[54] Emanuel Todorov and Michael I Jordan. 2002. Optimal feedback control as a theory of motor coordination. *Nature neuroscience* 5, 11 (2002), 1226–1235.
[55] Eric-Jan Wagenmakers and Scott Brown. 2007. On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological review* 114, 3 (2007), 830.
[56] 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.
[57] 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.
[58] 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.
[59] 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.
[60] Jun Yan. 2007. Enjoy the joy of copulas: with a package copula. *Journal of statistical software* 21 (2007), 1–21.
[61] 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.
[62] 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.
[63] 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.

A PROOF OF PROPOSITION 3.1

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

We define

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

We recall the following results:

- The Pearson correlation coefficient is defined as the normalized covariance $\mathrm{cov}(X, Y)$ between X and Y
$$r(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\sigma_X \sigma_Y}. \tag{38}$$
- The law of total expectation
$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X], \tag{39}$$
which, when applied to a product, yields
$$\mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|Y]]. \tag{40}$$

- The law of total variance

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X]). \tag{41}$$

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

$$\mathbb{E}[XY] = \mathbb{E}[Y\mathbb{E}[X|Y]] = \mathbb{E}[Yf(Y)] \tag{42}$$

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

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

$$= \mathbb{E}[g(Y)] + \mathrm{Var}(f(Y)) \tag{45}$$

The general practice of computing block averages described subsection 2.4 essentially considers $\overline{X} = \mathbb{E}[X|Y]$ instead of X:

$$r(\overline{X}, Y) = \frac{\mathbb{E}[\overline{X}Y] - \mathbb{E}[\overline{X}]\mathbb{E}[Y]}{\sigma_{\overline{X}}\sigma_Y}. \tag{46}$$

One notices from Equation 42 and Equation 43 that interestingly, the covariance between X and Y equals the covariance between $\overline{X}$ and Y. We then have

$$\sigma^2_{\overline{X}} = \mathrm{Var}(\mathbb{E}[X|Y]) = \mathrm{Var}(f(Y)). \tag{47}$$

When compared with Equation 44, one thus sees that the only difference between $r(X, Y)$ and $r(\overline{X}, Y)$ is that $\mathbb{E}[\mathrm{Var}(X|Y)]$ 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

$$\mathbb{E}[X|Y] = a + bY = f(Y). \tag{48}$$

We also consider any conditional variance model

$$\mathrm{Var}[X|Y] = g(Y). \tag{49}$$

The correlation coefficients, following the definition in Equation 38 can then be written as

$$r(X, Y) = \frac{\mathbb{E}[Y(a + bY)] - \mathbb{E}[a + bY]\mathbb{E}[Y]}{\sqrt{(\mathbb{E}[g(Y)] + \mathrm{Var}(a + bY))}\sigma_Y} \tag{50}$$

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

The covariance parts (*i.e.*, the denominators in Equation 50 and Equation 51) simplify as

$$\mathbb{E}[Y(a + bY)] - \mathbb{E}[a + bY]\mathbb{E}[Y] = a\mu_Y + b\mathbb{E}[Y^2] - a\mu_Y - b\mathbb{E}[Y]^2 \tag{52}$$

$$= b\sigma_Y^2. \tag{53}$$

Because

$$\mathrm{Var}(a + bY) = b^2\mathrm{Var}(Y) = b^2\sigma_Y^2, \tag{54}$$

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

$$r(X,Y) = \frac{b\sigma_y^2}{\sigma_y\sqrt{\mathbb{E}[g(y)] + b^2\sigma_y^2}} = \frac{1}{1 + \frac{\mathbb{E}[g(y)]}{b^2\sigma_y^2}}. \tag{55}$$

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 17)

$$g(Y) = s^2 + (\lambda_0 + \lambda_1 Y)^2 \tag{56}$$

which gives

$$\mathbb{E}[g(Y)] = s^2 + \lambda_0^2 + 2\lambda_0\lambda_1\mu_Y + \lambda_1^2(\mathrm{Var}(Y) - \mu_Y^2) \tag{57}$$

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

B FITTING AND COMPARING COPULAS

B.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[18] and the independent copula. We utilized the R `copula` package [60] to fit these candidate copulas; copulas are estimated based on maximum likelihood estimation, and the marginals are estimated with *empirical* maximum likelihood estimation; estimations are available by directly calling library functions, as illustrated in the code that comes with this paper.

B.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

$$AIC = 2(k - \mathcal{L}); \tag{58}$$

[18] 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.