

List of changes to the manuscript

Reviewer comments slanted, personal comments in plain text, **action points in bold**.

1. *I have carefully followed the proofs in Appendix B. In my understanding, they are (mostly) correct. One minor error is that Equation 35 should be $-bE[Y]^2$ instead of $+bE[Y]^2$. I assume it is a typo because the conclusion (Section 36) is correct.* **Corrected typo.**
2. *Another thing that I had trouble understanding is the linear standard deviation stated in Equation 32, that is, $\text{Var}[X|Y] = (c + dY)^2$. This literally means that the variance of MT given ID_e increases quadratically with ID_e . What is the supporting evidence for this assumption? I further wonder if this assumption is relevant to the proof at all. In my understanding, the proof should end at Line 1020. There is no need to discuss the extra term in $r(X, Y)$ and later stuff further, which means Equation 32 is not even necessary. It is true that the linear standard deviation assumption is not needed for the first part of the proof, we have made this clear. It is not a general assumption but is a property of the EMG model which is one of the focuses of this paper. Evidence for this property is provided by the good fits of the EMG model, and work by [?] who test various variance models, and find quadratic variance (i.e., linear standard deviation) fits best.* **Split the proof in two and clarified that the second part of the proof shows that the correlation between ID_e and MT can not be specified freely from the EMG parameters, but is instead dictated entirely by them.**
3. *The proof is difficult to read, partially because the authors omit the intermediate steps to derive one equation from another. I suggest the authors add the rationale of the derivation. For example, in Equation 25, $E[XY] = E[YE[X|Y]]$ (law of total expectation). And the law of expectation should be in another section of the Appendix for readers to refer to. Similarly, for Equation 26-28.* **Reminded the necessary ingredients to follow the proof, including law of total expectation, including product form, and law of total variance.**
4. *The proof also uses phrases like "one notices from Equation 25 and 26..." "and some further algebra shows..." Just present the equations and be clear about what they mean. For example, the authors could provide*

$COV(X, Y) = E[XY] - E[X]E[Y] = E[YE[X|Y]] - E[E[X|Y]]E[Y] \dots$ to show how the "notice" makes sense. The point is that a reader will likely spend significantly less time than the authors used to derive the equations and may be less familiar with the discussed methods, so it will be helpful to make their job easier. **Removed such phrases and made some steps, such as the definition of covariance, more explicit.**

5. *The Introduction and Application of Copula.* The authors claim to introduce copulas to HCI, but the explanations in the manuscript are too brief, and I had to consult other resources to understand them. To improve this, perhaps the authors could include a simple example that clearly demonstrates what copulas can achieve compared to other methods. Additionally, I believe Model 1 has the potential to do more than Models 2 and 3 because of the pros of using copulas. I would like to see an example that illustrates this. Here, we only partially agree with the reviewer. The goal of the paper is not to introduce copulas to HCI, but to introduce pointing models. It turns out that using copulas is the principled way to introduce dependencies between probability distributions, and so we use this to build one of the candidate models, and since it is likely an unknown tool for the HCI community we need to introduce it somewhat. It seems out of scope to expect from this paper that a reader will understand copulas by reading it. It does seem fair that the reader should be able to grasp what copulas are, and what they can be used for, and the previous version was indeed a little quick on copulas. As for the second point, model 1 using copulas is generic, whereas model 2 and 3 are more problem-specific, so it is not obvious that model 1 would outperform the two others. **Added an example to explain what copulas can do, a graph about how they do it, but also point readers towards relevant literature to learn more about copulas.**
6. *The Evaluation of the Models.* The model construction and evaluation seem mostly fine. One note is that the authors claim between Line 706 and Line 707 that "all models produce data that closely match the metrics..." How did the authors quantify the data to be closely matched? Through absolute values? How close is defined as close? The titles of the Figure have the scores embedded, and they are close by any reasonable metric, except for \bar{r} as explained. We don't think more precise quantification would be helpful, but believe perhaps the values themselves have been missed by this reviewer. **Add a title on top of each figure to make it clear there is a ground truth and 3 model evaluations, and clarify in text that the values are displayed in the Figure.**
7. *The paper claims ID_e can only be defined after the experiment has been conducted.* However, this completely ignores the related works that model spatial endpoint distributions from Grossman and Balakrishnan, Bi and Zhai to name a few. On the same note, more than 1/4 of the existing references are from the same group. This is somewhat questionable.

****TODO****

8. *The paper should explain more about EMG models. For example, I am not sure why $E[MT \mid ID_e]$ follows a simple condition mean where the exponential mean is $\lambda * x$, as in Equation 11. The paper should explain this in a self-contained manner. On the same note, I don't understand Equation 17. **We provided more context about EMG, in particular the conditional mean and equation 17 are now better explained.***
9. *The paper only considers the dependence between ID_e and MT when manipulating W . However, D could also influence the dependence, as indicated by Grossman and Balakrishnan. Why are they not considered?*
****TODO****
10. ** It is a bit strange to capture \overline{MT} and ID_e through a bivariate Gaussian distribution. First, movement time data itself is one-dimensional. Second, there is no clear evidence suggesting that it could be a bivariate Gaussian. Instead, the variance could be constant or skewed towards one side. Why not consider these alternative models? Furthermore, the authors claim that in the "Speed emphasis" condition, MT and ID seem to be uncorrelated. Have the authors used the correct distribution for these? Was it because of the calculated ID_e having a small range? Here, the reviewer seems to have misunderstood the work. The marginal distribution of MT is 1D. The marginal distribution of ID_e is 1D. The joint distribution of (MT, ID_e) is then necessarily 2D. Additionally, the use of this model is already justified in the manuscript: "We chose the bivariate model because it is simple, easy to interpret, it makes the dependence between MT and ID_e completely specified via Pearson's r , and is expected to fit reasonably well given the visual appearance of the clusters". The results in this work do not rely particularly on this assumption, any other bivariate distribution could have been used with perhaps a slightly better precision, at the cost of making the model more complex. Note that we use this model essentially for evaluation. In this context, Shmueli [?] discusses how models used for evaluation do not have to score high on prediction to be useful, but instead must prove to have good identifiability of parameters, which is the case of the bivariate Gaussian.*
11. *By analyzing existing datasets, the authors claimed to have successfully modeled the mean and standard deviation of MT, the mean and standard deviation of the effective task difficulty ID_e , and the throughput (TP) of Fitts' law. TP was used as an evaluation metric, and we claim to model the joint distribution of MT and ID_e for various user strategies, not just the mean and standard deviation of ID_e .*
12. *While it is somewhat acceptable for a conference paper to omit details or to say "refer to Chapter 1 of a textbook [XYZ]", the way the authors described how they used copulas for data processing is overly simple. **Answer to item 5. Also emphasized that the data processing for copulas***

is performed by calling functions of an R library, and does not require significant coding.

13. *I would like to point out that analyzing the existing datasets does not fully achieve the authors' initial goals. The experiment of JGP dataset did not ask participants to operate with shifting their speed-accuracy tradeoff. The GO dataset had targets with $W=1$ pixel, which is unlike typical tasks in the HCI field and extremely unrealistic situation that rarely occurs in our PC tasks.* Here, we disagree with the reviewer, and further believe the reviewer misunderstood how the GO dataset was produced. In a speed-accuracy tradeoff experiment, one has to make the participant tradeoff speed and accuracy in ways that can be controlled. One way, due to Fitts, and the one that is most used in HCI, is to vary W . As explained in the paper, the downside to that is that people tend to over and under utilize W depending on its actual value. Another is to simply ask the participants to modify how they trade off speed for accuracy, which is how the GO dataset is produced. Both are practically useful (pointing towards a target, vs pointing towards a point in space)**Introduced a new subsection in the background to clarify `**TODO**`**
14. *The proposed model introduces numerous constant values, which were determined experimentally. If these numerical values are necessary, would not an experiment be required as before? If so, then the initial goal has not been achieved. Alternatively, if the authors are aware that numerous constant values are necessary, the initial goal setting is mismatched, and the authors need to explain the research objectives in more detail.* The numerical values are constant parameters, just like a and b in Fitts' law. Sure they have to be estimated once, but they remain mostly constant over time for a given experimental setup. This is not at the same level as IDE, who changes based on user strategy and changes each block.
15. *However, if this approach is acceptable, could not we estimate MT using a simpler model (though a point estimate) without employing a complex method like the authors'? For example, how about a model like $MT=a+b*ID+c*S$, where S is the strategy (taking values from -1 to +1 as in Figure 4)? In other words, a model where MT decreases when S takes values like -1 or -0.5 under speed emphasis. Alternatively, considering that the effective movement amplitude hardly changes even if the speed-accuracy strategy changes, how about using a model like $MT=a+b*\log_2(D/(W-c*S)+1)$, which affects only the target size term in Fitts' law?* The initial goal, as explained at the head of the paper is to predict the entire distribution, not just a mean value. It is true that we may suggest a simpler model as an extension of Fitts' law as suggested by this reviewer, which may be enough for some cases. **`**TODO**`, see if achievable from the bivariate Gaussian MMT, IDE**
16. *If mapping the speed-accuracy strategy to numerical values addresses the*

authors' initial problem in this way, it would be a higher priority to verify just that effect. The current paper introduces unnecessarily complex analysis methods, and it is unclear what specifically solved the problem of not being able to introduce the speed-accuracy strategy into the MT estimation model beforehand (whether it was the numerical mapping of the speed-accuracy strategy, the introduction of copulas, or other factors). As explained in the introduction, the proposed model want to introduce the strategy in the MT model, and want to predict the entire distribution. **added a subsection in background to further emphasize why we want to consider distributions over means.**

17. Also, users unconsciously shift their bias toward speed or accuracy [<https://dl.acm.org/doi/10.1145/3373625.3416999>]; is the proposed method applicable to such a situation where S values cannot be predetermined? This is precisely a situation where the speed-accuracy strategy can only be obtained a posteriori. If the authors' method allows for estimating MT and ID_e even without prior instructions regarding strategy, I would like them to emphasize that more. ****TODO****
18. As mentioned earlier, I assume that the authors' research goal might also be achievable using reinforcement learning proposed in previous studies. However, in reinforcement learning, for example, Do et al's method took six days of processing time. Could the advantage of the authors' method be that processing time is shorter, or that estimation can be done with almost no time lag similar to Fitts' law analysis? If so, I would like the authors to emphasize that. One of the objectives being a parametric model, Reinforcement learning can not achieve that. Processing time is also very quick (almost immediate) for the typical sample size used in HCI experiments. **Comparison with RL treated in the new related work section.**
19. Finally, please ensure there are no undefined variables. I had to read the paper while guessing what they meant, and I could not read it smoothly. For example, regarding σ_i and σ_t in Equation 13, I guessed that i refers to the " ID_e " component and t refers to the " MT " component. It is the authors' responsibility to explain everything. **This notation has been clarified.**
20. For example, on page 2, some notations are defined; however, G and E on page 5 are not explained. Readers may not follow if they forgot the notations from previous work. **clarified**
21. In Section 3, details about the dataset used are unclear. While it is a "well-known" dataset, information about the number of unique ID s and total trials is necessary to give readers an overall impression of the analysis. **Added factor levels and number of trials of the experiment.**
22. Please explain why the log-likelihood is normalized by dividing by the maximum log-likelihood across all potential copulas.

23. *In Section 4, the statement "The previous section investigated the nature of the dependence between ID_e and MT for a given strategy" requires clarification. Please summarize the findings in Section 3 and clarify "nature" before making this claim to enhance clarity.***Contributions of the section now summarized.**
24. *In Figure 2, it is clear that rotGumbel performs best. However, in Figure 5 (left panel), the results are not straightforward: for p1 and p2, t seems best, while for p3, t-EV is better. Please clarify the criteria used to select the best strategy—specifically, whether it is the mean value of the normalized log-likelihood across all participants.***clarified and added details to the comparison.**
25. *In Section 4.4, why is the bivariate Gaussian model chosen? You mention it fits reasonably well based on the visual appearance of the clusters, but this is not rigorous. Was a normality test conducted to assess the fit?*
We completely disagree with this comment. First, normality tests are known to offer little practical insight: they usually have low power for small samples and are overly sensitive for large samples. Further, most normality tests are designed for univariate data, and do not translate well to multivariate cases; Second, the point of the bivariate Gaussian used here is not to achieve perfect prediction, but to summarize data efficiently. A visual inspection is enough here to see that the data looks unimodal and symmetric, in which case it is known that a Gaussian bivariate will fit well enough. Visual inspection is often considered superior to normality tests by many statisticians, for example see Gelman's view on that (in various papers/textbooks and blogposts). The idea that a normality test is more rigorous is False, and does not consider what the model is supposed to do. This idea likely stems from the perception that a test is objective whereas a visual inspection is subjective. This is forgetting about all the subjective choices one makes when testing, among which what significance level to consider, which normality test to consider.
26. *Page 11, lines 524-525 states that "for many cases, the t-copula actually reduces to a Gaussian copula." This raises a question: why not compare directly with the Gaussian copula, given that your results suggest it reduces to this in most cases, as seen in Figures 2 and 5?* Precisely because the t-copula can reduce to a Gaussian, and hence, it can only be better (at worst it is equivalent). There is thus nothing to gain by adding the Gaussian copula.
27. *The evaluation of the model is not clear. Since the model is derived using two datasets, it is important to address potential bias. Comparisons using the same dataset could lead to misleading conclusions. It is essential to compare the model against another publicly available dataset or to collect new data through a user study to validate the proposed models.*