Statistical dances

Why no statistical analysis is reliable and what to do about it

biovis 2016  Pierre Dragicevic, Inria, France
User studies

Image and video from (Le Goc et al, 2016)
Traditional statistics

In Study 2, the average search time was 7.71 seconds with the contrast slider as opposed to 4.72 seconds for the Color Lens, a 43% improvement for our new technique (significant, $t = 5.605, p < .001$).
Traditional statistics

“[significance testing] is based upon a fundamental misunderstanding of the nature of rational inference, and is seldom if ever appropriate to the aims of scientific research.”

(Rozeboom, 1960)

All references for this talk at www.aviz.fr/dances
Bad Stats are Miscommunicated Stats

Pierre Dragicevic, INRIA

BELIV 2014


www.aviz.fr/badstats
Bad Stats are
Miscommunicated Stats

Pierre Dragicevic, INRIA

Chapter 13
Fair Statistical Communication in HCI

Pierre Dragicevic

Abstract  Statistics are tools to help end users accomplish their task. In research, to be qualified as usable, statistical tools should help researchers advance scientific knowledge by supporting and promoting the effective communication of research findings. Yet areas such as human-computer interaction (HCI) have adopted tools—i.e., $p$-values and dichotomous testing procedures—that have proven to be poor at supporting these tasks. The abusive use of these procedures has been severely
It’s all about uncertainty

“Statistics has been described as the science of uncertainty. But, paradoxically, statistical methods are often used to create a sense of certainty where none should exist.”

(Gelman, 2016)

All references for this talk at www.aviz.fr/dances
Statistical dances

The dance of $p$-values (Cumming, 2009)
All references for this talk at [www.aviz.fr/dances](http://www.aviz.fr/dances)
Outline

• Statistical dances — why no statistical analysis is reliable
• what do to about it
  • False solutions
  • OK solutions
  • Real solutions
Dance of the sample means

Evaluating a new graph layout technique

Variant 1

Variant 2
Dance of the sample means

Evaluating a new graph layout technique

Variant 1 → Baseline → Variant 2
Dance of the sample means

- Variant 1 Improvement: 11.4 s
- Variant 2 Improvement: 2.1 s
Dance of the sample means

![Graph showing the comparison of improvement times between Variant 1 and Variant 2.]

- Variant 1 Improvement: 11.4 s
- Variant 2 Improvement: 2.1 s
Dance of the sample means
Dance of the sample means

- Variant 1 Improvement
  - 10 s

- Variant 2 Improvement
  - 0 s

- Variant 1 Improvement
  - 11.4 s

- Variant 2 Improvement
  - 2.1 s
Dance of the sample means
Dance of the sample means
Dance of the sample means

[Graph showing the distribution of sample means for Variant 1 and Variant 2 improvements]

- **Variant 1 Improvement**
- **Variant 2 Improvement**

[Histograms showing the time differences between the two variants, with Variant 1 showing an average improvement of 7.0 s and Variant 2 showing a slight decrease of -1.9 s.]
Dance of the sample means

- ** Variant 1 Improvement **
- ** Variant 2 Improvement **

- ** Variant 1 Improvement **: 16.4 s
- ** Variant 2 Improvement **: 8.5 s
Dance of the sample means

![Bar chart showing two variants with improvement times: Variant 1 with 16.4 s and Variant 2 with 8.5 s.]
Dance of the sample means

- Variant 1 Improvement: 16.4 s
- Variant 2 Improvement: 8.5 s
Dance of the sample means
Dance of the sample means
Dance of the sample means

Our data

Don’t trust sample means

0 s  10 s  20 s

Variant 1 Improvement  Variant 2 Improvement

11.4 s  2.1 s
p-values

Our data

Variant 1 Improvement

Variant 2 Improvement

p = 0.0076
11.4 s

p = 0.53
2.1 s
Dance of the p-values

- Variant 1 Improvement: $p = 0.0076$, 11.4 s
- Variant 2 Improvement: $p = 0.53$, 2.1 s
Dance of the p-values

My apologies to 3% of the audience
Dance of the p-values
Dance of the p-values

Variation 1

Variation 2

 improvement

improvement

p = 0.0011

p = 0.0089
Dance of the p-values

- **Variant 1 Improvement**: 16.4 s
- **Variant 2 Improvement**: 8.5 s
- **Variant 1 Improvement**: p = 0.0011
- **Variant 2 Improvement**: p = 0.0089
Our data

Don’t trust p-values

p-values

Our data

Don’t trust p-values
Standard errors

Our data
95% confidence intervals

Our data
95% confidence intervals

Our data
Dance of the confidence intervals
Dance of the confidence intervals
95% confidence intervals

Don’t trust confidence intervals

Our data
Bayesian credible intervals and posterior distributions

“Eye plots” made using (Kay, 2016)
Bayesian dance

“Eye plots” made using (Kay, 2016)
Bayesian dance

"Eye plots" made using (Kay, 2016)
Bayesian credible intervals and posterior distributions

Our data

Don’t trust credible intervals and posterior distributions

“Eye plots” made using (Kay, 2016)
p-values

![Bar chart showing comparison of improvement times and p-values for two variants.](image)
Bayes Factors

- Variant 1 Improvement: 11.4 s
- Variant 2 Improvement: 2.1 s

- BF = 170 (Better)
- BF = 2.6 (Worse)
Dance of the Bayes Factors

- Variant 1 Improvement: 16.4 s
- Variant 2 Improvement: 8.5 s

Bayes Factor (BF): 146
Bayes Factors

- Variant 1 Improvement: 11.4 s
- Variant 2 Improvement: 2.1 s

- BF = 170 for Better
- BF = 2.6 for Worse

Nope, don’t trust these either
User Study
Our study examines the benefits of adding haptic detents to touch sliders. We used 1-D target acquisition tasks involving both easy and hard targets (see Figure 1).

A repeated measure full-factorial within-subject design was used. The factors were Technique = \{S=slider, HS=haptic slider\}, and Difficulty = \{Easy, Hard\}. Twelve volunteers (2 female) familiar with touch devices, aged 22–36, participated in the study. We collected a total of 12 Participant \times 2 Technique \times 2 Difficulty \times 128 repetitions = 6144 trials with completion Time.

Hypotheses
(H1) Technique HS is faster than technique S overall.
(H2) Easy tasks are faster than Hard tasks overall.

Results
An ANOVA on Time with the model Technique\times Difficulty\times Rnd(Participant) reveals a highly significant effect of Technique but no significant effect of Difficulty and no Technique\times Difficulty interaction (see Table 1).

Table 1: ANOVA table.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>1,11</td>
<td>12.7336</td>
<td>0.0044***</td>
</tr>
<tr>
<td>Difficulty</td>
<td>1,11</td>
<td>2.7084</td>
<td>0.1281</td>
</tr>
<tr>
<td>Technique\times Difficulty</td>
<td>1,11</td>
<td>4.0402</td>
<td>0.0696</td>
</tr>
</tbody>
</table>

Our ANOVA analysis therefore confirms that technique HS yields significantly shorter completion times than technique S overall, i.e., all task difficulties confounded. The average Time is 1.09s for S, and 1.04s for HS (see Figure 3). This difference corresponds to a 4.8% increase in speed for technique HS compared to technique S.

Discussion
Our user study shows that subjects completed the tasks significantly faster in the presence of haptic feedback (4.8% faster). Our hypothesis (H1) is therefore confirmed.

The superiority of haptic feedback seems to hold for all target difficulties, as suggested by the lack of significant interaction between Technique and Difficulty. Even though large targets do not suffer from the "fat finger" problem, multimodal feedback still seems superior to visual-only feedback. This could be explained by the fact that the haptic channel is a sensory modality directly connected with kinesthetic and motor functions, and therefore capitalizes on our reflexive motor responses.

Surprisingly, we found no significant effect of Difficulty overall, so our hypothesis (H2) is not confirmed. This could be explained by the fact that differences in target difficulty were not large enough to significantly affect performance. We could have used different target sizes, but the limited input resolution of the device prevented us from using much smaller targets. Conversely, a very large target would occupy most of the slider range, which does not capture realistic slider tasks. Overall, it seems that for sliders, target size is not a crucial factor.

To summarize, our study provides strong evidence for the benefits of tactile feedback when operating sliders. Although moderate, the effect of technique was found to be highly significant. Tactile guidance provides additional proprioceptive cues when interacting with the glass surface of the device—otherwise uniformly flat. This allows users to maintain an accurate mental model of the slider thumb's location, speeding up the reaching of specific locations. Overall, based on our results, we recommend the use of sliders with haptic detents on touch devices, both for fine and for coarse control.

(Dragicevic, Chevalier and Huot, 2014)
User Study

Our study examines the benefits of adding haptic detents to touch sliders. We used 1-D target acquisition tasks involving both easy and hard targets (see Figure 1).

A repeated measure full-factorial within-subject design was used. The factors were Technique = \{S=slider, HS=haptic slider\}, and Difficulty = \{Easy, Hard\}. Twelve volunteers (3 female) familiar with touch devices, aged 20–37, participated in the study. We collected a total of 12 Participant x 2 Technique x 2 Difficulty x 128 repetitions = 6144 trials with completion Time.

Hypotheses

(H1) Technique HS is faster than technique S overall.

(H2) Easy tasks are faster than Hard tasks overall.

Results

An ANOVA on Time with the model Technique x Difficulty x Rnd(Participant) reveals a significant effect of both Technique and Difficulty, but no significant Technique x Difficulty interaction effect (see Table 1).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>1, 11</td>
<td>5.1139</td>
<td>0.0450*</td>
</tr>
<tr>
<td>Difficulty</td>
<td>1, 11</td>
<td>6.2892</td>
<td>0.0291*</td>
</tr>
<tr>
<td>Technique x Difficulty</td>
<td>1, 11</td>
<td>1.3669</td>
<td>0.2671</td>
</tr>
</tbody>
</table>

Our analysis therefore confirms that HS is faster than S overall, with an average Time of 1.16s for S vs. 1.10s for HS, a 5.5% increase in speed (see Figure 2). Our analysis also confirms the effects of task difficulty, with an average Time of 1.25s for Hard vs. 1.01s for Easy, corresponding to a 23.8% increase in speed (see Figure 3).

Discussion

Our user study shows that subjects completed the tasks significantly faster in the presence of haptic feedback (5.5% faster). Our hypothesis (H1) is therefore confirmed.

The superiority of haptic feedback seems to hold for all target difficulties, as suggested by the lack of significant interaction between Technique and Difficulty. Even though large targets do not suffer from the “fat finger” problem, multimodal feedback still seems superior to visual-only feedback. This could be explained by the fact that the haptic channel is a sensory modality directly connected with kinesthetic and motor functions, and therefore capitalizes on our reflexive motor responses.

Our analysis also shows a significant difference between the two levels of difficulty all techniques confounded, with Easy being as much as 23.8% faster than Hard. Therefore, our hypothesis (H2) is also supported. We derived our difficulty levels based on extensive pilot studies, so as not to favor any technique. Our results validate our experimental design and confirm that target size is an adequate metric for task difficulty. HS appears to perform comparably well under two widely different task difficulties, suggesting that its advantages may well generalize to other difficulty levels.

To summarize, our study confirms that adding tactile feedback in the form of simulated detents facilitates the operation of sliders. Tactile guidance provides additional proprioceptive cues when interacting with the glass surface of the device—otherwise uniformly flat. This likely allows users to maintain an accurate mental model of the slider thumb’s location, speeding up the reaching of specific locations. Overall, based on our results, we recommend the use of sliders with haptic detents on touch devices, both for fine and for coarse control.

(Dragicevic, Chevalier and Huot, 2014)
Universe 3

User Study
Our study examines the benefits of adding haptic detents to touch sliders. We used 1-D target acquisition tasks involving both easy and hard targets (see Figure 1).

A repeated measure full-factorial within-subject design was used. The factors were Technique = \{S=slider, HS=haptic slider\}, and Difficulty = \{Easy, Hard\}. Twelve volunteers (4 female) familiar with touch devices, aged 18–32, participated in the study. We collected a total of 12 Participant × 2 Technique × 2 Difficulty × 128 repetitions = 6144 trials with completion Time.

Hypotheses
(H1) Technique HS is faster than technique S.
(H2) Easy tasks are faster than Hard.

Results
An ANOVA on Time with the model Technique × Difficulty × RND(Participant) reveals no significant effect of Technique, but a highly significant effect of Difficulty with also a highly significant Technique × Difficulty interaction effect (see Table 1).

Discussion
While we did not observe a significant main effect of Technique, an analysis of simple effects reveals that HS significantly outperformed S in the Hard condition, with as much as 9.2% in speed improvement. Therefore, our hypothesis (H1) is only partially confirmed.

Although we did not find a significant difference between techniques in the Easy condition, Figure 3 exhibits an intriguing trend, raising the possibility of HS being worse than S under the Easy condition. This seems to be confirmed by the very strong interaction observed between Technique and Difficulty. A possible explanation could be that the regular bursts generated by the haptic detents is distracting to some users, which in turn slightly impairs their performance. Indeed, some participants expressed discomfort while interacting with HS.

In the Hard condition, however, the situation is very different: due to the “fat finger” problem, users are likely deprived of visual cues during the corrective phase of their movement. In this case, multimodal feedback likely alleviates this issue by providing non-visual guidance. In other terms, when the target is small, the benefits brought by haptic feedback largely outweigh discomfort issues, allowing users to acquire these targets much more easily.

To summarize, our study shows that adding tactile feedback in the form of simulated detents can be an effective solution to the “fat finger” problem when manipulating sliders on touch devices. However, haptic feedback can also be distracting and in some cases, impair performance when the task is easy (large 1-D targets). Overall, based on our results, we recommend the use haptic detents on touch sliders for tasks that require fine control, but not for tasks where coarse control is sufficient.

(Dragicevic, Chevalier and Huot, 2014)
Universe 4

User Study
Our study examines the benefits of adding haptic detents to touch sliders. We used 1-D target acquisition tasks involving both easy and hard targets (see Figure 1).

A repeated measure full-factorial within-subject design was used. The factors were Technique = \{S=slider, HS=haptic slider\}, and Difficulty = \{Easy, Hard\}. Twelve volunteers (5 female) familiar with touch devices, aged 21–50, participated in the study. We collected a total of 12 Participant × 2 Technique × 2 Difficulty × 128 repetitions = 6144 trials with completion Time.

Hypotheses
(H1) Technique HS is faster than technique S overall.
(H2) Easy tasks are faster than Hard tasks overall.

Results
An ANOVA on Time with the model Technique × Difficulty × Rnd(Participant) reveals a significant effect of Technique and a significant interaction Technique × Difficulty (see Table 1).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>1,1</td>
<td>7.2144</td>
<td>0.0212*</td>
</tr>
<tr>
<td>Difficulty</td>
<td>1,1</td>
<td>4.1479</td>
<td>0.0665</td>
</tr>
<tr>
<td>Technique × Difficulty</td>
<td>1,1</td>
<td>5.5941</td>
<td>0.0375*</td>
</tr>
</tbody>
</table>

Our analysis therefore confirms that HS is faster than S overall, with an average Time of 1.12s for S vs. 1.06s for HS, a 5.7% increase in speed (see Figure 2). Student’s t-tests reveal no significant difference between techniques for Easy (avg. Times: S=1.05s, HS=1.03s, \( p = 0.4065 \)), and a highly significant difference between techniques for Hard, with an 8.2% increase in speed with HS (avg. Times: S=1.19s, HS=1.10s, \( p = 0.0060 \)) (see Figure 3).

Discussion
Our user study shows that subjects completed the tasks significantly faster in the presence of haptic feedback (5.7% faster). Our hypothesis (H1) is therefore confirmed.

In addition, we found a significant interaction between technique and task difficulty, with a higher performance gain brought by HS for the Hard condition (8.2% faster). In contrast, the improvement was lower (1.9%) under the Easy condition (also see Figure 3). One explanation is that in the Hard condition, the “fat finger” problem interferes with the corrective phase of users’ movement. Multimodal feedback likely alleviates this by providing non-visual guidance. Under the Easy condition, the target was larger and the fat finger issue not as pronounced, making haptic feedback still useful but less critical.

Surprisingly, we were not able to find a significant effect of Difficulty overall, despite the trends visible in Figure 3. This could be explained by the fact that differences in the target difficulty were not large enough to significantly affect performance. In our pilot studies we considered tasks involving much smaller or much larger targets, but dismissed them as unrealistic. So it seems that overall, target size is not a crucial factor for sliders.

To summarize, our study confirms that adding tactile feedback in the form of simulated detents facilitates the operation of sliders. Tactile guidance provides additional proprioceptive cues when interacting with the glass surface of the device—otherwise uniformly flat. Operating sliders is hard on touch devices in general, but even more so when fine control is needed, due to the “fat finger” problem. We show that haptic guidance greatly facilitates this task. Overall, based on our results, we recommend the use of sliders with haptic detents on touch devices, especially when fine control is needed.

(Dragicevic, Chevalier and Huot, 2014)
Universe 5

User Study
Our study examines the benefits of adding haptic detents to touch sliders. We used 1-D target acquisition tasks involving both easy and hard targets (see Figure 1).

A repeated measure full-factorial within-subject design was used. The factors were Technique = \{S=slider, HS=haptic slider\}, and Difficulty = \{Easy, Hard\}. Twelve volunteers (4 female) familiar with touch devices, aged 18–39, participated in the study. We collected a total of 12 Participant \times 2 Technique \times 2 Difficulty \times 128 repetitions = 6144 trials with completion Time.

Hypotheses
(H1) Technique HS is faster than technique S overall.
(H2) Easy tasks are faster than Hard tasks overall.

Results
An ANOVA on Time with the model Technique \times Difficulty \times Rnd(Participant) reveals a significant effect of Technique and a significant interaction Technique \times Difficulty (see Table 1).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>1,11</td>
<td>6.0536</td>
<td>0.0317*</td>
</tr>
<tr>
<td>Difficulty</td>
<td>1,11</td>
<td>1.0392</td>
<td>0.3299</td>
</tr>
<tr>
<td>Technique \times Difficulty</td>
<td>1,11</td>
<td>9.4480</td>
<td>0.0106*</td>
</tr>
</tbody>
</table>

Our analysis therefore confirms that HS is faster than S overall, with an average Time of 1.08s for S vs. 1.01s for HS, a 6.9% increase in speed (see Figure 2). Student’s t-tests reveal no significant difference between techniques for Easy (avg. Times: S=1.01s, HS=1.01s, p = 0.9601), and a highly significant difference between techniques for Hard, with a 12.9% increase in speed with HS (avg. Times: S=1.14s, HS=1.01s, p = 0.0071) (see Figure 3).

Discussion
Our user study shows that subjects completed the tasks significantly faster in the presence of haptic feedback (6.9% faster). Our hypothesis (H1) is therefore confirmed.

In addition, we found a significant interaction between technique and task difficulty, with a higher performance gain brought by HS for the Hard condition (as much as 12.9% faster). In contrast, the two techniques seem to perform very similarly under the Easy condition (see Figure 3). One explanation is that in the Hard condition, users are deprived of visual cues during the corrective phase of their movement because of the “fat finger” problem. Multimodal feedback likely alleviates this by providing non-visual guidance. Under the Easy condition, the target may have been large enough for users to rely on visual feedback only, making haptic feedback superfluous.

Surprisingly, we were not able to find a significant effect of Difficulty overall. A tentative explanation can be found in Figure 3: while S seems to be affected by difficulty, HS exhibits a stable performance across difficulty levels. This suggests that with haptic feedback, all targets are equally easy. Although this seems to contradict Fitts’ Law, recall this law is about aimed movements with visual feedback. The haptic channel may not be as sensitive to target size, possibly due to the fact that it is a sensory modality directly connected with kinesthetic and motor functions.

To summarize, our study shows that adding tactile feedback in the form of simulated detents facilitates the precise manipulation of sliders. Precise control of sliders is challenging on touch devices, partly due to the “fat finger” problem. We show that with haptic guidance, it becomes practically as easy as coarse control. Overall, based on our results, we recommend the use of sliders with haptic detents on touch devices when fine control is needed.

(Dragicevic, Chevalier and Huot, 2014)
User Study
Our study examines the benefits of adding haptic detents to touch sliders. We used 1-D target acquisition tasks involving both easy and hard targets (see Figure 1).

A repeated measure full-factorial within-subject design was used. The factors were Technique = \{S=slider, HS=haptic slider\}, and Difficulty = \{Easy, Hard\}. Twelve volunteers (2 female) familiar with touch devices, aged 20–43, participated in the study. We collected a total of 12 Participant \times 2 Technique \times 2 Difficulty \times 128 repetitions = 6144 trials with completion Time.

Hypotheses
(H1) Technique HS is faster than technique S overall.
(H2) Easy tasks are faster than Hard tasks overall.

Results
An ANOVA on Time with the model Technique \times Difficulty \times Rnd(Participant) reveals a highly significant effect of Technique, and a very highly significant effect of Difficulty, and no Technique \times Difficulty interaction (see Table 1).

Our analysis therefore confirms that HS is faster than S overall, with an average Time of 1.17s for S vs. 1.10s for HS, a 6.4% increase in speed (see Figure 2). Our analysis also confirms the effects of task difficulty, with an average Time of 1.24s for Hard vs. 1.03s for Easy, corresponding to a 20.4% increase in speed (see Figure 3).

Discussion
Our user study shows that subjects completed the tasks significantly faster in the presence of haptic feedback (6.4% faster). Our hypothesis (H1) is therefore confirmed.

The superiority of haptic feedback seems to hold for all target difficulties, as suggested by the lack of significant interaction between Technique and Difficulty. Even though large targets do not suffer from the “fat finger” problem, multimodal feedback still seems superior to visual-only feedback. This could be explained by the fact that the haptic channel is a sensory modality directly connected with kinesthetic and motor functions, and therefore capitalizes on our reflexive motor responses.

Our analysis also shows a highly significant difference between the two levels of difficulty all techniques confounded, with Easy being as much as 20.4% faster than Hard. Therefore, our hypothesis (H2) is also supported. We derived our difficulty levels based on extensive pilot studies, so as not to favor any technique. Our results validate our experimental design and confirm that target size is an adequate metric for task difficulty. HS appears to perform comparably well under two widely different task difficulties, suggesting that its advantages may well generalize to other difficulty levels.

To summarize, our study confirms that adding tactile feedback in the form of simulated detents facilitates the operation of sliders. Tactile guidance provides additional proprioceptive cues when interacting with the glass surface of the device—otherwise uniformly flat. This likely allows users to maintain an accurate mental model of the slider thumb’s location, speeding up the reaching of specific locations. Overall, based on our results, we recommend the use of sliders with haptic detents on touch devices, both for fine and for coarse control.

(Dragicevic, Chevalier and Huot, 2014)
Universe 7

User Study
Our study examines the benefits of adding haptic detents to touch sliders. We used 1-D target acquisition tasks involving both easy and hard targets (see Figure 1).

A repeated measure full-factorial within-subject design was used. The factors were Technique = \{S=slider, HS=haptic slider\}, and Difficulty = \{Easy, Hard\}. Twelve volunteers (7 female) familiar with touch devices, aged 19–31, participated in the study. We collected a total of 12 Participant × 2 Technique × 2 Difficulty × 128 repetitions = 6144 trials with completion Time.

Hypotheses
(H1) Technique HS is faster than technique S overall.
(H2) Easy tasks are faster than Hard tasks overall.

Results
An ANOVA on Time with the model Technique × Difficulty × Rnd(Participant) reveals no significant effect of Technique, but a significant effect of Difficulty. Furthermore, the ANOVA analysis did not reveal any significant Technique × Difficulty interaction effect (see Table 1 below).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>1,11</td>
<td>4.6215</td>
<td>0.0547</td>
</tr>
<tr>
<td>Difficulty</td>
<td>1,11</td>
<td>4.8698</td>
<td>0.0495*</td>
</tr>
<tr>
<td>Technique × Difficulty</td>
<td>1,11</td>
<td>1.8322</td>
<td>0.2030</td>
</tr>
</tbody>
</table>

Our analysis confirms the effects of task difficulty, with an average Time of 1.29s for Hard vs. 1.02s for Easy, corresponding to a 26.5% increase in speed (see Figure 3). Thus our second hypothesis (H2) is confirmed.

Discussion
Our initial hypothesis was that haptic feedback would facilitate 1-D target acquisition tasks (H1). Our analyses failed to support this hypothesis. Yet, our results suggest that if haptic feedback may not help, it does not harm either. Indeed, HS was still on average 4% faster than S, although this difference was not statistically significant.

Participants' answers to our post-experiment questionnaire suggest that haptic feedback may provide qualitative benefits beyond pure task completion times. Many participants rated the technique high in hedonic value (a median of 4 on a 5-point Likert scale), and feedback on haptic detents was overall positive.

The feedback collected during our study also helped us identify directions for improvement for our current prototype. Some participants expressed discomfort while interacting with HS. One mentioned "a feeling similar as if the device was sending little electrical shocks to the finger", and thought the equipment was dysfunctional. We believe this could easily be fixed by allowing users to personalize the haptic signal. One participant commented that haptic feedback "feels weird. [She] would rather expect [her] finger to smoothly glide on the glass surface". Indeed, a flat screen provides conflicting affordances with haptic feedback. Visual techniques that emphasize physicality (e.g. shadow or cushion effects to convey holes and bumps) could address this problem.

In summary, while our study did not reveal significant quantitative benefits of haptic detents over the traditional touch slider, the qualitative feedback we received was very positive and encouraging. We were able to collect valuable insights that shed light on the limitations of current haptic interfaces. We hope that our results will inform and inspire further development in the area.

(Dragicevic, Chevalier and Huot, 2014)
Universe 8

User Study
Our study examines the benefits of adding haptic detents to touch sliders. We used 1-D target acquisition tasks involving both easy and hard targets (see Figure 1).

A repeated measure full-factorial within-subject design was used. The factors were Technique = \{Slider, Haptic Slider\}, and Difficulty = \{Easy, Hard\}. Twelve volunteers (5 female) familiar with touch devices, aged 19–35, participated in the study. We collected a total of 12 Participant × 2 Technique × 2 Difficulty × 128 repetitions = 6144 trials with completion Time.

Hypotheses
(H1) Technique H5 is faster than technique S.
(H2) Easy tasks are faster than Hard.

Results
An ANOVA on Time with the model Technique × Difficulty × Rnd(Participant) reveals no significant effect of Technique, but a significant effect of Difficulty with also a very highly significant Technique × Difficulty interaction effect (see Table 1).

Table 1: ANOVA table.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
<td>1, 11</td>
<td>2.1350</td>
<td>0.1719</td>
</tr>
<tr>
<td>Difficulty</td>
<td>1, 11</td>
<td>5.1621</td>
<td>0.0442**</td>
</tr>
<tr>
<td>Technique × Difficulty</td>
<td>1, 11</td>
<td>22.6791</td>
<td>0.0006***</td>
</tr>
</tbody>
</table>

Our analysis confirms the effect of difficulty (avg. Times: Easy = 1.02s, Hard = 1.19s, see Figure 2). Student’s t-tests reveal no significant difference between techniques for Easy (avg. Times: S = 1.01s, HS = 1.04s, p = 0.2757), and a very highly significant difference between techniques for Hard, with an 8.8% increase in speed with HS (avg. Times: S = 1.24s, HS = 1.14s, p = 0.0061) (see Figure 3).

Discussion

While we did not observe a significant main effect of Technique, an analysis of simple effects reveals that HS significantly outperformed S in the Hard condition, with as much as 8.8% in speed improvement. Therefore, our hypothesis (H1) is only partially confirmed.

Although we did not find a significant difference between techniques in the Easy condition, Figure 3 exhibits an intriguing trend, raising the possibility of HS being worse than S under the Easy condition. This seems to be confirmed by the very strong interaction observed between Technique and Difficulty. A possible explanation could be that the regular bursts generated by the haptic detents is distracting to some users, which in turn slightly impairs their performance. Indeed, some participants expressed discomfort while interacting with HS.

In the Hard condition, however, the situation is very different: due to the “fat finger” problem, users are likely deprived of visual cues during the corrective phase of their movement. In this case, multimodal feedback likely alleviates this issue by providing non-visual guidance. In other terms, when the target is small, the benefits brought by haptic feedback largely outweigh discomfort issues, allowing users to acquire these targets much more easily.

To summarize, our study shows that adding tactile feedback in the form of simulated detents can be an effective solution to the “fat finger” problem when manipulating sliders on touch devices. However, haptic feedback can also be distracting and in some cases, impair performance when the task is easy (large 1-D targets). Overall, based on our results, we recommend the use of haptic detents on touch sliders for tasks that require fine control, but not for tasks where coarse control is sufficient.

(Dragicevic, Chevalier and Huot, 2014)
Lessons learned so far

- Everything dances
- Descriptive and inferential statistics alike
- Dances propagate along the analysis pipeline
- There is no way around it
- But there must be ways!
False solution: 
discretize / dichotomize
False solution:
discretize / dichotomize
False solution: discretize / dichotomize

\[ \alpha = 0.05 \]
False solution: discretize / dichotomize
False solution: discretize / dichotomize
False solution: discretize / dichotomize

Don’t do this
False solution: discretize / dichotomize
False solution: discretize / dichotomize

Variant 1 Improvement: 4.0 s
Variant 2 Improvement: -5.1 s

Type I error: 5% of the time
False solution: discretize / dichotomize

Type I error: 5% of the time

Type II error: 30% of the time (power = .7)
False solution: discretize / dichotomize
False solution: discretize / dichotomize
False solution: discretize / dichotomize
False solution: 
discretize / dichotomize
False solution:
discretize / dichotomize

Don’t do this
False solution: discretize / dichotomize
False solution: 
discretize / dichotomize
False solution: discretize / dichotomize
False solution: discretize / dichotomize
False solution: discretize / dichotomize

Dance of the Bayes factors

You might have seen the ‘Dance of the p-values’ video by Geoff Cumming (if not, watch it here). Because p-values and the default Bayes factors (Rouder, Speckman, Sun, Morey, & Iverson, 2009) are both calculated directly from t-values and sample sizes, we might expect there is also a Dance of the Bayes factors. And indeed, there is. Bayes factors can vary widely over identical studies, just due to random variation.

If people would always correctly interpret Bayes factors, that would not be a problem. Bayes factors tell you how much data are in line with models, and quantify relative evidence in favor of one of these models. The data is what it is, even when it is misleading (i.e., supporting a hypothesis that is not true). So, you can conclude the null model is more likely than some other model, but purely based on a Bayes factor, you can't draw a conclusion such as “This Bayes factor allows us to conclude that there are no differences between conditions". Regrettably, researchers are massively starting to misinterpret Bayes factors (I won't provide references, though I have many). This is not surprising – people find statistical inferences difficult, whether these are about p-values, confidence intervals, or Bayes factors.

(Lakens, 2016)
False solution: discretize / dichotomize

Don't do this either
Better solution: use large samples
Better solution: use large samples

\[ n = 16 \]
Better solution: use large samples

\[ n = 32 \]
Better solution: use large samples

$n = 64$
Better solution: use large samples

\[ n = 128 \]
Better solution: use large samples

\[ n = 128 \]

Type I error: 0.1% of the time

Type II error: 0.003% of the time (power = 0.99997)
Better solution: use large samples

$n = 128$
Limits of large samples

- More participants stabilize dances only slowly
- Running participants can be costly
- Power can be increased by better measurement
- Researchers would still report secondary findings with lower power
- Still need ways of conveying uncertain results
Another solution: use informed priors
Another solution: use informed priors

Uninformed

Weakly informed

Strongly informed
Another solution: use informed priors

Uninformed  Weakly informed  Strongly informed
Limits of informed priors

• Need very strong priors to reduce the dance

• Good priors require replications (Kay et al, 2016), but replications are rare in our fields

• Many would probably consider the use of strongly informed priors as a form of “cheating”

• Again, still need ways of conveying uncertain results with little prior knowledge
The real problem: lack of awareness

• Most researchers are already familiar with the dance of the sample means

• But they overestimate the reliability of p-values, of statistical tests and of interval estimates

• We tend to believe that statistics “stabilize" noisy data. Very hard to overcome this wrong intuition

• Solutions: education, more willingness, new principles
Robustness principle

- A statistical analysis is robust to sampling variability if two similar datasets yield similar results.

- A plot is robust to sampling variability if two similar datasets yield visually similar plots.

- A way of interpreting results is robust to sampling variability if two similar datasets yield similar interpretations.
Examples

Plotting distributions

Dot plot
Bee swarm
Violin plot
Box plot
Examples

Plotting distributions

Dataset 1

Dataset 2
Examples

Sorting by effect size

- Integrated Shearing
- Device Tilt
- Top-Down Persp.
- Inclined Persp.
- Standard Panning

Accuracy
Examples

Interpreting multiple CIs

(Dragicevic, 2016)
Examples

Interpreting multiple CIs

(Dragicevic, 2016)
Examples

Interpreting multiple CIs

(Dragicevic, 2016)
Examples

Interpreting multiple CIs

(Dragicevic, 2016)
Examples

Interpreting multiple CIs

(Dragicevic, 2016)
Examples

Interpreting multiple CIs

(Dragicevic, 2016)
Examples

Interpreting multiple CIs

(Dragicevic, 2016)
Examples

“...It is best for individual researchers to present point estimates and confidence intervals and refrain from attempting to draw final conclusions about research hypotheses.”

(Schmidt and Hunter, 1997)
Examples

“ We have the duty of [...] communicating our conclusions in intelligible form, in recognition of the right of other free minds to utilize them in making their own decisions. ”

(Fisher, 1955)
Conclusions

- Embrace uncertainty (Giner-Sorolla, 2012)
- Convey it clearly, use plots
- Always keep the dances in mind, seek robustness
- Do not dichotomize results
- Be nuanced, use vague language (Van Deemter, 2010)
- Let your readers judge by themselves
Conclusions

• We need more research on this!
Conclusions

• We need more research on this!

(Gradient plots)

(Correll and Gleicher, 2014)
Conclusions

• We need more research on this!

Hypothetical outcome plots

(Hullman, Resnick and Adar, 2015)
Conclusions

• For more: www.aviz.fr/badstats
• Animated plots created by Pierre Dragicevic and Yvonne Jansen