ID:D	-0.011 **(p<0.001)**		0.009 (**p<0.001**)	
log-likelihood	-223	-476	-674	-754

decorrelate D from ID, includes a condition with a very small target size (less than 1 mm), employs a standard mouse with constant CD gain, and includes a large number of trials per condition (12 participants × 8 blocks × 102 trials).

8.4.1 Dataset Processing. We only considered movements in the left-to-right direction. The first 20 trials for each condition were also removed in the version of the PDD dataset that is publicly available. After removing spatial outliers (endpoint position more than 2 standard deviations away from the mean per condition), we ended up with a dataset of 3744 movements. For some blocks in the ID=2 conditions, we found ID_e levels below 1, which is abnormally low. We thus removed these blocks and in the end we were left with 3160 movements across 87 different conditions.

8.4.2 Model between ID_e and D, W, ID. We fitted the same mixed-effect linear models as in the first and third studies. Consistent with prior results, ID_e was largely determined by ID and strategy, with D having only a weak secondary effect when decorrelated from ID. W remains strongly correlated with ID in this dataset, so its separate contribution is harder to estimate reliably. Importantly, the model reproduces the classic target under- and over-utilization phenomenon: positive intercepts and slopes slightly below 1 indicate under-utilization for small IDs and over-utilization for large IDs. This confirms that the simple linear model in Equation 27 captures the essential structure of ID_e across datasets. Parameter estimates are summarized in Table 6.

8.4.3 Fitting copulas. We next assessed local and global dependencies between ID_e and MT using copulas. For local dependencies (within each W condition), no single copula stood out consistently. The Galambos copula provided the best fit for 4 of 8 W conditions and remains plausible for 2 others, with its parameter θ increasing steadily with W (from about 0.4 for the smallest target to about 1.2 for the largest, plot available in the supplementary materials), indicating a progressive increase in local dependence.

For global dependencies (pooling across W conditions per participant), results are consistent with previous studies: the strongest dependencies are captured either by the Gaussian/t-copula pair ($\rho \simeq 0.87$) or the rotated Gumbel copula ($\theta \simeq 3.1$), indicating very strong correlation between ID_e and MT (Figure 21). This confirms that while local dependencies are weak and variable across small targets, global dependencies reflecting the speed-accuracy tradeoff are robust across datasets and experimental designs.

Fig. 20. Model evidence ratio $\mathcal{R}$ for the PDD dataset, for 8 different conditions (labeled by W value).

Fig. 21. Model evidence ratio $\mathcal{R}$ for the PDD dataset, computed for each participant.

9 GENERATING POINTING DATA FOR DIFFERENT STRATEGIES: RATIONALE AND METHODS FOR THE DIFFERENT MODELS

As Shmueli [58] argues, statistical models can serve two distinct purposes: to *explain* observed data or to *predict* new data. These two goals are not the same, and success in one does not guarantee success in the other. In the preceding sections, our focus was on explanation: we used models to probe theoretical questions about the effects of ID, ID_e, strategy, and MT, and to understand the structure of pointing behavior. In this section, we shift to the predictive role. We ask whether the same models, when used generatively, can produce realistic pointing data, that can be used *e.g.*, in simulations or in optimal design processes [51, 55].

We exploit the parametric models developed before (Gaussian bivariate model for $(\overline{\mathrm{MT}}, \mathrm{ID}_e)$, the EMG model for MT and the copula models) to generate synthetic data under different strategy conditions. A key feature of these models is their explicit parametrization of strategy, which allows us to simulate data for both pure-strategy and Fitts-with-strategy protocols. Note that we present only methods for the pure strategy protocol, but these methods generalize easily to the Fitts-with-strategy protocol. As with any generative approach, there is a trade-off between simplicity and realism. Simpler models tend to be more robust and easier to generalize, but may miss "second-order" effects; here, we prefer simplicity.

Throughout this section, Θ denotes a vector of model parameters, including strategy s. $p_\Theta(\mathrm{ID}_e, \mathrm{MT})$ refers to the joint distribution between ID_e and MT parametrized by Θ, and $p_\Theta(\mathrm{MT}|\mathrm{ID}_e)$ refers to the conditional distribution of MT given ID_e. The three methods we present illustrate different ways to generate realistic synthetic pointing data, while highlighting trade-offs in flexibility, interpretability, and computational tractability. We use parameter values estimated from the three studies as concrete examples, giving researchers a realistic starting point; if the context differs, they can easily re-estimate them on their own dataset.

9.1 Model 1: Joint distributions using Copulas.

The model. The first method defines the global dependency $p_\Theta(\mathrm{ID}_e, \mathrm{MT})$ in terms of copulas. As explained in subsection 3.2, the copula C_Θ links the marginals $F_\Theta(\mathrm{ID}_e)$ and $G_\Theta(\mathrm{MT})$ into the joint $p_\Theta(\mathrm{ID}_e, \mathrm{MT})$

$$p_\Theta(\mathrm{ID}_e, \mathrm{MT}) = C_\Theta(F_\Theta(\mathrm{ID}_e), G_\Theta(\mathrm{MT})). \tag{29}$$

Pointing data can then be obtained by sampling directly from the joint. This method assumes access to the marginals, which can be estimated from empirical data, or given/assumed.

Accounting for different strategies. To have the resulting joint distribution depend on s even though the copula does not, we sample ID_e values from the $p(\mathrm{ID}_e|s)$ model and use them together with the *conditional* version of the aforementioned copula. For the $\mathrm{ID}_e|s$ model, we can directly consider the marginal distribution of the bivariate gaussian model Equation 14, which reads

$$\mathrm{ID}_e \sim \mathcal{N}(4.72 + 2.3\,s, (1.06 + 0.29\,s)^2). \tag{30}$$

Inputs of the model. To operationalize this model, one needs

(1) The parameters of the t-copula. We found $\rho_1 = .69$, $\nu = 16.9$ by estimating the t-copula on the entire GO dataset; the simpler Gaussian copula model would also be appropriate.
(2) The parameters of the marginal distributions for MT and ID_e, which can be estimated using any popular statistical package. For MT, several asymmetric distributions may be valid; we prefer the EMG distribution as

explained subsubsection 2.3.2. For the GO dataset, we found $\beta = 0.53$, $\sigma = 0.19$, and $\lambda = .75$. Note that we use the "pure" EMG distribution here, and not the EMG model with linear conditional mean Equation 8.

Pro's and con's. Copulas offer a clean way to specify the dependency between ID_e and MT independently of their marginals. This means one can mix and match marginals for ID_e (*e.g.*, to fit a specific task) and MT (*e.g.*, using another asymmetric distribution) while keeping the same t-copula to specify their dependence. In short, copulas provide a principled and flexible framework for generating dependent pointing data. On the downside, copulas have been used mostly to capture tail dependencies, but in pointing tasks, the lower tail of MT is linked to the entire range of ID_e values, not just its extremes [30], so their generative performance needs evaluation.

9.2 Model 2: Conditional distribution using EMG.

The model. The second method directly leverages the EMG model with quadratic variance. For each point, a value of ID_e is selected by sampling from the marginal $F_\theta(\mathrm{ID}_e)$; then a value of MT is selected by sampling from the conditional distribution

$$p_\Theta(\mathrm{MT}|\mathrm{ID}_e) = \mathrm{EMG}(\mathrm{MT}; \mathrm{ID}_e, \Theta). \tag{31}$$

Accounting for different strategies. Similar to the previous model, we can directly draw ID_e values from the Gaussian Equation 30.

Inputs of the model. To operationalize this model, one needs the parameters of the EMG distribution described Equation 8. We estimated these values with MLE for the GO dataset as $a = 0.08$, $b = .14$, $\sigma = 0.18$, $\lambda_0 = .17$, $\lambda_1 = .08$.

Pro's and con's. This method will produce data that resembles empirical data by construction, since it relies directly on the EMG distribution which was validated. One of the main weaknesses of this method is that it does not allow controlling the association between ID_e and MT independently of the EMG parameters. As demonstrated in Appendix B, the Pearson correlation r is entirely determined by the EMG model's parameters. Additionally, as explained subsection 3.1 it will produce values of $r(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ much closer to 1 than what we found empirically in Figure 7 and inFigure 8.

9.3 Model 3: Joint distribution of (ID_e, $\overline{\mathrm{MT}}$).

The model. This third model directly controls the association between $\overline{\mathrm{MT}}$ and ID_e by sampling a pair of $(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ values from the Gaussian bivariate model (Equation 14). Then, because the EMG model parameters are related to $\overline{\mathrm{MT}}$ by

$$\overline{\mathrm{MT}} = \beta\mathbf{x} + \lambda\mathbf{x} = a + b\mathrm{ID}_e + \lambda_0 + \lambda_1\mathrm{ID}_e, \tag{32}$$

we "correct" λ_1[15], while also ensuring it remains positive to keep conditional mean and variance that increase across ID_e.

$$\lambda_1 = \max(0, \frac{\overline{\mathrm{MT}} - a - b\,\mathrm{ID}_e}{\mathrm{ID}_e}). \tag{33}$$

[15]We don't correct β since this would alter the low movement time values, which we believe are an important feature of the dataset, and don't alter λ_0 to ensure a sufficiently large variance.

Accounting for different strategies. Here, we draw samples directly from the bivariate Gaussian Equation 14 with

$$\mu_i = 4.72 + 2.3\,s \tag{34}$$

$$\mu_t = 1.3 + .64\,s \tag{35}$$

$$\sigma_i = 1.06 + 0.29\,s \tag{36}$$

$$\sigma_t = .39 + .08\,s \tag{37}$$

$$r = \begin{cases} 0 \text{ if } s = -1 \\ .44 \text{ else} \end{cases} \tag{38}$$

Inputs of the model. Same as for Model 2.

Pro's and con's. Like the previous method, this approach generates realistic data by relying on the EMG model. Its main advantage is that, unlike Method 2, it explicitly controls the association between $\overline{\text{MT}}$ and ID_e. The drawback is that the correction in Equation 33 is not always fully applied: when $\lambda_1 = 0$, part of $\overline{\text{MT}}$ is left unaccounted for.

10 MODEL COMPARISON

We now evaluate the three generative models, asking how well they reproduce the empirical distributions and features relevant to Fitts' law users. Two criteria are considered: (i) *data consistency*, the stability of common data metrics when re-estimated from generated data; and (ii) *consistency with the bivariate Gaussian model*, since it provides a simple way to assess the effect of strategy. To reduce random sampling effects, we generated large datasets[16] using the exact ID_e values from the original data.[17]

10.1 Data Consistency

We compared the three generated and the ground truth datasets on four type of features that are practically useful for Fitts' law users:

(1) association measures: Pearson's $r(\text{MT}, \text{ID}_e)$, Spearman's $\rho(\text{MT}, \text{ID}_e)$, Kendall's $\tau(\text{MT}, \text{ID}_e)$,
(2) The β parameter of the EMG model, and Fitts' law parameters a and b. Indeed, Gori and colleagues [29, 30] have argued that there exist two version of Fitts' law: an average Fitts' law, computed via linear regression on $\overline{\text{MT}}$ and a minimum Fitts' law, which represents the best performance possible, and corresponds to β in the EMG model [27].
(3) The "classical" criterion $r(\overline{\text{MT}}, \text{ID}_e)$,
(4) computation of ISO-throughput, also known as the "mean-of-means" throughput [59].

The results are shown in Figure 22. Overall, differences between the models are modest and pass the "eye test" when compared to the ground truth data; The association measures also show little to no differences. Regarding the other features:

- All models slightly underestimate ISO throughput, with Model 1 deviating the most (by up to 0.25 bit/s).

[16] The data were generated with a fixed seed (1234). Replications with two other seeds (777, 999) are available in the supplementary materials. The conclusions for these replications remain the same as for the results reported in the paper, indicating the sample size of the dataset was indeed large enough.
[17] That is, ID_e was fixed to empirical values and only MT was generated. This is feasible in all three models: via conditional copula sampling (Model 1), conditional Gaussian sampling (Model 3), or trivially in Model 2.

- Model 1 preserves the average Fitts' law (OLS) but distorts the minimum law (EMG), producing a higher intercept and shallower slope, while Model 2 shows the opposite pattern. Model 3 most faithfully reproduces the original data, with only a small intercept bias in the EMG fit.
- $\overline{r} = r(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ is inflated for all models. For Model 2, this is structural and due to EMG. For Model 1, it likely reflects the symmetry of the t-copula. Model 3 reduces $\overline{r}$, but not up to ground truth levels, presumably because of the enforced positive λ_1.[18]
- All models underestimate MT at low ID_e values, with Models 2 and 3 even producing occasional negative MTs.

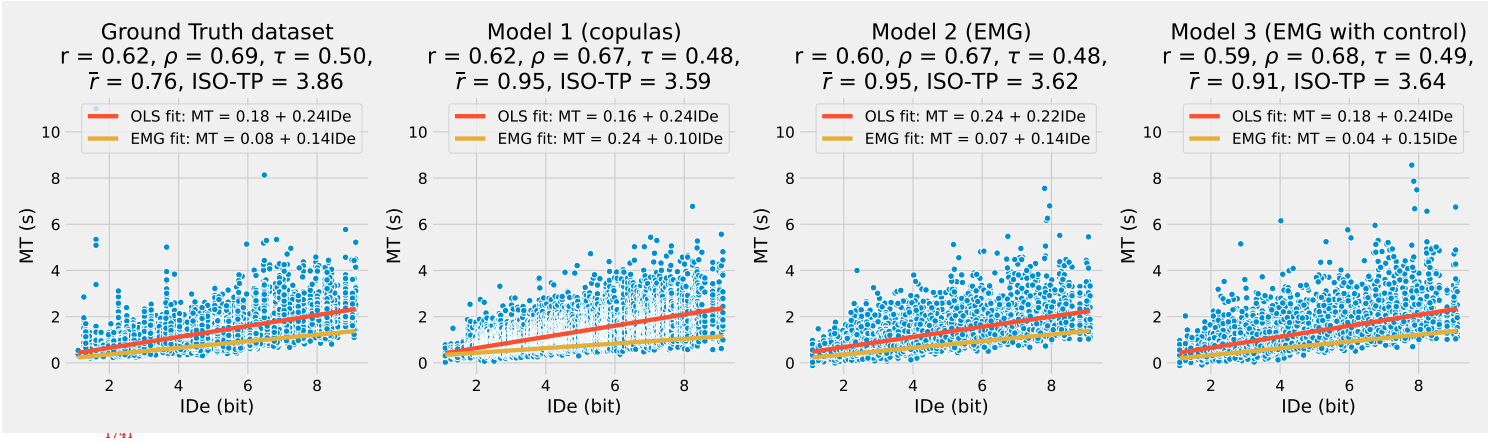

Fig. 22. Consistency as evaluated by the criterion of subsection 10.1. The original dataset (left panel) is compared to data produced by Model 1 (middle left panel), Model 2 (middle right panel) and Model 3 (right panel).

10.2 Consistency of the bivariate Gaussian model

Because our goal is to fit a Gaussian model parametrized by strategy, we next asked whether the generated datasets are also consistent with this model. For each method, we generated data under five strategy levels and refit the Gaussian model as in Figure 8.

Results are shown in Figure 23. Model 1 reproduces the ID_e dimension well but fails on $\overline{\mathrm{MT}}$ and the correlation structure. Models 2 and 3 are much closer overall, with Model 3 providing the best alignment with the Gaussian parametrization.

[18] Analyses reported in the supplementary materials confirm that correcting a instead of λ_1 recovers the target $\overline{r}$, but at the cost of much worse performance on other metrics.

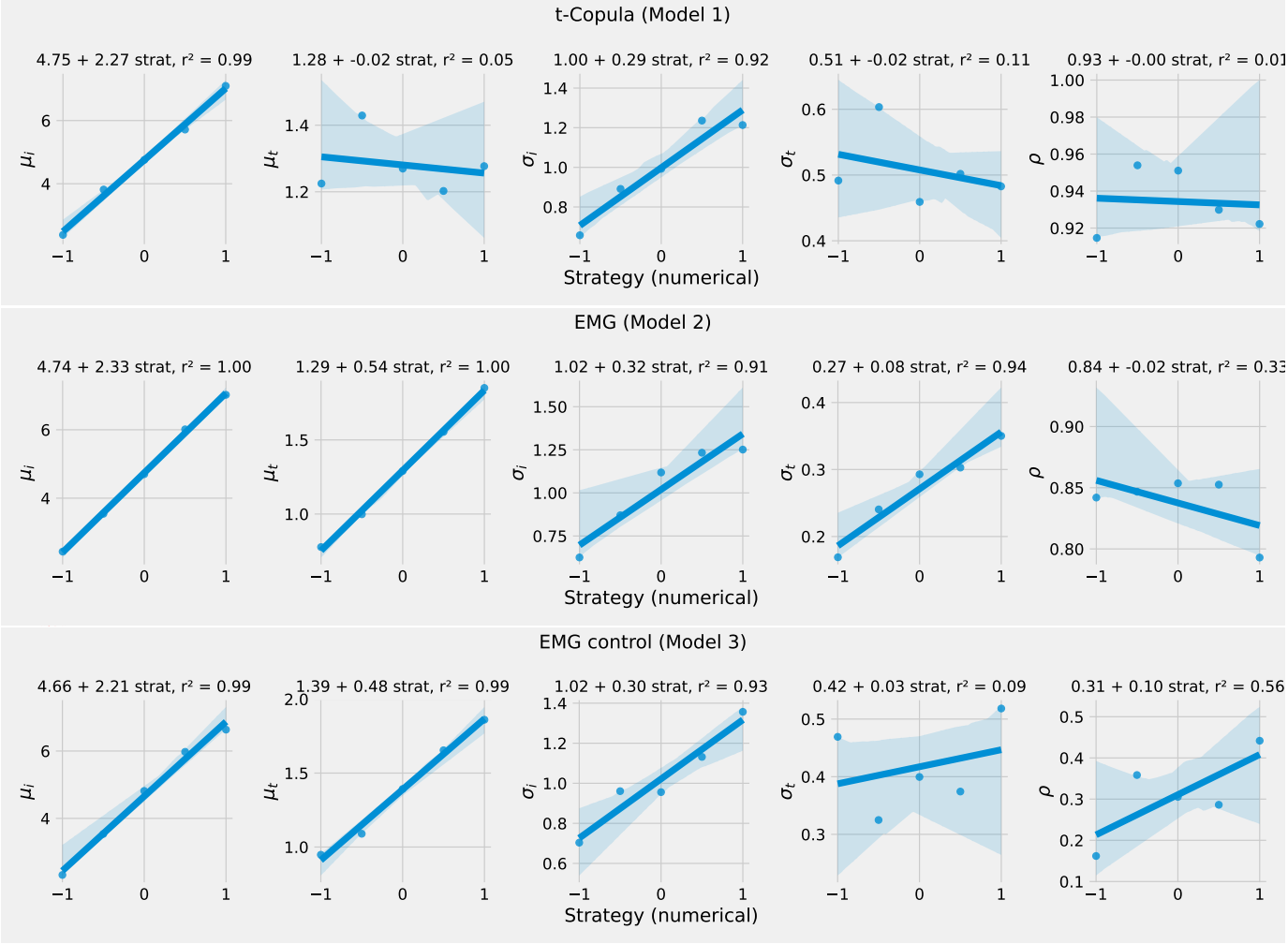

Fig. 23. An evaluation of the linear parametrization of the bivariate Gaussian model of Equation 14. Each subpanel reads like those of Figure 8. Top panel: evaluation for Model 1. Middle panel: evaluation for Model 2. Bottom panel: evaluation for Model 3.

11 DISCUSSION

11.1 Models at a glance

Across this work, we introduced several models, which can complement or extend the traditional use of Fitts' law.

- The probabilistic model for ID_e in Equation 28 can be used to generate ID_e values from nominal values of ID and strategy.
- The bivariate Gaussian model, described in Equation 14 together with its parameters relationships to ID and strategy described in Figure 14, allows generating ID_e and $\overline{MT}$ values from nominal values. It provides a relatively simple description, while still capturing effects of strategy and task geometry and part of the variability of human movements.

- The three generative models described section 9, with their strengths and weaknesses previously discussed.

Often, researchers can start with our reported parameters values if the goal is to predict movement times *e.g.*, to design an intelligent system where movement times play a role. As we have used datasets acquired over a variety of conditions (pointing with a mouse, with the limb, or with gestures), we expecet our results to be quite robust. However, if the context differs substantially, or to obtain more accurate predictions, these parameters can be re-estimated from new data using standard statistical methods.

11.2 On the local and global dependency

In this work, we examined both local and global dependencies between ID_e and MT, corresponding to two levels of the speed-accuracy tradeoff (SAT): a local SAT operating within conditions and a global SAT across conditions. Our finding of minimal to no local dependency suggests two possible explanations. One is that movement time variability from internal motor noise [19, 62] dwarfs the variability induced by changes in ID_e within a single condition. Another is that movements executed under a fixed condition are functionally equivalent, leaving little room for systematic tradeoffs to appear locally.

To our knowledge, this local-global distinction has not been recognized in pointing models before. Yet, SAT effects operating at multiple scales have been examined in other domains, most notably reaction time research, where they are referred to as micro and macro SATs [13, 16]. Interestingly, the asymmetric, long-tailed distributions we found to be critical for modeling movement times are also a hallmark of reaction times [44]. A proposed explanation for the micro SAT is that individuals adjust their speed-accuracy tradeoffs while responding on a trial-by-trial basis [16], which suggests that the local tradeoff may have been obscured by large movement time variability. Whether such mechanisms translate from psychophysical tasks to motor control remains an open question. What our results do show clearly is that the dependence between MT and ID_e is much stronger at the global scale, across conditions.

11.3 Applicability of copulas

We introduced copulas as a way to model the joint distribution of ID_e and MT. To our knowledge, this is the first application of copulas in HCI. The main advantage of copulas is that they separate the dependency structure from the marginals: one can represent the dependency with a compact parametric form (often just one or two parameters) without having to impose restrictive assumptions about the marginal distributions.

Admittedly, the case we studied is not the most compelling for copulas; the dependency is already well captured by the EMG model, and our goal was not to position copulas as a central contribution of this paper, but rather to use it as a tool to study dependencies with the least amount of assumptions. Nonetheless, we see strong potential for their use in HCI. Interaction, at its core, is about dependencies between variables[19], and in many situations these dependencies may be nonlinear or dominated by tail behavior. Copulas are designed for precisely such cases. We see promising opportunitiesin any setting where tail dependencies are expected *i.e.*, where variable effects become mutually reinforcing or diminishing at the extremes of their ranges.

From a practical standpoint, copula tooling is still limited. The most mature implementations exist in R, and we could not find full-feature equivalents in Python. In many cases however, the most common copulas are widely available; many copula's parameters can even be recovered from simple statistics without any dedicated software. For example,

[19]Two variables that are independent can not describe a useful interaction scenario, since they can be understood one without the other.

for a t-copula, ρ can be deduced from Kendall's τ via $\rho = \sin(\tau\pi/2)$, while ν can be obtained from tail dependence coefficients using established formulas [12, Section 3].

12 CONCLUSION

In this work, we presented models of pointing that serve two distinct but complementary roles. As explanatory tools, they highlight the existence of two levels of the speed-accuracy tradeoff. As predictive tools, they allow realistic data generation. Together, these findings clarify the nature of the speed-accuracy tradeoff of pointing, and provide new models for researchers who wish to use them in simulation and optimisation contexts. For future research, we recommend exploring hierarchical pointing models as proposed in [73]. Notably, the GO dataset reveals considerable inter-participant differences in ID_e distributions across strategies (see supplementary materials): some participants display highly consistent distributions, indicating strong internal strategy adherence, while others perform faster conditions more precisely than accurate ones. Similarly, there are notable differences among participants in the variability of MT.

ACKNOWLEDGMENTS

The author thanks Shota Yamanaka, Homei Miyashita, Taiki Kinoshita, Yosuke Oba and Ryuto Tomihari for sharing the YKORM dataset and providing assistance in using it.

REFERENCES

[1] Johnny Accot and Shumin Zhai. 1997. Beyond Fitts' law: models for trajectory-based HCI tasks. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. 295–302.

[2] Johnny Accot and Shumin Zhai. 2003. Refining Fitts' law models for bivariate pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 193–200.

[3] D Anderson and K Burnham. 2004. Model selection and multi-model inference. *Second. NY: Springer-Verlag* 63, 2020 (2004), 10.

[4] Anil Ufuk Batmaz and Wolfgang Stuerzlinger. 2022. Effective Throughput Analysis of Different Task Execution Strategies for Mid-Air Fitts' Tasks in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3939–3947.

[5] Bastien Berret and Frédéric Jean. 2016. Why don't we move slower? The value of time in the neural control of action. *Journal of neuroscience* 36, 4 (2016), 1056–1070.

[6] Xiaojun Bi and Shumin Zhai. 2013. Bayesian touch: a statistical criterion of target selection with finger touch. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 51–60.

[7] Stuart K Card, William K English, and Betty 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] Géry Casiez and Nicolas Roussel. 2011. No more bricolage! Methods and tools to characterize, replicate and compare pointing transfer functions. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 603–614.

[9] Olivier Chapuis, Renaud Blanch, and Michel Beaudouin-Lafon. 2007. Fitts' law in the wild: A field study of aimed movements. (2007).

[10] Jhon PF Charaja, Isabell Wochner, Pierre Schumacher, Winfried Ilg, Martin Giese, Christophe Maufroy, Andreas Bulling, Syn Schmitt, Georg Martius, and Daniel FB Haeufle. 2024. Generating Realistic Arm Movements in Reinforcement Learning: A Quantitative Comparison of Reward Terms and Task Requirements. In *2024 10th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 562–568.

[11] ERFW Crossman. 1957. The speed and accuracy of simple hand movements. *The nature and acquisition of industrial skills* (1957).

[12] Stefano Demarta and Alexander J McNeil. 2005. The t copula and related copulas. *International statistical review* 73, 1 (2005), 111–129.

[13] Ian Dennis and Jonathan St BT Evans. 1996. The speed–error trade-off problem in psychometric testing. *British Journal of Psychology* 87, 1 (1996), 105–129.

[14] Paul R Dewick and Shuangzhe Liu. 2022. Copula modelling to analyse financial data. *Journal of Risk and Financial Management* 15, 3 (2022), 104.

[15] Seungwon Do, Minsuk Chang, and Byungjoo Lee. 2021. A simulation model of intermittently controlled point-and-click behaviour. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.

[16] Benjamin W Domingue, Klint Kanopka, Ben Stenhaug, Michael J Sulik, Tanesia Beverly, Matthieu Brinkhuis, Ruhan Circi, Jessica Faul, Dandan Liao, Bruce McCandliss, et al. 2022. Speed–accuracy trade-off? Not so fast: Marginal changes in speed have inconsistent relationships with accuracy in real-world settings. *Journal of Educational and Behavioral Statistics* 47, 5 (2022), 576–602.

[17] Heiko Drewes. 2013. A lecture on Fitts' law. *Retrieved January* 5 (2013), 2017.

[18] Digby Elliott, James Lyons, Spencer J Hayes, James J Burkitt, James W Roberts, Lawrence EM Grierson, Steve Hansen, and Simon J Bennett. 2017. The multiple process model of goal-directed reaching revisited. *Neuroscience & Biobehavioral Reviews* 72 (2017), 95–110.

[19] A Aldo Faisal, Luc PJ Selen, and Daniel M Wolpert. 2008. Noise in the nervous system. *Nature reviews neuroscience* 9, 4 (2008), 292–303.

[20] Florian Fischer, Arthur Fleig, Markus Klar, and Jörg Müller. 2022. Optimal feedback control for modeling human–computer interaction. *ACM Transactions on Computer-Human Interaction* 29, 6 (2022), 1–70.

[21] Paul 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.

[22] Khai-Chung Gan and Errol R Hoffmann. 1988. Geometrical conditions for ballistic and visually controlled movements. *Ergonomics* 31, 5 (1988), 829–839.

[23] Christian Genest and Johanna G Nešlehová. 2017. When Gumbel met Galambos. In *Copulas and Dependence Models with Applications: Contributions in Honor of Roger B. Nelsen*. Springer, 83–93.

[24] Julien Gori. 2018. *Modeling the speed-accuracy tradeoff using the tools of information theory*. Ph. D. Dissertation. Université Paris-Saclay (ComUE).

[25] Julien Gori and Quentin Bellut. 2023. Positional Variance Profiles (PVPs): A New Take on the Speed-Accuracy Trade-off. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.

[26] Julien Gori, Bruno Fruchard, and Gilles Bailly. 2024. Model-based Evaluation of Recall-based Interaction Techniques. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.

[27] Julien Gori and Olivier Rioul. 2019. Regression to a linear lower bound with outliers: An exponentially modified Gaussian noise model. In *2019 27th European Signal Processing Conference (EUSIPCO)*. IEEE, 1–5.

[28] Julien Gori and Olivier Rioul. 2020. A feedback information-theoretic transmission scheme (FITTS) for modeling trajectory variability in aimed movements. *Biological Cybernetics* 114, 6 (2020), 621–641.

[29] Julien Gori, Olivier Rioul, and Yves Guiard. 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*. 260–264.

[30] Julien Gori, Olivier Rioul, and Yves Guiard. 2018. Speed-accuracy tradeoff: A formal information-theoretic transmission scheme (fitts). *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 5 (2018), 1–33.

[31] Julien Gori, Olivier Rioul, Yves Guiard, and Michel Beaudouin-Lafon. 2017. One fitts' law, two metrics. In *IFIP Conference on Human-Computer Interaction*. Springer, 525–533.

[32] Julien Gori, Olivier Rioul, Yves Guiard, and Michel Beaudouin-Lafon. 2018. The perils of confounding factors: How Fitts' law experiments can lead to false conclusions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–10.

[33] Tovi Grossman and Ravin Balakrishnan. 2005. A probabilistic approach to modeling two-dimensional pointing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 3 (2005), 435–459.

[34] Marvin Gruber. 2017. *Improving efficiency by shrinkage: The James–Stein and Ridge regression estimators*. Routledge.

[35] Yves Guiard. 2009. The problem of consistency in the design of Fitts' law experiments: Consider either target distance and width or movement form and scale. In *Proceedings of the sigchi conference on human factors in computing systems*. 1809–1818.

[36] Yves Guiard and Halla B Olafsdottir. 2011. On the measurement of movement difficulty in the standard approach to Fitts' law. *PLoS one* 6, 10 (2011), e24389.

[37] Yves Guiard, Halla B Olafsdottir, and Simon 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*. 1619–1628.

[38] Yves Guiard and Olivier Rioul. 2015. A mathematical description of the speed/accuracy trade-off of aimed movement. In *Proceedings of the 2015 British HCI conference*. 91–100.

[39] Xavier A Harrison, Lynda Donaldson, Maria Eugenia Correa-Cano, Julian Evans, David N Fisher, Cecily ED Goodwin, Beth S Robinson, David J Hodgson, and Richard Inger. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6 (2018), e4794.

[40] Aleksi Ikkala, Florian Fischer, Markus Klar, Miroslav Bachinski, Arthur Fleig, Andrew Howes, Perttu Hämäläinen, Jörg Müller, Roderick Murray-Smith, and Antti Oulasvirta. 2022. Breathing life into biomechanical user models. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[41] Alvin Jude, Darren Guinness, and G Michael Poor. 2016. Reporting and Visualizing Fitts's Law: Dataset, Tools and Methodologies. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2519–2525.

[42] Markus Klar, Florian Fischer, Arthur Fleig, Miroslav Bachinski, and Jörg Müller. 2023. Simulating Interaction Movements via Model Predictive Control. *ACM Transactions on Computer-Human Interaction* 30, 3 (2023), 1–50.

[43] Yanxi Li, Derek S Young, Julien Gori, and Olivier Rioul. 2024. A novel mixture model for characterizing human aiming performance data. *Statistical Modelling* (2024), 1471082X241234139.

[44] R Duncan Luce. 1991. *Response times: Their role in inferring elementary mental organization*. Oxford University Press.

[45] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139.

[46] I Scott MacKenzie and Poika Isokoski. 2008. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1633–1636.

[47] Hee-Seung Moon, Yi-Chi Liao, Chenyu Li, Byungjoo Lee, and Antti Oulasvirta. 2024. Real-time 3D Target Inference via Biomechanical Simulation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[48] Jörg Müller, Antti Oulasvirta, and Roderick Murray-Smith. 2017. Control theoretic models of pointing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 4 (2017), 1–36.
[49] Atsuo Murata and Hirokazu Iwase. 2001. Extending Fitts' law to a three-dimensional pointing task. *Human movement science* 20, 6 (2001), 791–805.
[50] Sohad Murrar and Markus Brauer. 2018. Mixed model analysis of variance. *The SAGE encyclopedia of educational research, measurement, and evaluation* 1 (2018), 1075–1078.
[51] Roderick Murray-Smith, Antti Oulasvirta, Andrew Howes, Jörg Müller, Aleksi Ikkala, Miroslav Bachinski, Arthur Fleig, Florian Fischer, and Markus Klar. 2022. What simulation can do for HCI research. *Interactions* 29, 6 (2022), 48–53.
[52] Roger B Nelsen. 2006. *An introduction to copulas*. Springer.
[53] Karin Nieuwenhuizen and Jean-Bernard Martens. 2016. Advanced modeling of selection and steering data: beyond Fitts' law. *International Journal of Human-Computer Studies* 94 (2016), 35–52.
[54] Halla Olafsdottir, Yves Guiard, Olivier Rioul, and Simon Perrault. 2012. A new test of throughput invariance in Fitts' Law: Role of the intercept and of Jensen's Inequality. (2012).
[55] Antti Oulasvirta, Per Ola Kristensson, Xiaojun Bi, and Andrew Howes. 2018. *Computational interaction*. Oxford University Press.
[56] Ning Qian, Yu Jiang, Zhong-Ping Jiang, and Pietro Mazzoni. 2013. Movement duration, Fitts's law, and an infinite-horizon optimal feedback control model for biological motor systems. *Neural computation* 25, 3 (2013), 697–724.
[57] Richard A Schmidt, Howard Zelaznik, Brian Hawkins, James S Frank, and John T Quinn Jr. 1979. Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychological review* 86, 5 (1979), 415.
[58] Galit Shmueli. 2010. To explain or to predict? *Statistical science* (2010), 289–310.
[59] R William Soukoreff and I Scott 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.
[60] Emanuel Todorov and Michael I Jordan. 2002. Optimal feedback control as a theory of motor coordination. *Nature neuroscience* 5, 11 (2002), 1226–1235.
[61] Emanuel Vassilev Todrov. 1998. *Studies of goal directed movements*. Ph. D. Dissertation. Massachusetts Institute of Technology.
[62] Robert J Van Beers, Patrick Haggard, and Daniel M Wolpert. 2004. The role of execution noise in movement variability. *Journal of neurophysiology* 91, 2 (2004), 1050–1063.
[63] Eric-Jan Wagenmakers and Simon Farrell. 2004. AIC model selection using Akaike weights. *Psychonomic bulletin & review* 11 (2004), 192–196.
[64] Michael Wang, Hang Zhao, Xiaolei Zhou, Xiangshi Ren, and Xiaojun Bi. 2021. Variance and Distribution Models for Steering Tasks. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 1122–1143.
[65] Robert C Wilson and Anne GE Collins. 2019. Ten simple rules for the computational modeling of behavioral data. *Elife* 8 (2019), e49547.
[66] Jacob O Wobbrock, Kristen Shinohara, and Alex Jansen. 2011. The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1639–1648.
[67] Robert Sessions Woodworth. 1899. Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements* 3, 3 (1899), i.
[68] Shota Yamanaka, Taiki Kinoshita, Yosuke Oba, Ryuto Tomihari, and Homei Miyashita. 2023. Varying Subjective Speed-accuracy Biases to Evaluate the Generalizability of Experimental Conclusions on Pointing-facilitation Techniques. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.
[69] Shota Yamanaka and Hiroki Usuba. 2020. Rethinking the dual gaussian distribution model for predicting touch accuracy in on-screen-start pointing tasks. *Proceedings of the ACM on Human-Computer Interaction* 4, ISS (2020), 1–20.
[70] Jun Yan. 2007. Enjoy the joy of copulas: with a package copula. *Journal of statistical software* 21 (2007), 1–21.
[71] Shumin Zhai, Jing Kong, and Xiangshi Ren. 2004. Speed–accuracy tradeoff in Fitts' law tasks—on the equivalency of actual and nominal pointing precision. *International journal of human-computer studies* 61, 6 (2004), 823–856.
[72] Hao Zhang, Jin Huang, Huawei Tu, and Feng Tian. 2023. Shape-Adaptive Ternary-Gaussian Model: Modeling Pointing Uncertainty for Moving Targets of Arbitrary Shapes. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
[73] Hang Zhao, Sophia Gu, Chun Yu, and Xiaojun Bi. 2022. Bayesian hierarchical pointing models. In *Proceedings of the 35th annual acm symposium on user interface software and technology*. 1–13.
[74] Alain F Zuur, Elena N Ieno, Neil J Walker, Anatoly A Saveliev, Graham M Smith, et al. 2009. *Mixed effects models and extensions in ecology with R*. Vol. 574. Springer.

A PROTOCOLS FOR MANIPULATING THE SPEED ACCURACY TRADE-OFF

The SAT is well known in HCI: in most tasks, to achieve greater precision, users generally need to reduce their speed. However, reliably manipulating speed and accuracy conditions in controlled experiments is challenging; in pointing, at least four distinct experimental protocols have been proposed.

A.1 Fitts' protocol

The most widely used protocol is Fitts', where accuracy is dictated by the width of the target: the independent variables are target distance (D) and width (W), while movement time (MT) serves as the dependent variable. This paradigm closely aligns with meaningful HCI tasks, like icon selection, since target geometry is explicitly part of the task. However, participants often fail to fully adhere to target width constraints, either overusing or underusing the available target space [36, 71].

A.2 Schmidt's protocol

A second paradigm, due to Schmidt *et al.* [57], controls MT (speed) through a timing procedure, making MT and D the independent variables, and σ the dependent variable. This approach has mainly been applied to rapid movement studies, which are of lesser importance to HCI, and we do not consider this protocol further.

A.3 Pure strategy protocol

In the two aforementioned protocols, either speed or accuracy are considered independent variables, the other being the dependent one, but in reality both these variables are outcomes of human behavior. A third protocol, which we call the pure strategy protocol, thus uses D and strategic instructions (strategy) as independent variables, and MT (speed) and σ (accuracy) as the dependent variables [37]. Strategies are enforced by verbal instructions before each block, which participants quickly learn to associate after some training. Note that there is no target width defined a priori in the pure strategy protocol.

A.4 Fitts-with-strategy protocol

A fourth protocol [4, 46, 68] has the experimenter give strategic instructions to participants while they perform a classical Fitts task. Independent variables are strategy, D and W, while the dependent variables are MT and σ. Empirical results show that participants are in fact able to modify their strategies as in the pure strategy protocol, even in presence of a prescribed width, as reflected *e.g.*, by changes in ID_e for a given nominal ID level. For example in Mackenzie's [46] experiment, on average participants erred up to 20% in the speed condition, but erred close to 0% in the accuracy condition. This protocol is the most complete one, since it accounts both for task properties (D and W) and user behavior (strategy). Note that even when fitted on data from this protocol, and even adjusted with ID_e, Fitts' law does not link strategy, ID_e and MT *a priori* since ID_e can only be defined *after* the experiment has been conducted.

B PROOF OF PROPOSITION 3.1

In this proof, we compute correlation coefficients between MT and ID_e, and $\overline{\mathrm{MT}}$ and ID_e. Interestingly, we show any distribution with a linear conditional expectation leads to $r^2(\overline{\mathrm{MT}}, \mathrm{ID}_e) = 1$ whatever its conditional variance. Note that in this proof, we use the mathematical expectation (*i.e.*, population averages) rather than sample averages — in practice this means these results are asymptotic (the more precise the larger the sample size).

We define

- $Y = \mathrm{ID}_e$ and $X = \mathrm{MT}$,
- $\mu_X = \mathbb{E}[X]$ the mean of the random variable X,
- $\sigma_X = \sqrt{\mathrm{Var}(X)}$ the standard deviation of the random variable X.

We recall the following results:

- The Pearson correlation coefficient is defined as the normalized covariance $\text{cov}(X, Y)$ between X and Y

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sigma_X \sigma_Y}. \tag{39}$$

- The law of total expectation

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X], \tag{40}$$

  which, when applied to a product, yields

$$\mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|Y]]. \tag{41}$$

- The law of total variance

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X)] + \text{Var}(\mathbb{E}[Y|X]). \tag{42}$$

The proof works by writing Pearson's r in terms of the conditional distribution using the laws of total expectations and variances. We write $\mathbb{E}[X|Y] = f(Y), \text{Var}[X|Y] = g(Y)$. We have that

$$\mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|Y]] = \mathbb{E}[Yf(Y)] \tag{43}$$

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[f(Y)] \tag{44}$$

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y]) \tag{45}$$

$$= \mathbb{E}[g(Y)] + \text{Var}(f(Y)) \tag{46}$$

The general practice of computing block averages described subsubsection 2.3.1 essentially considers $\overline{X} = \mathbb{E}[X|Y]$ instead of X:

$$r(\overline{X}, Y) = \frac{\mathbb{E}[\overline{X}Y] - \mathbb{E}[\overline{X}]\mathbb{E}[Y]}{\sigma_{\overline{X}} \sigma_Y}. \tag{47}$$

One notices from Equation 43 and Equation 44 that interestingly, the covariance between X and Y equals the covariance between $\overline{X}$ and Y. We then have

$$\sigma^2_{\overline{X}} = \text{Var}(\mathbb{E}[X|Y]) = \text{Var}(f(Y)). \tag{48}$$

When compared with Equation 45, one thus sees that the only difference between $r(X, Y)$ and $r(\overline{X}, Y)$ is that $\mathbb{E}[\text{Var}(X|Y)]$ is not present in the denominator. This shows $r(\overline{X}, Y)$ is always larger than $r(X, Y)$, and explains why considering block averages will lead to "better" results.

Linear conditional expectation leads to perfect correlation. It is known that pointing data reaches very high values of $r(\overline{X}, Y)$, often above .9, and sometimes even above .99. Here, we show that a linear model of conditional expectation will reach $r(\overline{X}, Y) = 1$. A linear conditional expectation reads

$$\mathbb{E}[X|Y] = a + bY = f(Y). \tag{49}$$

We also consider any conditional variance model

$$\text{Var}[X|Y] = g(Y). \tag{50}$$

The correlation coefficients, following the definition in Equation 39 can then be written as

$$r(X,Y) = \frac{\mathbb{E}[Y(a+bY)] - \mathbb{E}[a+bY]\mathbb{E}[Y]}{\sqrt{(\mathbb{E}[g(Y)] + \mathrm{Var}(a+bY))}\sigma_Y} \tag{51}$$

$$r(\overline{X},Y) = \frac{\mathbb{E}[Y(a+bY)] - \mathbb{E}[a+bY]\mathbb{E}[Y]}{\sqrt{\mathrm{Var}(a+bY)}\sigma_Y}. \tag{52}$$

The covariance parts (*i.e.*, the denominators in Equation 51 and Equation 52) simplify as

$$\mathbb{E}[Y(a+bY)] - \mathbb{E}[a+bY]\mathbb{E}[Y] = a\mu_Y + b\mathbb{E}[Y^2] - a\mu_Y - b\mathbb{E}[Y]^2 \tag{53}$$

$$= b\sigma_Y^2. \tag{54}$$

Because

$$\mathrm{Var}(a+bY) = b^2\mathrm{Var}(Y) = b^2\sigma_Y^2, \tag{55}$$

we obtain $r(\overline{X},Y) = 1$. On the other hand, we obtain a smaller value for $r(X,Y)$:

$$r(X,Y) = \frac{b\sigma_y^2}{\sigma_y\sqrt{\mathbb{E}[g(y)] + b^2\sigma_y^2}} = \frac{1}{1 + \frac{\mathbb{E}[g(y)]}{b^2\sigma_y^2}}. \tag{56}$$

The reason that $g(Y)$ does not play a role in the r^2 value is that the variance of X for a given Y is nullified when considering $\overline{X}$.

The correlation between ID_e and MT can not be specified independently of the EMG parameters in an EMG model. The EMG model assumes a quadratic variance model (Equation 9)

$$g(Y) = s^2 + (\lambda_0 + \lambda_1 Y)^2 \tag{57}$$

which gives

$$\mathbb{E}[g(Y)] = s^2 + \lambda_0^2 + 2\lambda_0\lambda_1\mu_Y + \lambda_1^2(\mathrm{Var}(Y) - \mu_Y^2) \tag{58}$$

which does not offer further simplification. This result shows that the dependence between ID_e and MT can not be specified independently of the EMG parameters: for a given ID_e distribution, the correlation between MT and ID_e are fully determined by the EMG parameters

C FITTING AND COMPARING COPULAS

C.1 Fitting procedure

We considered copulas from the most widely recognized families: elliptical (*e.g.*, the Gaussian and *t* copulas), Archimedean (*e.g.*, Clayton, Gumbel), and extreme value (*e.g.*, HR, Galambos, *t*-EV), as well as their rotated variants[20] and the independent copula. We utilized the R `copula` package [70] to fit these candidate copulas; copulas are estimated based on maximum likelihood estimation, and the marginals are estimated with *empirical* maximum likelihood estimation; estimations are available by directly calling library functions, as illustrated in the code that comes with this paper.

[20] Copulas can exhibit dependencies in the lower tails (low values) or upper tails (high values) of distributions. Given that pointing data in our dataset displays high variance at higher ID levels, copulas with upper tail dependence, such as the Gumbel copula, tend to perform poorly. By rotating the copula, we switch the dependence from the upper tail to the lower tail, providing additional copula candidates for consideration.

C.2 Comparison procedure

To compare copulas, we use the model evidence ratio $\mathcal{R}$, which builds on AIC. A full exposition of model comparison based on AIC and $\mathcal{R}$ can be found in [3] but is summarized here for convenience. In the maximum likelihood estimation approach, the combination of model and parameter values that reaches the maximum log-likelihood $\mathcal{L}$ is preferred. This leads to overfitting, as a nested model with fewer parameters can only do worse than the full model. One solution to that is AIC, a score that penalizes $\mathcal{L}$ with the number of parameters. A model's AIC reads

$$AIC = 2(k - \mathcal{L}); \tag{59}$$

a lower AIC indicates better fitting power. Models can then be compared based on their AIC difference, generally, an AIC difference of 10 is considered very strong support *i.e.*, the competing model is implausible. An interesting transformation of AIC differences is the relative likelihood ratio or model evidence ratio $\mathcal{R}$

$$\mathcal{R} = \exp\left(-\frac{\mathrm{AIC}_1 - \mathrm{AIC}_2}{2}\right) \tag{60}$$

which can directly be interpreted as odds (see also [63]). For example an $R = 0.125$ means that model 2 is about 8 times as likely as model 1, given current data.

One difficulty is how to aggregate $\mathcal{R}$ from different datasets; for example, for different participants and replications. Existing literatures lists three strategies:

- Pool the data before, then compute the loglikelihood $\mathcal{L}$ and $\mathcal{R}$. This is justified when the data is considered homogeneous across the datasets.
- Compute the loglikelihoods $\mathcal{L}$ for each dataset, and sum them to get a single $\mathcal{L}$. This is essentially equivalent to pooling the data $\mathcal{L}$, but we allow each dataset to be fitted by a model (copula) with different parameters. It is thus best suited if one expects some heterogeneity between datasets.
- Compute the loglikelihoods $\mathcal{L}$ for each dataset, and average them to get a single $\mathcal{L}$. This is equivalent to the previous method, except that when summing $\mathcal{L}$, each dataset is "weighted" by its sample size, whereas when averaging $\mathcal{L}$ each dataset is equally weighted.

Since we expected differences between experimental conditions, we did not pool the data. We considered the sum of the likelihoods rather than the averages to deal with unsuccessful fits. Indeed, an unsuccessful fit results in a NaN (not a number). When averaging, NaNs are not taken into account, but they are when summed (in which case their value is 0). Therefore, averaging likelihoods does not penalize fits that did not succeed.

D VISUALIZATION OF COPULAS USED IN THIS WORK

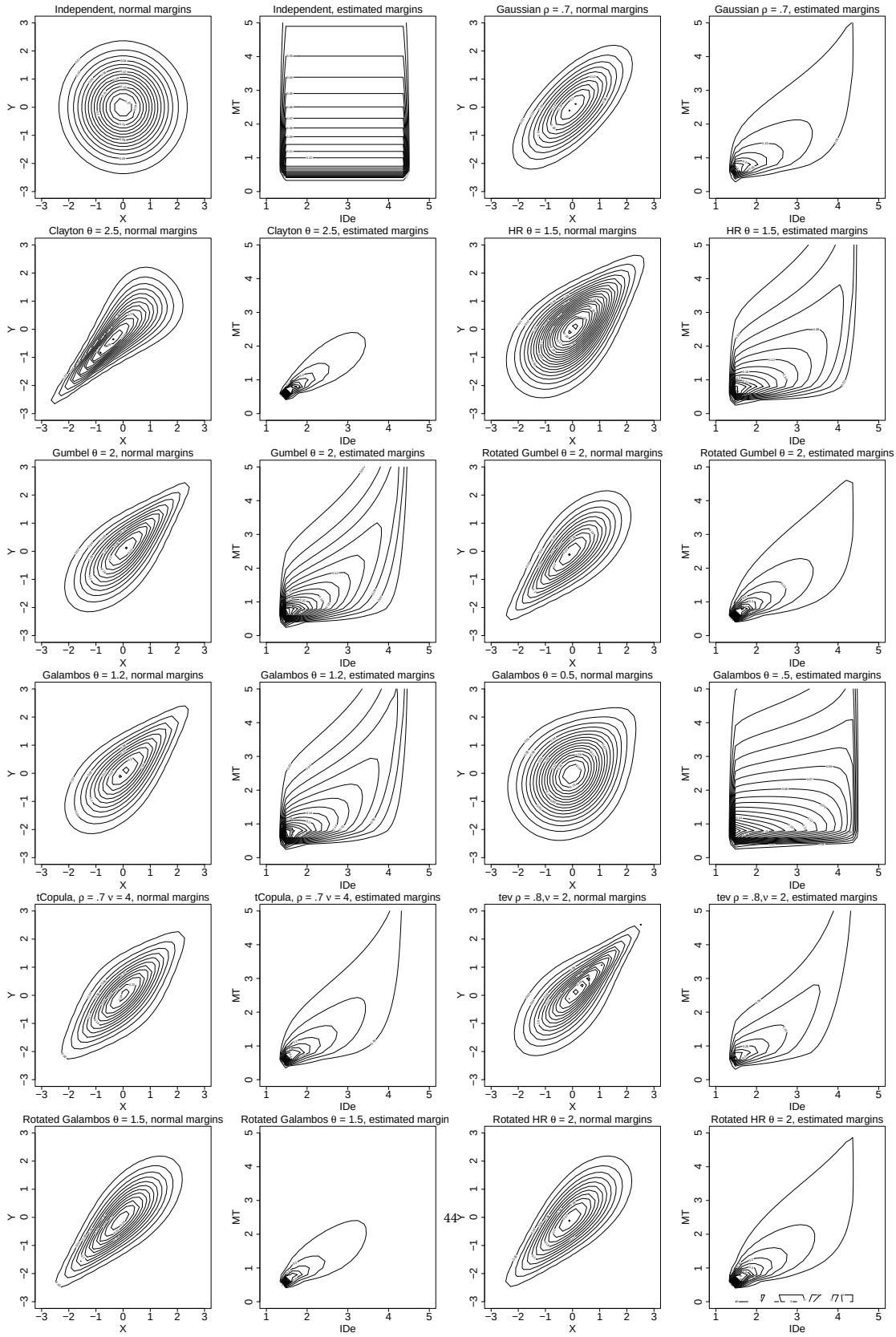

Fig. 24. Visualization of the copulas used in this work. We represent the densities of the joint distribution as coupled by 10 different copulas. For the first and third columns, the copula joins standard normal marginal distributions, which is the usual way to visualize copulas. For the second and fourth columns, the same copulas with the same parameters joins realistic marginals for MT and ID_e, estimated from the JGP dataset. The parameters values are chosen among the set of typical parameters obtained by fitting data from the JGP dataset per participant.

E METHODS: LINEAR MIXED-EFFECT MODELS

Linear Mixed-effect Models (LMMs) are useful when dealing with hierarchical or clustered data, such as repeated measurements on the same participants as in this work[39].[21] The advantage of using LMMs over traditional linear models in an experiment with repeated measurements is that they better estimate the parameters variances, thereby reducing type-1 and type-2 errors and avoiding the problem of "pseudo-replication" [39].

Scaling values. The typical values of D and W is given in pixel and can reach values beyond 1000, but regression coefficients are more easy to interpret when the dependent variables have similar scales. So, we scale D by 1/1000 and W by 1/100 for each model. In the JGP dataset, we computed that 40px = 1cm, while for the YROMK dataset, 37px = 1cm, and for the PDD dataset, 36px = 1cm. Hence, the three scales are almost equivalent. However, considering we have no information about the transfer function used in the JGP and the YROKM datasets, we do not know if these measures are truely comparable. Strategy is converted to the numerical scale [-1, 0.5, 0, 0.5, 1], and treated as a continuous variable.

Model selection. For model selection, we follow Zuur et al. [74], who suggest selecting the best model based on comparing likelihoods obtained with maximum likelihood (ML) estimation on the classical linear model, but report the parameters of the LMM obtained with restricted maximum likelihood estimation (MERL). Hence, the tables feature the parameter estimated obtained via MERL and the likelihoods obtained with ML. Fitting with ML was more prone to failures to converge, but cycling through different optimization methods always resolved this.

F IDENTIFIABILITY OF THE ROTATED GUMBEL COPULA

One hypothesis to explain the discrepency between the best fitting copulas for the global dependency is that the rotated Gumbel could be difficult to identify: perhaps we sometimes observe the Gaussian copula "by default'. We thus performed a simulation to identify how easily the Gaussian and rotated Gumbel copulas can be discriminated. We considered two cases

- first, we generated a joint distribution using the rotated Gumbel copula with typical marginals from the JGP dataset. We then identified the rotated Gumbel and Gaussian copulas on this generated dataset, and computed the model evidence ratio $\mathcal{R}$. This was performed for 100 repetitions for increasing sample sizes. This condition is labeled *continuous* in Figure 25.
- second, we wanted to evaluate the effect of blocking on identifiability (see [26] to see how a choice of stimulus level can modify the identifiability of computational models). To do so, we would have had to sample ID_e levels, and then sample from the conditional form of the rotated Gumbel copula. Unfortunately, at the time of writing, the conditional form of rotated copulas was not available in the library that we used. Therefore, we performed a bootstrap analysis instead, where we select a random dataset out of the 15 participants $\times$ 4 replications, from which we sampled data with replacement. Then, the procedure is identical to the first case; the resulting $\mathcal{R}$ is labeled *blocks* in Figure 25.

Figure 25 displays $\log \mathcal{R}$, which implies that values below 0 indicate support for the rotated Gumbel copula. This analysis shows that with the size of the datasets used (240 movements per dataset), and the replications (4 per participant), the rotated Gumbel copula should be properly identifiable. Hence, the observed difference between the two experiments does not appear spurious.

[21] Just like the repeated measures ANOVA, often used in HCI, extends ANOVA, LMM extend traditional linear regression for repeated measurements; the repeated measures ANOVA is actually a particular LMM [50].

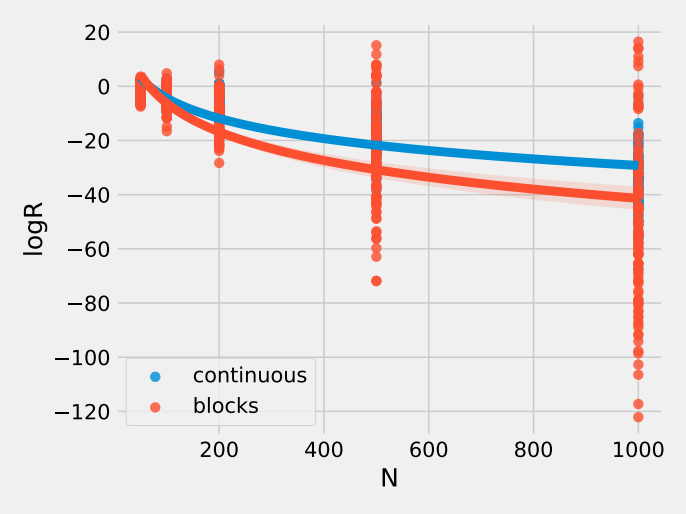

Fig. 25. Identifiavility of the rotated Gumbel copula. Data (size= N) was sampled from a rotated Gumbel copula and combined with typical marginals to form a realistic joint distribution for (ID_e, MT). A Gaussian and a rotated Gumbel copula where then fit, and the model evidence ratio for the rotated Gumbel copula against the Gaussian copula was computed ($\mathcal{R}$). Values of $\log \mathcal{R}$ below 0 indicate support for the rotated Gumbel copula. This was repeated 100 times, and fit with a quadratic model. The procedure was performed by sampling randomly from ID_e (continuous) or by simulating repeated measurements *i.e.*, several MT values for one ID_e value (blocks).