Varying Subjective Speed-accuracy Biases to Evaluate the Generalizability of Experimental Conclusions on Pointing-facilitation Techniques

Shota Yamanaka
syamanak@yahoo-corp.jp
Yahoo Japan Corporation
Tokyo, Japan

Taiki Kinoshita
kinop.tofu@gmail.com
Meiji University
Tokyo, Japan

Yosuke Oba
bonscow@gmail.com
Meiji University
Tokyo, Japan

Ryuto Tomihari
ryuutomi1110@gmail.com
Meiji University
Tokyo, Japan

Homei Miyashita
homei@homei.com
Meiji University
Tokyo, Japan

ABSTRACT

In typical experiments to evaluate novel pointing-facilitation techniques, participants are asked to perform a task as rapidly and accurately as possible. However, the balance can differ among participants, and the techniques' effectiveness would change if the majority of participants give weight to either speed or accuracy. We investigated the effects of three subjective biases (emphasizing speed, neutral, and emphasizing accuracy) on the evaluation results of pointing-facilitation techniques, namely Bubble Cursor and Bayesian Touch Criterion (*BTC*). The results indicate that Bubble Cursor outperformed the baseline in terms of movement time and error rate under all bias conditions, while *BTC* underperformed a simpler target-prediction technique, which was an inconsistent outcome to the original study. Examining multiple biases enables researchers to discuss the (dis)advantages of novel or existing techniques more precisely, which can be beneficial to reach a more reliable conclusion.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **Pointing**; **Empirical studies in HCI**.

KEYWORDS

Human motor performance, target prediction, Fitts' law, area cursor

ACM Reference Format:
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 (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3544548.3580740

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580740

1 INTRODUCTION

1.1 Background

Since pointing to a target is one of the most fundamental operations on PCs and touch devices, reducing operation times and error rates is beneficial to users. HCI researchers have thus proposed numerous pointing-facilitation techniques over at least the past three decades. When they evaluate the effectiveness of a novel technique, a user experiment to compare it with a baseline is commonly conducted.

A key factor on which we focus is the instruction to participants. The most typical instruction is "Select a target as rapidly and accurately as possible" [56], or variations such as "Work as fast as possible while still maintaining high accuracy." [24]. These instructions are reasonable, because in realistic computer usage, users would like to select a target without spending an unnecessarily long time and avoid missing targets.

However, two concerns arise. First, balancing the speed-accuracy tradeoff is just one possible realistic condition [7, 60, 65, 69]. For example, when there are unwanted hyperlinks (i.e., distractors) around the intended one, users have to carefully aim for the target; otherwise, an additional time cost to go back to the previous page then point to the target again is needed [68]. Second, the interpretation of "operating as rapidly and accurately as possible" could differ among participants. In a previous study, when the participants were unsure how to weight the balance between speed and accuracy, the researchers answered "It is up to you" [43]. Therefore, it is possible that a group of participants in an experiment be biased towards either speed or accuracy, which changes the task outcomes such as movement time *MT* and error rate *ER*.

This speed-accuracy imbalance could affect research conclusions when evaluating pointing-facilitation techniques. Suppose that an error-reduction technique exhibits a mean *ER* of 3% while a baseline technique exhibits an *ER* of 9%. Researchers then run a statistical test to claim that the proposed technique was significantly better than the baseline. If the participant group is biased towards accuracy, however, the *ER*s would be (say) 1 and 1.5% for the proposed and baseline techniques, respectively, and no significant difference would be found. Such a low *ER* is possible according to previous studies. Zhai et al., for example, reported that *ER* without a facilitation technique was 0% when the instruction emphasized accuracy [75].

Assuming that conclusions on the effectiveness of a novel technique is affected by the participants' speed-accuracy balances, we should thus consider the importance of replication studies that has been pointed out in the HCI field [17, 36]. In addition to the established methodology of such a direct replication, we examined how subjective speed-accuracy biases affect the effectiveness of pointing-facilitation techniques. We conducted two experiments to evaluate two pointing-facilitation techniques, i.e., Bubble Cursor [32] and Bayesian Touch Criterion (*BTC*) [11], under three speed-accuracy bias conditions. Our findings are as follows.

- In Experiment 1, Bubble Cursor outperformed the baseline in terms of *MT* and *ER* under all three bias conditions. This result enhances the claim in the original study [32] that this technique achieves high performance regardless of users operating rapidly or carefully, resulting in increasing the generalizability of its effectiveness.
- In Experiment 2, *BTC* underperformed a simpler technique under all three bias conditions. This result supports the lack of reproducibility in *BTC*'s effectiveness more strongly than testing only a single bias (i.e., balancing speed and accuracy).

1.2 Contribution Statement

We offer the following two main contributions, which have never been conducted and discussed in the literature as an application of varying speed-accuracy biases.

- We conducted two replication studies to evaluate the superiority of Bubble Cursor and *BTC* to the baselines under the three subjective bias conditions. These two experiments were designed carefully for fair comparisons with the conclusions of original studies, e.g., using the same task difficulties, same number of participants, and same statistical tests.
- Our results will enable researchers to introduce this methodology of intentionally varying speed-accuracy biases to strengthen the research conclusions when (1) testing their novel techniques' effectiveness, (2) replicating an evaluation study for a previously proposed technique, and (3) validating the applicability of a technique to other devices. This methodology will enrich their understanding of the techniques, ensure the validity of research conclusions, and indicate future directions to resolve limitations, e.g., by reducing *ER* even when users are biased towards accuracy.

Because varying speed-accuracy biases can be used with only oral instructions, our work provides an easy-to-apply methodology when researchers conduct user studies. While we limited our focus to target-pointing tasks in this paper, this methodology could be introduced to other operation paradigms and facilitation techniques, such as those for goal-crossing tasks [2, 4, 71] and path-steering ones [1, 22, 66]. Our work thus opens up a new research space, which informs potential future work.

2 RELATED WORK

2.1 Fitts' Law and Target Selection-facilitation Techniques

The *MT* to point to a target using mice and touchscreens is accurately modeled using Fitts' law [26, 47]:

$$MT = a + b \cdot \log_2 \left(\frac{A}{W} + 1 \right), \tag{1}$$

where A is the distance from the initial position of the cursor or finger to the target center, W is the target width, and a and b are empirical constants. The logarithmic term is called the index of difficulty *ID* in bits.

Fitts' law expresses that *MT* can be reduced by shortening A or expanding W. Example techniques to shorten A include Delphian Desktop [5], Drag-and-pop [9], and Ninja Cursors [41], and those to widen W include Area Cursor [40], Expanding Targets [49, 73], and Sticky Icons [64]. Techniques for touch-based operations require certain considerations such as no mouse cursor can be used. Examples of such techniques include LinearDragger [6], Shift [59], Escape [70], and 2D-BayesPointer [45].

For all the abovementioned techniques, researchers conducted user experiments to evaluate how well they outperform the baseline technique, i.e., using a mouse cursor (or a finger) without any facilitation. Several researchers also examined other previously proposed techniques to claim the novelty and effectiveness of their newly proposed ones. For example, in a user study on LinearDragger [6], the researchers compared it with four other techniques: baselines of one- and two-handed tapping, Shift [59], and Escape [70].

2.2 Varying Speed-accuracy Biases

In typical pointing experiments including those in the abovementioned studies, participants are instructed to "point to a target as rapidly and accurately as possible," which emphasizes balancing speed and accuracy [56]. However, it is common that participants unintentionally weight either speed or accuracy; sometimes they show short *MT*s and high *ER*s, while in other cases they show long *MT*s and low *ER*s [55, 67].

To compare user performance measured in such different speed-accuracy balances, using throughput *TP* [bits/s] is recommended [18, 47, 56]. Theoretically, *TP* is invariant even if the speed-accuracy balance changes [47, 56]. For example, MacKenzie and Isokoski compared *TP*s under three bias instructions: emphasizing speed, neutral, and emphasizing accuracy, which we respectively call Fast, Neutral, and Accurate. The *TP*s were 5.67, 5.73, and 5.70 bits/s, respectively [48] (< 1% difference); thus the researchers claimed that *TP* is an invariant metric.

In contrast, Olafsdottir et al. examined five instructions, i.e., *max speed* (asking to just minimize *MT*), *max accuracy* (asking to point to a 1-pixel line without any error), Fast, Neutral, and Accurate [51]. The results indicated that *TP*s ranged from approximately 6 to 10 bits/s (a 42% difference); thus, they questioned the invariance of *TP*.

3 METHODOLOGY OF VARYING SPEED-ACCURACY BIASES

3.1 Candidate Methods to Control Speed-accuracy Tradeoffs

Varying speed-accuracy biases has been used to validate the invariance of *TP* for mouse-based pointing [48, 51], as well as pen-based pointing [75] and keyboard typing [76]. We assume that this methodology is also beneficial to researchers who would like to evaluate the effectiveness of their novel pointing-facilitation techniques. This is because users sometimes cannot maintain the speed-accuracy balance and in realistic situations would change the balance depending on the density of distractors [13, 65] and time costs when missing targets [7, 69].

Regarding the relevance of previous work on the *TP* invariance to our current study, researchers have established several methods to appropriately control speed-accuracy tradeoffs; thus, we have only to choose one of several possible approaches. For example, MacKenzie and Isokoski gave participants a goal *MT* and feedback of each bias's result [48]. That is, a mean *MT* in a pre-test session with the Neutral instruction was recorded for each participant, and this result was written on a piece of paper. In the subsequent data-collection session under the Fast (or Accurate) condition, the participant was then asked to make the mean *MT* at least 10% shorter (or longer) than in the pre-test session.

While the objective was not to validate the *TP* invariance, researchers have also used other approaches to enforce a specific speed-accuracy tradeoff. For example, in Wobbrock et al.'s experiment, the participants had to point to the target before the time limit informed with a metronome [60]. *ER*-based control has also been used. For example, in an experiment to evaluate the target-expansion technique, Zhai et al. controlled the *ER* to 4% by means of the experimental system displaying the resultant *ER* after each block as well as suggesting to slow-down or speed-up the movement speed [74].

Another way to control the speed-accuracy tradeoff is to give only verbal instructions to emphasize either speed or accuracy. For example, *TP*-invariance studies, including Guiard et al.'s work on pointing [33] and Zhang et al.'s text-entry experiment [76], mentioned that the speed-accuracy biases were instructed verbally. Studies on model-fit improvements by using the *effective width* for target pointing [75] and path steering [78] also gave verbal instructions to emphasize speed or accuracy. Note that, according to their papers, feedback on the resultant *MT* and *ER* was not given to the participants during or after the experiments.

We were concerned about the verbal instruction that a group of participants could not perform the task differently for two or more bias conditions, e.g., they would exhibit almost identical *MT*s under the Neutral and Accurate conditions. However, these studies using verbal instructions have shown that their participants appropriately differentiated *MT*s and *ER*s depending on the instructed bias. We thus followed this approach to use the verbal instructions, which were also printed on a piece of paper and noted on the screen as a reminder, and gave no feedback on the results to the participants.

3.2 Biases used in Our Experiments

In this study, we used three biases: Fast, Neutral, and Accurate. To determine if the participants could adequately follow the instruction, we examined the statistical differences among the biases in terms of *MT* and *ER*. Previous studies used more aggressive instructions, such as max speed and max accuracy [51], or extremely fast and extremely accurate [75]. Because our question is whether the empirical conclusions would change when the participants are biased towards speed or accuracy (unintentionally [55, 67] or due to external situations [7, 13]), such drastic instructions, such as "always click on the same one pixel" [51], are outside our focus.

In our experiments, we gave the instructions in two ways.

- An experimenter read aloud the three bias conditions that were written on a piece of paper at the beginning of the experiment. This sheet was placed in front of the participants, and they could check the three biases anytime.
- The current bias was noted at the top of the screen as a reminder.

The three biases noted on the sheet were as follows.

- **Fast**: Perform the task as rapidly as possible. It is okay to make errors. However, please avoid repeatedly pressing the mouse button without aiming for the target.
- **Neutral**: Perform the task as rapidly and as accurately as possible.
- **Accurate**: Perform the task as accurately as possible. It is okay to spend some time to reduce errors. However, please avoid spending an unnecessarily long time.

On the experiment screen, only the first sentence was displayed as a reminder, i.e., the sentences with underlines for each bias.

3.3 Benefits for Researchers to Test Multiple Biases

Researchers are aware that examining multiple conditions in evaluating pointing techniques is preferable to claim the generalizability of their conclusions. For example, testing a wide range of *ID*s by using small to large targets is recommended [56], and using 1D bar-shaped targets, such as hyperlinks, as well as 2D targets is suggested for simulating more realistic user interfaces [10, 62, 63, 68]. Similarly, it was worth varying speed-accuracy biases in their experiments as another factor to enrich the task variations and increase realism.

If we find significant differences in task outcomes (e.g., *MT* and *ER*) between a facilitation technique and baseline under any bias condition, it enhances the effectiveness of the technique regardless of users being careful or in a hurry. In contrast, if the facilitation technique is significantly better than the baseline in limited biases (e.g., only for Neutral), it constrains the claim on the high performance of the facilitation technique to when users appropriately balance speed and accuracy. In both cases, however, testing multiple biases can deepen our understanding of when a novel technique is beneficial to users.

In our study, we summarize our research hypotheses (**H**s) as follows. Because Bubble Cursor is assumed to reduce both *MT* and *ER* [32] while *BTC* was designed to reduce only *ER* [11], **H1** and **H2** have variations noted as (a) and (b).

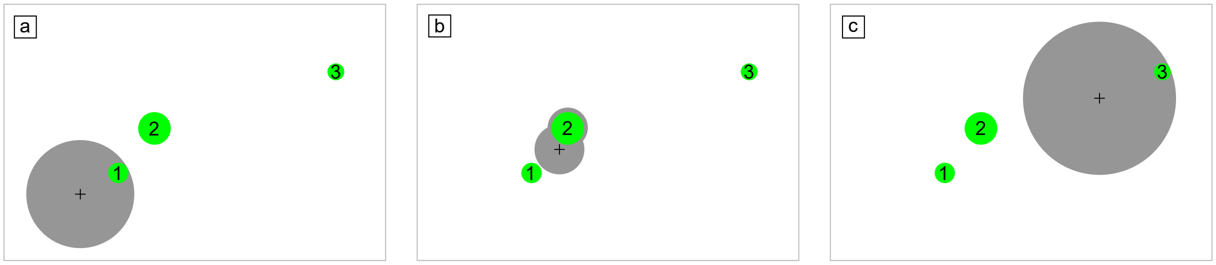

Figure 1: Bubble Cursor updates the radius of its activation area (gray circle) to capture only the closest target. When the mouse button is pressed, the target labeled "1" is selected in (a), "2" in (b), and "3" in (c).

- **H1**: For the Neutral condition, we reproduce the same conclusion on the superiority of the facilitation techniques to the baselines in terms of (a) both *MT* and *ER* for Bubble Cursor and (b) *ER* for *BTC*.
- **H2**: When the speed-accuracy balance is biased towards speed or accuracy, the superiority of facilitation techniques is not statistically confirmed in terms of (a) either *MT* or *ER* for Bubble Cursor and (b) *ER* for *BTC*.

For H2, on the basis of the report by Zhai et al. who found a 0% *ER* for the baseline technique under the Accurate condition [75], we suspect that Bubble Cursor and *BTC* would not outperform the baselines in terms of *ER* when we instruct the participants to perform the tasks accurately.

Although we examined Bubble Cursor [32] and *BTC* [11], any other technique is worth being tested, such as extended versions of Bubble Cursor [15, 29, 34, 35, 44, 46, 57, 58]. Our choices were determined on the basis of simplicity of implementation and variety of input devices. Varying subjective biases for tasks other than target pointing would also be of great significance, which will be included in our future work.

4 EXPERIMENT 1: BUBBLE CURSOR

4.1 Mechanism of Bubble Cursor

Bubble Cursor has a wide area to invoke a click event, rather than a single hotspot with the baseline technique [32]. The radius of Bubble Cursor is dynamically updated so that the activation area captures only one closest target (Figure 1). This mechanism practically shortens the distance to the target and widens the target width, which should reduce *MT* in accordance with Fitts' law.

In the original study [32], Bubble Cursor outperformed the baseline technique in terms of *MT* and *ER*. Thanks to its easy-to-implement algorithm without tuning environment-specific parameters (e.g., those for the cursor-expansion function in DynaSpot [15]), evaluation of Bubble Cursor has been replicated several times [13, 15, 42], and its superiority to the baseline has been consistently confirmed.

4.2 Participants

Twelve volunteer students (the same number as in the main experiment of Bubble Cursor [32]) from a local university participated in this study; ten men, two women; ages: M = 20.6, SD = 1.11 years. All were right-handed and daily mouse users.

4.3 Apparatus

We used a desktop PC (Intel Core i9-12900KF, GeForce RTX 3070 Ti, 32.00 GB RAM, Windows 10 Home). The display was manufactured by ASUS (model: VZ249HR; 23.8 inches, 1920×1080 pixels; 5-msec response time) and had a 60-Hz refresh rate. The experimental system was implemented with Unity 2019.2.19f1 and used in full-screen mode. The wired optical mouse was Logitech G300s (800 dpi). To increase the ecological validity, we set the cursor-speed slider in the Control Panel to default (middle) and turn on the cursor acceleration function ("Enhance pointer precision"), which is the default of the Windows OS.

4.4 Task

We mostly followed the experimental task used in previous studies to evaluate Bubble Cursor [15, 32]. As shown in Figure 2a, the participants aimed for the red target.

The order of the 23 targets is shown in Figure 2b, which referred to ISO 9241-9 [15, 39]. A *set* consisted of 23 successive selections (excluding the starting click) with a fixed $A \times W \times EW/W$ condition, where EW is the effectively available width for Bubble Cursor defined by the distance to four surrounding distractors (Figure 2c).

When a click was successful, a bell sound was played, while a beep was played when a target was missed. The participants had to re-aim for the current target until success.

4.5 Design

This study was a $3 \times 3 \times 2 \times 3 \times 3$ repeated-measures design with the following independent variables and levels: three subjective biases (Bias = Fast, Neutral, and Accurate), two cursor types (Cursor = Point and Bubble), two As (400 and 770 pixels), three Ws (8, 24, and 70 pixels), and three EW/W ratios (1.33, 2, and 3). The As and Ws were determined by referring to the original study [32] so that the ID ranges were the same to maintain the task difficulty for the baseline technique (Cursor = Point). The three EW/Ws were also the same as in that study [32].

For Bubble, the density of intermediate targets that the cursor passes over (called distractor density DD) had only a small effect on MTs and no significant effect on ERs in the original study [32]. Therefore, a replication study to compare Point and Bubble that focused on other conditions (laboratory vs. crowdsourcing) used only DD = 0.5 [42], that is, the density is in the middle of *no distractor* and *fully tiled distractors*. We followed this design to use DD = 0.5.

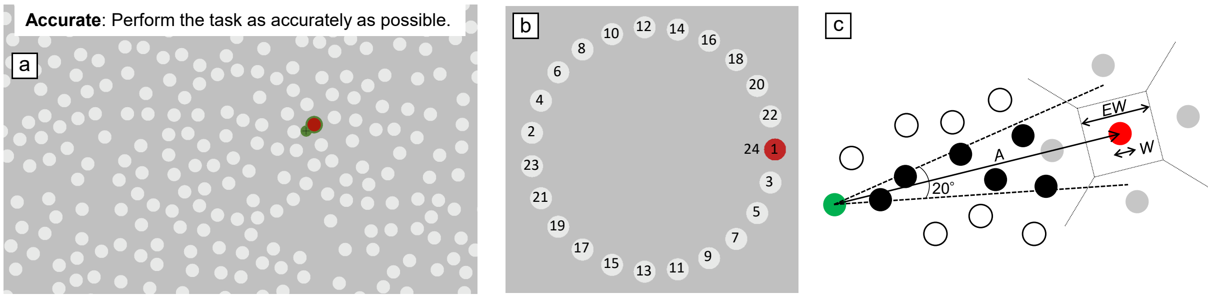

Figure 2: Setup in Experiment 1. (a) Display layout. (b) The order of targets to be clicked. (c) Task conditions based on [15]. Red circle is the current target. Green circle is the previous target (i.e., current start position). Four gray circles are arranged to control the EW/W ratio. Black circles are placed to control the distractor density. White circles are randomly placed distractors through which the cursor would not pass. Colors in (c) are for illustration; the actual colors used in the system are shown in (a).

The offset of distractors in the direction perpendicular to the line of movement was a pseudo-random length, such that each one remained within the 20° slice where the cursor would pass over (see Figure 2c). The remaining space outside the 20° slice was also filled with distractors, the density of which was close to those within the slice. The distractor positions were reset in every trial depending on the new start and goal target locations [32].

4.6 Procedure

We set Neutral to be the first condition so that the participants could speed-up or slow-down their cursor operation in the second and third Biases compared with the first experienced condition. Our experimental results would be affected by this arbitrary choice of ordering, e.g., the participants needed a long time to get used to the task, thus exhibited a high *ER* in the first Neutral condition. Yet, according to our survey, there is no consensus on the order of Bias, e.g., balanced among the participants [48, 75], Neutral to be the last condition [76], and slow-to-fast order for half the participants and fast-to-slow order for the remaining half (thus Neutral to be the center) [33]. Testing if our conclusions would hold for counter-balanced orders will be included in our future work.

For each of the three Biases, the participants performed the task with both Cursors, and the order of the two Cursors was fixed for each participant. We had four possible combinations regarding the orders of second and third Bias conditions (Accurate or Fast) and the two Cursors (Point or Bubble), which were balanced among the 12 participants.

For each of the six Bias×Cursor conditions, the participants first accomplished a practice set with a fixed target condition ($A = 500$, $W = 45$, and $EW/W = 1.6$) that was not used in the data-collection sets. They then performed 18 sets ($2_A \times 3_W \times 3_{EW/W}$) in a random order. We recorded the data from $12_{\text{participants}} \times 3_{\text{Bias}} \times 2_{\text{Cursor}} \times 2_A \times 3_W \times 3_{EW/W} \times 23_{\text{selections}} = 29{,}808$ trials. This experiment took approximately one hour for each participant.

5 RESULTS OF EXPERIMENT 1

We removed 53 outlier trials (0.18%) where the movement distance for the first click point was shorter than $A/2$ or the distance to the target center was longer than $2W$ for Point and $2EW$ for Bubble [7, 48, 60].

Regardless of the data type (*MT* or *ER*), we used RM-ANOVA, although more appropriate analysis methods have been proposed such as the aligned rank transform [61]. This decision was made to maintain consistency with the original study [32] and justified by the fact that ANOVA is robust against data distribution [23, 50]. Bonferroni correction was used for the p-value adjustment method in pairwise comparisons.

As we examined five independent variables, reporting all interaction effects and pairwise tests (e.g., statistical significance values of pairwise tests for Cursor $\times A \times EW/W$) would take up too much space and not relate to the contributions of this paper. Thus, we describe the results related to our main objective, i.e., the bias effects on pointing-technique comparison.

We did not use *TP* because it is inappropriate for Bubble Cursor. In Chapuis et al.'s study, the *MT*s and *ER*s of Bubble were smaller than those for Point under all target conditions, but the *TP* of Bubble was worse than that of Point [15]. Because Bubble Cursor enables users to click when the cursor center is away from the target, the endpoint distribution became wider, which resulted in a worse *TP*. This led to an inappropriate conclusion that a technique with smaller *MT* and *ER* performs worse. This decision not to use *TP* also holds for Experiment 2.

5.1 Movement Time

We analyzed successful (i.e., error-free) *MT*s. Previous studies on Bubble Cursor showed that using the error-free or error-inclusive *MT*s did not change the overall conclusions [15, 32]. Our results indicate that the *MT*s were 0.9734 and 0.9728 sec on average, respectively: a 0.0006-sec difference. This was not sensible in our system (60 fps; a 0.0167-sec loop), thus would not affect our conclusion.

We found significant main effects of Bias ($F_{2,22} = 44.67$, $p < 0.001$, $\eta_p^2 = 0.80$), Cursor ($F_{1,11} = 237.5$, $p < 0.001$, $\eta_p^2 = 0.96$), A ($F_{1,11} = 381.9$, $p < 0.001$, $\eta_p^2 = 0.97$), W ($F_{2,22} = 624.0$, $p < 0.001$, $\eta_p^2 = 0.98$), and EW/W ($F_{2,22} = 211.9$, $p < 0.001$, $\eta_p^2 = 0.95$). The *MT*s for Fast, Neutral, and Accurate were 0.825, 1.001, and 1.094 sec, respectively. Pairwise tests showed significant differences: $p < 0.001$ for (Fast, Neutral) and (Fast, Accurate), and $p < 0.05$ for (Neutral, Accurate). These results indicate that the participants adequately changed their speed-accuracy balances in accordance with the given Bias instructions in terms of operational speed.

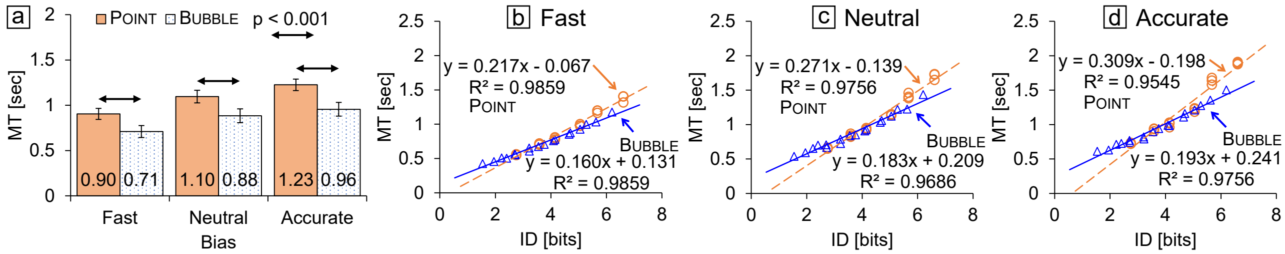

Figure 3: (a) Effects of Cursor on MT, and (b–d) Fitts' law regressions for each Bias.[1] Throughout this paper, error bars indicate 95% CIs.

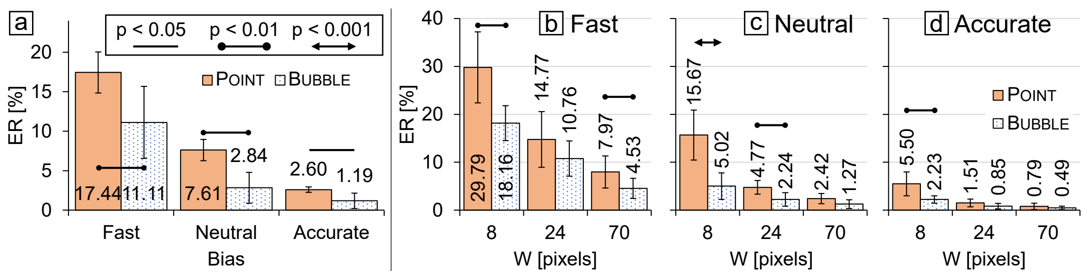

Figure 4: Effects of Cursor on ER (a) for each Bias and (b–d) for each Bias × W.

With respect to our interest on whether Bubble is significantly better than Point throughout the three Biases, we found significant interaction effect of Bias × Cursor ($F_{2,22} = 4.063$, $p < 0.05$, $\eta_p^2 = 0.27$). As shown in Figure 3a, for any Bias, Bubble had significantly shorter MTs than Point ($p < 0.001$ for all pairs).

5.2 Fitts' Law Fitting

To test whether our participants' behaviors followed typical pointing movements in any Bias, we show Fitts' law regressions in Figure 3b–d. Consistent with the original study [32], we used W for Point and EW for Bubble as the target size in Fitts' law. When we separately analyzed the fits for each Cursor, we obtained R^2s above 0.95, which was a bit lower than that in the original study (0.96) [32]. The authors also reported the fit when the two Cursors' data were combined ($R^2 = 0.966$). Thus, they claimed that, regardless of the Cursor, MTs can be predicted using the finally available target sizes. In our data that combined the two Cursors, we obtained $R^2 = 0.9649$, 0.9445, and 0.9201 for Fast, Neutral and Accurate, respectively. The prediction accuracy decreased as the Bias became more accurate.

5.3 Error Rate

We found significant main effects of Bias ($F_{2,22} = 46.73$, $p < 0.001$, $\eta_p^2 = 0.81$), Cursor ($F_{1,11} = 37.22$, $p < 0.001$, $\eta_p^2 = 0.77$), A ($F_{1,11} = 10.03$, $p < 0.01$, $\eta_p^2 = 0.48$), W ($F_{2,22} = 81.87$, $p < 0.001$, $\eta_p^2 = 0.88$), and EW/W ($F_{2,22} = 22.36$, $p < 0.001$, $\eta_p^2 = 0.67$). The ERs for Fast, Neutral, and Accurate were 14.33, 5.231, and 1.893%, respectively. Pairwise tests showed significant differences: $p < 0.001$ for (Fast, Neutral) and (Fast, Accurate), and $p < 0.05$ for (Neutral, Accurate).

We found a significant interaction effect of Bias×Cursor ($F_{2,22} = 7.065$, $p < 0.01$, $\eta_p^2 = 0.39$). As shown in Figure 4a, for any Bias, Bubble had significantly lower ERs than Point (at least $p < 0.05$). The ER for the baseline condition (Neutral × Point) was 7.61%, which was higher than that in the original study of 2.98% and typical pointing studies of 4% [56]. However, our result is not problematic because (1) the 4% criterion is arbitrary and ERs should increase for smaller W [30], and (2) the original and our studies used different apparatuses and OS configurations such as cursor speeds.

Figure 4b–d illustrates the validation of our assumption that the differences in ERs between the two Cursors decrease as the Bias becomes more accurate. The interaction of Bias × Cursor × W was significant ($F_{4,44} = 3.982$, $p < 0.01$, $\eta_p^2 = 0.27$). The results supported our assumption that the absolute ER differences decreased. However, there still remains a significantly different pair in Figure 4d. For Accurate, the participants could spend a long time if needed, but they still made significantly more errors when using Point for the smallest target of $W = 8$ pixels.

5.4 Discussion of Experiment 1

Our first hypothesis **H1** was that Bubble outperforms Point in terms of MT and ER under the Neutral condition. This was statistically supported, as shown in Figures 3a and 4a. Thus, we appropriately reproduced the superiority of Bubble Cursor to the baseline as reported in the original study [32].

In contrast, our second hypothesis **H2** that Bubble does not outperform Point under the Fast or Accurate condition was rejected; under both conditions, Bubble exhibited significantly shorter MTs (Figure 3a) and lower ERs (Figure 4a) than Point. Because the MTs

[1] We followed a common procedure to analyze Fitts' law using the mean MT data. However, Gori et al. later pointed out that Fitts' law is about extreme performance; thus, using minimum MT data is also valid [31]. We applied this novel method to our data, and the results are included in the supplementary materials.

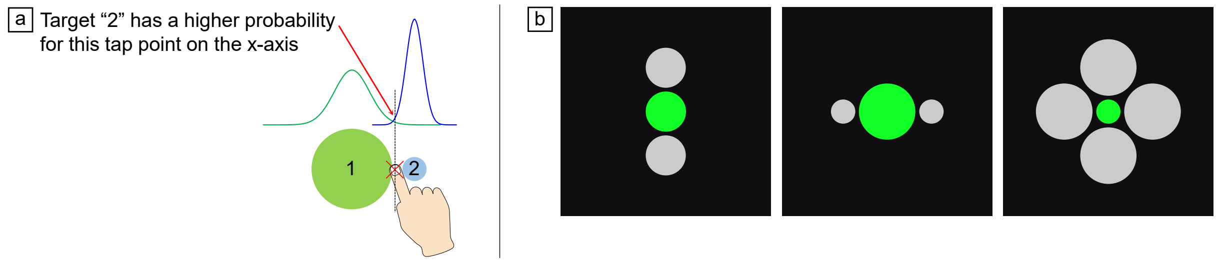

Figure 5: (a) Mechanism of *BTC* in a simplified version with 1D task axis. Tap points form normal distributions for the two targets labeled “1” and “2”, depicted as curved lines above. When a user's tap falls at the red “×” mark position, which is slightly closer to the boundary of target “1”, *BTC* selects the target “2” because its probability for this tap point is higher than that for “1”. Thus, although the geometric distance from the tap point is shorter for “1”, the probabilistic distance is shorter for “2”, which is what is aimed for with *BTC*. (b) Three distractor layouts used in Experiment 2.

and *ER*s significantly changed in accordance with the given Bias with at least $p < 0.05$, we empirically enhanced the effectiveness of Bubble Cursor compared with that reported in the original study [32]. That is, even when users would like to select a target more rapidly or more carefully than usual, Bubble Cursor is beneficial in terms of time and accuracy.

Bubble had significantly lower *ER*s than Point even for the Accurate condition (Figure 4a), but this conclusion would change depending on the Ws and EWs used in an experiment. If we had not used the extremely small target of $W = 8$ pixels, the participants would have selected targets with Point more accurately, and we would not find significant differences between Bubble and Point (Figure 4d). However, to demonstrate the effectiveness of a novel technique, it is fair to design a task so that users may face a problem when using the baseline technique then show that the novel one resolved the issue. Therefore, using a target of $W = 8$ pixels to maintain consistency with the original study's task difficulty was a reasonable choice to examine replicability.

6 EXPERIMENT 2: BAYESIAN TOUCH CRITERION

6.1 Mechanism of Bayesian Touch Criterion

BTC predicts a target that a user tries to select on the basis of the tapped position and touch-point variability parameters [11]. If users repeatedly tap targets with several sizes, the squares of observed variability in terms of the standard deviation (σ) and target size (W) are assumed linearly related for both x- and y-axes on the screen:

$$\sigma_x^2 = \sigma_{ax}^2 + \alpha_x W^2 \quad \text{and} \quad \sigma_y^2 = \sigma_{ay}^2 + \alpha_y W^2, \tag{2}$$

where the intercepts (σ_{ax}^2 and σ_{ay}^2) express the absolute precision of the finger, and the slopes (α_x and α_y) are affected by users' speed-accuracy bias. Hereafter, subscripts x and y indicate the axes on the screen.

Bi and Zhai derived a Bayesian touch distance *BTD* for circular targets as follows.

$$BTD = \frac{1}{2}\left[\frac{(s_x - c_x)^2}{\alpha_x W^2 + \sigma_{ax}^2} + \frac{(s_y - c_y)^2}{\alpha_y W^2 + \sigma_{ay}^2} + \ln\left(\alpha_x W^2 + \sigma_{ax}^2\right) + \ln\left(\alpha_y W^2 + \sigma_{ay}^2\right)\right], \tag{3}$$

where s is touch-point location, c is the target center, and the remaining parameters (σ_{ax}, α_x, σ_{ay}, and α_y) are computed from Equation 2. When a user taps the screen, each target's *BTD* is calculated, then the target having the shortest *BTD* is selected, as illustrated in Figure 5a.

In the original paper [11], the four parameter values in Equation 2 were obtained from the data produced from 18 participants, and different 18 participants joined another experiment to evaluate Equation 3. Because *BTC* predicted the intended target more accurately than the other candidate prediction techniques, Bi and Zhai claimed the external validity of the four parameter values, which can be generally used for other participant groups.

BTC can be extended by applying the prior probability, e.g., a certain target is selected more often than others [79]. This is a special case in which developers have released an app and the frequencies of selection for each button is known. In our experiment, however, we used the most conservative case having no prior probabilities for every target (i.e., all are equally probable), which is the same condition as in the original paper [11].

6.2 Participants

Eighteen participants (same as in [11]) joined this study; 11 men, seven women; ages: $M = 26.9$, $SD = 11.3$ years. All were right-handed and use touch device daily.

6.3 Apparatus

We used an iPad Pro (Gen 2); 12.9 inches, 2732 × 2048 pixels; 264 pixel-per-inch resolution. The experimental system was developed with JavaScript, and the web page was viewed using Safari browser. The refresh rate was set to 120 Hz. The device was put flat on a table in portrait orientation.

6.4 Task

In the beginning of a *session*, the participants tapped a 6-mm-diameter green start target shown at the screen center. The first target and gray distractors then appeared, and tapping the target revealed the next set of target and distractors.

If a tap fell inside the target, it turned red for a moment, while tapping a distractor made it yellow as visual feedback. Tapping the empty area gave no feedback. Judgment of tapping the target, distractor, or empty area depended on their visual boundaries. We used the coordinate of Touch-Up event as the tap position [11]. When tapping outside the target, the participants had to immediately re-aim for it until success.

6.5 Design and Procedure

This study had $3 \times 3 \times 3 \times 3$ conditions. We used three Biases: Fast, Neutral, and Accurate. The three Ws, three distractor widths, and three distractor layouts were the same as in the original study [11]. Both Ws and distractor widths were 3, 5, and 7 mm, which were independently chosen. Three distractor layouts were (1) top and bottom of the target, (2) left and right, and (3) these four positions (see Figure 5b). The gap between the target edge and the distractor edge was 0.5 mm.

To check whether the participants changed their speed-accuracy balance depending on the given Bias, we ensured that the distance from the previous target to the current one was fixed to 20 mm, which enabled us to compare the *MT*s for all Biases while the angle was randomized. The target and any distractor was at least 0.2 mm from the screen edges.

For each Bias, the participants performed $3_W \times 3_{\text{distractor widths}} \times 3_{\text{distractor layouts}} \times 20_{\text{repetitions}} = 540$ successful taps. A session consisted of two repetitions; thus, they accomplished ten sessions for each Bias. They could take a short break between sessions. The first session (i.e., two repetitions) for each Bias was discarded as practice, and data from the remaining nine sessions were analyzed. This experiment involved $3_W \times 3_{\text{distractor widths}} \times 3_{\text{distractor layouts}} \times 18_{\text{repetitions}} \times 3_{\text{Bias}} \times 18_{\text{participants}} = 26{,}244$ data-collection trials. This experiment took approximately 35 minutes for each participant.

6.6 Target Selection Criteria

We used the following four criteria, which are the same as in the original paper [11].

- Visual Boundary (*VB*): A circle is selected if and only if a tap falls inside it. This is the baseline criterion used in common user interfaces.
- Visual Boundary or Shortest Distance to Circle Boundary (*VB*/*SDB*): It applies the *VB* rule first. If a tap falls on the empty space, the selected target is the circle with the shortest distance from its boundary to the tap point.
- Visual Boundary or Shortest Distance to Circle Center (*VB*/*SDC*): It applies the *VB* rule first. If a tap falls on the empty space, the selected target is the circle with the shortest distance from its center to the tap point.
- Bayesian Touch Criterion (*BTC*): A circle with the shortest *BTD* is selected.

In the original study [11], Bi and Zhai measured the four parameter values in Equation 2 with the thumb and index finger separately then used the parameter values after merging the two finger-type data, called *generic parameters*. They confirmed that whether using the parameter values of each finger or the generic parameters did not affect the target-prediction accuracy ($p \approx 1$). Following this procedure, we compared the *ER* difference when using their generic and index-finger parameters, which we call *BTC*/*Gen* and *BTC*/*Idx*. We also examined if using the parameter values computed from our data for each Bias would improve prediction accuracy. We call this *BTC* using bias-specific parameters as *BTC*/*Spc*.

7 RESULTS OF EXPERIMENT 2

We recorded 32,129 taps including successes and failures. Outlier trials were removed if the tap position was ≥15 mm from the target center [11]. We detected 137 outliers (0.43%). According to the experimenters' observation, such outliers occurred mainly because the participants' little finger accidentally contacted the screen.

7.1 Movement Time

Although *MT* was not a main objective in this experiment, the result supported that the participants appropriately tried following the given Bias. The mean *MT*s for the first tap in each trial for Fast, Neutral and Accurate were 566, 651, and 764 ms, respectively. An RM-ANOVA showed the main effect of Bias ($F_{2,34} = 35.42$, $p < 0.001$, $\eta_p^2 = 0.68$), and pairwise tests showed significant differences between Fast and Neutral ($p < 0.001$), Fast and Accurate ($p < 0.001$), and Neutral and Accurate ($p < 0.01$).

This finding holds even if we use *MT* for every tap, i.e., we analyzed all *MT*s for the first and subsequent taps if errors occurred. The mean *MT*s for Fast, Neutral and Accurate were 549, 635, and 755 ms, respectively. An RM-ANOVA showed the main effect of Bias ($F_{2,34} = 42.18$, $p < 0.001$, $\eta_p^2 = 0.71$), and pairwise tests showed $p < 0.001$ for all Bias pairs. Thus, the participants appropriately followed the given Bias and changed their tapping speed.

7.2 Regressions of Tap-point Variability

Regarding the *BTC*/*Spc* criterion, we obtained the following parameter values for Equation 2.

$$\text{Fast}: \ \sigma_x^2 = 1.351 + 0.009001W^2 \quad (R^2 = 0.9516) \tag{4}$$
$$\sigma_y^2 = 1.252 + 0.01200W^2 \quad (R^2 = 0.9502) \tag{5}$$
$$\text{Neutral}: \ \sigma_x^2 = 0.8936 + 0.009057W^2 \quad (R^2 = 0.9986) \tag{6}$$
$$\sigma_y^2 = 0.7990 + 0.01004W^2 \quad (R^2 = 0.9999) \tag{7}$$
$$\text{Accurate}: \ \sigma_x^2 = 0.5672 + 0.01109W^2 \quad (R^2 = 0.9978) \tag{8}$$
$$\sigma_y^2 = 0.5814 + 0.009249W^2 \quad (R^2 = 0.9753) \tag{9}$$

The number of data points was three (i.e., $W = 3$, 5, and 7 mm) for each regression, and other independent variables were merged. For the *BTC*/*Gen* and *BTC*/*Idx* criteria, we used the following parameter values, as reported in the original study [11][2].

$$BTC/Gen: \ \sigma_x^2 = 1.680 + 0.0075W^2, \quad \sigma_y^2 = 1.329 + 0.0108W^2 \tag{10}$$
$$BTC/Idx: \ \sigma_x^2 = 1.540 + 0.0075W^2, \quad \sigma_y^2 = 1.250 + 0.0104W^2 \tag{11}$$

7.3 Error Rate

Figure 6 shows the *ER*s for all Biases. In the original paper, the *ER* for the baseline *VB* was 19.6% [11], which was higher than the *ER* under our Neutral condition (14.29%) and slightly less than that of Fast (19.83%). This indicates that our participants were relatively more accurate when instructed to operate as rapidly and accurately as possible.

An RM-ANOVA showed that *ER* was significantly affected by the selection criteria ($F_{5,85} = 305.3$, $p < 0.001$, $\eta_p^2 = 0.95$) and Bias ($F_{2,34} = 52.86$, $p < 0.001$, $\eta_p^2 = 0.76$). The *ER*s for Fast, Neutral, and

[2] We used more precise values with a larger number of digits reported in Bi and Zhai's other paper [12], which referred to [11].

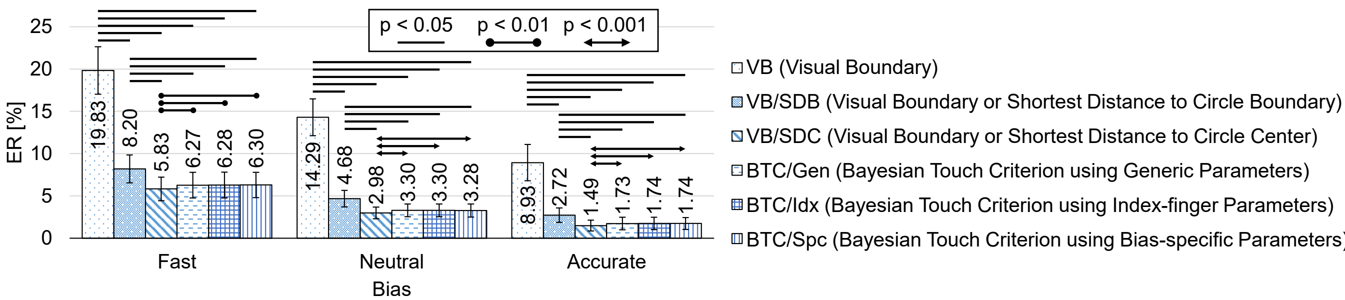

Figure 6: *ER*s per selection criterion for each Bias.[3]

Accurate were 8.785, 5.304, and 3.061%, respectively, and we found $p < 0.001$ for all pairs. This indicates that, in addition to the *MT* results, the participants appropriately changed their speed-accuracy balance in accordance with the given Bias.

The interaction of selection criteria × Bias was significant ($F_{10,170} = 45.82$, $p < 0.001$, $\eta_p^2 = 0.73$). As shown in Figure 6, the results of pairwise tests indicate the following points.

- Using *VB/SDC* was the best throughout the three Biases. All pairwise tests to compare *VB/SDC* and the other five criteria showed significant differences with $p < 0.05$ at least.
- For the three variations of *BTC*, we found no significant differences. This was consistent with Bi and Zhai's report that using generic or each-finger parameter values did not affect prediction accuracy. We also found that using the parameter values computed for each Bias (*BTC/Spc*) did not improve prediction accuracy over the other *BTC* variants.

7.4 Discussion of Experiment 2

Our first hypothesis **H1** was that *BTC* outperforms the other selection-criterion candidates in terms of *ER* under the Neutral condition. This was statistically rejected, as *VB/SDC* exhibited the best prediction accuracy (Figure 6). Therefore, we could not reproduce the superiority of *BTC* to the baselines contrary to the report in the original study [11]. Moreover, under any Bias condition, our conclusion was that using *VB/SDC* was the best, which supported our second hypothesis **H2** that *BTC* does not outperform baselines under the Fast or Accurate condition. The results were not affected by the choice of four parameter values in Equation 2; the *ER* difference fell within 0.03 points depending on the use of *BTC/Gen*, *BTC/Idx*, or *BTC/Spc*.

If we had used only the Neutral condition, the conclusion that *BTC* did not achieve the best prediction accuracy might be simply considered a failure of replication because, for example, our participants were unintentionally biased towards speed or accuracy. However, our results rejected an assumption that *VB/SDC* achieved the best performance by chance; rather, *VB/SDC* is helpful in selecting desired targets regardless of the users being careful or in a hurry. Our empirical results are thus more insightful than a simple replication study with a single subjective bias.

Because there is always a chance of reaching a different result on statistical significance tests when conducting ANOVA, it is possible to observe such a reversal in which a state-of-the-art technique underperforms other simple ones. Also, our result that *BTC* was not the best is likely due to the condition differences between the original study and ours, including apparatuses or participant groups. However, a pointing-facilitation technique is hopefully effective even if these conditions change, as we observed in Experiment 1; otherwise, the generalizability of the technique considerably degrades.

8 GENERAL DISCUSSION

8.1 Methodology of Varying Subjective Speed-accuracy Biases

The results indicate that Bubble Cursor was effective in reducing both *MT* and *ER* under all three Bias conditions (Experiment 1), while we could not confirm the superiority of *BTC* to *VB/SDC* (Experiment 2). Our hypothesis was that Bubble Cursor and *BTC* would be significantly better for Neutral in accordance with the original studies while their effectiveness would degrade under other Bias conditions, particularly for Accurate because too low *ER*s would be observed. This hypothesis was, however, ultimately rejected in both experiments.

The results of Experiment 1 indicate that, when researchers propose a pointing-facilitation technique and evaluate it, they can claim that the technique is effective in reducing *MT* and *ER* even if users are biased towards speed or accuracy. In addition to the factors examined in previous studies such as ensuring a wide range of task difficulty [56], varying subjective biases is another helpful methodology when researchers would like to validate the effectiveness of their new technique.

In Experiment 2, *BTC* underperformed *VB/SDC* under all Bias conditions. The *ER* observed under the Neutral condition was lower than that in the original study (14.29 vs. 19.6%), but that under the Fast condition was higher (19.83%). Thus, using the three Biases enabled us to sufficiently encompass the *ER* observed in the original study. If we had replicated Experiment 2 with only the Neutral condition, our conclusion that *BTC* is not the best would be due to the fault in instruction; our participants were biased towards accuracy more than in the original study. However, testing three Biases enabled us to reject such a concern.

Varying speed-accuracy biases has a history in target-pointing research [27, 33, 48, 75], but the objective was mainly on the invariance of throughput. In contrast, we empirically demonstrated

[3] We observed several significantly different pairs although their 95% CIs overlap, which is statistically acceptable [19, 25].

that using multiple subjective biases enables researchers to state additional claims, such as consistent effectiveness under different speed-accuracy biases that users unintentionally have or the robustness of a finding that a previously proposed technique has a limitation that has never been pointed out. This methodology is easy for researchers to introduce in their user studies, because no additional development for their experimental systems is needed; just repeating the same experiment under multiple biases with oral instruction suffices.

8.2 Implications for Future Research

Varying subjective speed-accuracy biases will work well for other comparison studies on novel pointing-facilitation techniques, including target selection for smartwatches [79] and 3D pointing tasks in virtual reality [8]. Tasks outside the target pointing are also promising applications, such as evaluating novel techniques for goal-crossing [38] and path-steering tasks [3].

Given that *ER*s change depending on the Bias, evaluating the generalizability of *ER* prediction models [12, 37, 52] is another possible scope. Although we mentioned that the choice of 4% *ER* is arbitrary (Section 5.3), some models presuppose that participants exhibit a 4% *ER* and that the spread of endpoints follow a normal distribution under the Neutral condition (e.g., [60]). It is currently unclear if endpoints are normally distributed when the bias shifts towards speed or accuracy, which is worth investigating in the future.

While we did not focus on the difference in devices in Experiment 2 (smartphone in the original study vs. tablet in ours), we can interpret this as an evaluation of the effectiveness of a previously proposed technique with a new device. A similar approach has been extensively evaluated, e.g., evaluating Bubble Cursor with 3D input devices [46, 58] and eye trackers [16]. Because the speed-accuracy balances change depending on the device even for the same participant group performing the same task [14, 28, 53], if a previously proposed technique was not effective for a new device with only a single Bias (typically Neutral), the result does not necessarily mean the ineffectiveness of the technique. It might be due to the participants' unintentional bias, but testing multiple biases partially decreases this concern.

On the basis of the discussions thus far, we list the following uses of varying speed-accuracy biases for future research.

- **Evaluating a novel interaction technique**:
  When researchers propose a novel interaction technique, testing multiple subjective biases can strengthen the claim of effectiveness over baselines. It also enables a thorough evaluation and can reveal more insightful findings than using a single bias, such as understanding when users do not gain the benefit from the technique (e.g., when they slowly operate the interface).
- **Replicating a user experiment to evaluate a previously proposed technique**:
  When researchers conduct a replication study to confirm whether a *good* technique would be superior to baselines, testing multiple biases can more strongly support the finding, regardless of the result supporting the goodness of the technique (as in our Experiment 1) or lack of goodness over a baseline (Experiment 2).
- **Validating the applicability of a technique to other devices**:
  As shown in Experiment 2, a previously proposed technique may not be effective in certain devices that are different from those in the original study. In this case, researchers cannot distinguish the reason behind the lack of goodness if using only one instruction (typically Neutral); it is probably due to the differences in participant group, their skills for the newly tested device, their unintentional speed-accuracy bias (e.g., they are accidentally biased towards accuracy), and so on. Examining multiple biases can reduce the possibility that the ineffectiveness comes from the unintentional bias and supports the generalizability of their new finding on the limited applicability of the technique.

8.3 Limitations and Future Work

We confirmed that *MT* and *ER* were significantly affected by Bias, but it is possible that more intense instructions could change the results. For example, under the *extremely accurate* condition in which participants try to avoid any error [75], we might not observe a significant positive effect of Bubble Cursor on *MT* or *ER*.

This assumption is partially suggested from the results in Experiment 1; we found significant interaction of Bias × Cursor × W on *ER* ($p < 0.01$, $\eta_p^2 = 0.27$). As the bias became more accurate and as W increased, the differences between Point and Bubble disappeared (see Figure 4d). If this tendency holds for instructions asking for more accuracy, we may reach a conclusion that Bubble Cursor is not effective for a certain bias particularly when users are careful, which is inconsistent with our current conclusion. Such a conclusion, however, still supports our main recommendation that researchers obtain insights from varying speed-accuracy biases.

We are concerned about the participants' interpretations of the balances (weights) for each bias. While we confirmed that our oral instructions adequately changed their biases with large effect sizes, another possible approach is to give quantitative biases externally such as monetary incentives and penalties for speed and accuracy [21, 27] and time penalties for error clicks [7, 65]. Multiple time limits has also been used to vary speed-accuracy biases [52, 54, 60, 72, 77]. Comparing subjective vs. objective biases is included in our future work.

Another concern is that the participants' bias might shift towards accuracy immediately after an operation error occurs [20]. To omit such an effect, one possible strategy is to discard several trials after an error from data analysis, as well as increasing the same number of trials to fairly compare the *ER*s. In our study, however, we did not adopt this approach to remain consistent with the original studies, i.e., the participants re-aimed for the current target again and we did not discard any data after an error selection.

Finally, throughout this paper, our viewpoint was from researchers who want to conduct thorough and informative user experiments. However, as a negative perspective, varying speed-accuracy biases directly lengthens the duration of experiments with more bias levels. This would increase the fatigue of participants and affect their performance. Also, if researchers have to test more task factors (e.g., multiple distractor densities in Experiment 1), the priority of using multiple biases might be low; whether varying biases depends on the main objective of the research.

9 CONCLUSION

We investigated the effect of varying subjective speed-accuracy biases to evaluate two previously proposed pointing-facilitation techniques, Bubble Cursor and Bayesian Touch Criterion *BTC*. Bubble Cursor reduced *MT* and *ER* compared with the baseline pointing technique under all bias conditions (confirming the results of the original study), while *BTC* underperformed a simple target-prediction method inconsistently compared with the original study. We demonstrated that testing multiple bias conditions enabled us to discuss the effectiveness of novel techniques in more detail than using only a single (Neutral) instruction. This methodology has been used to evaluate throughput metrics, but can enrich researchers' user studies on evaluating novel interaction techniques, replications of previously proposed techniques, and validations of the applicability of techniques on future devices.

ACKNOWLEDGMENTS

We thank the reviewers of CHI 2023 for their valuable feedback.

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* (Atlanta, Georgia, USA) *(CHI '97)*. Association for Computing Machinery, New York, NY, USA, 295–302. https://doi.org/10.1145/258549.258760

[2] Johnny Accot and Shumin Zhai. 2002. More than dotting the i's — foundations for crossing-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 73–80. https://doi.org/10.1145/503376.503390

[3] David Ahlström. 2005. Modeling and Improving Selection in Cascading Pull-down Menus Using Fitts' Law, the Steering Law and Force Fields. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 61–70. https://doi.org/10.1145/1054972.1054982

[4] Georg Apitz and François Guimbretière. 2004. CrossY: A Crossing-based Drawing Application. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology* (Santa Fe, NM, USA) *(UIST '04)*. ACM, New York, NY, USA, 3–12. https://doi.org/10.1145/1029632.1029635

[5] Takeshi Asano, Ehud Sharlin, Yoshifumi Kitamura, Kazuki Takashima, and Fumio Kishino. 2005. Predictive Interaction Using the Delphian Desktop. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology* (Seattle, WA, USA) *(UIST '05)*. Association for Computing Machinery, New York, NY, USA, 133–141. https://doi.org/10.1145/1095034.1095058

[6] Oscar Kin-Chung Au, Xiaojun Su, and Rynson W.H. Lau. 2014. LinearDragger: A Linear Selector for One-finger Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 2607–2616. https://doi.org/10.1145/2556288.2557096

[7] Nikola Banovic, Tovi Grossman, and George Fitzmaurice. 2013. The Effect of Time-based Cost of Error in Target-directed Pointing Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 1373–1382. https://doi.org/10.1145/2470654.2466181

[8] Anil Ufuk Batmaz, Mayra Donaji Barrera Machuca, Junwei Sun, and Wolfgang Stuerzlinger. 2022. The Effect of the Vergence-Accommodation Conflict on Virtual Hand Pointing in Immersive Displays. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 633, 15 pages. https://doi.org/10.1145/3491102.3502067

[9] Patrick Baudisch, Ed Cutrell, Daniel Robbins, Mary Czerwinski, Peter Tandler, Benjamin Bederson, and Alex Zierlinger. 2003. Drag-and-Pop and Drag-and-Pick: Techniques for Accessing Remote Screen Content on Touch- and Pen-Operated Systems. In *Human-Computer Interaction–INTERACT '03* (human-computer interaction–interact '03 ed.). IOS Press, Amsterdam, Netherlands, 57–64. https://www.microsoft.com/en-us/research/publication/drag-and-pop-and-drag-and-pick-techniques-for-accessing-remote-screen-content-on-touch-and-pen-operated-systems/

[10] Xiaojun Bi, Yang Li, and Shumin Zhai. 2013. FFitts Law: Modeling Finger Touch with Fitts' Law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 1363–1372. https://doi.org/10.1145/2470654.2466180

[11] Xiaojun Bi and Shumin Zhai. 2013. Bayesian touch: a statistical criterion of target selection with finger touch. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 51–60. https://doi.org/10.1145/2501988.2502058

[12] Xiaojun Bi and Shumin Zhai. 2016. Predicting Finger-Touch Accuracy Based on the Dual Gaussian Distribution Model. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. ACM, New York, NY, USA, 313–319. https://doi.org/10.1145/2984511.2984546

[13] Renaud Blanch and Michael Ortega. 2011. Benchmarking Pointing Techniques with Distractors: Adding a Density Factor to Fitts' Pointing Paradigm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 1629–1638. https://doi.org/10.1145/1978942.1979180

[14] 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. https://doi.org/10.1080/00140137808931762

[15] Olivier Chapuis, Jean-Baptiste Labrune, and Emmanuel Pietriga. 2009. DynaSpot: Speed-dependent Area Cursor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. ACM, New York, NY, USA, 1391–1400. https://doi.org/10.1145/1518701.1518911

[16] Myungguen Choi, Daisuke Sakamoto, and Tetsuo Ono. 2020. Bubble Gaze Cursor + Bubble Gaze Lens: Applying Area Cursor Technique to Eye-Gaze Interface. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 11, 10 pages. https://doi.org/10.1145/3379155.3391322

[17] Andy Cockburn, Pierre Dragicevic, Lonni Besançon, and Carl Gutwin. 2020. Threats of a Replication Crisis in Empirical Computer Science. *Commun. ACM* 63, 8 (July 2020), 70–79. https://doi.org/10.1145/3360311

[18] Edward R.F.W. Crossman. 1956. *The measurement of perceptual load in manual operations.* Ph. D. Dissertation. University of Birmingham.

[19] Geoff Cumming, Fiona Fidler, and David L. Vaux. 2007. Error bars in experimental biology . *Journal of Cell Biology* 177, 1 (04 2007), 7–11. https://doi.org/10.1083/jcb.200611141

[20] Claudia Danielmeier and Markus Ullsperger. 2011. Post-Error Adjustments. *Frontiers in Psychology* 2 (2011), 10 pages. https://doi.org/10.3389/fpsyg.2011.00233

[21] Mark Dean, Shih-Wei Wu, and Laurence T. Maloney. 2007. Trading off speed and accuracy in rapid, goal-directed movements. *Journal of Vision* 7, 5 (07 2007), 10–10. https://doi.org/10.1167/7.5.10

[22] Jack Tigh Dennerlein, David B. Martin, and Christopher Hasser. 2000. Force-feedback Improves Performance for Steering and Combined Steering-targeting Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. ACM, New York, NY, USA, 423–429. https://doi.org/10.1145/332040.332469

[23] Peter Dixon. 2008. Models of accuracy in repeated-measures designs. *Journal of Memory and Language* 59, 4 (2008), 447–456.

[24] Sarah A. Douglas, Arthur E. Kirkpatrick, and I. Scott MacKenzie. 1999. Testing Pointing Device Performance and User Assessment with the ISO 9241, Part 9 Standard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 215–222. https://doi.org/10.1145/302979.303042

[25] Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*. Springer, Gewerbestr, Switzerland, 291 – 330. https://doi.org/10.1007/978-3-319-26633-6_13

[26] 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), 381–391. https://doi.org/10.1037/h0055392

[27] P. M. Fitts and B. K. Radford. 1966. Information capacity of discrete motor responses under different cognitive sets. *Journal of experimental psychology* 71, 4 (1966), 475–482.

[28] Clifton Forlines and Ravin Balakrishnan. 2008. Evaluating Tactile Feedback and Direct vs. Indirect Stylus Input in Pointing and Crossing Selection Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 1563–1572. https://doi.org/10.1145/1357054.1357299

[29] Kentaro Go and Yuki Endo. 2007. CATKey: Customizable and Adaptable Touchscreen Keyboard with Bubble Cursor-like Visual Feedback. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-Computer Interaction* (Rio de Janeiro, Brazil) *(INTERACT'07)*. Springer-Verlag, Berlin, Heidelberg, 493–496.

[30] Julien Gori, Olivier Rioul, and Yves Guiard. 2018. Speed-Accuracy Tradeoff: A Formal Information-Theoretic Transmission Scheme (FITTS). *ACM Trans.*

Comput.-Hum. Interact. 25, 5, Article 27 (Sept. 2018), 33 pages. https://doi.org/10.1145/3231595

[31] Julien Gori, Olivier Rioul, Yves Guiard, and Michel Beaudouin-Lafon. 2017. One Fitts' Law, Two Metrics. In *Human-Computer Interaction – INTERACT 2017*. Springer International Publishing, Cham, 525–533.

[32] Tovi Grossman and Ravin Balakrishnan. 2005. The Bubble Cursor: Enhancing Target Acquisition by Dynamic Resizing of the Cursor's Activation Area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 281–290. https://doi.org/10.1145/1054972.1055012

[33] Yves Guiard, Halla B. Olafsdottir, and Simon T. Perrault. 2011. Fitts' Law as an Explicit Time/Error Trade-Off. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1619–1628. https://doi.org/10.1145/1978942.1979179

[34] Maxime Guillon, François Leitner, and Laurence Nigay. 2014. Static Voronoi-Based Target Expansion Technique for Distant Pointing. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (Como, Italy) *(AVI '14)*. Association for Computing Machinery, New York, NY, USA, 41–48. https://doi.org/10.1145/2598153.2598178

[35] Morten Hertzum and Kasper Hornbæk. 2007. Input techniques that dynamically change their cursor activation area: A comparison of bubble and cell cursors. *International Journal of Human-Computer Studies* 65, 10 (2007), 833–851. https://doi.org/10.1016/j.ijhcs.2007.05.001

[36] Kasper Hornbæk, Søren S. Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is Once Enough? On the Extent and Content of Replications in Human-Computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3523–3532. https://doi.org/10.1145/2556288.2557004

[37] Jin Huang, Feng Tian, Xiangmin Fan, Xiaolong (Luke) Zhang, and Shumin Zhai. 2018. Understanding the Uncertainty in 1D Unidirectional Moving Target Selection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article 237, 12 pages. https://doi.org/10.1145/3173574.3173811

[38] Susu Huang, Daqing Qi, Jiabin Yuan, and Huawei Tu. 2019. Review of studies on target acquisition in virtual reality based on the crossing paradigm. *Virtual Reality & Intelligent Hardware* 1, 3 (2019), 251–264. https://doi.org/10.3724/SP.J.2096-5796.2019.0006

[39] ISO. 2000. ISO 9241-9. International standard: ergonomic requirements for office work with visual display terminals (VDTs)–Part 9: requirements for non-keyboard input devices, International Organization for Standardization.

[40] Paul Kabbash and William A. S. Buxton. 1995. The "Prince" Technique: Fitts' Law and Selection Using Area Cursors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 273–279. https://doi.org/10.1145/223904.223939

[41] Masatomo Kobayashi and Takeo Igarashi. 2008. Ninja Cursors: Using Multiple Cursors to Assist Target Acquisition on Large Screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 949–958. https://doi.org/10.1145/1357054.1357201

[42] Steven Komarov, Katharina Reinecke, and Krzysztof Z. Gajos. 2013. Crowdsourcing Performance Evaluations of User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 207–216. https://doi.org/10.1145/2470654.2470684

[43] Sergey Kulikov, I. Scott MacKenzie, and Wolfgang Stuerzlinger. 2005. Measuring the Effective Parameters of Steering Motions. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. ACM, New York, NY, USA, 1569–1572. https://doi.org/10.1145/1056808.1056968

[44] Joona Laukkanen, Poika Isokoski, and Kari-Jouko Räihä. 2008. The Cone and the Lazy Bubble: Two Efficient Alternatives between the Point Cursor and the Bubble Cursor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 309–312. https://doi.org/10.1145/1357054.1357107

[45] Nianlong Li, Feng Tian, Jin Huang, Xiangmin Fan, and Hongan Wang. 2018. 2D-BayesPointer: An Implicit Moving Target Selection Technique Enabled by Human Performance Modeling. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188520

[46] Yiqin Lu, Chun Yu, and Yuanchun Shi. 2020. Investigating Bubble Mechanism for Ray-Casting to Improve 3D Target Acquisition in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, USA, 35–43. https://doi.org/10.1109/VR46266.2020.00021

[47] I. Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction* 7, 1 (1992), 91–139. https://doi.org/10.1207/s15327051hci0701_3

[48] 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* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 1633–1636. https://doi.org/10.1145/1357054.1357308

[49] Michael McGuffin and Ravin Balakrishnan. 2002. Acquisition of Expanding Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI '02)*. Association for Computing Machinery, New York, NY, USA, 57–64. https://doi.org/10.1145/503376.503388

[50] Blanca Mena, M José, Rafael Alarcón, Jaume Arnau Gras, Roser Bono Cabré, and Rebecca Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema* 29, 4 (2017), 552–557.

[51] Halla B. Olafsdottir, Yves Guiard, Olivier Rioul, and Simon T. Perrault. 2012. A New Test of Throughput Invariance in Fitts' Law: Role of the Intercept and of Jensen's Inequality. In *Proceedings of BCS-HCI '12* (Birmingham, United Kingdom) *(BCS-HCI '12)*. BCS Learning & Development Ltd., Swindon, GBR, 119–126.

[52] Eunji Park and Byungjoo Lee. 2020. *An Intermittent Click Planning Model*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376725

[53] Vijay Rajanna and Tracy Hammond. 2018. A Fitts' Law Evaluation of Gaze Input on Large Displays Compared to Touch and Mouse Inputs. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 8, 5 pages. https://doi.org/10.1145/3206343.3206348

[54] Richard A. Schmidt, Howard Zelaznik, Brian Hawkins, James S. Frank, and Jr John T. Quinn. 1979. Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review* 86, 5 (1979), 415–451. https://doi.org/10.1037/0033-295X.86.5.415

[55] Ather Sharif, Victoria Pao, Katharina Reinecke, and Jacob O. Wobbrock. 2020. The Reliability of Fitts's Law as a Movement Model for People with and without Limited Fine Motor Function. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 16, 15 pages. https://doi.org/10.1145/3373625.3416999

[56] 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. https://doi.org/10.1016/j.ijhcs.2004.09.001

[57] Edward Tse, Mark Hancock, and Saul Greenberg. 2007. Speech-Filtered Bubble Ray: Improving Target Acquisition on Display Walls. In *Proceedings of ICMI 2017* (Nagoya, Aichi, Japan) *(ICMI '07)*. Association for Computing Machinery, New York, NY, USA, 307–314. https://doi.org/10.1145/1322192.1322245

[58] Lode Vanacken, Tovi Grossman, and Karin Coninx. 2007. Exploring the Effects of Environment Density and Target Visibility on Object Selection in 3D Virtual Environments. In *2007 IEEE Symposium on 3D User Interfaces*. IEEE, USA. https://doi.org/10.1109/3DUI.2007.340783

[59] Daniel Vogel and Patrick Baudisch. 2007. Shift: A Technique for Operating Pen-based Interfaces Using Touch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. ACM, New York, NY, USA, 657–666. https://doi.org/10.1145/1240624.1240727

[60] Jacob O. Wobbrock, Edward Cutrell, Susumu Harada, and I. Scott MacKenzie. 2008. An Error Model for Pointing Based on Fitts' Law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 1613–1622. https://doi.org/10.1145/1357054.1357306

[61] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[62] Jacob O. Wobbrock, Alex Jansen, and Kristen Shinohara. 2011. Modeling and Predicting Pointing Errors in Two Dimensions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 1653–1656. https://doi.org/10.1145/1978942.1979183

[63] 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* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 1639–1648. https://doi.org/10.1145/1978942.1979181

[64] Aileen Worden, Nef Walker, Krishna Bharat, and Scott Hudson. 1997. Making Computers Easier for Older Adults to Use: Area Cursors and Sticky Icons. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '97)*. Association for Computing Machinery, New York, NY, USA, 266–271. https://doi.org/10.1145/258549.258724

[65] Shota Yamanaka. 2018. Effect of Gaps with Penal Distractors Imposing Time Penalty in Touch-pointing Tasks. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.

1145/3229434.3229435
[66] Shota Yamanaka. 2019. Steering Performance with Error-accepting Delays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 570, 9 pages. https://doi.org/10.1145/3290605.3300800
[67] Shota Yamanaka. 2022. Test-Retest Reliability on Movement Times and Error Rates in Target Pointing. In *Designing Interactive Systems Conference* (Virtual Event, Australia) *(DIS '22)*. Association for Computing Machinery, New York, NY, USA, 178–188. https://doi.org/10.1145/3532106.3533450
[68] Shota Yamanaka, Hiroaki Shimono, and Homei Miyashita. 2019. Towards More Practical Spacing for Smartphone Touch GUI Objects Accompanied by Distractors. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces* (Daejeon, Republic of Korea) *(ISS '19)*. Association for Computing Machinery, New York, NY, USA, 157–169. https://doi.org/10.1145/3343055.3359698
[69] Shota Yamanaka, Keisuke Yokota, and Takanori Komatsu. 2021. Time-Penalty Impact on Effective Index of Difficulty and Throughputs in Pointing Tasks. In *Human-Computer Interaction – INTERACT 2021*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer International Publishing, Cham, 100–121.
[70] Koji Yatani, Kurt Partridge, Marshall Bern, and Mark W. Newman. 2008. Escape: A Target Selection Technique Using Visually-cued Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 285–294. https://doi.org/10.1145/1357054.1357104
[71] Chuang-Wen You, Yung-Huan Hsieh, Wen-Huang Cheng, and Yi-Hsuan Hsieh. 2014. AttachedShock: Design of a crossing-based target selection technique on augmented reality devices and its implications. *International Journal of Human-Computer Studies* 72, 7 (2014), 606–626. https://doi.org/10.1016/j.ijhcs.2014.03.001
[72] Howard N. Zelaznik, Susan Mone, George P. McCabe, and Christopher Thaman. 1988. Role of temporal and spatial precision in determining the nature of the speed-accuracy trade-off in aimed-hand movements. *Journal of Experimental Psychology: Human Perception and Performance* 14, 2 (1988), 221–230. https://doi.org/10.1037/0096-1523.14.2.221
[73] Shumin Zhai, Stéphane Conversy, Michel Beaudouin-Lafon, and Yves Guiard. 2003. Human On-Line Response to Target Expansion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 177–184. https://doi.org/10.1145/642611.642644
[74] Shumin Zhai, Stéphane Conversy, Michel Beaudouin-Lafon, and Yves Guiard. 2003. Human On-Line Response to Target Expansion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 177–184. https://doi.org/10.1145/642611.642644
[75] 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. https://doi.org/10.1016/j.ijhcs.2004.09.007
[76] Mingrui Ray Zhang, Shumin Zhai, and Jacob O. Wobbrock. 2019. Text Entry Throughput: Towards Unifying Speed and Accuracy in a Single Performance Metric. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300866
[77] Xiaolei Zhou, Xiang Cao, and Xiangshi Ren. 2009. Speed-Accuracy Tradeoff in Trajectory-Based Tasks with Temporal Constraint. In *Human-Computer Interaction – INTERACT 2009*. Springer Berlin Heidelberg, Berlin, Heidelberg, 906–919.
[78] Xiaolei Zhou and Xiangshi Ren. 2010. An investigation of subjective operational biases in steering tasks evaluation. *Behaviour & Information Technology* 29, 2 (2010), 125–135. https://doi.org/10.1080/01449290701773701
[79] Suwen Zhu, Yoonsang Kim, Jingjie Zheng, Jennifer Yi Luo, Ryan Qin, Liuping Wang, Xiangmin Fan, Feng Tian, and Xiaojun Bi. 2020. *Using Bayes' Theorem for Command Input: Principle, Models, and Applications.* Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376771