# Pointing Models for Users Operating Under Different Speed Accuracy Strategies

JULIEN GORI, Sorbonne Université, CNRS, ISIR, France

This work advances our understanding of the speed-accuracy trade-off (SAT) in human pointing movements, moving beyond the classical Fitts' law, which models movement time (MT) as a function of target distance and size but overlooks individual strategy and variability in movement times. We propose a statistical framework that incorporates user strategy and probabilistic descriptions of MT distributions. Using data from three experimental protocols and four empirical studies, we analyze the influence of strategy on SAT. We model joint distributions of speed and accuracy with copulas, a novel statistical tool in HCI. Our contributions include a theoretical analysis of correlation as a measure of goodness of fit for MT models, parametric models linking MT to difficulty and strategy, and evaluation of data generation methods for realistic pointing simulations. This work offers a richer, distribution-based perspective on pointing performance. Code and parameters are provided for reproducibility and future research.

**ACM Reference Format:**
Julien Gori. 2024. Pointing Models for Users Operating Under Different Speed Accuracy Strategies. 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, 52 pages. https://doi.org/XXXXXXXXXXXX

## 1 INTRODUCTION

Pointing towards a target, *e.g.*, selecting an icon with a mouse, is a fundamental interaction in HCI [8, 61]. The time to point towards a target of size W located at a distance D — movement time MT — is traditionally modeled using Fitts' law [22, 31, 61]

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

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

$\mathrm{ID} = \log_2(1 + \mathrm{D}/\mathrm{W})$ is called the index of difficulty; the parameters $a$ and $b$ are derived from empirical data [22, 61]. Fitts' law is a quantitative model of the speed-accuracy trade-off (SAT) of human movement [22, 61], which states that being more precise in motor tasks comes at the expense of execution speed.

Fitts' law is considered valid across diverse populations, devices conditions and input modalities [61], and is commonly used to evaluate input techniques [61] and to design interfaces [11]. However, it provides only a partial account of the SAT:

- It only describes the SAT as a central tendency, stating that a higher accuracy (smaller W) requires longer pointing times (higher MT). However, the variability of MT is significant, meaning some movements can be both faster and more precise than others — or conversely, slower and less accurate.


Manuscript submitted to ACM

Fitts' law is a deterministic description of the SAT that does not account for the distribution of movement times.

- Users can choose their precision level for a given task[1] [39, 49, 71]. While the SAT conveys one can trade off speed for accuracy, Fitts' law does not explicitly incorporate user *strategy* — a bias towards speed or accuracy — in its formulation. Instead, it enforces a task-defined accuracy via W.

This work aims to provide a more complete description of the SAT in human movement than Fitts' law, that includes user strategy explicitly and describes distribution of movement times. The suggested models have the following characteristics:

- They describe the entire distribution of movement times, and not just the means as in Fitts' law. Because of this focus on distributions, MT is treated throughout this work as a random variable. We thus suggest probabilistic descriptions of the SAT.
- They are parametric. While we could also use generative models to learn these distributions (see *e.g.* [41] for a review on the topic), parametric models offer greater adaptability to new contexts (*e.g.*, interaction devices, populations) and facilitate re-use by other researchers. Additionally, parametric models with relatively few parameters are generally more interpretable.
- One of the model's parameters maps to the concept of strategy.

Given the success of Fitts' law, why focus on aspects like movement time distributions and strategy? First, Guiard and Rioul [39] make a strong case for considering strategy. In short, their argument is that a trade-off between two variables, such as $x$ and $y$, involves both the conservation of a quantity (*e.g.*, $x \times y = k$) and changes in the contributions of $x$ and $y$ (*e.g.*, if $x$ increases, $y$ decreases to keep the product $x \times y = k$ constant). While Fitts' law addresses conservation, it does not address how users allocate their effort towards being faster or more precise (change). Second, there are practical advantages to examining distributions rather than just mean movement times that will be shortly presented (subsection 3.1). At a high level, the argument is that "who can do more can do less": distributions are inherently more desirable than means which involve information loss. While we describe in section 2 several existing models that account for either user strategy of movement time distributions, to our knowledge there is no single model that unifies both aspects.

Moving beyond Fitts' law and its deterministic view of the SAT that focuses on means implies that we adopt a *statistical* approach with a probabilistic interpretation, focused on distributions. Instead of describing the relationship between speed and accuracy in absolute terms (if one increases, the other decreases), we use statistical measures of association that quantify how *likely* it is that one variable takes on a certain value given the value of another — conditional distributions. We also use copulas, a statistical tool that to our knowledge has never been used in HCI, to describe how likely certain speed and accuracy measures are to appear together — their joint distribution.

To better understand how strategy influences the SAT, we model pointing data collected under three distinct experimental protocols (described in Appendix A) which differ in how accuracy is defined. In the standard *Fitts* protocol, accuracy is explicitly defined by the task via the target width W. In the *pure strategy* protocol, participants are required to balance speed and accuracy, according to qualitative instructions of

[1]For instance, participants of a Fitts' law experiment are often instructed to "be as fast as possible while not making too many errors" [61]

the experimenter, while aiming at a single point in space. Finally, the *Fitts-with-strategy* protocol merges both aforementioned protocols, where accuracy is set via W, but where participants are also provided strategical instructions. These three protocols provide complementary insight into how users behave when both task and strategy constraints are at play. We then suggest three different models that are based on the aforementioned empirical analysis, which are compared on their ability to generate realistic pointing data.

The main contributions of this work are:

- A theoretical analysis of correlation between the index of difficulty and the average movement time in pointing models, and what it implies for evaluating pointing models.
- Advancing knowledge of pointing, by studying the dependency between MT and the effective index of difficulty $\mathrm{ID}_e$ at the global and local level in the three aforementioned protocols.
- We demonstrate how copulas, a novel method in the field of HCI, can be used to describe dependencies between two variables. Given their flexibility, we believe copulas have the potential to model dependencies parametrically across various types of interaction data.
- Linear models for the relation between $\mathrm{ID}_e$ and D, W, ID and strategy.
- A strategy-dependent bivariate Gaussian model that relates mean movement time and $\mathrm{ID}_e$ in the pure strategy protocol
- Empirical support for the EMG model for movement time distributions in the pure strategy and Fitts-with-strategy protocols
- Three methods to generate realistic pointing data in a pure-strategy task, and their technical evaluation.

To ensure a better dissemination of these ideas, we have highlighted the parameters used to generate realistic data available in the paper that could be reused directly by other researchers; the code used for this paper is available at the following anonymized repository.

*Notations.* $\mathcal{N}(\mu, \Sigma)$ refers to the (multivariate) Gaussian distribution with mean $\mu$ and covariance matrix $\Sigma$, $\mathcal{E}(\lambda)$ refers to the exponential distribution with mean $\lambda$, $\mathbb{E}[X]$ refers to the mathematical expectation (population mean) of $X$, $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ refers to the (population) variance of $X$. We use $X \sim \mathcal{D}$ to indicate that the random variable $X$ is distributed following the distribution $\mathcal{D}$. We use R-style formulas to describe statistical models for example, $A \sim X * Y + (1|Z)$ is a linear mixed effect model with random intercepts for groups of $Z$, with main effects $X$ and $Y$, and with an interaction effect between $X$ and $Y$.

## 2 RELATED WORK

The existing work on pointing models is vast. Indeed, pointing is one type of goal-directed movement, which have been extensively described in the motor control literature (see *e.g.* [63, Section 2] for an introduction). In HCI, pointing is one of the fundamental elementary interactions of users with GUI; the topics has and continues to attract attention.

### 2.1 Extending Fitts' law

Many studies extend Fitts' law — namely its applicability to novel paradigms, for example 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 [34], while Murata and Iwase [52] extend the

framework to trivariate pointing. These models however generally do not provide movement time (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 [7, 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. Finally, Grossman and Balakrishnan [34] define ID in terms of the probability of hitting the target — a concept that may initially seem tied to user strategy. However, as the authors clarify, this probability is influenced by geometric factors, specifically the likelihood of hitting the target in a ballistic, open-loop movement, but as noted by Gori *et al.* [29], strategy differences are more pronounced during the corrective phase of movement[2].

## 2.2 Pointing with different strategies

Various work have investigated the effect of strategy on pointing performance. This includes work on the invariance of throughput for different user strategies [5, 49, 57] and the comparison of evaluations of different interaction techniques [68]. The inclusion of strategy in pointing models has lead researchers to define variations on Fitts' protocol described Appendix A.

*2.2.1 The effective index of difficulty $ID_e$.* Because participants do not always follow the accuracy constraint as enforced by W, Crossman [12] suggested a correction that was popularized in HCI by Mackenzie [31, 61], called the effective index of difficulty[3] $\mathrm{ID}_e$, which is a function of the actual measured standard deviation of endpoints $\sigma$

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

This correction can only be computed a posteriori given that it relies on $\sigma$.

*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 [31, Appendix 3], when the spatial distribution of endpoints is Gaussian — which is a reasonable assumption that holds well in practice [29, 61] — $W_e = \frac{2.066}{\sqrt{2}\mathrm{erf}^{-1}(1-\varepsilon)}W$, where $\varepsilon$ is the rate at which participants miss the target. Thus, $I_u$ is a non-linear transformation of the miss rate $\varepsilon$. 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 $\mathrm{ID}_e$.

*2.2.3 The WHo model.* To our knowledge, the WHo model [39], which follows from the ressource allocation model described [38], is the only parametric model that describes pointing data that conforms to the SAT with parameters standing for participant strategy. Guiard and Rioul [39] listed 6 axioms that any *deterministic* theory of the SAT should respect

[2]This phase, crucial for precise tasks, can last up to four times longer than the initial ballistic phase, see [29].
[3]The index generalizes to multiple dimensions, but not without controversies, see *e.g.* [26, 48, 67].

axiom 1 Each participant has a minimum average movement time $x_0$ that they can reach

axiom 2 With relative precision $y = \sigma/D$ with $\sigma$ the standard deviation of endpoints, each participant has a minimum relative precision $y_0$ that they can reach.

axiom 3 The function $f$ that links x to $y$, $y = f(x)$ is decreasing and convex

axiom 4 For a given participant effort $k$, there is an operation such that $x \odot y = k$

axiom 5 Effort invested is never total: participants may routinely perform worse than ideally possible.

axiom 6 Participants can modulate how they allocate resources to speed versus accuracy.

The WHo model

$$(y - y_0)^{1-\alpha}(x - x_0)^{\alpha} = k \tag{5}$$

satisfies axioms 1–4, and is fit on the convex hull of the scatterplot in the $(x, y)$ space to satisfy axiom 5. A $\beta \in [0, 1]$ parameter for strategy (axiom 6) is defined as

$$\beta = \frac{\lambda}{1 + \lambda}, \text{ with } \lambda = -\frac{dy}{dx}. \tag{6}$$

However, the WHo model does not describe a distribution of movement times. Instead it characterizes the "best average" samples (axioms 1, 2 and 5), 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. 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[4]:

$$\mathrm{MT} = a + b\mathrm{ID}_e + \text{noise (additive)}, \quad \mathbb{E}[\text{noise}] = 0 \text{ (centered)}. \tag{7}$$

Pearson's $r$ is commonly used as a goodness-of-fit measure. However, as shown by Drewes [19] and Gori *et al.* [33], the Pearson correlation between MT and $\mathrm{ID}_e$, $r(\mathrm{MT}, \mathrm{ID}_e)$, often yields moderate to low values in practice on pointing datasets, with some cases showing correlations close to zero [31].

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 $\overline{\mathrm{MT}}$[5], before calculating $r$. Consequently, the model actually considered by many researchers is:

$$\overline{\mathrm{MT}} = a + b\mathrm{ID}_e + \text{noise (additive)}, \quad \mathbb{E}[\text{noise}] = 0 \text{ (centered)}, \tag{8}$$

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 8 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 $\mathrm{ID}_e$ level

[4]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.

[5]See [32] for further discussion on considering the sample average.

is distributed according to an asymmetric, positive (right) skewed [10, 25, 28, 44, 47, 56, 73] distribution. While several distributions have been suggested, including gamma, Gumbel and lognormal, we follow Gori and colleagues [25, 28, 47], and Zhao *et al.* [73] who considered an Exponentially Modified Gaussian (EMG) noise model: it is positive, is assymetric (except when $1/\lambda << \sigma^2$), has parameters that are easy to interpret, has shown good performance (including when modeling steering data [65]), and there exist tools to aid in its parameter estimation [47].

The EMG model for movement time defines MT as the sum of a Gaussian ($G$) and an Exponential ($E$) random variable [28, 47]

$$\mathrm{MT} = G + E, \tag{9}$$

$$G \sim \mathcal{N}(\beta\mathbf{x}, s) = a + b\,\mathrm{ID}_e + \mathcal{N}(0, s), \text{ with } \beta = [a, b] \text{ and } \mathbf{x} = [1, \mathrm{ID}_e]^t \tag{10}$$

$$E \sim \mathcal{E}(\lambda) \tag{11}$$

The $G$ part models pointing as the standard Fitts' law with Gaussian noise, in line with Equation 7, as reflected by the right hand side of Equation 10. $E$, being an exponential random variable, is always positive meaning that the EMG model only allows deviations from Fitts' model that increase movement times. This idea was justified for example in [25, 30, 33] and builds on arguments of the WHo modelsubsubsection 2.2.3. That the positive random variable is exponential can be justified by a maximum entropy [42] argument, see [28].

Gori [25] further showed that a model where the standard deviation of the MT distribution increases linearly with its mean provided a good fit to empirical data:

$$E \sim \mathcal{E}(\lambda\mathbf{x}), \;\; \lambda = [\lambda_0, \lambda_1], \tag{12}$$

which is in line with results by Zhao *et al.* [73] who find that quadratic variance models provide the best fit among the models they tested.

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

$$p(y) = \frac{1}{2\lambda\mathbf{x}} \exp\left(\frac{1}{2\lambda\mathbf{x}}(2\beta\mathbf{x} + \frac{s^2}{\lambda\mathbf{x}} - 2y)\right) \mathrm{erfc}\left(\frac{\beta\mathbf{x} + \frac{s^2}{\lambda\mathbf{x}} - y}{\sqrt{2}s}\right), \tag{13}$$

where erfc is the complimentary error function. Although this density is analytically quite complex, it has very simple conditional mean

$$\mathbb{E}[\mathrm{MT}|\mathrm{ID}_e] = \beta\mathbf{x} + \lambda\mathbf{x} = a + \lambda_0 + (b + \lambda_1)\mathrm{ID}_e, \tag{14}$$

and conditional variance

$$\mathrm{Var}(\mathrm{MT}|\mathrm{ID}_e) = s^2 + (\lambda\mathbf{x})^2 = s^2 + \lambda_0^2 + 2\lambda_0\lambda_1\mathrm{ID}_e + \lambda_1^2{\mathrm{ID}_e}^2. \tag{15}$$

To fit the EMG distribution, one can use standard methods such as maximum likelihood estimation[6].

[6]The code associated with this paper provides examples fits.

### 2.4 Two Fitts' laws

Gori and colleagues [30, 31] have argued that there exist two version of Fitts' law: an average Fitts' law — the *mean* law — computed via linear regression on $\overline{\text{MT}}$ as in Equation 8, and a minimum Fitts' law — the *min* law. Fitts' min law represents the best performance possible, whereas Fitts' mean law represents a more central tendency.[7] The $\beta$ parameters of the EMG model identify the min law [28], while the $a$ and $b$ of linear regression identify the mean law. In this work, we use both version of Fitts' law to compare our models.

### 2.5 Models that simulate entire trajectories

*2.5.1 Optimal control models.* Müller and colleagues have applied several optimal control models for pointing in HCI [21, 45] that follow from the seminal work by Todorov and Jordan [62]. These predict complete movement trajectories as well as their spatial variability. Unfortunately, these models are not well suited to predict movement time. Indeed, operationalizing optimal control models requires introducing a stopping condition to terminate movements[8]. For instance, Fischer [20, pp. 36–37] truncated movements at 0.97 seconds to observe endpoint standard deviation. The rationale behind this specific cut-off is unclear, and more critically, this approach assumes that all movements have identical durations. This assumption contrasts sharply with empirical pointing data, where MT variability is substantial — particularly in less controlled environments [10]. While several solutions to this cut-off problem can be considered, to our knowledge, no existing optimal control model currently satisfactorily predicts MT distributions. While Qian et al. [58], and Berret and Jean [6] propose interesting steps in this direction, further developments are needed to address this gap.

*2.5.2 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 — referred to here as simulation models — have also been proposed. For instance, Do et al. [17] 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. [50] 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.

[7] In fact, it is probably possible to go even further: as demonstrated by Chapuis *et al.* [10, Fig. 9], a Fitts' law applies at the quantile level of MT.
[8] Or the use of a fixed horizon, which also has to be chosen.

# 3 PRELIMINARIES ON MODELS OF MOVEMENT TIME

## 3.1 Why consider distributions over mean values?

One objective of this work is to generate joint distributions of $\mathrm{ID}_e$ and MT, rather than some linear relation between $\mathrm{ID}_e$ and mean movement time $\overline{\mathrm{MT}}$. One reason is that a model that produces entire distributions is inherently more informative than one focused on generating mean values. However, beyond this generic argument, we further identify three practical reasons to prioritize distributions over means when modeling movement times:

- When averaging function values across multiple points, the results generally differ from evaluating the same function at the mean of those points; the two coincide only if the function is linear. This point has been made before in the context of throughput [57].[9] In HCI research, simulations are increasingly used to predict task completion times [54], sometimes relying on pointing models as core elements. Recent examples include [43, 46, 60], which use the WHo model as a low-level component in a computational-rationality framework. Given many simulations include non-linear functions, simulations using MT distributions will lead to different outcomes than simulations using an average movement time.[10]
- In some cases, means lack sufficient detail. For instance, simulating the sensitivity of various throughput measures introduced in [57] to specific changes would require a generative model that produces distributions, not just means. In fact, exploring this specific question is what motivated the present work.
- Models of movement time poorly fit pointing data from non controlled studies, such as web-based or field experiments (*e.g.*, $r^2$ between MT and ID is close to 0 in [31]), in contrast to models that describe the full distribution [10, 25] and which can be fitted to the variability inherent to observational/field studies, which are increasingly recognized as complimentary to controlled studies [16]. Thus, pointing models that describe distributions are required to model data acquired ouf of the lab.

## 3.2 Theoretical analysis of the $r^2(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ correlation

The goodness of fit of Fitts' law is usually justified via Pearson correlation $r^2(\overline{\mathrm{MT}}, \mathrm{ID}_e)$, which is close to 1 in many experiments where accuracy is manipulated by W [61]. It turns out that many MT distributions can achieve $r^2(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ 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 delayed to Appendix B, but essentially relies on the fact that Pearson correlation is a function of the covariance, which measures linear deviations from the mean. Below, we give some example distributions that have been used in modeling pointing times that fulfill this condition

- The EMG distribution [25, 28, 47, 73] Equation 14 has linear conditional mean.

[9] For convex functions, the mean of the function values is greater than the function evaluated at the mean; for concave functions, the reverse is true — a result known as Jensen's inequality [57].

[10] One application of these simulations is in optimizing designs [54]: simulations are conducted with different design parameters, and the most effective parameter set is selected. To our knowledge, no studies have yet assessed how optimizing based on distributions, rather than means of pointing models (or any other model for that matter), might impact design outcomes, which forms an interesting research avenue.

- A gamma [44] distribution with shape $k$ and scale $\theta$ has mean value $k\theta$. Having $k$ or $\theta$ increase linearly with $\mathrm{ID}_e$ would fulfill the condition.
- A generalized extreme value [10] distribution $\mathrm{GEV}(\mu, \sigma, s)$ has a mean value that increases linearly with $\mu$. Having $\mu$ as a linear function of $\mathrm{ID}_e$ 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. Hence, while having a high $r^2(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ is a necessary condition, it is not a sufficient condition to evaluate model fit of Fitts' law, since it is easily attainable by imposing a linear conditional mean. Hence, $r^2(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ does not characterize the dependency between MT and $\mathrm{ID}_e$ well enough. There are two important takeaways from this result. First, it supports the linear conditional mean models for MT described subsubsection 2.3.2. Second it motivates the need to turn to measures of dependence that are more sophisticated than Pearson correlation; these will be introduced shortly.

### 3.3 Several $\overline{\mathrm{MT}}$'s, one notation.

A typical Fitts' law experiment has several participants perform pointing movements under various (D,W) conditions. Sometimes these conditions may be repeated. It is usually unclear when researchers talk about the mean movement time which quantity they actually consider; there are numerous ways in which the data can be aggregated before computing the mean [57]. For example, the aggregate can be on all different (D,W) combinations, as suggested in [61], or on all different ID levels[11] [1]. Sometimes, researchers will aggregate over all participants, like in [1], sometimes not like in [49]. When some blocks are repeated, aggregation may again be performed over repetitions [44] or not [37]. Each extra layer of aggregation will lead to a higher $R^2$; the same mechanism as the one described in the proof of Proposition 3.1 applies. In practice, $\overline{\mathrm{MT}}$ may thus refer to different quantities; high $r^2(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ usually result from computing $\overline{\mathrm{MT}}$ over several aggregated layers.

## 4 MEASURING THE DEPENDENCE BETWEEN TWO RANDOM VARIABLES WITH BIVARIATE COPULAS

A (bivariate) joint distribution describes the probability of two variables simultaneously; a marginal distribution represents the probability of a single variable when the other variable(s) are averaged out. (Bivariate) copulas are a statistical function that uniquely describe the dependence between two random variables: any joint (bivariate) distribution can be expressed in terms of its marginal distributions and a copula — a result known as Sklar's theorem [55]. Thus, copulas "couple" [55, Chapter 1] two marginals into a joint distribution; they thus capture how variables depend on each other irrespective of their actual values. Copulas are parametric and describe how the two variables are associated, for example imposing a prescribed (rank) correlation, everywhere, or only at their tails.

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

[11] Because of the usual geometric progression of W and D, multiple W and D pairs may lead to the same ID [33].

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 [25].

Different types of copulas introduce different types of dependencies; for example a Gumbel copula introduces so-called upper tail dependence [55], 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 for example used when modeling extreme financial events [15]. 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. The copulas used in this work have all been represented Figure 27.

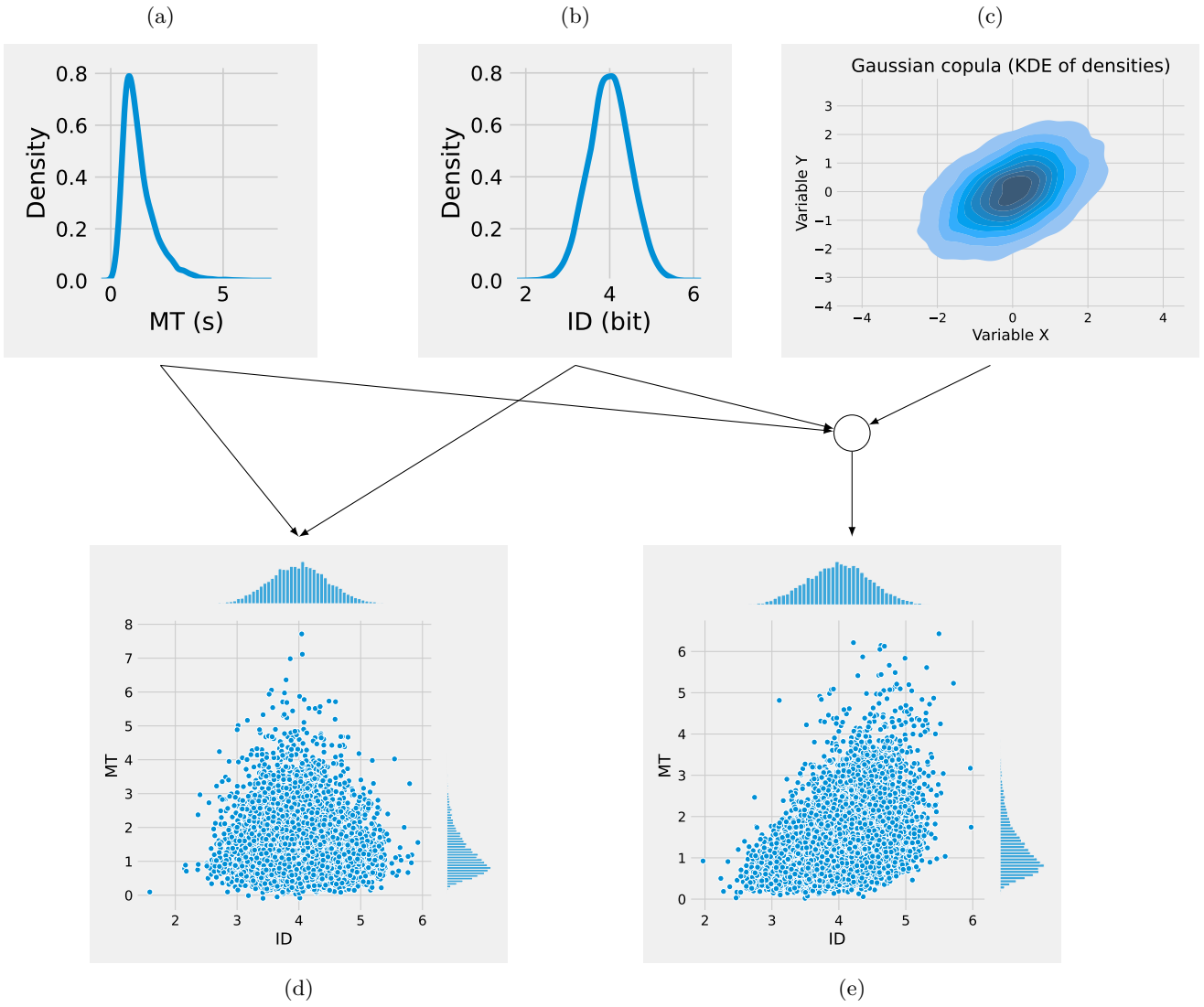


Fig. 1. How copulas couple two marginals. (a) represents the marginal distribution for MT, here an EMG model. (b) represents the marginal distribution for ID, here a Gaussian model. One can sample independently from MT and ID to obtain a joint (MT, ID) distribution as in (d), however, it does not reflect 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.

## 5 OVERVIEW OF THE RESULTS

We characterize how the dependent variables MT and $\sigma$ (speed and accuracy) vary with the independent task variables D and W, and strategy $s$. For simplicity, and to build on existing results, we consider $ID_e$ rather than $\sigma$, which is just a transformation of $\sigma$. Hence, the two models of interest are the joint distribution

of MT and $ID_e$

$$p(\mathrm{MT}, \mathrm{ID}_e), \tag{16}$$

and the same joint distribution conditional on s, and D and W.

$$p(\mathrm{MT}, \mathrm{ID}_e | s, \mathrm{D}, \mathrm{W}). \tag{17}$$

The conditional distribution Equation 17 describes MT and $ID_e$ for given values of D, W and s *i.e.*, one experimental condition. Because spatially the associated data is usually clustered in a small area, we call this a "local" type of dependence. On the other hand, Equation 16 describes MT and $ID_e$ for data aggregated over all values D, W and s of the dataset, and represents what we call a "global" type of dependence. The local and global dependencies may turn out to be quite different (and they will); Figure 2 illustrates data which is locally uncorrelated, but which has mean values and a variability that increases with $ID_e$ levels. The result is that there is a global dependency structure, where high values of MT are associated with high values of $ID_e$, as expected from the SAT, even if locally there may be none.

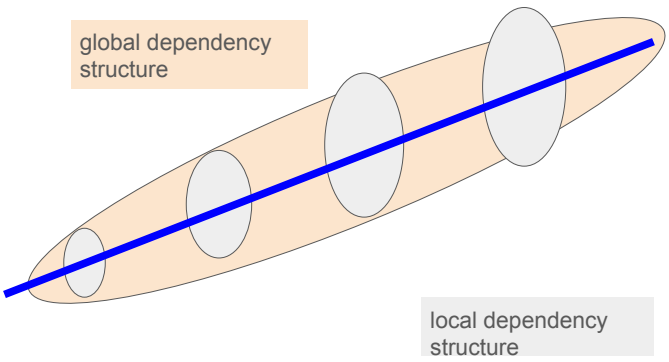


Fig. 2. The difference between the local and global dependency structure. The blue ellipses, which represent 95% confidence ellipses for data associated with one block, display an uncorrelated structure between MT and $ID_e$, since they are oriented vertically. The yellow ellipsis, which represents the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

To learn these joint distributions, we suggest two approaches:

(1) In the first approach, we use the chain rule of probability to write

$$p(\mathrm{MT}, \mathrm{ID}_e | \mathrm{D}, \mathrm{W}, s) = p(\mathrm{MT} | \mathrm{ID}_e, \mathrm{D}, \mathrm{W}, s) \times p(\mathrm{ID}_e | \mathrm{D}, \mathrm{W}, s) \tag{18}$$

We then make the assumption that the movement time distribution only depends on W, D and s through $ID_e$:

$$p(\mathrm{MT} | \mathrm{ID}_e, \mathrm{D}, \mathrm{W}, s) = p(\mathrm{MT} | \mathrm{ID}_e). \tag{19}$$

We then leverage the existing EMG model of $p(\mathrm{MT}|\mathrm{ID}_e)$ described in subsubsection 2.3.2 on one hand, and what remains to be modeled is $p(\mathrm{ID}_e|\mathrm{D}, \mathrm{W}, s)$.

(2) The first approach relies on the quality of the EMG model and its quadratic variance model. A competing model could be obtained by comparing several variance models and determine the best model *e.g.*, based on maximum likelihood estimation or bayesian estimation as performed by [73]. In this second approach, we instead take a more principled route, and try to directly match the structure of the dependence between $\mathrm{ID}_e$ and MT with existing copulas.

Learning the global dependence $p(MT, \mathrm{ID}_e)$ follows equivalently.

*Dependence between $ID_e$ and MT in the Fitts protocol.* In the first study (section 6), we investigate $p(\mathrm{MT}, \mathrm{ID}_e)$ and $p(\mathrm{MT}, \mathrm{ID}_e|D, W)$ when accuracy is manipulated by W, using an existing dataset produced with gestural interaction and the standard Fitts protocol described subsection A.1. We find that $p(\mathrm{MT}, \mathrm{ID}_e)$ is best modeled by a rotated Gumbel copula, which is characterized by a lower tail dependency *i.e.*, strong correlation for low values of $\mathrm{ID}_e$ and MT. On the other hand, $p(\mathrm{MT}, \mathrm{ID}_e|D, W)$ is best modeled by a low correlation Gaussian copula. We suggest a linear model relating $\mathrm{ID}_e$ to ID, used to form a statistical model for the conditional distribution $p(\mathrm{ID}_e|\mathrm{ID}, \mathrm{D}, \mathrm{W})$.

*Dependence between $ID_e$ and MT in the pure strategy protocol.* In the second study (section 7), we model data from the pure strategy protocol described subsection A.3, where accuracy is manipulated by strategy $s$. We first investigate the joint distribution $(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ of mean movement times and effective index of difficulty, and propose a bivariate Gaussian model for $(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ parametrized by $s$. We then show that the EMG distribution with quadratic variance is also suited to model data from the pure strategy protocol. Finally, we find that Gaussian / t-copula pair provides the best fit for $p(\mathrm{MT}, \mathrm{ID}_e|s)$ and $p(\mathrm{MT}, \mathrm{ID}_e)$.

*Dependence between $ID_e$ and MT when accuracy is manipulated by W and strategic instructions.* In the third study (section 8), we investigate $p(\mathrm{MT}, \mathrm{ID}_e|s, D, W)$ when accuracy is manipulated by W and strategic instructions, using an existing dataset produced with the *Fitts-with-strategy* protocol described subsection A.4. We also revisit the model relating $\mathrm{ID}_e$ to ID previously obtained, by adding a strategy term, and, similar to the first study, find that the combination of Gaussian / t-copula provides the best fit to empirical data.

## 6 FIRST STUDY: DEPENDENCE BETWEEN $\mathrm{ID}_e$ AND MT IN FITTS' PROTOCOL

### 6.1 Dataset presentation

We used an existing dataset by Jude *et al.* [44] (JGP). The JGP dataset was generated using the well-known 2D multi-directional tapping task [61] via the FittsStudy software [67], 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, with $\mathrm{ID}_e$ and MT already calculated.

### 6.2 Fitting EMG on MT given $ID_e$ values

We fitted the EMG model with linear mean Equation 9–Equation 12 for each participant and for each replication. Without surprise, and in line with previous works discussed subsubsection 2.3.2, the EMG fit was largely superior to the Gaussian fit obtained via linear regression with AIC differences routinely above 300 (a difference of 10 is usually considered good support). A pair-plot of the parameters obtained after fitting can be found in the supplementary materials.

### 6.3 Fitting copulas on (MT, $ID_e$)

*6.3.1* $p(MT, ID_e|D, W)$. We first examine $p(\mathrm{MT}, \mathrm{ID}_e|D, W)$ *i.e.*, the relationship between MT and $\mathrm{ID}_e$ within each D and W condition individually — what we previously called the local dependency.

We fitted copulas to the pointing data for each participant and (D, W) pair separately as detailed in subsection C.1. The comparison of copulas was based on the model evidence ratio $\mathcal{R}$ explained subsection C.2. The interpretation of $\mathcal{R}$ is as follows: when comparing two models with an evidence ratio of $x$, there is 1 chance out of $1/x$ that the worst model is actually better than the other. The t-EV, t, rotated Gumbel, rotated Galambos, and rotated HR copulas often failed to converge.[12] Because these failures to converge were spread equally across (D,W) pairs (see the supplementary materials), we simply excluded these copulas from the analysis.

The $\mathcal{R}$ values for the remaining copulas and D×W condition pairs are shown in Figure 3. The Gaussian copula emerged as the best fit according to $\mathcal{R}$, while the Galambos, HR, and Clayton copulas also provided reasonable dependence models under certain conditions. We did not observe any relationship between the D and W conditions and the best-fitting copula, which challenges the hypothesis that the structure of the dependence between $\mathrm{ID}_e$ and MT is influenced by D or W. However, these findings should be interpreted with caution due to the limited size of the datasets. Specifically, for each participant and each D,W pair, there are only four distinct $\mathrm{ID}_e$ levels, which we believe is insufficient to yield robust results.

Therefore, we pooled the participant data together before fitting the copulas and re-analyzing them for different (D,W) conditions — the interpretation is that we look at population version of $p(\mathrm{MT}, \mathrm{ID}_e|D, W)$. Again, we had to remove the following copulas from the analysis: Galambos, HR, t-EV, rotated Galambos, and rotated HR. Surprisingly, as illustrated Figure 4 we found that the independent copula was the best fitting copula, except for the W = 64 condition. This means that except for the smallest target, there is no evidence for a local dependence between MT and $\mathrm{ID}_e$ *i.e.*, for a given (D,W) pair, a larger $\mathrm{ID}_e$ is not associated with larger MT values. For completeness, we computed rank correlations for each condition; for the 4 different conditions where W = 64, both Spearman's and Kendall's rank correlation between $\mathrm{ID}_e$ and MT were slightly negative, which is opposite to the speed-accuracy tradeoff.

*6.3.2* $p(MT, ID_e)$. The previous analysis suggest that there is little to no local dependency structure between MT and $\mathrm{ID}_e$. We thus examined the global dependency between $\mathrm{ID}_e$ and MT by fitting $p(\mathrm{MT}, \mathrm{ID}_e)$ with the same method as before.

[12] This can happen when a copula is made to fit a dependency structure it can't represent, *e.g.*, fitting a copula which can only model positive rank correlation to a dataset with negative rank correlation.

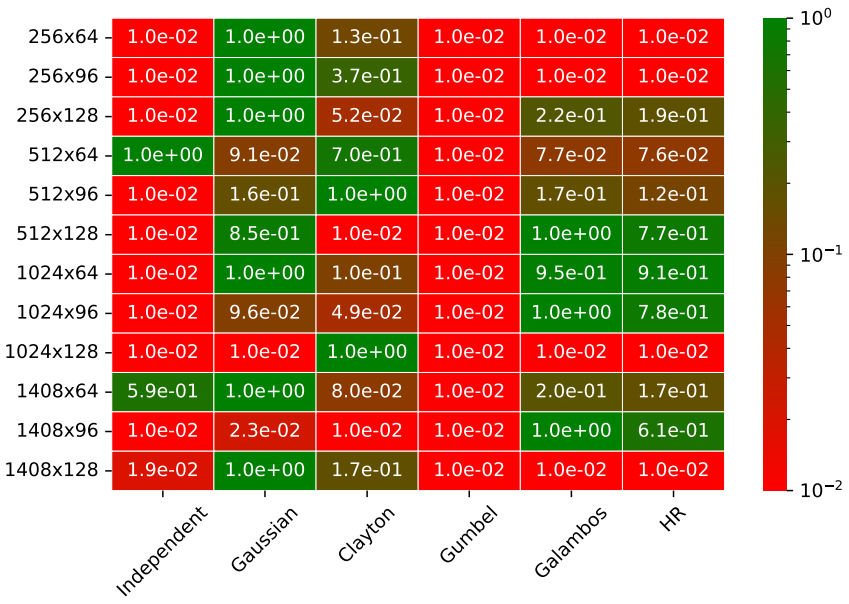


Fig. 3. Model evidence ratio $\mathcal{R}$ computed against the best copula for each D×W pair and each participant.

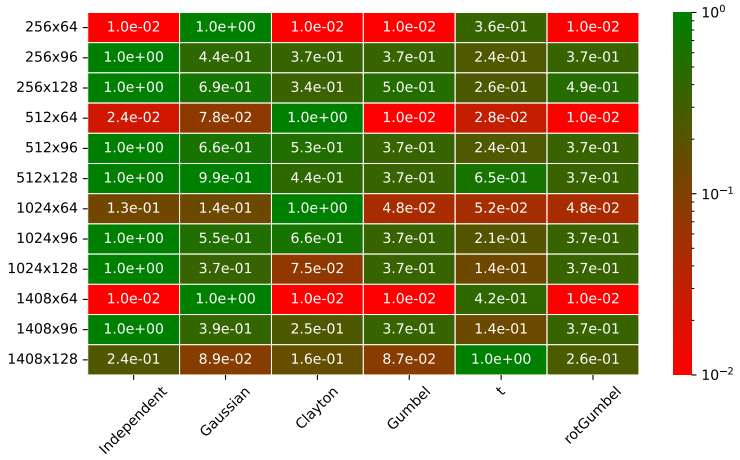


Fig. 4. Model evidence ratio $\mathcal{R}$ computed against the best copula for each D×W pair, pooled over participants.

The results displayed Figure 5 show that the rotated Gumbel copula consistently outperformed other models. The rotated Galambos copula followed closely — under certain conditions the Gumbel and Galambos copulas are known to be very similar [24] which explains this observation.

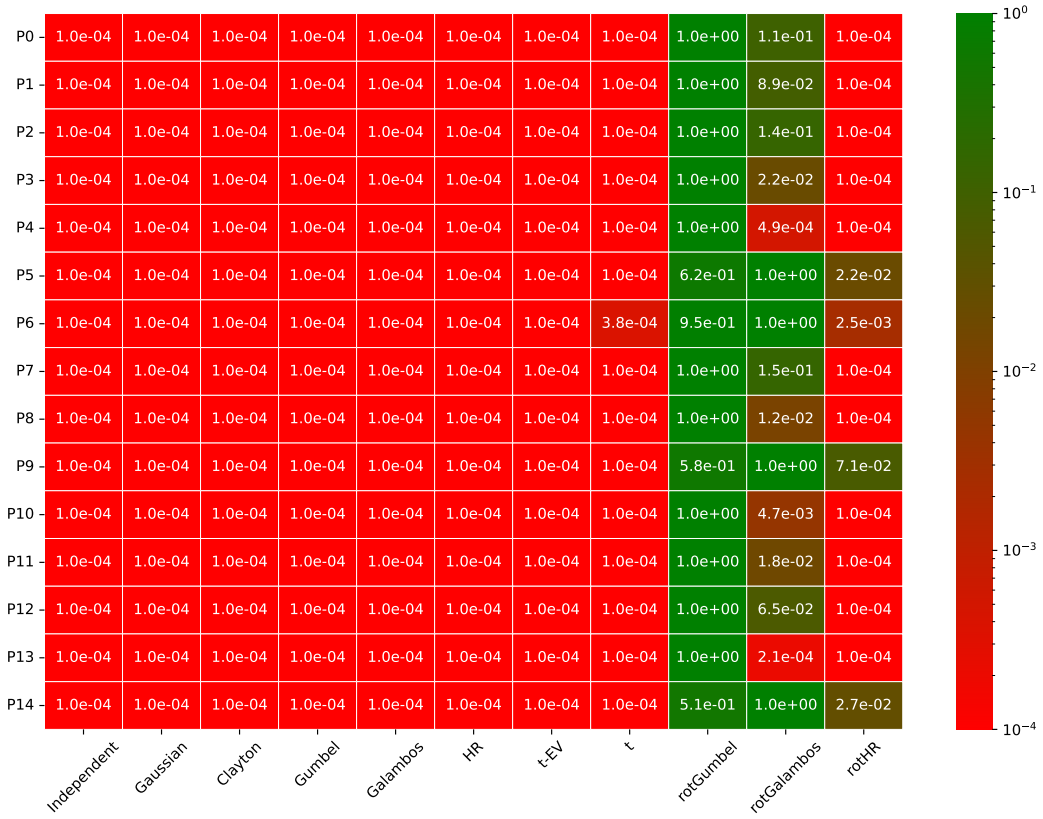


Fig. 5. Model evidence ratio $\mathcal{R}$ computed against the best copula for each participant, with a threshold at $\mathcal{R} = 1e^{-4}$ for a better use of the color palette. The interpretation of $\mathcal{R}$ is as odds against the best fitting copula: For example, a model with $\mathcal{R} = 0.11$ indicates that the best model is about $1/0.11 \simeq 9$ times more likely.

### 6.4 Model between $ID_e$ and D, W, ID

To evaluate the effects W and D can have on $ID_e$ other than through ID, we fitted[13] four mixed-effects linear models incorporating main and interaction terms, each with a random intercept for participants:

$$ID_e \sim ID * W * D + (1|Participant) \tag{20}$$

$$ID_e \sim ID * W + (1|Participant) \tag{21}$$

$$ID_e \sim ID * D + (1|Participant) \tag{22}$$

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

The full model (Equation 20) includes seven terms: three main effects, three pairwise interactions, and one three-way interaction. However, since ID is a function of both D and W, the model has highly correlated

[13] See Appendix E

predictors, which complicates interpretation of the 8 coefficients. Therefore, we also examined simpler models (Equation 21 and Equation 22). The fits of the simpler models are shown in Figure 6, where we observe:

- **W's Effect:** A slight slope reduction as W decreases, suggesting a positive interaction between W and ID. This aligns with subsubsection 2.2.2, where smaller W values prompt over-utilization of the target, leading to smaller $\mathrm{ID}_e$ values.
- **D's Effect:** Both slope and intercept increase with larger D, indicating higher overall $\mathrm{ID}_e$ values.

Detailed summaries of the fitted models are provided in the supplementary materials and summarized Table 1. The key findings are:

- In the full model (Equation 20), only ID has a significant effect ($\alpha = 0.05$). However, multicollinearity likely hampers the separate evaluation of D, W, and ID effects. Further, model comparison based on likelihood does not support this model (the simpler models achieve the same likelihood with less free parameters).
- In the simpler models:
  - D and its interaction with ID are significant ($\alpha = 0.05$) contributors to $\mathrm{ID}_e$.
  - W and its interaction with ID are not significant ($\alpha = 0.05$) contributors to $\mathrm{ID}_e$.

Log-likelihood comparisons show that the full model's log-likelihood is within one point of the two simpler models (Equation 21 and Equation 22), despite the former having twice as many fixed effect parameters. Consequently, complexity-penalizing criteria (e.g., AIC, BIC) favor the simpler models. The model with only ID as main effect scores much lower on likelihood, and is therefore not supported by the data.

In conclusion, while the visual fits suggested a significant effect of W on the slope of the ($\mathrm{ID}_e$, ID) relationship, this was not confirmed by statistical analysis. On the other hand, the effect of D was found to be significant, and the D model Equation 22 was equivalent from a likelihood perspective. This finding diverges from earlier results by Zhai *et al.* [71]. Two possible explanations are:

- The values of W in this experiment (64, 96, 128)px were larger than in [71] (12, 36, 72)px[14]; The interaction between W and ID likely would likely manifest itself more for small values of W.
- Evaluating a model like Equation 21 would benefit from a fully crossed ID and W design, such as a shape × scale design [33, 36] to break the collinearity between ID and W (similar reasoning can be applied to the collinearity between ID and D).

### 6.5 Interpretation of the first study

This study examined the dependency between $\mathrm{ID}_e$ and MT in pointing data collected using Fitts' protocol, where accuracy is specified a priori through W. As already found in the literature, and as further supported in our theoretical analysis subsection 3.2, the conditional distribution $p(\mathrm{MT}|\mathrm{ID}_e)$ can be well fit with the EMG distribution with a quadratic variance model.

Fitting copulas to the joint distribution $p(\mathrm{MT}, \mathrm{ID}_e|D, W)$ showed that there was little to no correlation between $\mathrm{ID}_e$ and MT for a given (D,W) pair; the condition with the smallest width ($\mathrm{W} = 64$) even being slightly negatively correlated. In our opinion, this suggests that, from the perspective of the dependency of $\mathrm{ID}_e$ and MT, it is the task variables D and W defined a priori that matter, rather than $\mathrm{ID}_e$ defined a

[14]Comparisons based on pixel sizes only are not necessarily meaningful, as they are only equivalent under identical pixel densities, and from a motor control perspective, also depend on various properties of the input device [9].

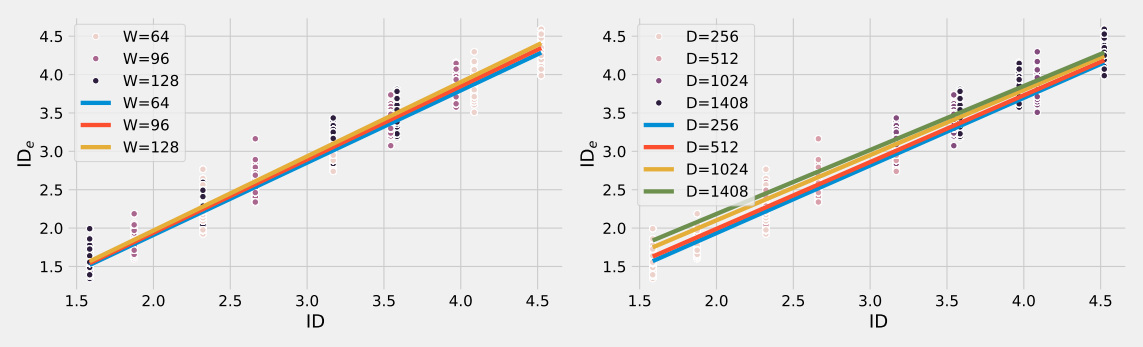


Fig. 6. The relationship between $ID_e$ and ID depending on W values (left) and D values (left). The filled circles represent the observed $ID_e$ values for prescribed ID levels for varying W or D values, while the lines represent the linear model fits for varying W or D values.

Table 1. Mixed effect model fit results for Equation 20, Equation 21 and Equation 22. 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 (Equation 20) | W model (Equation 21) | D model (Equation 22) | only ID (Equation 23) |
|---|---|---|---|---|
| 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 |

posteriori. In other words, when accuracy is imposed by W, movements produced within a (D,W) pair are equivalent — actual differences in execution are likely due to inherent variability of movements rather than planning and do not adhere to a local speed-accuracy trade-off.

We then pooled D and W conditions to model $p(\mathrm{MT}, \mathrm{ID}_e)$. This time, the rotated Gumbel copula clearly emerged as the best fit. The rotated Gumbel copula models lower tail dependence *i.e.*, the fact that low values of $ID_e$ are associated with low values of MT. This finding is in line with the results from subsection 6.2, since lower tail dependence implies strong correlation between low values of $ID_e$ and MT, and hence a small spread of the MT data for low values of $ID_e$.

Finally, we investigated the relationship between $ID_e$ and ID. While the linear models tested provided good fits, we could not establish statistical significance for the effect of W on $ID_e$, contrary to previous findings in the literature: indeed [71] reports that W impacts $I_u$ far more than D. For completeness, and to address potential collinearity among D, W, and ID, we also fit the linear models with ridge regression with cross-validation[15]. Although detailed results are not reported here, the estimated coefficient for ID was 0.94,

[15]Ridge regression is a technique for fitting linear models when collinearity exists among predictors, as it produces more stable and reliable estimates of effects [35]. The method introduces a regularization term controlled by a hyperparameter

closer to 1 than the estimates from the presented linear models. This suggests that our linear models may have overestimated the effects of D and W, which already appeared weak.

## 7 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 various strategies — favoring speed or accuracy. In this section we assess whether we obtain different results using the pure strategy protocol.

### 7.1 The GO dataset

We investigate the effect of participant strategy on movement time and $ID_e$, using the Guiard-Olafsdottir (GO) dataset [37]. 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. The balanced condition represents a somewhat natural strategy *i.e.*, how the participant would operate had no instructions been given. For comparison, the instruction in a 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 [29]. 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.

### 7.2 Correlation between $ID_e$ and $\overline{MT}$ and conformity with Fitts' law

A summary of the GO dataset, shown in Figure 7 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. The correlation between $\overline{MT}$ computed per block and $ID_e$ is $R^2 = 0.58$, but reaches up to $R^2 = 0.97$ when $\overline{MT}$ is computed by first aggregating all data per strategy (see subsection 3.3).

### 7.3 A bivariate Gaussian model for $\overline{MT}$ and $ID_e$ for each strategy

The regression between $ID_e$ and $\overline{MT}$ does not describe how $ID_e$ and $\overline{MT}$ are distributed for different strategies, and does not describe the variability of $\overline{MT}$. Figure 8 presents a different analysis, based on fitting Gaussian bivariate distributions to the observed data. For each of the five strategies of the GO dataset, we fitted

$$\begin{pmatrix} ID_e \\ \overline{MT} \end{pmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mu_i \\ \mu_t \end{bmatrix}, \begin{bmatrix} \sigma_i^2 & r\sigma_i\sigma_t \\ r\sigma_i\sigma_t & \sigma_t^2 \end{bmatrix}, \right). \tag{24}$$

with $\mu_i$ and $\sigma_i^2$ the mean and variance of $ID_e$, $\mu_t$ and $\sigma_t^2$ the mean and variance of $\overline{MT}$, and $r$ the Pearson correlation between $ID_e$ and $\overline{MT}$. We chose the bivariate Gaussian model because it is simple, easy to

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.

interpret, it makes the dependence between $\overline{\text{MT}}$ and $\text{ID}_e$ completely specified via a single parameter (Pearson's $r$), and is expected to fit well given the clusters' regular and symmetric shapes. Furthermore, using the well-known formula for the conditional expectation of a bivariate Gaussian, the average of the $\overline{\text{MT}}$'s reads

$$\mathbb{E}[\overline{\text{MT}}|\text{ID}_e] = \mu_t + r\frac{\sigma_i}{\sigma_t}(\text{ID}_e - \mu_i). \tag{25}$$

Hence, the model has linear conditional mean, and thus according to Theorem 3.1 and considering subsection 3.3, will reach $r^2(\text{ID}_e\overline{\text{MT}})$ close to 1. In other words, the bivariate Gaussian is (locally) compatible with Fitts' law Equation 3 for a given strategy.

The fit of the Gaussian model for each strategy is displayed Figure 8, where the associated 95% prediction ellipses are drawn. The separation between strategies shows effective, yet sometimes inconsistent self-regulation across participants, indicated by considerable overlaps of clusters. On average, $\overline{\text{MT}}$ increases with $\text{ID}_e$ across strategies, but also within, as indicated by the orientation[16] of the 95% prediction ellipses, suggesting a *local* dependency. When the strategy emphasizes precision, the mean of the Gaussian vector increases, as well as the size of the prediction ellipsis; however, except for the *speed emphasis* condition, its orientation stays relatively constant, suggesting that the correlation between the $\text{ID}_e$ and $\overline{\text{MT}}$ may also be constant. The *speed emphasis* condition thus appears to be an exception where $\overline{\text{MT}}$ and $\text{ID}_e$ are uncorrelated: when participants emphasize speed above anything else, accuracy actually becomes independent of local changes in speed, which is in line with purely feedforward movements [12, 23].

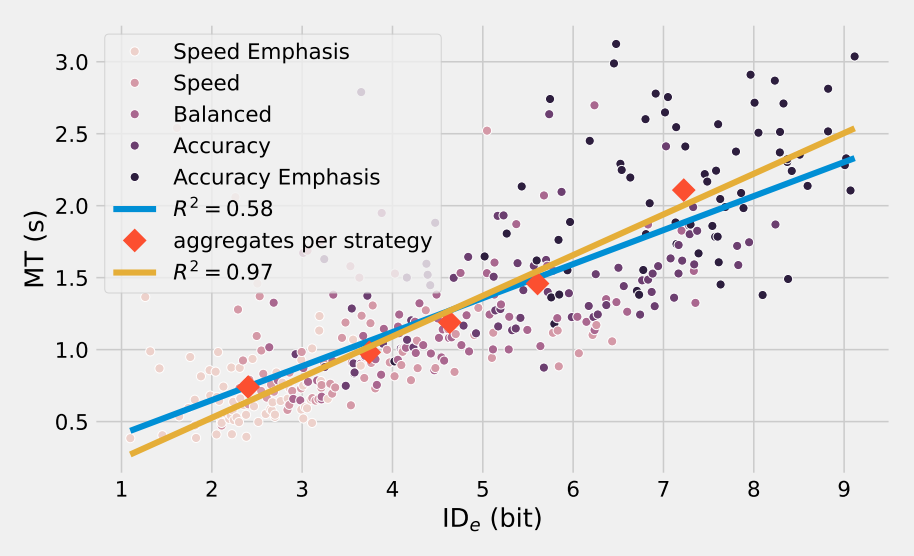


Fig. 7. ¡caption¿

[16]The covariance matrix of a Gaussian bivariate is the rightmost term of Equation 24 where $\rho$ is the correlation between the two components. The angle of the prediction ellipsis is given by $\frac{1}{2}\arctan\left(\frac{2\rho\sigma_i\sigma_t}{\sigma_i^2-\sigma_t^2}\right)$.

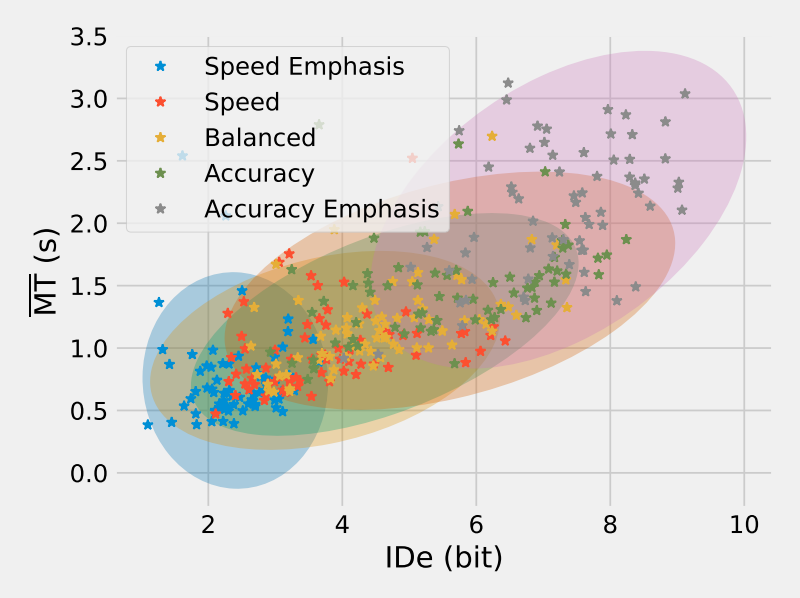


Fig. 8. $\overline{\text{MT}}$ against $\text{ID}_e$ for all participants and strategies for the GO dataset where the color indicates strategy. A bivariate Gaussian fit per strategy is also displayed through the associated 95% prediction ellipsis.

### 7.4 Strategy Dependent Bivariate Gaussian model for $\overline{\text{MT}}$ and $\text{ID}_e$

We can unite the five models into a single one by representing strategy numerically rather than ordinally. We thus map the five instructions to equally spaced points on the interval [-1,1][17], with 0 as the balanced strategy, -1 for speed emphasis, and 1 for precision emphasis. We then assessed how the five parameters of the bivariate Gaussian model Equation 24 aligned with this numerical strategy, by fitting a linear model on each parameter. First, we determined whether to consider the linear model

$$\text{X} \sim \text{strategy (strategy model)} \tag{26}$$

or the simpler constant model ($\text{X} \sim 1$) based on AIC comparisons reported Table 2, where $X$ represents any of the 5 parameters of Equation 24.

Table 2. Model comparison for $\mu_i$, $\mu_t$, $\sigma_i$, $\sigma_t$, $\rho$. 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 24 | strategy model AIC | constant model AIC |
|---|---|---|
| $\mu_i$ | **1** | 21 |
| $\mu_t$ | **-4** | 9 |
| $\sigma_i$ | 1 | **3** |
| $\sigma_t$ | **-17** | -10 |
| $\rho$ | -3 | **1** |

We then fitted the models, displayed Figure 9. We observe that:

[17]The choice here may seem naïve, and we could have mapped strategy to the set of real values via $\text{ID}_e$. However, our choice has two advantages: the mapping can be defined before movements are performed, and does not depend on each participant.

- The linear increase of the mean ($\mu_i$, $\mu_t$) with numerical strategy is visually clear, and supported by high goodness of fit ($r^2$) and large AIC differences.
- The increase of $\sigma_i$ and $\sigma_t$ with numerical strategy could also reasonably be described by a linear model, although evidence is less strong than for the mean. For $\sigma_i$, AIC comparison suggested the simpler constant model. Note that the value of $\sigma_i$ for the *speed emphasis* condition differs strongly from the others; the mean $\sigma_i$ without the *speed emphasis* condition is $\sigma_i = 1.19$. For $\sigma_t$, the AIC difference between the linear and constant model was slightly below 10 (7); we decided to consider the linear model to have an estimate of the effect size of strategy.
- For $\rho$, model comparison suggested the constant model. Still, one sees that the speed emphasis condition is quite different from the others. As alternate model we considered $r = 0$ for the *speed emphasis* ($s = -1$) condition, and a constant correlation for the other conditions; we computed the mean $r$ for the 4 other conditions and obtained $r = 0.44$.

The linear relationship of $\mu_i$ and $\mu_t$ on $s$ makes the obtained model consistent with Fitts' law (globally) across strategies since

$$\mu_i = \alpha' + \beta' s \longrightarrow s = \frac{\mu_i - \alpha'}{\beta'} \tag{27}$$

$$\mu_t = \alpha + \beta s = \alpha + \frac{\beta}{\beta'}(\mu_i - \alpha') = \alpha - \frac{\beta}{\beta'}\alpha' + \frac{\beta}{\beta'}\mu_i, \tag{28}$$

which conforms with our finding in Figure 7. Thus, the Gaussian bivariate model is consistent with the local Fitts' law (within a strategy) and the linear model for $\mu_i$ and $\mu_t$ is consistent with the global Fitts' law (across strategies). Note that the slopes and intercepts of each local Fitts' law are different, and different from the slopes and intercepts of the global Fitts' law, as on can guess in Figure 8, and since the parameters of Equation 25 depend on $s$.

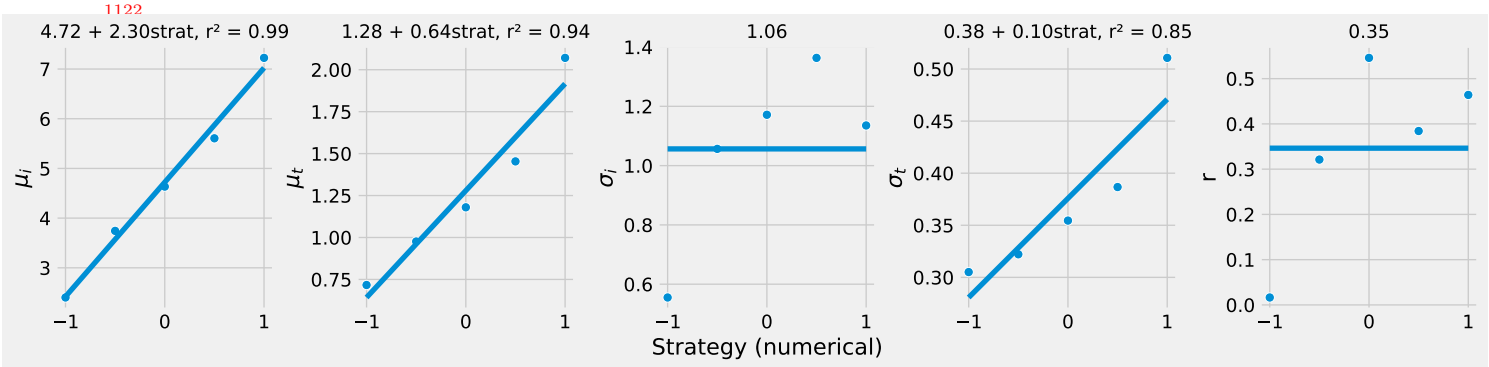


Fig. 9. The parameters of the mean and covariance of the bivariate Gaussian model defined in Equation 24 plotted against the five numerical strategies. The associated linear fits and $r^2$ are displayed above each plot when applicable.

### 7.5 EMG fitting

To our knowledge the EMG model was never fitted to data from the pure strategy protocol; however, we expect similar results to Fitts' protocol: the properties that make the EMG model relevant — the long

tails of the movement time distribution and the increasing variance as a function of difficulty — are likely properties of movement unrelated to any one experimental protocol.

To confirm this, we performed two model comparisons:

- First we compared a classical linear regression with a Gaussian model with quadratic variance. Linear regression is equivalent to fitting a Gaussian model with constant variance, so this first comparison determines whether the quadratic variance model fits better than the constant variance model.
- Second, we compared the Gaussian model with quadratic variance to the EMG model with quadratic variance. The point of that comparison was to evaluate the impact of the shape of the distribution — symmetric (Gaussian) versus asymmetric with long tails (EMG).

These models were compared for each participant. A violin plot of the AIC differences is reported Figure 10, which shows that both the EMG and the quadratic variance model are strongly supported by the data. A pairplot of the EMG parameters is available in the supplementary materials; with one exception, the same correlations as with the JGP dataset emerge.

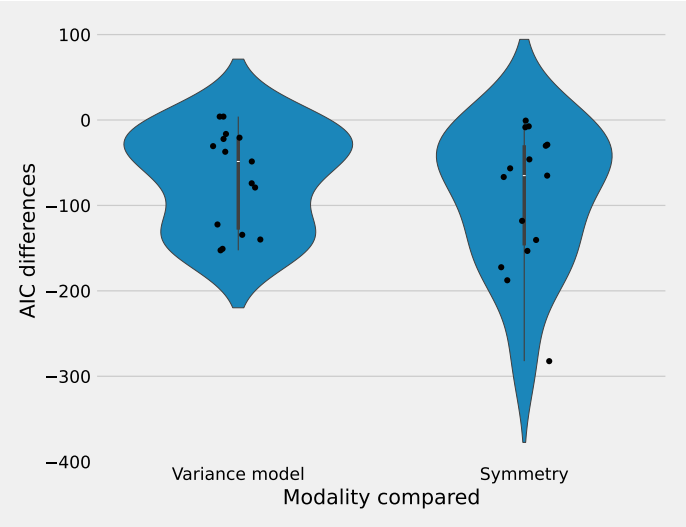


Fig. 10. Violin plot of AIC differences when comparing the variance and symmetry models. The actual AIC differences are superimposed in black dots. Left (variance model): We computed the AIC difference of a Gaussian model with quadratic variance with a Gaussian model with constant variance *i.e.*, classical linear regression. A negative AIC difference means more support for the quadratic variance model. Right (symmetry): We computed the AIC difference of an EMG model with quadratic variance with a Gaussian model with quadratic variance. A negative AIC difference means more support for the EMG model.

### 7.6 Fitting copulas

Similarly to subsection 6.3, we fit copulas on the GO dataset, per strategy and per participant with model evidence ratio $\mathcal{R}$ to compare copulas ($5 \times 15 = 75$ fits) on their ability to model the *local* dependence. The $\mathcal{R}$ are displayed Figure 11; the Gaussian copula provides the best fit for all strategies except for the balanced one, where the Galambos copula is the best fit. This finding may seem surprising based on Figure 27 since

the Galambos copula has strong upper tail dependence which suggests a decrease of variance for high $ID_e$ values. However, the Galambos copula only has upper tail dependence for $\theta \in [1, +\infty]$. In the GO dataset, it averages $\theta = 0.5$ ; as is apparent from Figure 27, the Galambos copula with $\theta = .5$ is close to the independent copula. A violinplot of the Galambos copula's parameters can be found in the Supplementary materials.

For completeness, we looked at the Gaussian copula's parameter see Table 3, and found small positive correlations, which are, as expected, very close to the Pearson correlations. Only strategies 3 (balanced) and 5 (accuracy emphasis) were found statistically significanty different from 0 ($p < 0.01$) with a one-sample t-test, with strategy 2 being highly significant and showing the strongest correlation.

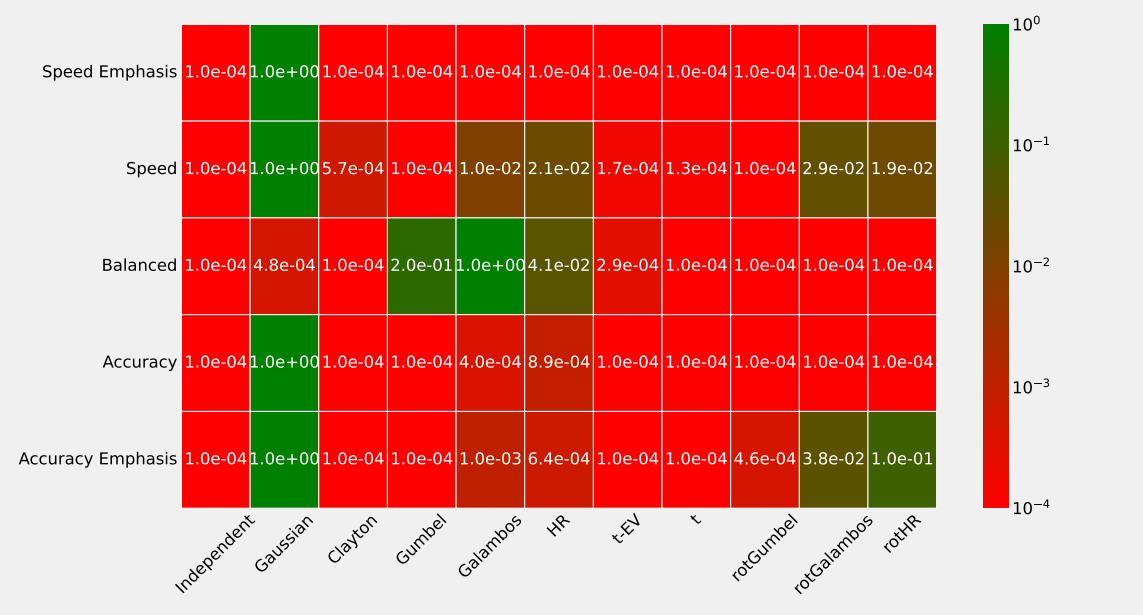


Fig. 11. Model evidence ratio $\mathcal{R}$ of the tested copulas against the best copula, for different strategies.

Table 3. Mean $\rho$ 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 |

We also performed an analysis per participant, with strategy pooled to investigate the *global* dependence between $ID_e$ and MT. The results Figure 12 show the combination Gaussian / t-copula provided the best fit for 10/15 participants. On average, we found a value of the correlation parameter of the Gaussian / t-copula of $\rho = .71$, and a large disparity in $\nu$ values (see Supplementary information for the graphs). For 4 our of the other 5 participants, the best fitting copulas was the Galambos copula (average $\theta = 0.6$)

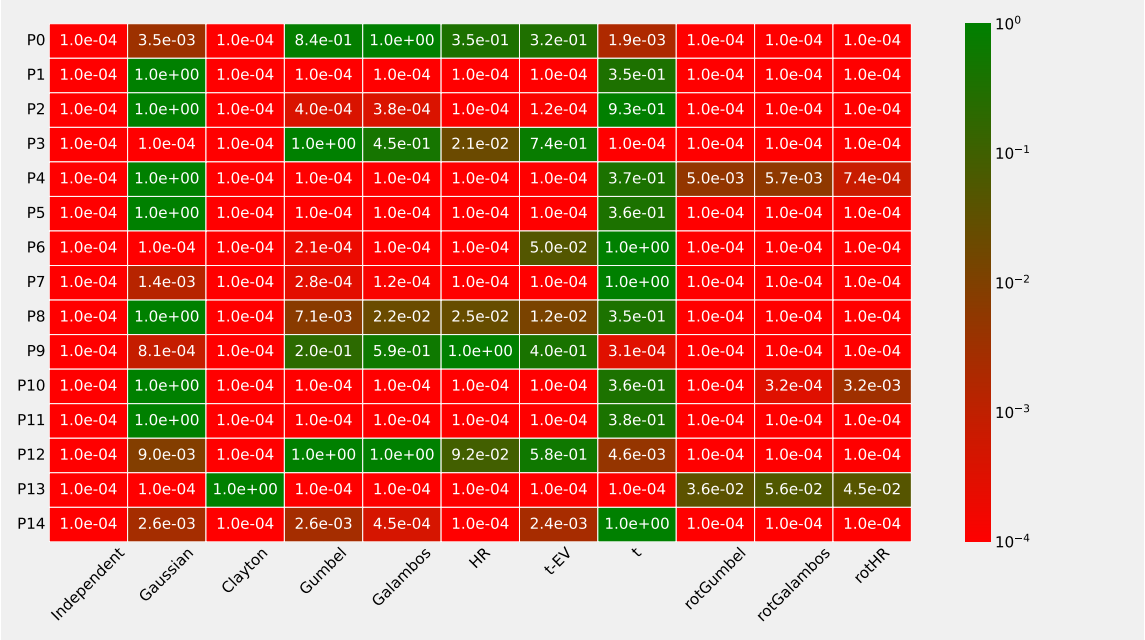


Fig. 12. Model evidence ratio $\mathcal{R}$ of the tested copulas against the best copula, for different participants.

### 7.7 Interpretation of the second study

This second study examined the dependency between $\mathrm{ID}_e$ and MT in pointing data collected using the pure strategy protocol, where accuracy is specified a priori by instructions. As in the first study, this protocol leads to strong *global* dependence between $\mathrm{ID}_e$ and MT when pooling conditions (in study 1 (D,W) pairs, in study 2 strategy), but less so within conditions, indicating weak *local* dependence.

We first verified that data from the pure strategy protocol was compatible with Fitts' law, and indeed found that correlation computed on aggregate values lead to very strong Pearson correlation ($\rho = .97$) between $\mathrm{ID}_e$ and $\overline{\mathrm{MT}}$. We then suggested a bivariate Gaussian model between $\mathrm{ID}_e$ and $\overline{\mathrm{MT}}$, parametrized by strategy. The model is consistent with Fitts' law and its parameters are simple to estimate and easy to interpret. We found that $\overline{\mathrm{MT}}$ and $\mathrm{ID}_e$ were positively correlated within strategies — which suggests a local "average" speed-accuracy trade-off — except for speed emphasis; We believe this to be in line with the "ballistic" nature of feed-forward movements [12, 23] in the speed emphasis condition.

We also confirmed the appropriateness of the EMG model for MT distributions; we did this by showing that both the quadratic variance model and the asymmetric, long tail model provided better fitting power than the constant variance and symmetric models. Note that we did not exclude that other asymmetric models may better fit the current data; there are far more candidate combinations of variance models and distributions than can be tested. Instead, like in the previous section, we assessed the MT and $\mathrm{ID}_e$ dependence more systematically with copulas.

For the local dependence (conditional on strategy), we found the Gaussian copula with positive correlation — although technically not all conditions has correlations different from 0 at $\alpha = 0.05$ level — to be the

most appropriate model, except for the balanced condition, where the Galambos was the best fit. What is the difference between these two copulas? The Gaussian copula implies a symmetric form of dependence between $ID_e$ and MT, with no specific upper or lower tail dependence *i.e.*, for the largest and smallest values, the co-occurence of $ID_e$ and MT is the same as if they were independently distributed. For the balanced condition, the Galambos ($\theta = 0.5$) provided the best fit, which is close to the independent copula, see Figure 27. Overall, this paints a picture of low correlated $ID_e$ and MT.

When pooling data from all strategies together to assess the global dependency between $ID_e$ and MT, the Gaussian and t-copula were also the best models for 10/15 participants, while the Galambos copula with $\theta < 1$ was the best model for 4/5 participants. We consider the pair Gaussian / t-copula because the t-copula with parameter $\rho$, $\nu$ is equivalent to the Gaussian copula with parameter $\rho$ for large $\nu$[18]. The Gaussian copula parameter averaged $\rho = .71$, indicating strong correlation between $ID_e$ and MT when data from different strategies are pooled.

## 8 THIRD STUDY: DEPENDENCE BETWEEN $ID_e$ AND MT IN THE FULL PROTOCOL

### 8.1 The YKORM dataset

The role of this study is to further support but also combine the results of the two previous studies targeting the Fitts and the pure strategy protocols. We considered the YKORM dataset [68] obtained using the *Fitts-with-strategy* protocol (see subsection A.4). 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). In this re-analysis, we only use regular mouse-based pointing data, for comparison with the two previous studies. Also, since our goal is not to account for distractors, we aggregated data across distractor density levels. With twelve participants recruited, and 23 trials performed per condition originally, this makes 12 participants $\times$ 3 W $\times$ 2 D $\times$ 3 strategy $\times$ 69 movements = 216 sets $\times$ 69 movements = 14904 trials. The first 3 trial of each set were discarded (648), and some outliers (244), leaving us with 14012 data points.

### 8.2 Conformity with Fitts' law

We first fit Fitts' law (effective and nominal form) with strategy and interaction terms using a mixed effect linear model with a random intercept for participants

$$\mathrm{MT} \sim \mathrm{ID} * \mathrm{strategy} + (1|\mathrm{Participant}), \text{ and} \tag{29}$$

$$\mathrm{MT} \sim \mathrm{ID}_e * \mathrm{strategy} + (1|\mathrm{Participant}). \tag{30}$$

We plotted the full dataset, as well as the mixed effect model adjusted on the block means Figure 13; the fits are summarized Table 4. The nominal version of Fitts' law Equation 1 is strongly preferred over the effective version Equation 3 according to likelihood scores (105 for ID versus 86 for $ID_e$, same number of parameters).

[18] Just like the t-distribution is equivalent to the Gaussian distribution for large degrees of freedom. In theory, the t-copula should always fit better than the Gaussian copula since it has an extra degree of freedom. However, the fit of the t-copula can appear to be less good for two reasons. First, the extra parameter of the t-copula is penalized by AIC and $\mathcal{R}$, second, the extra parameter makes the copula harder to identify since the optimization of the likelihood has more chance of getting stuck in a local minimum when the dimension of the search space increases.

For both models, the change in slope due to strategy is about 0.08, meaning one unit of strategy augments the slope by 0.08 s/bit. Since one unit of strategy separates the speed and accuracy conditions, this amounts to a change of 40 ms/bit for each condition. This is consistent with the study from Zhai *et al.* [71] who found differences of about 30ms/bit in a similar study.

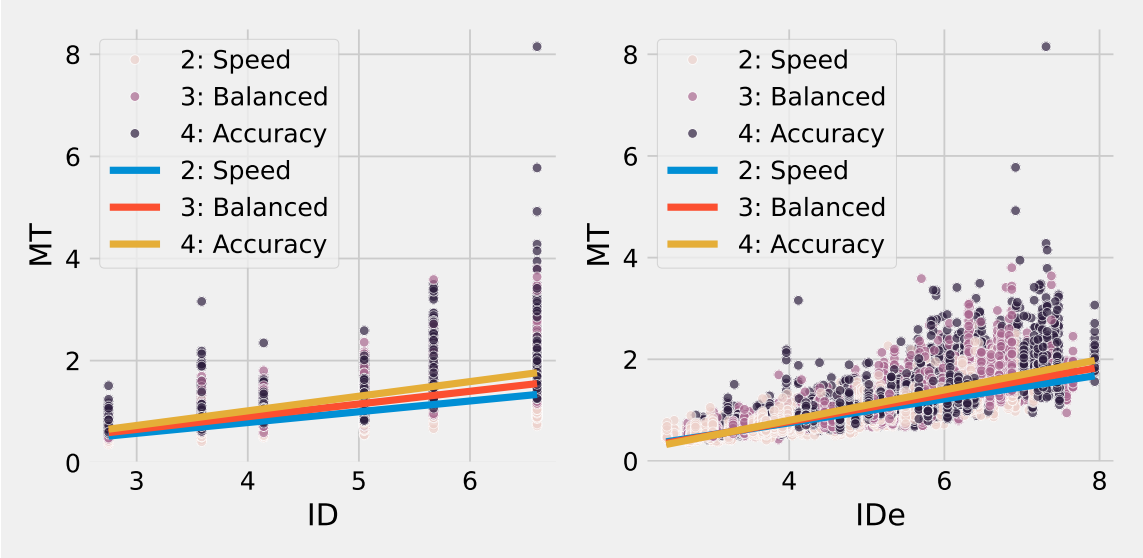


Fig. 13. Data of the YKORM dataset and adjusted Fitts' law. The data is the MT value associated with ID (left) or $ID_e$ (right). Fitts' law was adjusted by computing linear regression on $\overline{MT}$.

Table 4. Fitts' law fitted on the YKORM dataset. Fitts' model was fitted on $\overline{MT}$ and not MT.

| | ID model. | $ID_e$ model |
|---|---|---|
| Intercept | -0.090 **(p = 0.029)** | -0.24 **(p <0.001)** |
| ID | 0.247 **(p <0.001)** | |
| $ID_e$ | | 0.266 **(p <0.001)** |
| strategy_num | -0.091 (p=0.204) | -0.244 **(p = 0.003)** |
| ID:strategy_num | 0.078 **(p <0.001)** | |
| $ID_e$:strategy_num | | 0.073 **(p <0.001)** |
| log-likelihood | 105 | 85.8 |

### 8.3 Model between $ID_e$ and $\overline{MT}$

We identified the bivariate Gaussian model Equation 24 for each combination of strategy (3) and ID level (6). The 18 resulting 95% ellipses are displayed Figure 14. Compared with the results from section 7, the spread of the Gaussians is smaller; this is not surprising considering the Fitts with strategy protocol enforces a width; the strategy only modulates how close one follows that width. As in the previous section, we then

analyzed how the parameters of the bivariate Gaussian evolve with strategy, also including the ID factor. We first compared models of decreasing complexity

$$\mathrm{X} \sim \mathrm{ID} + \mathrm{strategy}\ \text{(full model)} \tag{31}$$

$$\mathrm{X} \sim \mathrm{ID}\ \text{(ID model)} \tag{32}$$

$$\mathrm{X} \sim \mathrm{strategy}\ \text{(strategy model)} \tag{33}$$

$$\mathrm{X} \sim 1\ \text{(constant model)} \tag{34}$$

where X is one of $\mu_i$, $\mu_t$, $\sigma_i$, $\sigma_t$, $\rho$ defined Equation 24. The model comparisons based on AIC are given in Table 5; for $\mu_i$ and $\mu_t$ the full model was preferred, with large AIC differences (>10). For $\sigma_i$ and $\rho$, the AIC scores were all within 2 points, and we therefore selected the simplest model *i.e.*, the constant model. For $\sigma_t$, we found that the ID model had similar AIC to the full model, so we selected the simplest ID model.

The $\overline{\mathrm{MT}}$ data and accompanying fits are visualized Figure 15, with parameter fits and goodness of fit (when applicable) in the title.

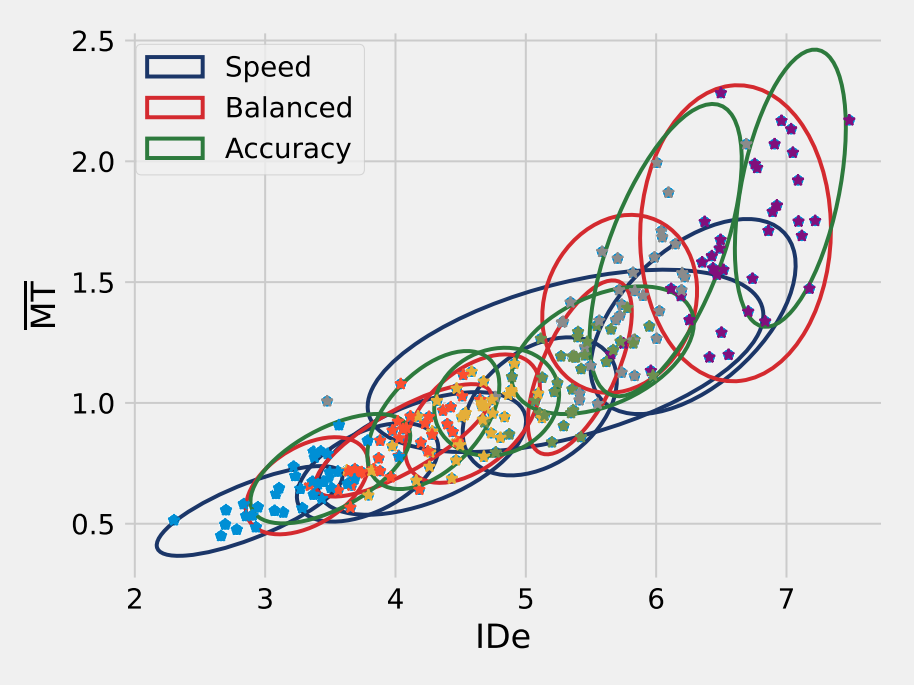


Fig. 14. 95% prediction ellipses of the bivariate Gaussian model fitted to the YKORM dataset. Each model is fit to ($\overline{\mathrm{MT}}$,$\mathrm{ID}_e$) data for a given strategy and ID level. The different colors of the points indicate different ID levels.

### 8.4 Model between $\mathrm{ID}_e$, W, D ID, and strategy

As in subsection 6.4, we evaluated the effect of W, D, ID, and strategy on $\mathrm{ID}_e$ using a mixed effect linear model. We did not test the full model with all interactions since this model would be unneededly complex,

Table 5. Model comparison for $\mu_i$, $\mu_t$, $\sigma_i$, $\sigma_t$, $\rho$. 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 24 parameters | Full model AIC | ID model AIC | strategy model AIC | constant model AIC |
|---|---|---|---|---|
| $\mu_i$ | **-30** | 7 | 59 | 58 |
| $\mu_t$ | **-34** | -14 | 17 | 17 |
| $\sigma_i$ | -40 | -38 | -40 | **-39** |
| $\sigma_t$ | -107 | **-106** | -90 | -91 |
| $\rho$ | -6 | -7.5 | -4 | **-5.7** |

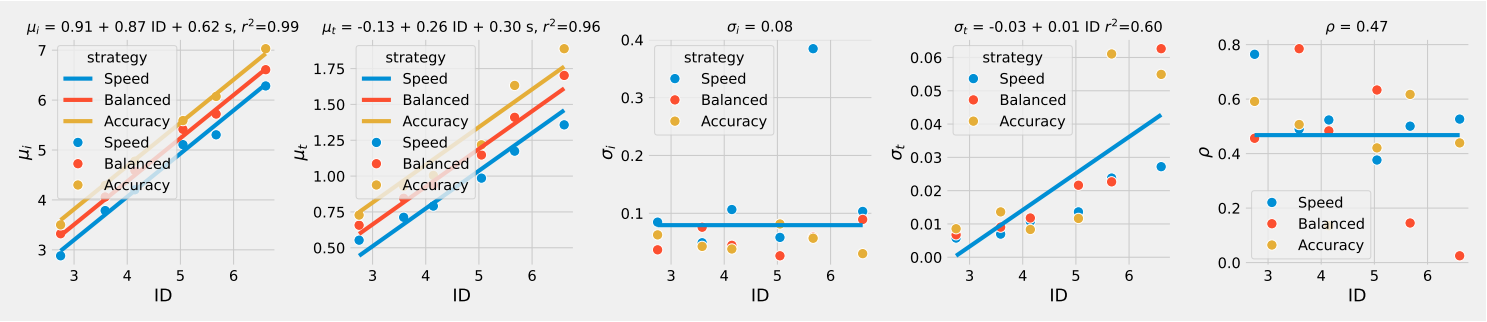


Fig. 15. Model of the bivariate Gaussian model fitted to ID and strategy levels. The model parameters are visible in the title, as well as the goodness of fit when applicable ($r^2$).

but compared the following simplified models

$$\mathrm{ID}_e \sim \text{strategy} + \mathrm{ID} * \mathrm{W} + \mathrm{ID} * \mathrm{D} + (1|\text{Participant}) \tag{35}$$

$$\mathrm{ID}_e \sim \mathrm{W} * \mathrm{ID} + \mathrm{W} * \text{strategy} + (1|\text{Participant}) \tag{36}$$

$$\mathrm{ID}_e \sim \mathrm{D} * \mathrm{ID} + \mathrm{D} * \text{strategy} + (1|\text{Participant}) \tag{37}$$

$$\mathrm{ID}_e \sim \mathrm{ID} + \text{strategy} + (1|\text{Participant}) \tag{38}$$

Instead of considering an extension of the full model Equation 20 with strategy as main effect, we used a simplified form in Equation 35 without the three way interaction and the interaction between W and D. The other models are extensions of the models tested section 6 with main and interaction effects of strategy. Their estimated parameters are reported Table 6. As in subsection 6.4, the more complex model does not lead to a significant effect of main or interaction effects of W or D, but unsurprisingly, leads to strong significant effects of ID and strategy. The simpler model Equation 36 does lead to a significant effect of W, but from a likelihood perspective can not be favored[19].

The simplest model fit is plotted Figure 16. One sees the strong effects of strategy and ID, with a very small interaction effect (slightly different slopes for the three strategies, found to be not significant ($p = 0.101$)).

*Comparison to the first study and synthesis.* In the first study, we were not able to show any significant effect of W on $\mathrm{ID}_e$ other than through ID, contrary to Zhai *et al.*'s finding [71], but we showed a significant

[19] It is only 6 points better, but with 2 extra parameters, this makes a difference of AIC of just 2, which is not enough to favor a more complex model [3]

Table 6. Estimation and comparison of the 4 models for $ID_e$

| | Equation 35 | Equation 36 | Equation 37 | Equation 38 |
|---|---|---|---|---|
| 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 |

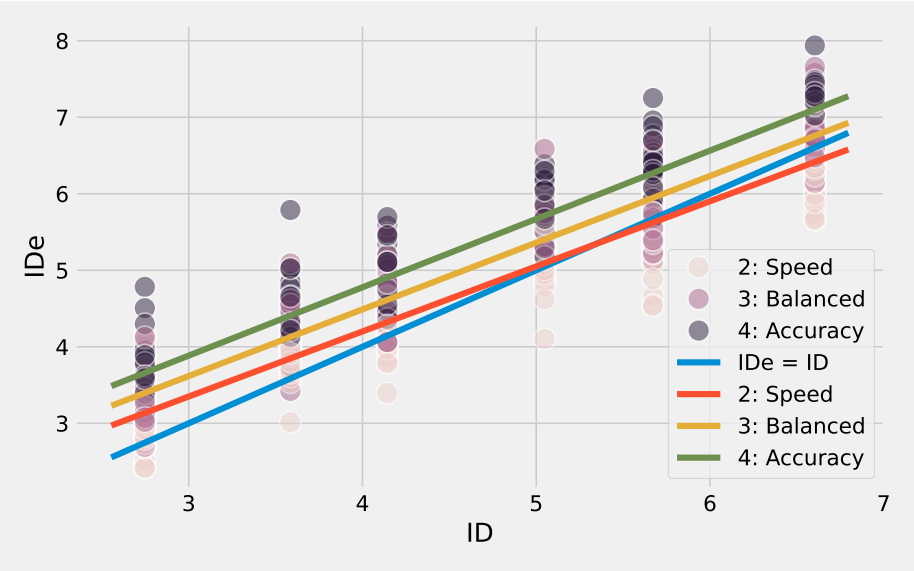


Fig. 16. The simple additive model of Equation 38 displayed on top of the YKORM data. The identity line ($ID_e = ID$) is also plotted.

effect of D. In this study, we could show a significant effect of W on $ID_e$ other than through ID, but model selection based on likelihood comparisons favored the simpler model without D and W effects. Note that in contrast to the first study, this dataset does have a condition with a very small target.

It appears that the effect of W and D on $ID_e$ other than through ID is difficult to show. An experiment specifically designed to measure these effects would certainly perform better, namely by adopting *shape* × *scale* designs [33, 36], which would decorrelate D and W with ID. More levels of D and W with very small and large values should also be included.

In our opinion, the effects of W and D may largely be considered of second order importance. We thus suggest the following simple linear model for $\mathrm{ID}_e$:

$$\mathrm{ID}_e = \alpha + \beta\ \mathrm{ID} + \gamma\ \text{strategy}. \tag{39}$$

Comparing the estimated parameters of this model for the YKORM and the JGP datasets, we find

$$\alpha = 1.003\ \text{(YKORM) vs } 0.08\ \text{(JGP)} \tag{40}$$

$$\beta = 0.872\ \text{(YKORM) vs } 0.937\ \text{(JGP)} \tag{41}$$

$$\gamma = 0.604\ \text{(YKORM)} \tag{42}$$

which suggest $\beta \simeq .9$ and positive intercepts ($\alpha$). It is unclear why we have such a disparity for $\alpha$. Interestingly, these conditions ensure that the $\mathrm{ID}_e$ line will cross the ID line, which explains target under-and-over-utilization: for values of ID before the cross-over point, $\mathrm{ID}_e$ is greater than ID, signaling under-utilization of the target, while for values beyond the cross-over point, $\mathrm{ID}_e$ is less than ID, signaling target over-utilization. This is consistent with Zhai *et al.*'s [71] results. Since usually W correlates (negatively) strongly with ID (high values of ID requires small targets) [33, 36], this simple model is able on its own to represent the fact that small targets are over-utilized and large targets are under-utilized. Strategy, by shifting the intercept of the $\mathrm{ID}_e$ vs ID relationship, moves the cross-over point, which occurs around ID=5 for the speed condition.

### 8.5 EMG

Similar to the previous section, we assessed the validity of the EMG model's quadratic variance and asymmetrical components. EMG fits were computed for each participant and each strategy, pooling over (D,W) conditions. AIC differences were then computed as in subsection 7.5 and represented in a violinplot Figure 17. Similar to the results on the GOP dataset, we found strong support for the quadratic variance and asymmetrical models.

### 8.6 Fitting copulas

We fitted copulas on the YKORM dataset, per strategy, (D,W) condition and participant with model evidence ratio $\mathcal{R}$ to compare copulas.

*Effect of (D,W) condition.* We first investigated the dependence between $\mathrm{ID}_e$ and MT for different (D,W) conditions. Because we only had 3 different replications per (D,W) pair *i.e.*, only three $\mathrm{ID}_e$ levels, we pooled data from the three strategies together, which meant we had $12 \times 2 \times 3 = 72$ datasets to fit copulas on. The resulting $\mathcal{R}$ is given Figure 18, which shows that the Gaussian copula fits the data best, except for the (D=400, W=8) condition, where the rotated Galambos was the best fit. Interestingly, the other condition with small width (W=8) was also well fitted by a rotated Galambos; for the W=8 condition the rotated HR also scores well. The Gaussian copula's average $\rho$ value is $\rho = 0.41$, which, contrary to the two previous studies, signals a moderate *local* dependency between $\mathrm{ID}_e$ and MT; this is likely due to aggregating strategies together. For completeness, we looked if D and W had an effect on $\rho$, but none of these factors turned out significant at the $\alpha = .1$ level. A barplot of the $\rho$ values and the fit details can be found in the supplementary materials.

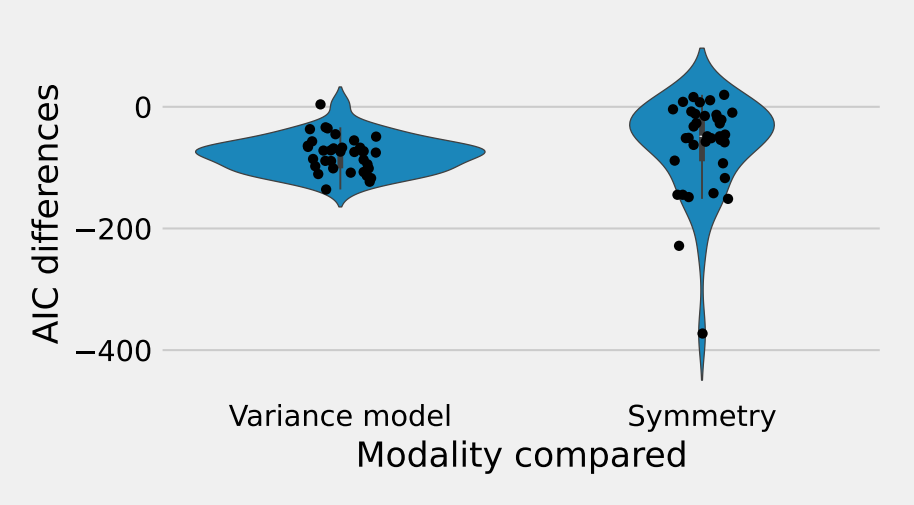


Fig. 17. Violin plots for the AIC differences when comparing variance and symmetry models. Refer to the caption of Figure 10 for more details.

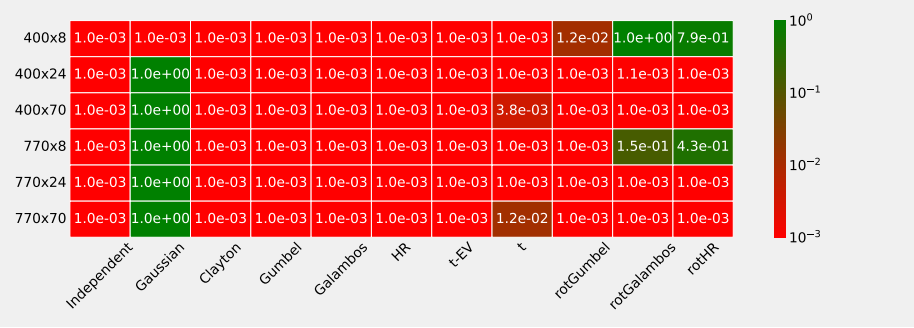


Fig. 18. Model evidence ratio $\mathcal{R}$ of the tested copulas against the best copula for different (D, W) pairs of the YKORM dataset.

*Effect of strategy.* We pooled the data from the different (D,W) conditions together and fit the copulas on the resulting $12 \times 3 = 36$ datasets. The associated $\mathcal{R}$ is given Figure 19. It shows again the Gaussian copula and the t-copulas to be the best fits on most cases. We verified for effects on strategy for the copula parameters; given the $\rho$ is quasi identical for both copulas, we only display parameters for the t-copula Figure 20, which shows no particular effect of strategy on copula parameters.

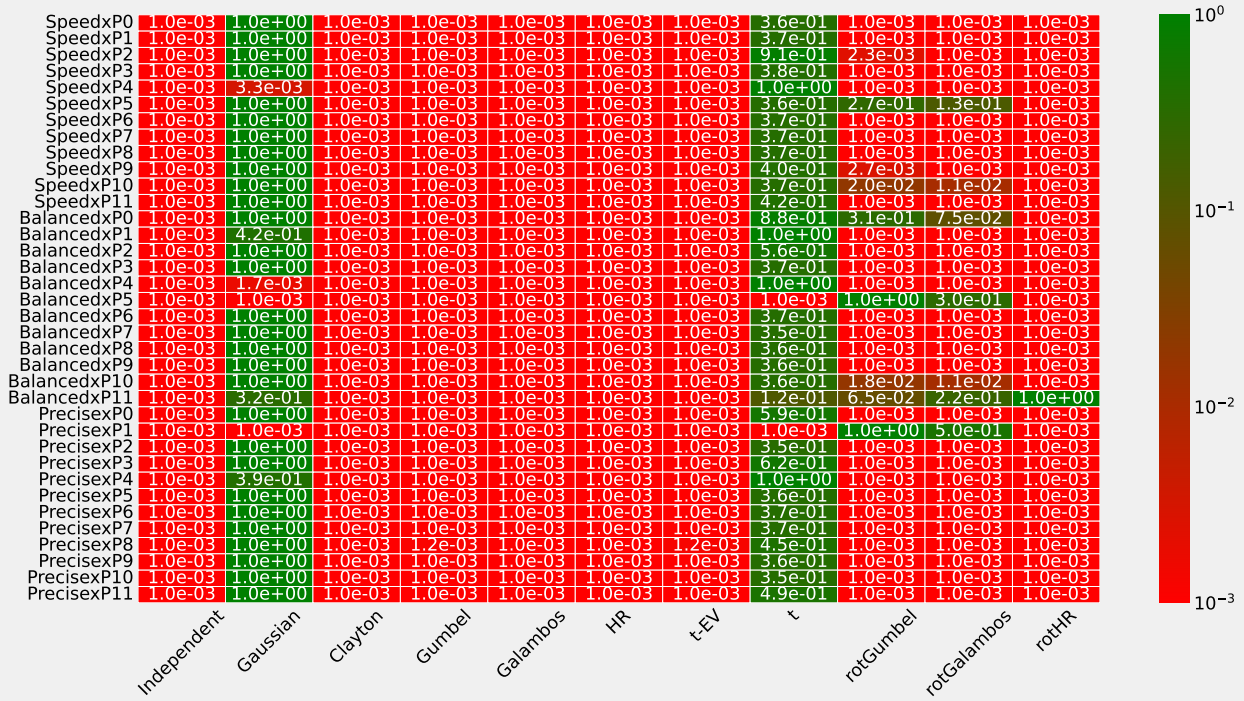


Fig. 19. Model evidence ratio $\mathcal{R}$ of the tested copulas against the best copula, for different strategies and participants of the YKORM dataset.

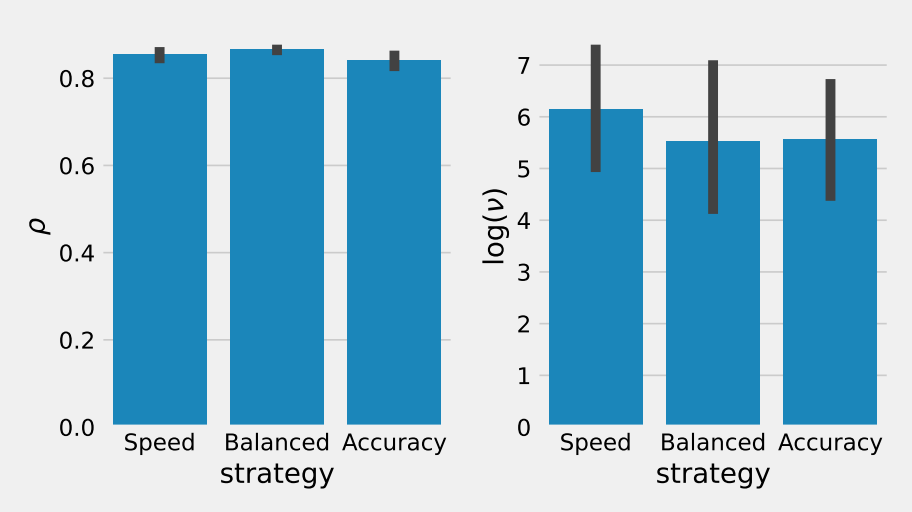


Fig. 20. Parameters of the t-copula for different strategies.

# 9 SYNTHESIS OF THE THREE STUDIES: COPULA FITS AND $ID_e$ MODEL

## 9.1 Copula fits

We used copulas to characterize the dependency between $ID_e$ and MT. This work suggests there is very little local dependency between $ID_e$ and MT *i.e.*, for a given (D,W) pair: $ID_e$ and MT be considered close to independent. When it comes to strategy, the local dependencies are marginally stronger. We interpret this as meaning that within a (D,W) condition, movements are essentially equivalent: due to variability, people may be both more precise and faster, and the speed-accuracy tradeoff does not necessarily apply. This is less true within a given strategy: one hypothesis is that participants do not "aim" for the same strategy within blocks. In particular, what is likely is that they aim relatively to the previous block *i.e.*, if the last block was a speed condition, the participant will aim to be a bit slower than before if the new block is a balanced condition. Hence, data within the same strategy condition may actually be performed with different intent by the participants.

On the other hand, the global dependence of $ID_e$ and MT across conditions was strong, in line with the speed-accuracy tradeoff, and best modeled by the rotated Gumbel copula (mean $\theta \simeq 2$) in the first study and the Gaussian/t-copula pair ($\rho$ around 0.8) in the remaining studies. What could explain this discrepancy? We first note that these three copulas lead to visually similar contour plots when coupled with the marginals observed in the datasets (see Figure 27).

One hypothesis then is that the rotated Gumbel is not easy to identify [27, 66] so that 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 21.
- second, we wanted to evaluate the effect of blocking on identifiability (see [27] 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 21.

Figure 21 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.

Hence, the difference is plausibly the result of differences between the experimental protocols, or may reflect the fact that different input modalities were considered in the studies: study 1 used gestural interaction, study 2 used real physical movements, while study 3 used regular mouse interaction. We return to this point in subsection 9.3

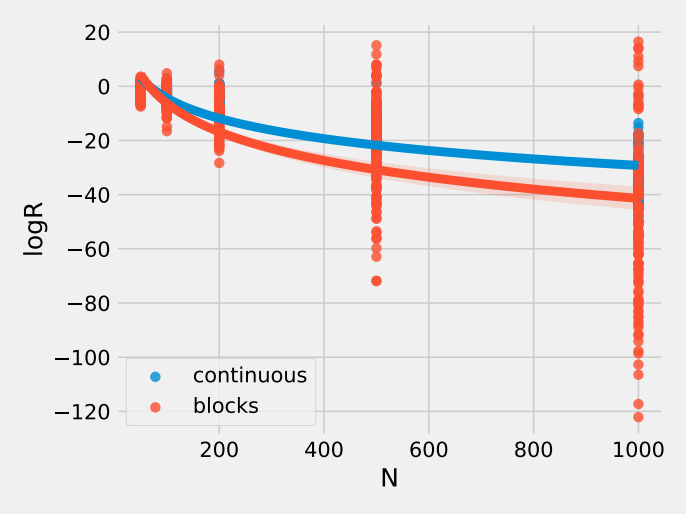


Fig. 21. Identifiavility of the rotated Gumbel copula. Data (size= $N$) was sampled from a rotated Gumbel copula and combined with typical marginals to form a realistic joint distribution for ($ID_e$, MT). A Gaussian and a rotated Gumbel copula where then fit, and the model evidence ratio for the rotated 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).

### 9.2 Model between $ID_e$ and ID

In both study 1 and 3, we were not able to demonstrate a reliable effect of D or W on $ID_e$ other than through their ratio in ID. We argued that a shape × scale design would be needed to better characterize the separate effect of D and W, since crossing D and W leads to them being highly correlated with ID, which in turns makes the estimation of the model's parameters difficult.

It should first be noted that the simple linear model suggested Equation 39 is sufficient to account for target over-and-under-utilization, which in our opinion is the most important phenomenom to capture. As a probabilistic model, we thus first suggest a simple Gaussian noise model (in line with the noise model usually assumed when fitting linear models):

$$\mathrm{ID}_e|\mathrm{ID}, \mathrm{strategy} \sim \mathcal{N}\left(\alpha + \beta\,\mathrm{ID} + \gamma\,\mathrm{strategy}, \sigma^2\right). \tag{43}$$

To give an idea of the variation of $ID_e$, we estimated $\sigma = 0.3$ for the YKORM dataset and $\sigma = 0.15$ for the JGP dataset. We return to this point in subsection 9.3.

### 9.3 Revisiting results with the PDD dataset

We revisit some of the previous results under the light of another dataset, called the Pointing Dynamics Dataset (PDD) by Müller *et al.* [51]. This dataset is interesting since:

- it has a shape × scale design, with D (212, 353) mm and ID (2,4,6,8) fully crossed
- the input device was a mouse with constant CD gain (4.36), and the task was a reciprocal 1D Fitts task

Table 7. Mixed effect model fit results for Equation 20, Equation 21 and Equation 22. 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 (Equation 20) | W model (Equation 21) | D model (Equation 22) | only ID (Equation 23) |
|---|---|---|---|---|
| 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 |

- there are more trials per condition than in the other datasets (12 participants × 8 blocks × 102 trials)
- it has conditions with very small targets (less than 1 mm)

*9.3.1 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 end 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 are left with 3160 movements across 87 different conditions.

*9.3.2 Model between $ID_e$ and D, W, ID.* We fitted the models Equation 20 — Equation 23 on the PDD dataset. As in study 1 and 3, we find that the conditions for target over and under utilization are fulfilled ($\alpha$ positive, $\beta$ slightly below 1 in Equation 39), and this for any variation of the model. The shape × scale of the PDD dataset effectively decorrelates D from ID, which, as shown by the model with only ID and D, provides strong support for a weak effect of D. This design however still strongly correlates W and ID, which makes the estimation of the parameters unreliable when W and ID both appear. (full model, W model).

*9.3.3 Fitting copulas.* We performed the same copula fits as in the previous studies; the $\mathcal{R}$ values for the copulas for the 8 different (local dependency) W conditions are presented Figure 22; the $\mathcal{R}$ values for each participant with W conditions aggregated (global dependency) are presented Figure 23. The results are mostly in line with results from study 1 and 3, namely:

- as in the other studies, a clear copula does not stand out as the best one for the local dependency. The Galambos copula is the best copula for 4/8 conditions, and remains a plausible model for 2/8 conditions, and thus appears to be the most plausible model overall. The value of the Galambos parameter $\theta$ is plotted against W values in Figure 24; interestingly, the value of $\theta$ increases steadily with W, from values close to 0.4 for the smallest target sizes, which correspond to very low dependence, almost identical to the independent copula, to values close to 1.2 for the largest target sizes.
- The global dependency is strong: The most plausible models alternate between the Gaussian/t-copula like in studies 2 and 3, and the rotated Gumbel like in study 1. The average $\rho$ parameter for the Gaussian / t-copula is $\rho = .87$, while the average $\theta$ parameter for the rotated Gumbel is $\theta = 3.1$,

which is a little stronger than what we found for the JGP dataset; both are indicative of very strong dependence.

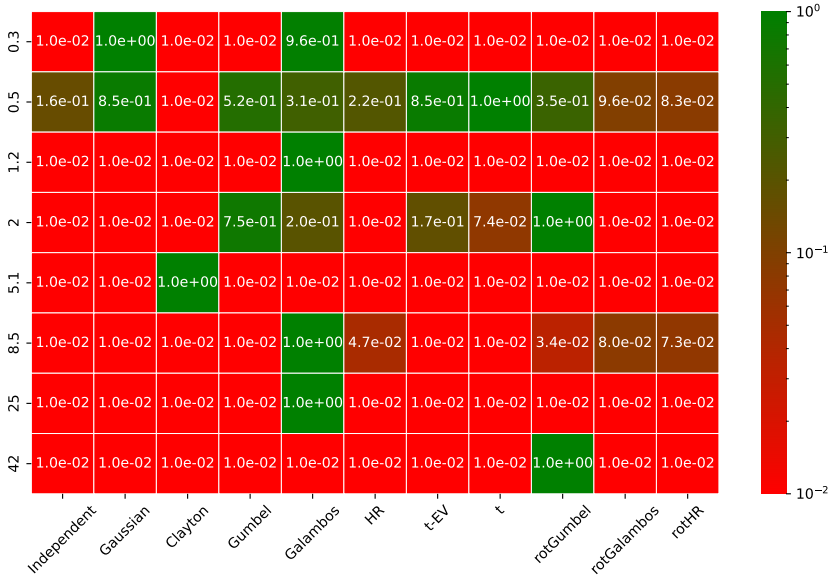


Fig. 22. Model evidence ratio $\mathcal{R}$ for the PDD dataset, for 8 different conditions (labeled by W value).


Fig. 23. Model evidence ratio $\mathcal{R}$ for the PDD dataset, computed for each participant.

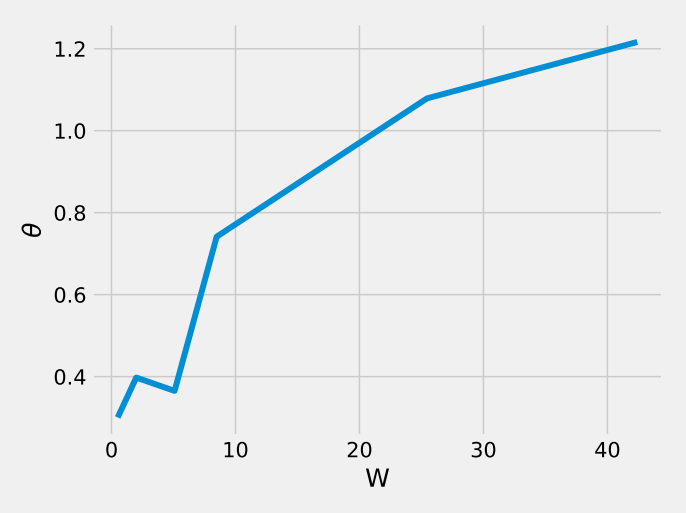


Fig. 24. The Galambos copula's parameter $\theta$ plotted against $W$.

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

We can exploit the previous models to generate pointing data for various protocols using a parametric model. This includes a strategy-dependent parametric model for pure strategy or Fitts-with-strategy protocols

which we motivated at the start of this paper. We suggest three methods for the pure strategy protocol; these methods generalize easily to the other protocols and will not be presented. We illustrate these methods by using data estimated on the GO dataset.

Note that as with any generative model, there is a trade-off between simplicity of the generating process and accuracy. We choose to walk towards the "simple" side of the trade-off because usually this leads to more generalizable and robust results, but more precise models could be achieved.

Everywhere, $\Theta$ refers to a vector of parameters which includes but is not limited to strategy $s$, $p_\Theta(\mathrm{ID}_e, \mathrm{MT})$ refers to the joint distribution between $\mathrm{ID}_e$ and MT parametrized by $\Theta$, and $p_\Theta(\mathrm{MT}|\mathrm{ID}_e)$ refers to the conditional distribution of MT given $\mathrm{ID}_e$.

### 10.1 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 section 4, the copula $C_\Theta$ links the marginals $F_\Theta(\mathrm{ID}_e)$ and $G_\Theta(\mathrm{MT})$ into the joint $p_\Theta(\mathrm{ID}_e, \mathrm{MT})$

$$p_\Theta(\mathrm{ID}_e, \mathrm{MT}) = C_\Theta(F_\Theta(\mathrm{ID}_e), G_\Theta(\mathrm{MT})). \tag{44}$$

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 $\mathrm{ID}_e$ 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 24, which reads

$$\mathrm{ID}_e \sim \mathcal{N}(4.72 + 2.3\,s, (1.06 + 0.29\,s)^2). \tag{45}$$

*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. Without access to routines to estimate copula parameters, $\rho_1$ can be deduced from Kendall's $\tau$ using $\rho_1 = \sin(\tau\frac{\pi}{2})$, and $\nu$ can be deduced from the upper and lower tail dependence coefficients using a more complex formula [13, Section 3]. The simpler Gaussian copula model would also be appropriate.
(2) The parameters of the marginal distributions for MT and $\mathrm{ID}_e$. These can be simply 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 one with linear conditional mean Equation 14.
(3) The levels and number of repetitions may also be given to generate repeated measures.

*Pro's and con's.* As explained section 4, using copulas allows specifying directly the level of association between variables independently of the marginals. Hence, one can couple arbitrary marginals for $\mathrm{ID}_e$ (*e.g.*, to fit a particular task) and MT (*e.g.*, using another asymmetric distribution for MT) with the same t-copula once the latter has been fitted. This first model based on copulas thus provides a principled way of identifying

dependencies between $\mathrm{ID}_e$ and MT, and enough flexibility in exploiting that dependency for generating data, including repeated measurements.

However, two negative points have to be highlighted. First, copulas have a good track record in modeling tail dependencies, but for pointing data the lower tail of the MT distribution should associate strongly with the entire $\mathrm{ID}_e$ distribution [31] —not only its tail—. Second, the tools to work with copulas remain somewhat underdeveloped, which may add barriers for some researchers. As an example, in writing this paper, we have used an R library for copulas, whereas the rest of the work was performed using Python.

### 10.2 Model 2: Conditional distribution using EMG.

*The model.* The second method directly leverages the EMG model of subsubsection 2.3.2. For each point, a value of $\mathrm{ID}_e$ 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

$$p_\Theta(\mathrm{MT}|\mathrm{ID}_e) = \mathrm{EMG}(\mathrm{MT}; \mathrm{ID}_e, \Theta). \tag{46}$$

*Accounting for different strategies.* Similar to the previous model, we can directly draw $\mathrm{ID}_e$ values from the Gaussian Equation 45.

*Inputs of the model.* To operationalize this model, one needs the parameters of the EMG distribution described Equation 9 – Equation 14. We estimated these values for the GO dataset to be $a = 0.08$, $b = .14$, $\sigma = 0.18$, $\lambda_0 = .17$, $\lambda_1 = .08$. The levels and number of repetitions may also be given to generate repeated measures.

*Pro's and con's.* This method will produce data that resembles empirical data by construction, since it relies directly on the EMG distribution. One of the main weaknesses of this method is that it does not allow controlling the association between $\mathrm{ID}_e$ and MT independently of the EMG parameters, as shown in Appendix B: the Pearson correlation $r$ is entirely determined by the EMG parameters. Additionally, as explained subsection 3.2 it will produce values of $r(\overline{\mathrm{MT}}, \mathrm{ID}_e)$ much closer to 1 than what we found Figure 8 and Figure 9.

### 10.3 Model 3: Joint distribution of ($\mathrm{ID}_e$, $\overline{\mathrm{MT}}$).

*The model.* This method exploits the Gaussian bivariate model Equation 24. A pair ($\mathrm{ID}_e$, $\overline{\mathrm{MT}}$) is first sampled from the joint. Then, because the EMG distribution parameters are related to $\overline{\mathrm{MT}}$ by

$$\overline{\mathrm{MT}} = \beta\mathbf{x} + \lambda\mathbf{x} = a + b\mathrm{ID}_e + \lambda_0 + \lambda_1\mathrm{ID}_e, \tag{47}$$

we "correct" $\lambda_1$, ensuring it remains positive to keep an increasing conditional mean and variance.

$$\lambda_1 = \max(0, \frac{\overline{\mathrm{MT}} - a - b\,\mathrm{ID}_e}{\mathrm{ID}_e}). \tag{48}$$

We don't correct $\beta$ since this would alter Fitts' min law parameters, which we believe is an important feature of the pointing dataset, and don't alter $\lambda_0$ to ensure a sufficiently large variance. Movement times are then sampled from the EMG with corrected parameters. Notice in Equation 48 that we force $\lambda_1$ positive to keep an increasing variance, which comes at the cost of occasional departures from the correction.

*Accounting for different strategies.* Here, we draw samples directly from the bivariate Gaussian Equation 24 with

$$\mu_i = 4.72 + 2.3\,s \tag{49}$$

$$\mu_t = 1.3 + .64\,s \tag{50}$$

$$\sigma_i = 1.06 + 0.29\,s \tag{51}$$

$$\sigma_t = .39 + .08\,s \tag{52}$$

$$r = \begin{cases} 0 \text{ if } s = -1 \\ .44 \text{ else} \end{cases} \tag{53}$$

*Inputs of the model.* Same as for Model 2.

*Pro's and con's.* Similarly to the previous method, this method will produce data according to the EMG distribution, making it realistic. Contrary to method 2, we do control the association between $\overline{\text{MT}}$ and $\text{ID}_e$ in this third method. However, the needed correction is sometimes not fully applied (when $\lambda_1 = 0$ in Equation 48, part of the $\overline{\text{MT}}$ has not been absorbed into $\lambda_1$); this method also does not really control how the movement time distribution is modified beyond the ability to fit $\overline{\text{MT}}$.

## 11 MODEL COMPARISON

### 11.1 Consistency

We first start by investigating the consistency of the different methods *i.e.*, how close the data produced by the generative models are to the reference dataset. We consider a dataset with large sample size so that our measures don't depend much on random sampling and are interpretable from a single simulation[20]. The generated datasets are compared based on the following metrics:

(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) Fitts' law measures: parameters of the "mean" and the "min" law (see subsection 2.4),
(3) The "classical" criterion $r(\overline{\text{MT}}, \text{ID}_e)$,
(4) computation of ISO-throughput.

We use ISO-throughput [61], also known as the "mean-of-means" throughput, since it is used in practice by some HCI researchers as a one dimensional score to evaluate overall pointing performance. Hence, it gives a concise way of measuring differences between the GO dataset and the generated ones, which are expected to be interpretable and meaningful to HCI researchers. It should be noted that, for maximum consistency, we consider the same $\text{ID}_e$ values as those from the original dataset for the generated data.[21]

The results are summarized Figure 25, with all 4 metrics displayed in each panel's title and legend. Overall, the differences between the models are minimal, as all models produce data that closely match the metrics of the original dataset. When considering measures of association, all models align well with the original

[20]The data were generated with a fixed seed (1234). Replications with two other seeds (777, 999) and no seeds 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

[21]In other words there is no $\text{ID}_e$ variability, only MT variability in these generated datasets. All 3 models can accommodate this: for Model 1 by sampling from the conditional copula, for Model 3 by sampling from the conditional Gaussian bivariate, and for Model 2 trivially.

dataset. All models lead to slightly underestimated ISO throughput, with a maximum difference of 0.25 bit/s for Model 1. In terms of Fitts' law, Model 1 maintains the mean law (OLS) but distorts the minimum law (EMG), resulting in a higher intercept and a smaller slope. Model 2, on the other hand, preserves the minimum law but slightly skews the mean law, yielding a higher intercept and lower slope. Model 3 comes closest to replicating the original dataset, with both laws nearly aligned; only the intercept in the EMG fit is slightly underestimated. All models lead to inflated values of $\overline{r} = r(\overline{\mathrm{MT}}, \mathrm{ID}_e)$. For Model 2, this is expected due to its construction based on a linear conditional expectation. For Model 1, the high values of $\overline{r}$ are more surprising, but likely due to the symmetric nature of the t-copula. Model 3 is the only one that reduces $\overline{r}$, although it still does not reach the target $\overline{r}$, presumably due to the enforced positive $\lambda_1$.[22]

A final observation concerns the lower tail of the joint distribution. For low $\mathrm{ID}_e$ values, all models tend to generate MT values that are too low. Models 2 and 3 are particularly problematic in this regard, as they even predict some negative MT values.

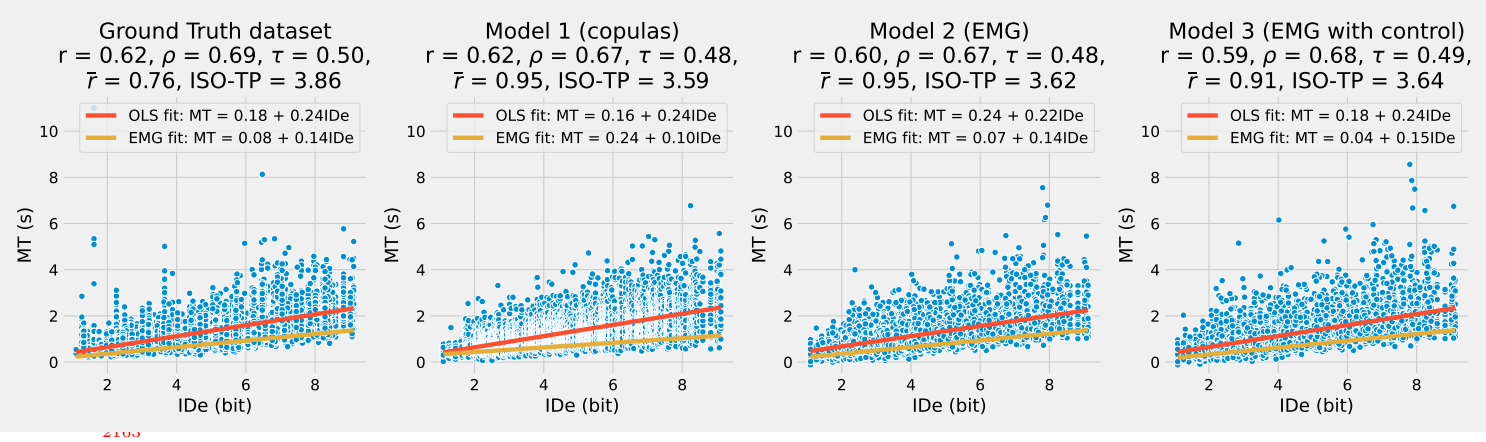


Fig. 25. Consistency as evaluated by the criterion of subsection 11.1. The original dataset (left panel) is compared to data produced by Model 1 (middle left panel), Model 2 (middle right panel) and Model 3 (right panel).

### 11.2 Comparisons of models parametrized by strategy

We have evaluated the consistency of the models for copying an existing dataset, which for all models was good. Hence, one approach that will work to generate pointing data per strategy is to simply fit one model per strategy. However, as stated in our initial goal, we want a single model parametrized by strategy. We thus used the method explained in each *Accounting for different strategies* paragraph in section 10 to generate datasets from for each strategy parameter. These five datasets were then aggregated to evaluate the fit of the linear models described Figure 9. The resulting fits are shown Figure 26. Model 1 (top panel) fits well on the $\mathrm{ID}_e$ dimension $(\mu_i, \sigma_i)$, but fails on the $\overline{\mathrm{MT}}$ dimension $(\mu_t, \sigma_t)$, and the Pearson r. In contrast, Model 2 and 3 (middle and bottom panels) are a lot closer to Figure 9, with Model 3 on the whole superior in about every metric.

[22] Indeed, the supplementary materials demonstrate an "$a$" correction instead of a $\lambda_1$ correction that is able to achieve the same $\overline{r}$ as the original dataset, but this comes at the expense of significantly poorer performance in other metrics.

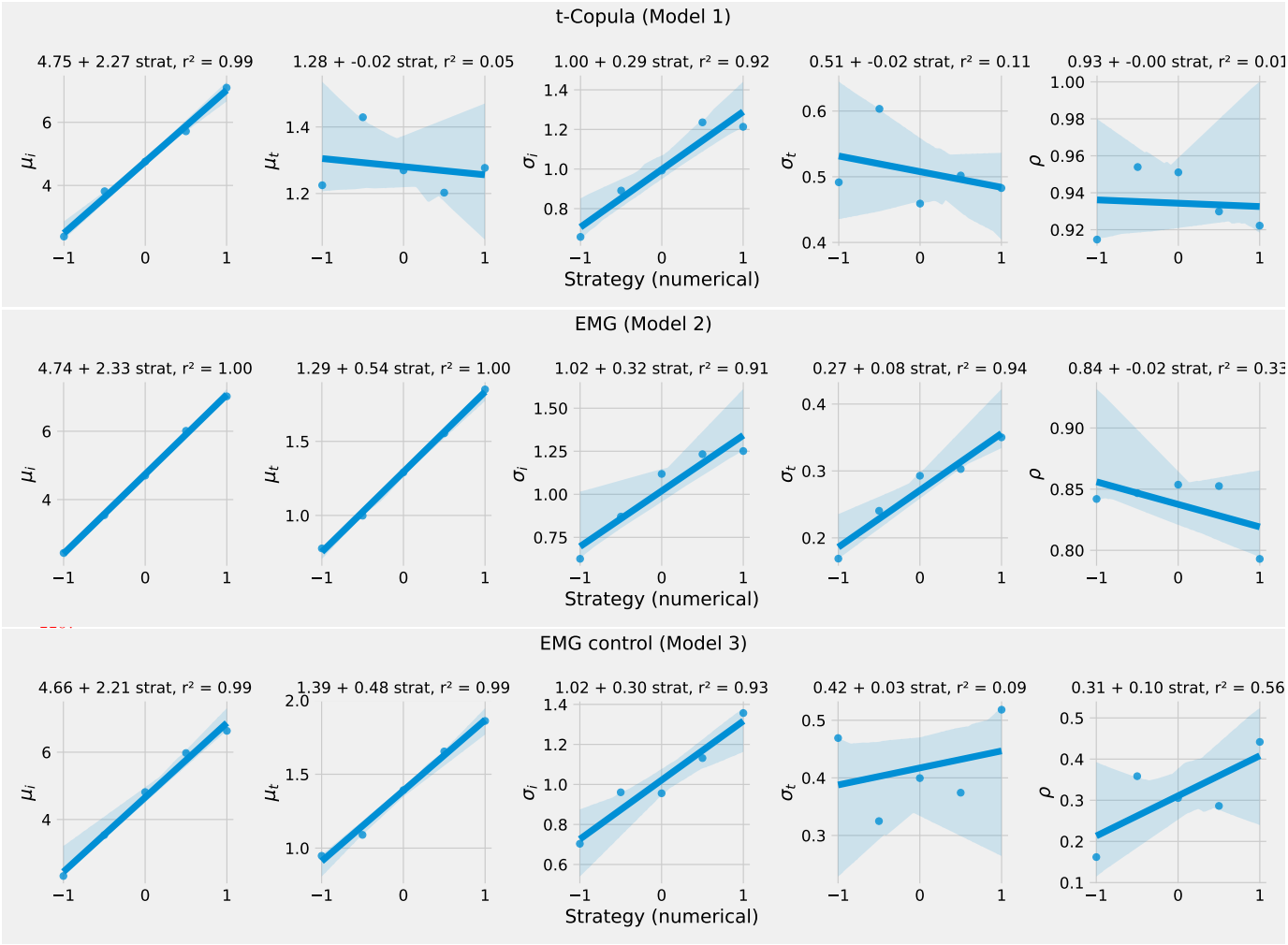


Fig. 26. An evaluation of the linear parametrization of the bivariate Gaussian model of Equation 24. Each subpanel reads like those of Figure 9. Top panel: evaluation for Model 1. Middle panel: evaluation for Model 2. Bottom panel: evaluation for Model 3.

## 12 CONCLUSION

We presented a comprehensive analysis of pointing behaviors in three different protocols, leveraging four existing, diverse datasets. We demonstrated how this analysis could be exploited to generate joint distributions of movement time MT and $\mathrm{ID}_e$ that are parametrized by strategy. We finish with two discussion points and avenues for future work.

### 12.1 The bivariate Gaussian model for $\overline{\mathrm{MT}}$

As with any model, there is a trade-off in pointing models between complexity and precision in movement time models, from the simplest deterministic Fitts' law Equation 1 to the more complex models that generate

entire trajectories described subsection 2.5. The models we suggest in this work produce an entire joint distribution of movement times and $\mathrm{ID}_e$. However, the simpler bivariate Gaussian model of section 8 may be preferred when modeling $\overline{\mathrm{MT}}$ only is sufficient.

$$\begin{pmatrix} \mathrm{ID}_e \\ \overline{\mathrm{MT}} \end{pmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mu_i \\ \mu_t \end{bmatrix}, \begin{bmatrix} \sigma_i^2 & r\sigma_i\sigma_t \\ r\sigma_i\sigma_t & \sigma_t^2 \end{bmatrix}, \right) \tag{54}$$

$$\text{with } \mu_i = \alpha_1 + \alpha_2\,\mathrm{ID} + \alpha_3\,s \tag{55}$$

$$\mu_t = \beta_1 + \beta_2\,\mathrm{ID} + \beta_3\,s \tag{56}$$

$$\sigma_i = \sigma_0 \tag{57}$$

$$\sigma_t = \theta_{t0} + \theta_{t1}\,\mathrm{ID} \tag{58}$$

$$r = \rho_0 \tag{59}$$

$$\tag{60}$$

Values for the parameters of this model can be estimated from an empirical dataset; the parameters of the YKORM dataset, where the input modality was the regular mouse, can be read off Figure 15. We believe this bivariate Gaussian model may represent a sweet spot in the tradeoff between model complexity and precision of the MT,$\mathrm{ID}_e$ representation.

### 12.2 On the local and global dependency

In this work, we studied local and global dependencies between $\mathrm{ID}_e$ and MT, or in other words two forms of the speed-accuracy tradeoff: a local SAT and a global SAT. The fact that we found little to no local dependency can suggest two things: either that the variability of movement time is much larger than $\mathrm{ID}_e$'s, or that from a motor control perspective, movements produced for one condition are equivalent. The idea of having two different scales of SAT is familiar in the study of reaction time (RT) in psychophysical tasks, and is referred to as the micro and macro SAT [14, 18]. One explanation for the micro SAT is that individuals are continuously making choices about trade-offs between speed and accuracy when responding [18]. While it is unclear whether results on RT in psychophysical tasks would transfer to movement tasks, our results are in line: the dependencies between MT and $\mathrm{ID}_e$ are stronger in the tasks with strategy, where participants are invited to exercise the trade-off.

### 12.3 Future work

For future research, we recommend exploring hierarchical pointing models as proposed in [73]. Notably, our supplementary materials reveal considerable inter-participant differences in $\mathrm{ID}_e$ distributions across strategies: 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] Douglas Bates, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015. Parsimonious mixed models. *arXiv preprint arXiv:1506.04967* (2015).

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] Andy Cockburn, Carl Gutwin, and Saul Greenberg. 2007. A predictive model of menu performance. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 627–636.

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

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

[14] 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.

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

[16] Ed Diener, Robert Northcott, Michael J Zyphur, and Stephen G West. 2022. Beyond experiments. *Perspectives on Psychological Science* 17, 4 (2022), 1101–1119.

[17] 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.

[18] 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.

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

[20] Florian Fischer, Miroslav Bachinski, Markus Klar, Arthur Fleig, and Jörg Müller. 2021. Reinforcement learning control of a biomechanical model of the upper extremity. *Scientific Reports* 11, 1 (2021), 14445.

[21] 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.

[22] 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.

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

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

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

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] GM Harshvardhan, Mahendra Kumar Gourisaria, Manjusha Pandey, and Siddharth Swarup Rautaray. 2020. A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review* 38 (2020), 100285.

[42] Edwin T Jaynes. 1957. Information theory and statistical mechanics. *Physical review* 106, 4 (1957), 620.

[43] Jussi Jokinen, Aditya Acharya, Mohammad Uzair, Xinhui Jiang, and Antti Oulasvirta. 2021. Touchscreen typing as optimal supervisory control. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–14.

[44] 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.

[45] 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.

[46] Thomas Langerak, Sammy Christen, Mert Albaba, Christoph Gebhardt, and Otmar Hilliges. 2022. MARLUI: Multi-Agent Reinforcement Learning for Adaptive UIs. *arXiv preprint arXiv:2209.12660* (2022).

[47] 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.

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

[49] 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.

[50] 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.

[51] 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.

[52] Atsuo Murata and Hirokazu Iwase. 2001. Extending Fitts' law to a three-dimensional pointing task. *Human movement science* 20, 6 (2001), 791–805.
[53] Sohad Murrar and Markus Brauer. 2018. Mixed model analysis of variance. *The SAGE encyclopedia of educational research, measurement, and evaluation* 1 (2018), 1075–1078.
[54] 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.
[55] Roger B Nelsen. 2006. *An introduction to copulas*. Springer.
[56] 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.
[57] 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).
[58] 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.
[59] 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.
[60] Danqing Shi, Yujun Zhu, Francisco Erivaldo Fernandes Junior, Shumin Zhai, and Antti Oulasvirta. 2025. Simulating Errors in Touchscreen Typing. *arXiv preprint arXiv:2502.03560* (2025).
[61] 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.
[62] Emanuel Todorov and Michael I Jordan. 2002. Optimal feedback control as a theory of motor coordination. *Nature neuroscience* 5, 11 (2002), 1226–1235.
[63] Emanuel Vassilev Todrov. 1998. *Studies of goal directed movements*. Ph. D. Dissertation. Massachusetts Institute of Technology.
[64] Eric-Jan Wagenmakers and Simon Farrell. 2004. AIC model selection using Akaike weights. *Psychonomic bulletin & review* 11 (2004), 192–196.
[65] 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.
[66] Robert C Wilson and Anne GE Collins. 2019. Ten simple rules for the computational modeling of behavioral data. *Elife* 8 (2019), e49547.
[67] 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.
[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 [37, 71].

### A.2 Schmidt's protocol

A second paradigm, due to Schmidt *et al.* [59], controls MT (speed) through a timing procedure, making MT and D the independent variables, and $\sigma$ 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 $\sigma$ (accuracy) as the dependent variables [38]. 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 [5, 49, 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 $\sigma$. 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 $\mathrm{ID}_e$ for a given nominal ID level. For example in Mackenzie's [49] 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 $\mathrm{ID}_e$, Fitts' law does not link strategy, $\mathrm{ID}_e$ and MT *a priori* since $\mathrm{ID}_e$ 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 $\mathrm{ID}_e$, and $\overline{\mathrm{MT}}$ and $\mathrm{ID}_e$. 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 $X$,
- $\sigma_X = \sqrt{\text{Var}(X)}$ 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 $Y$

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sigma_X \sigma_Y}. \tag{61}$$

- The law of total expectation

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X], \tag{62}$$

  which, when applied to a product, yields

$$\mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|Y]]. \tag{63}$$

- The law of total variance

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X)] + \text{Var}(\mathbb{E}[Y|X]). \tag{64}$$

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

$$\mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|Y]] = \mathbb{E}[Y f(Y)] \tag{65}$$

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[f(Y)] \tag{66}$$

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y]) \tag{67}$$

$$= \mathbb{E}[g(Y)] + \text{Var}(f(Y)) \tag{68}$$

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

$$r(\overline{X}, Y) = \frac{\mathbb{E}[\overline{X}Y] - \mathbb{E}[\overline{X}]\mathbb{E}[Y]}{\sigma_{\overline{X}} \sigma_Y}. \tag{69}$$

One notices from Equation 65 and Equation 66 that interestingly, the covariance between $X$ and $Y$ equals the covariance between $\overline{X}$ and $Y$. We then have

$$\sigma^2_{\overline{X}} = \text{Var}(\mathbb{E}[X|Y]) = \text{Var}(f(Y)). \tag{70}$$

When compared with Equation 67, 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

$$\mathbb{E}[X|Y] = a + bY = f(Y). \tag{71}$$

We also consider any conditional variance model

$$\mathrm{Var}[X|Y] = g(Y). \tag{72}$$

The correlation coefficients, following the definition in Equation 61 can then be written as

$$r(X,Y) = \frac{\mathbb{E}[Y(a+bY)] - \mathbb{E}[a+bY]\mathbb{E}[Y]}{\sqrt{(\mathbb{E}[g(Y)] + \mathrm{Var}(a+bY))}\sigma_Y} \tag{73}$$

$$r(\overline{X},Y) = \frac{\mathbb{E}[Y(a+bY)] - \mathbb{E}[a+bY]\mathbb{E}[Y]}{\sqrt{\mathrm{Var}(a+bY)}\sigma_Y}. \tag{74}$$

The covariance parts (*i.e.*, the denominators in Equation 73 and Equation 74) simplify as

$$\mathbb{E}[Y(a+bY)] - \mathbb{E}[a+bY]\mathbb{E}[Y] = a\mu_Y + b\mathbb{E}[Y^2] - a\mu_Y - b\mathbb{E}[Y]^2 \tag{75}$$

$$= b\sigma_Y^2. \tag{76}$$

Because

$$\mathrm{Var}(a+bY) = b^2\mathrm{Var}(Y) = b^2\sigma_Y^2, \tag{77}$$

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

$$r(X,Y) = \frac{b\sigma_y^2}{\sigma_y\sqrt{\mathbb{E}[g(y)] + b^2\sigma_y^2}} = \frac{1}{1 + \frac{\mathbb{E}[g(y)]}{b^2\sigma_y^2}}. \tag{78}$$

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 15)

$$g(Y) = s^2 + (\lambda_0 + \lambda_1 Y)^2 \tag{79}$$

which gives

$$\mathbb{E}[g(Y)] = s^2 + \lambda_0^2 + 2\lambda_0\lambda_1\mu_Y + \lambda_1^2(\mathrm{Var}(Y) - \mu_Y^2) \tag{80}$$

which does not offer further simplification. This result shows that the dependence between $\mathrm{ID}_e$ and MT can not be specified independently of the EMG parameters: for a given $\mathrm{ID}_e$ distribution, the correlation between MT and $\mathrm{ID}_e$ 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[23] and the independent copula. We utilized the R `copula` package [70] to fit these candidate copulas;

[23] 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.

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.

### 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

$$AIC = 2(k - \mathcal{L}); \tag{81}$$

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 $\mathcal{R}$

$$\mathcal{R} = \exp\left(-\frac{\mathrm{AIC}_1 - \mathrm{AIC}_2}{2}\right) \tag{82}$$

which can directly be interpreted as odds (see also [64]). 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

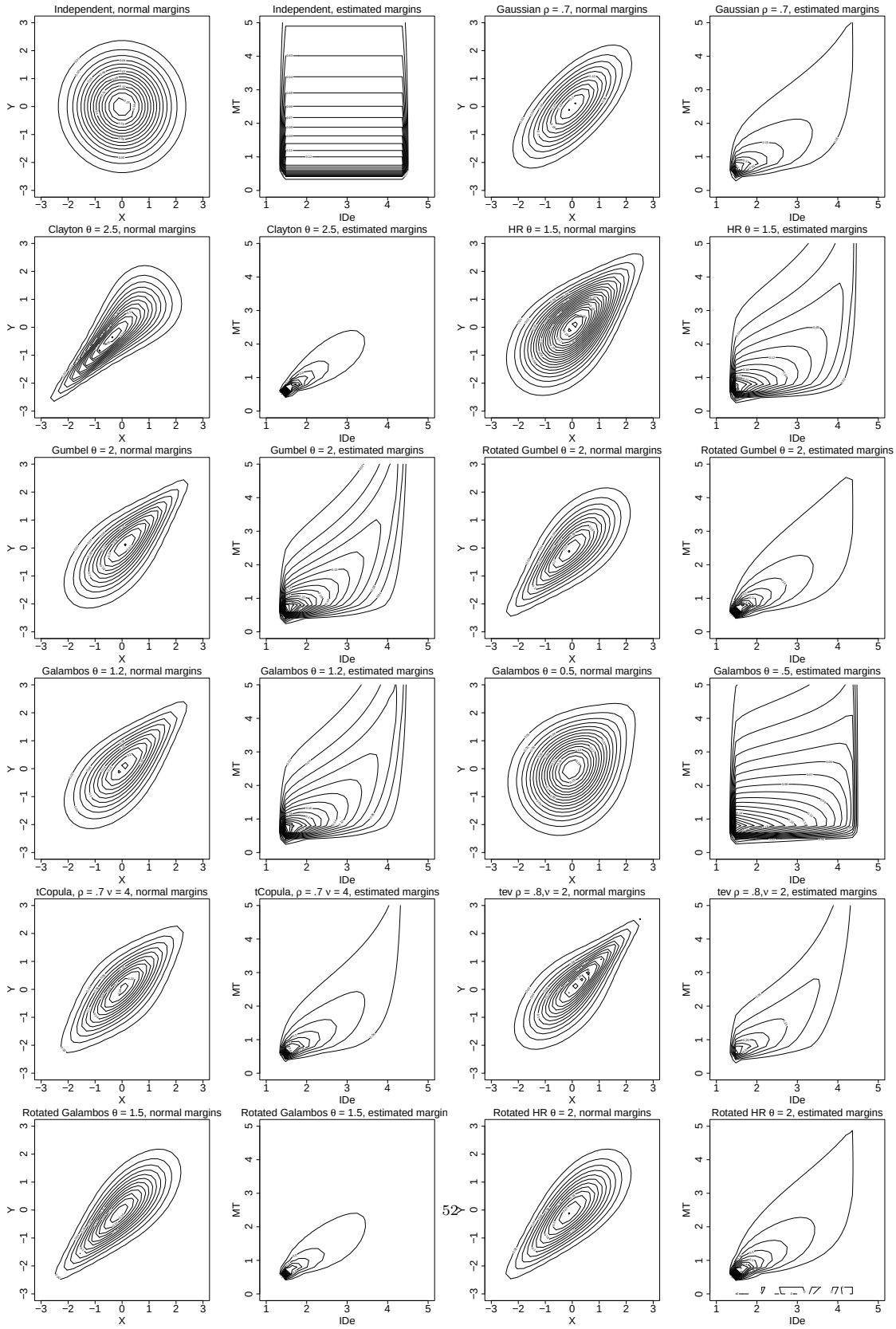


Fig. 27. Visualization of the copulas used in this work. We represent the densities of the joint distribution coupled with 10 different copulas. For the first and third row, the copula joins standard normal marginal distributions. For the second and fourth row, the same copulas with the same parameters joins the marginals as 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 FITTING LINEAR MIXED-EFFECT MODELS

In this work we fit Linear Mixed-effect Models (LMMs). They are useful when dealing with hierarchical or clustered data, such as repeated measurements on the same participants as in this work[40]. 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 [53]. 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" [40].

*Random intercept only models.* Throughout this work, we only consider LMM with random intercepts. This choice limits the complexity of the fitted models and associated risks[4] while controlling for the non-independent nature of repeated measurements [40].

*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. We found fitting with ML was more prone to errors of convergence, but changing the optimizer method in our case always succeeded.