Statistical dances

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

biovis **2016**

Pierre Dragicevic, Inria, France

User studies

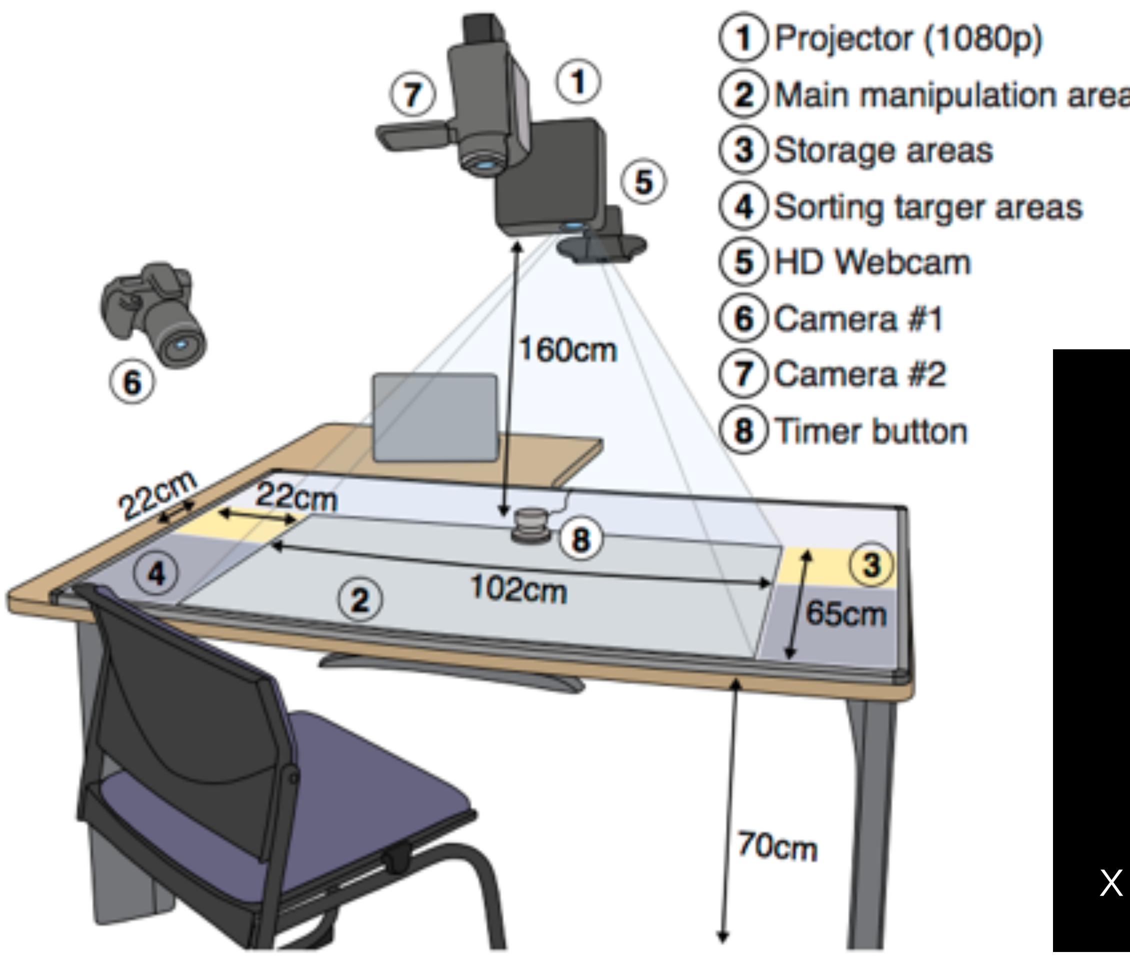


Image and video from (Le Goc et al, 2016)

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, $t = 5.605$, $\boxed{p < .001}$).

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)

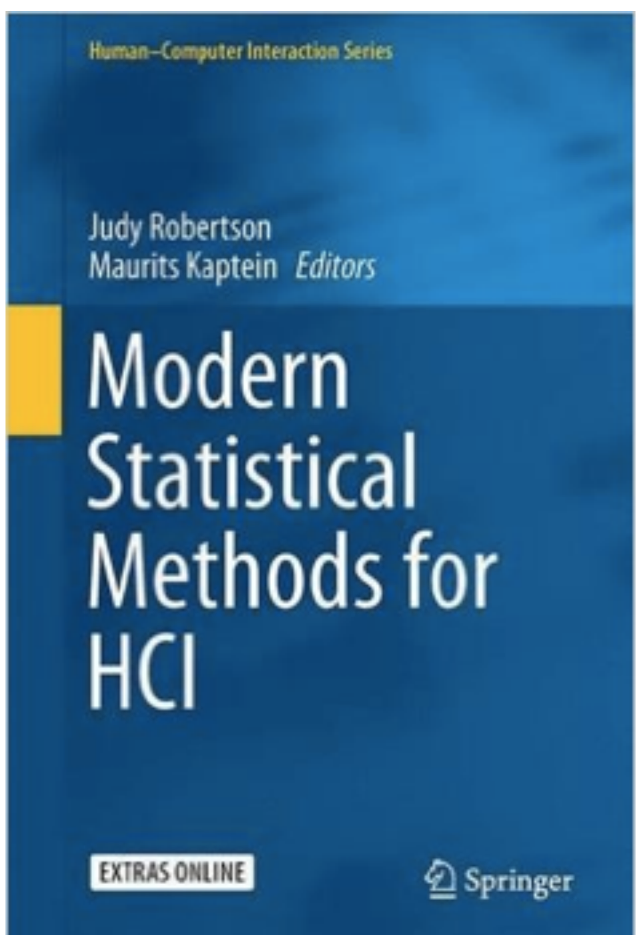
All references for this talk at www.aviz.fr/dances

Bad Stats are Miscommunicated Stats

Pierre Dragicevic, INRIA

BELIV 2014
November 10th, 2014, Paris, France.

www.aviz.fr/badstats



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

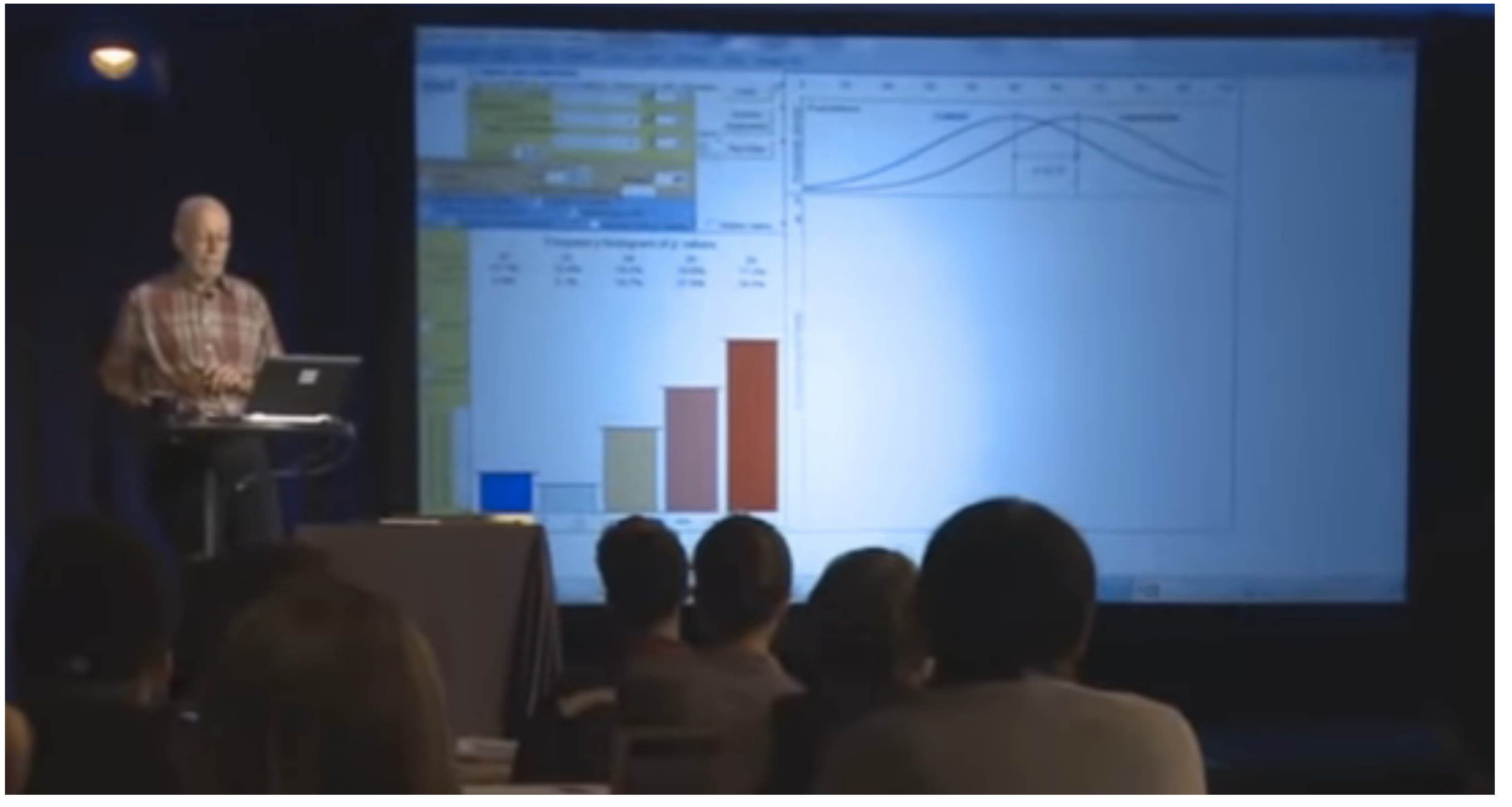
It's all about uncertainty

"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)

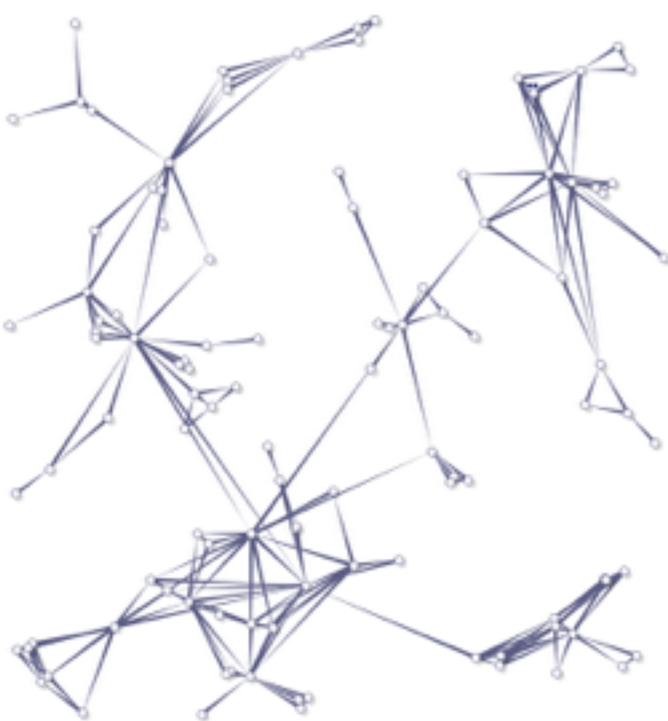
All references for this talk at www.aviz.fr/dances

Outline

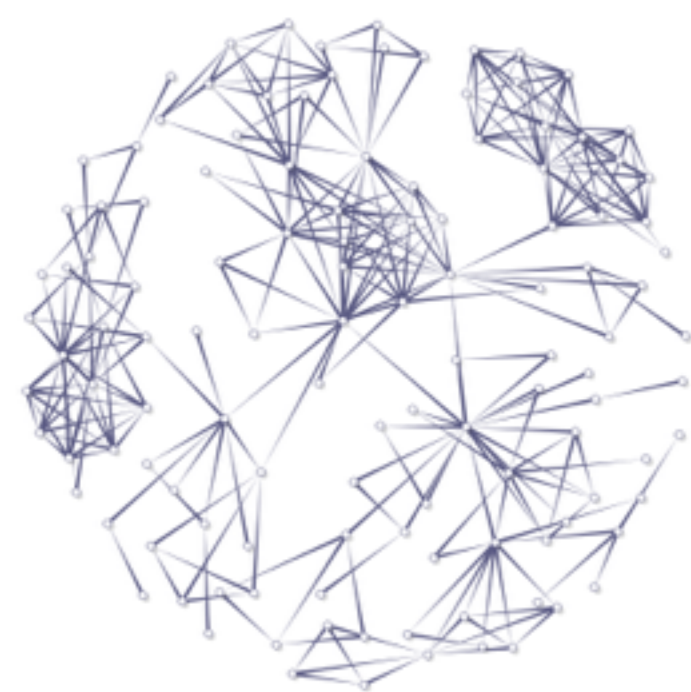
- Statistical dances — why no statistical analysis is reliable
- what do to about it
  - False solutions
  - OK solutions
  - Real solutions

Dance of the sample means

Evaluating a new graph layout technique



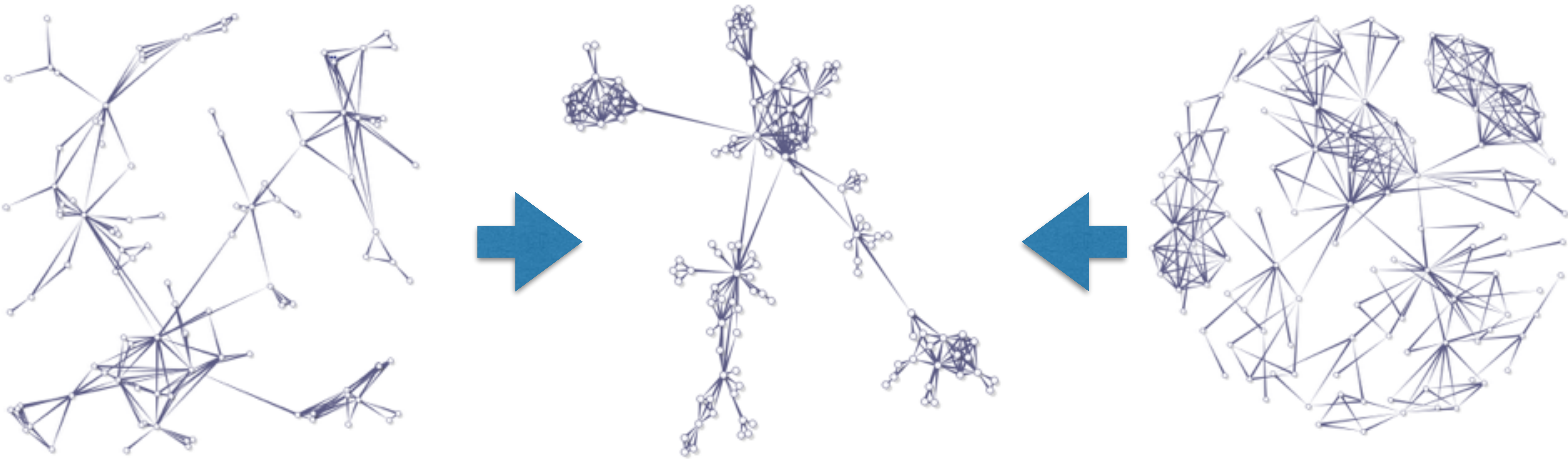
Variant 1



Variant 2

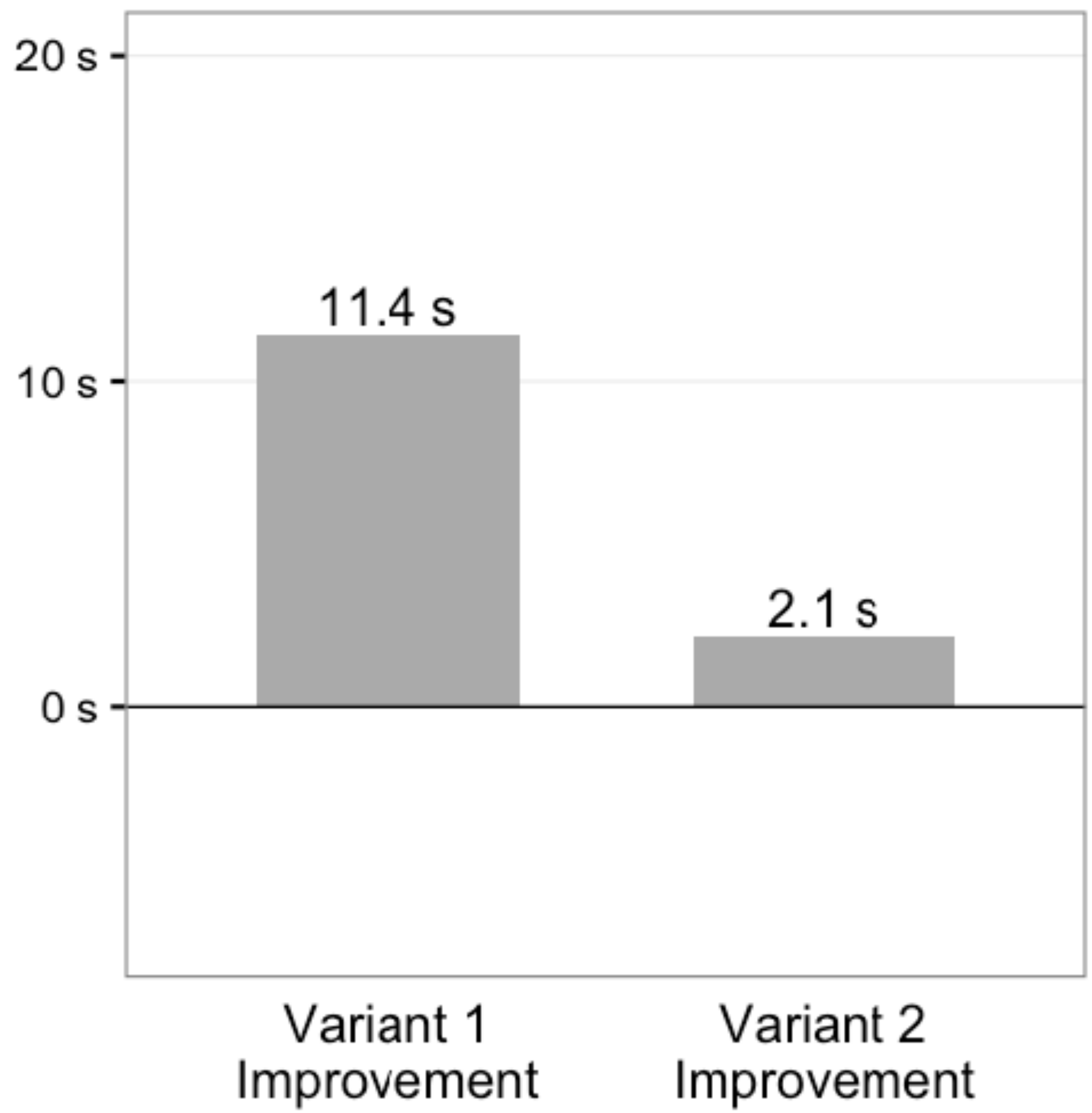
Dance of the sample means

Evaluating a new graph layout technique

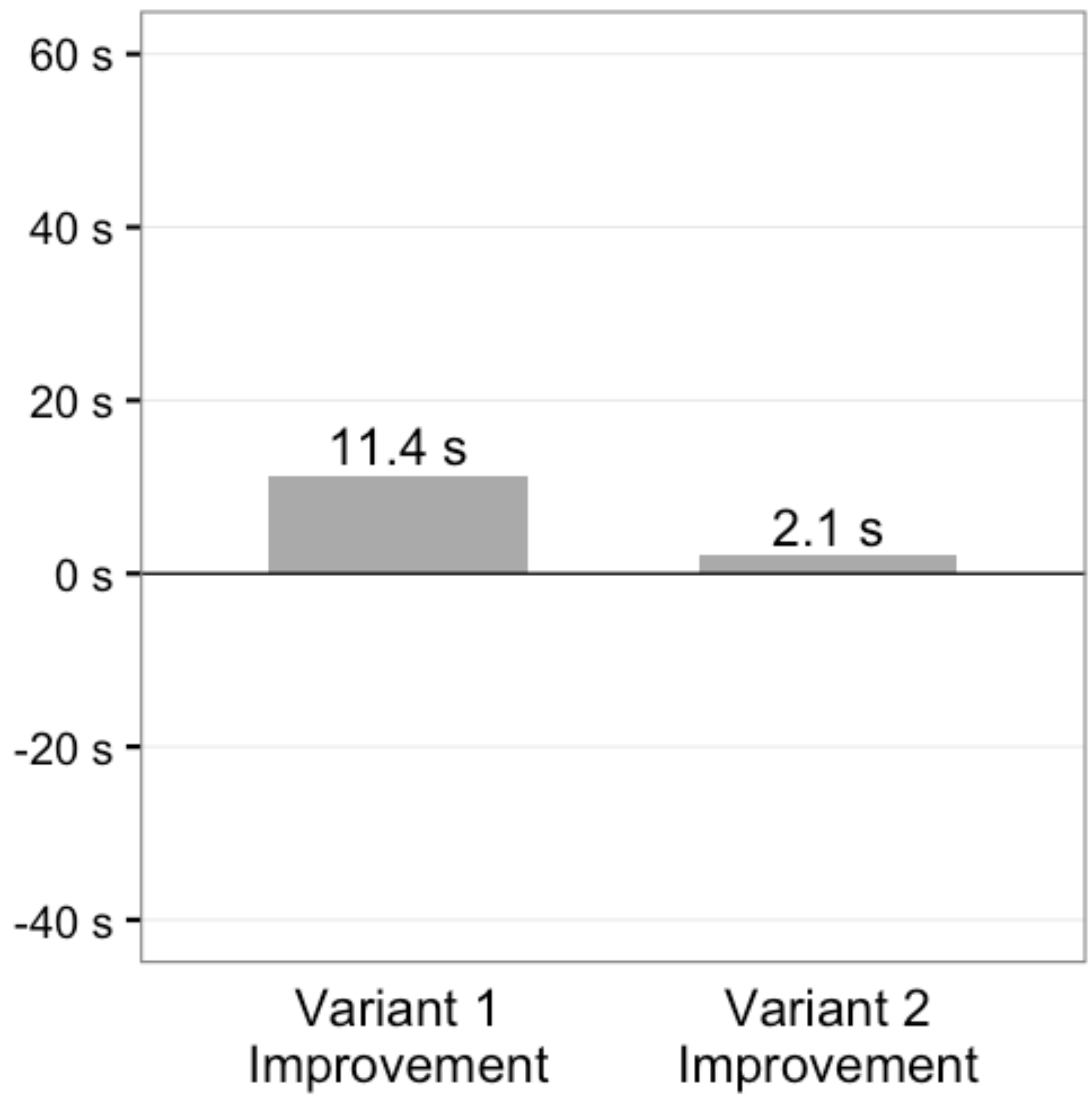


Variant 1 Baseline Variant 2

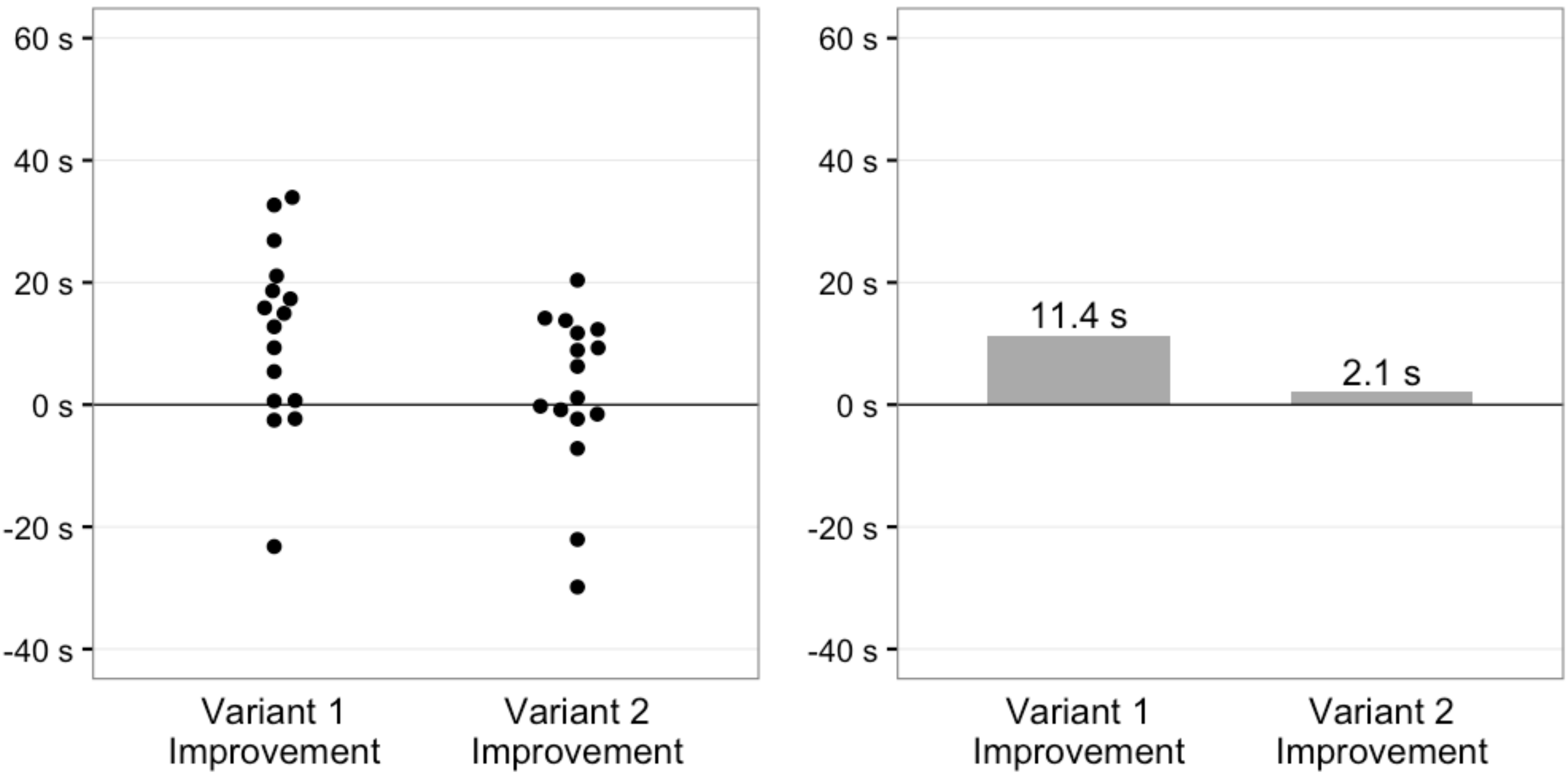
Dance of the sample means



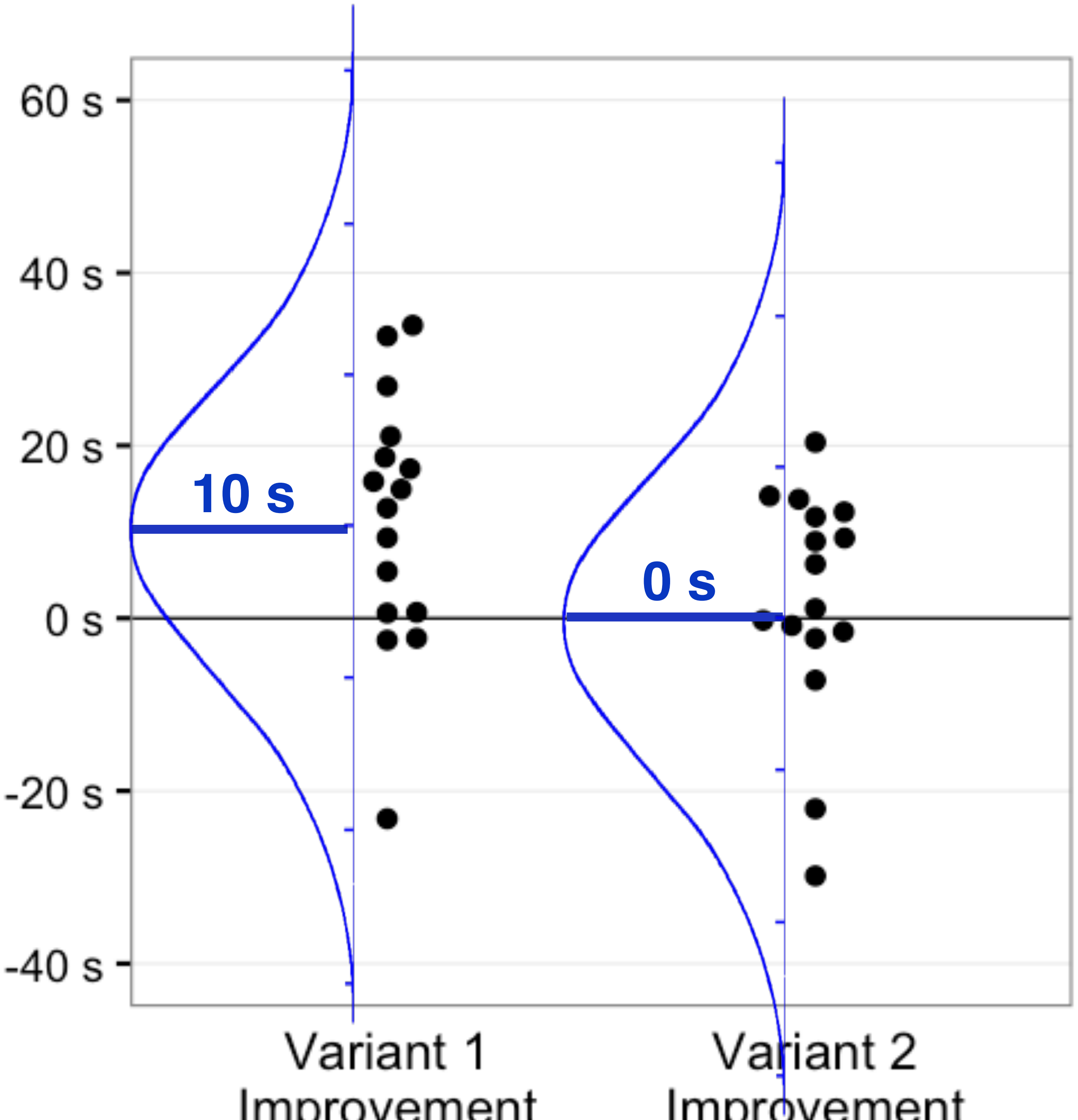
Dance of the sample means

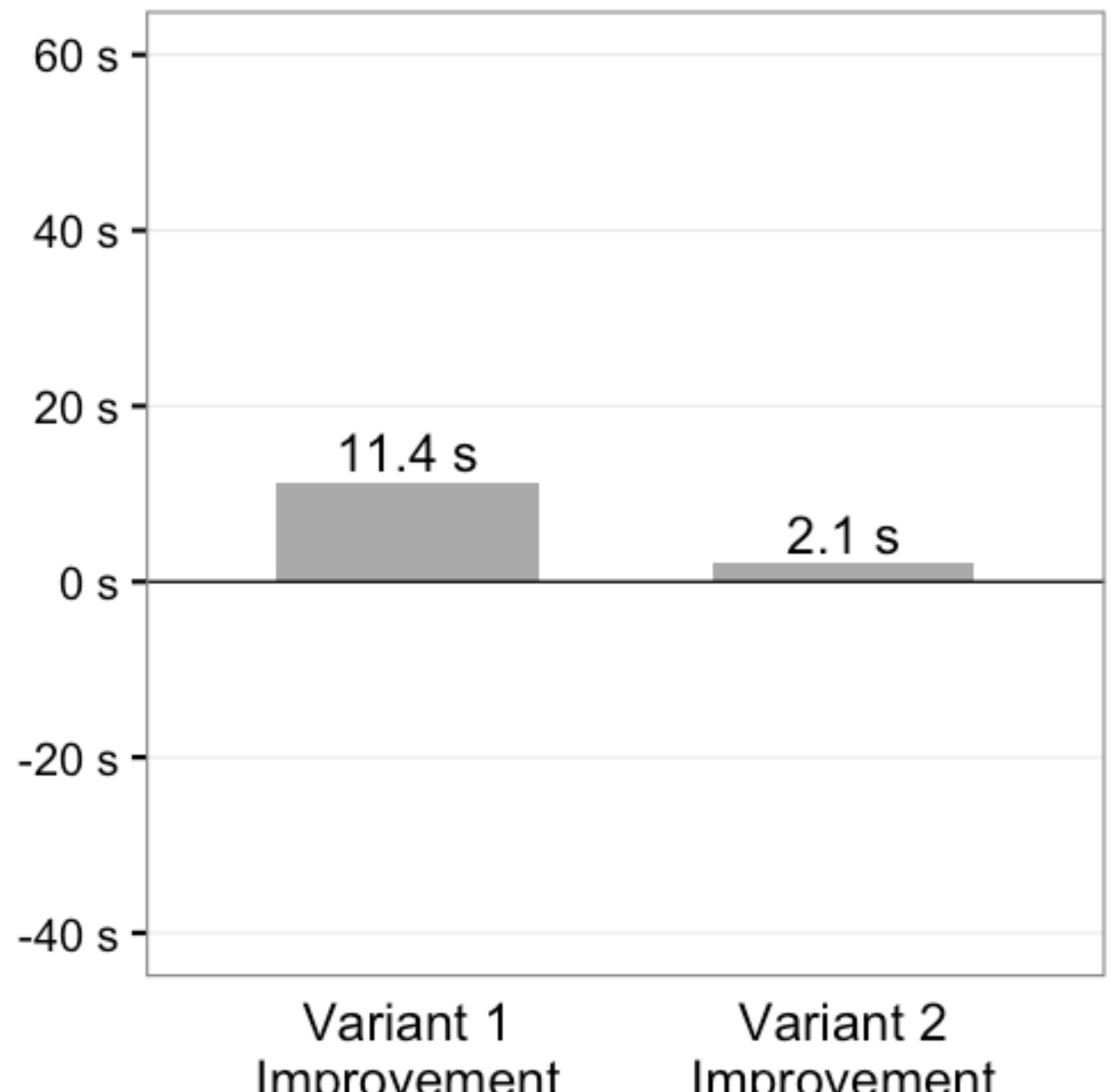


Dance of the sample means

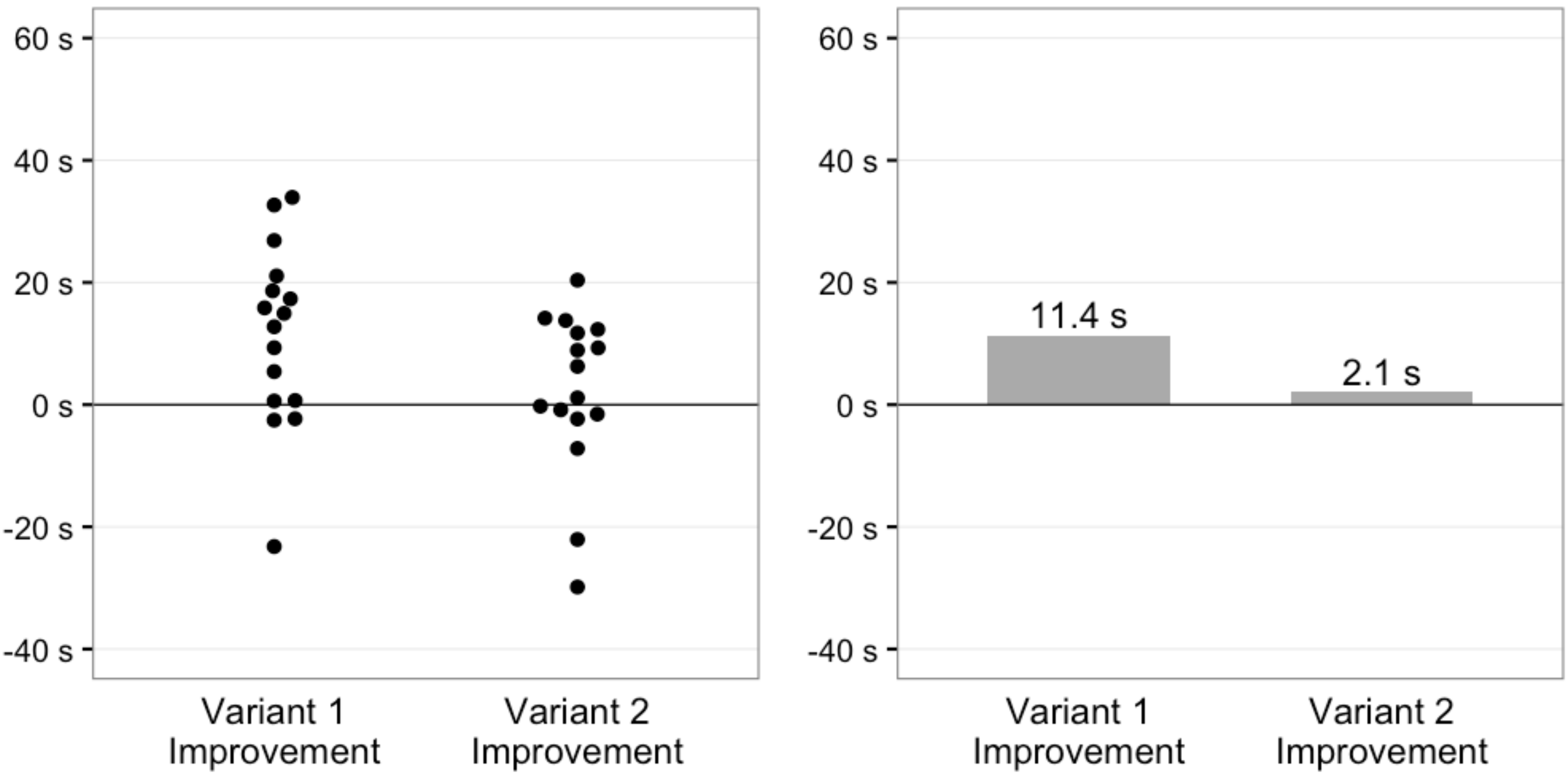


Dance of the sample means

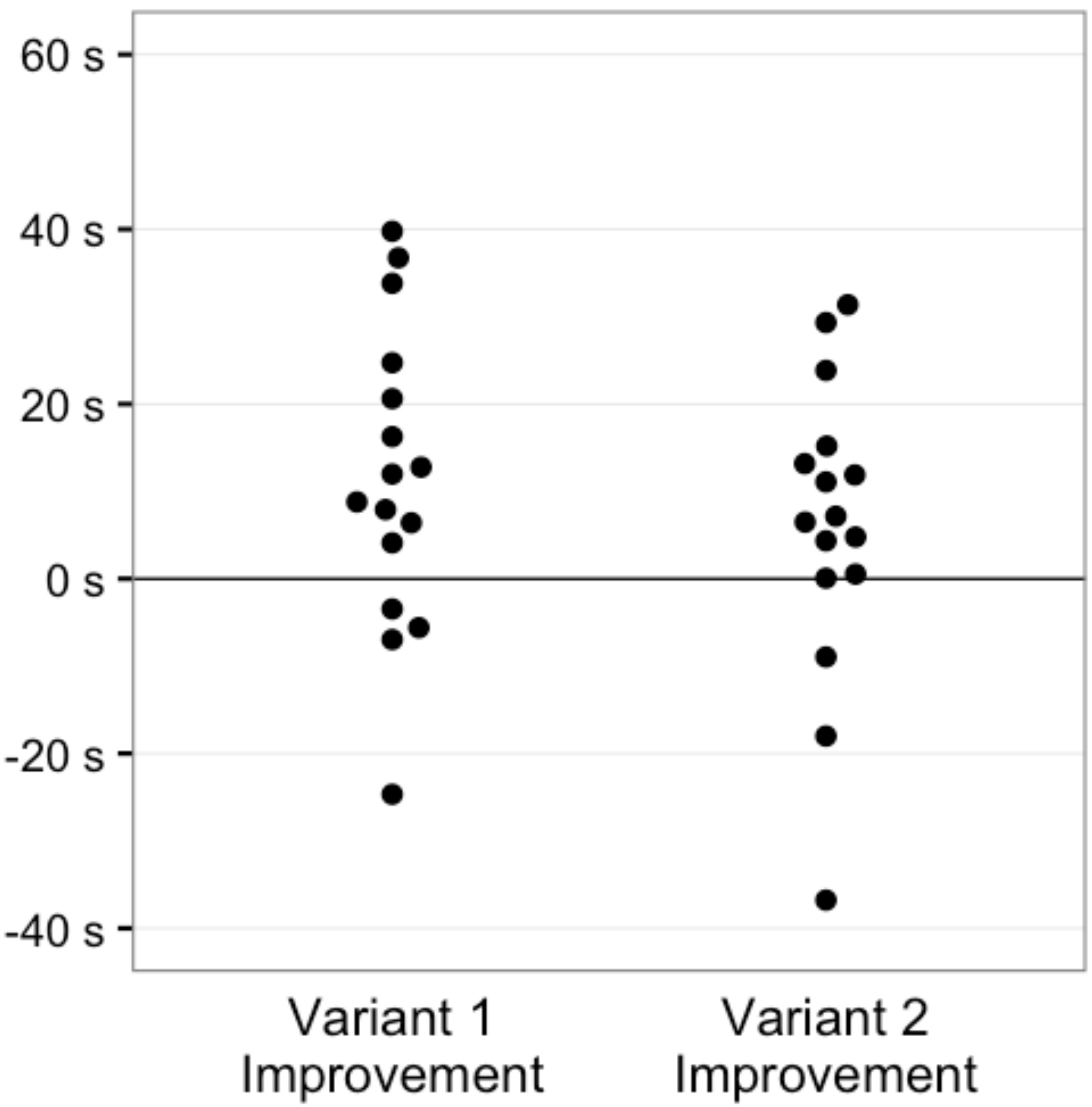


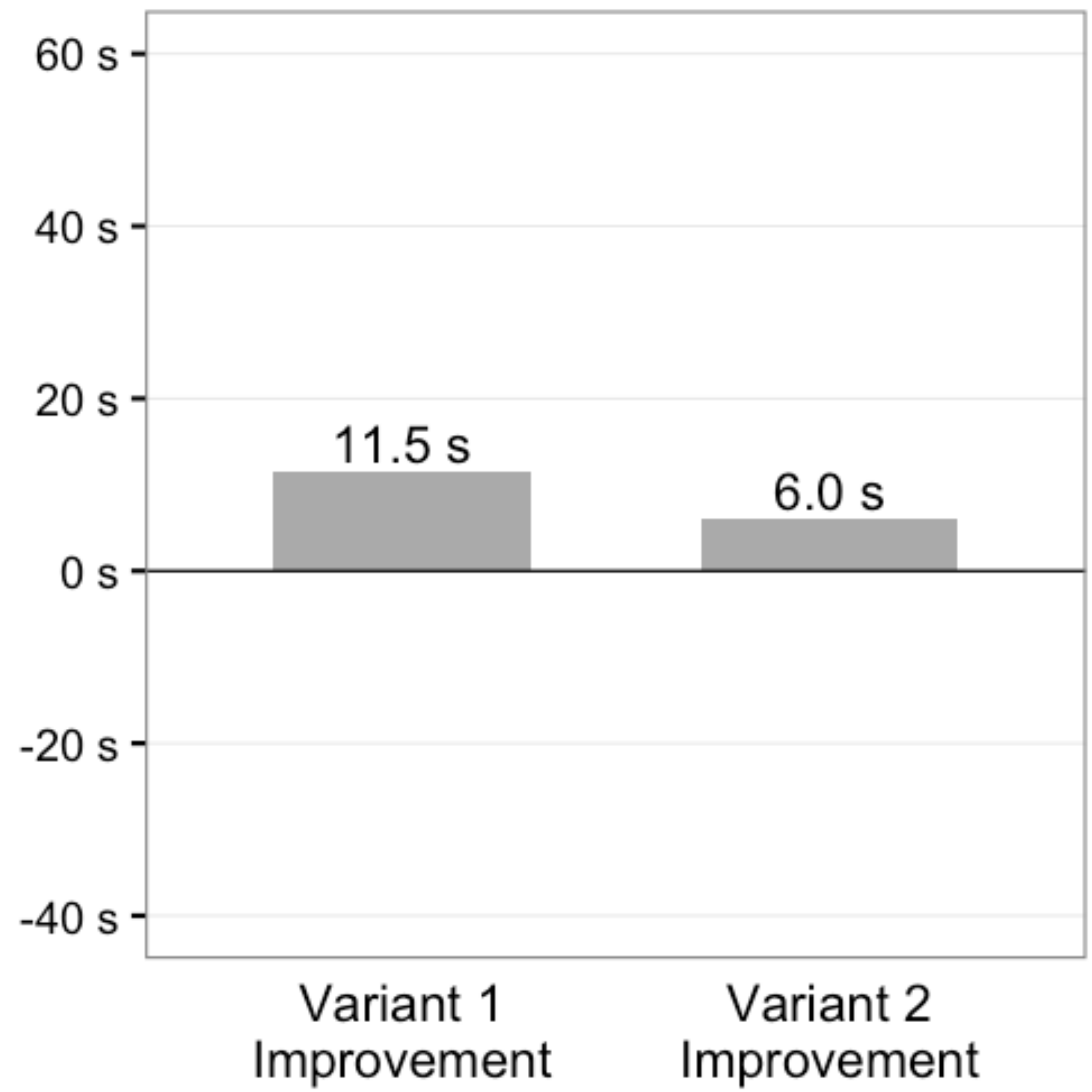


Dance of the sample means

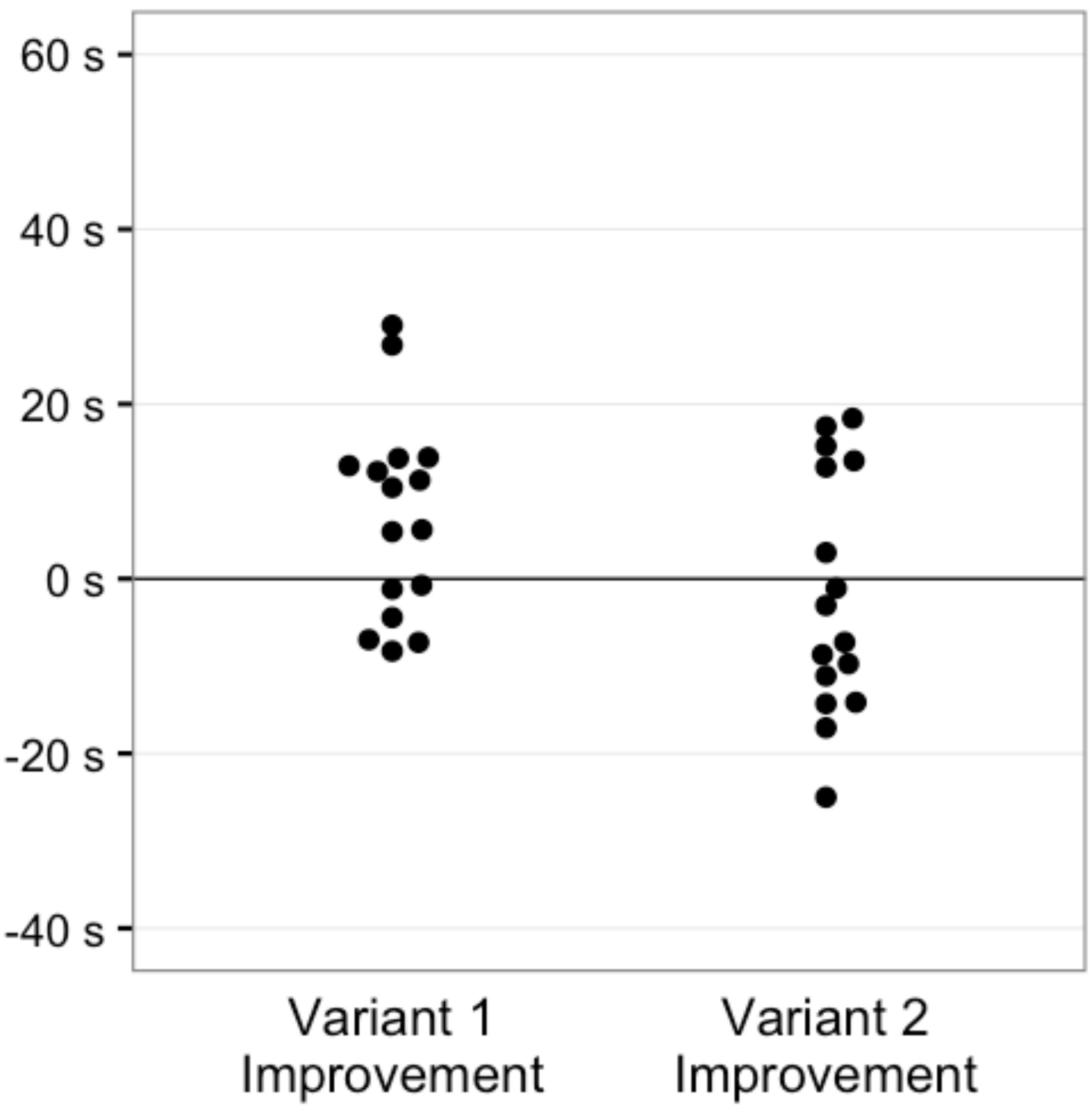


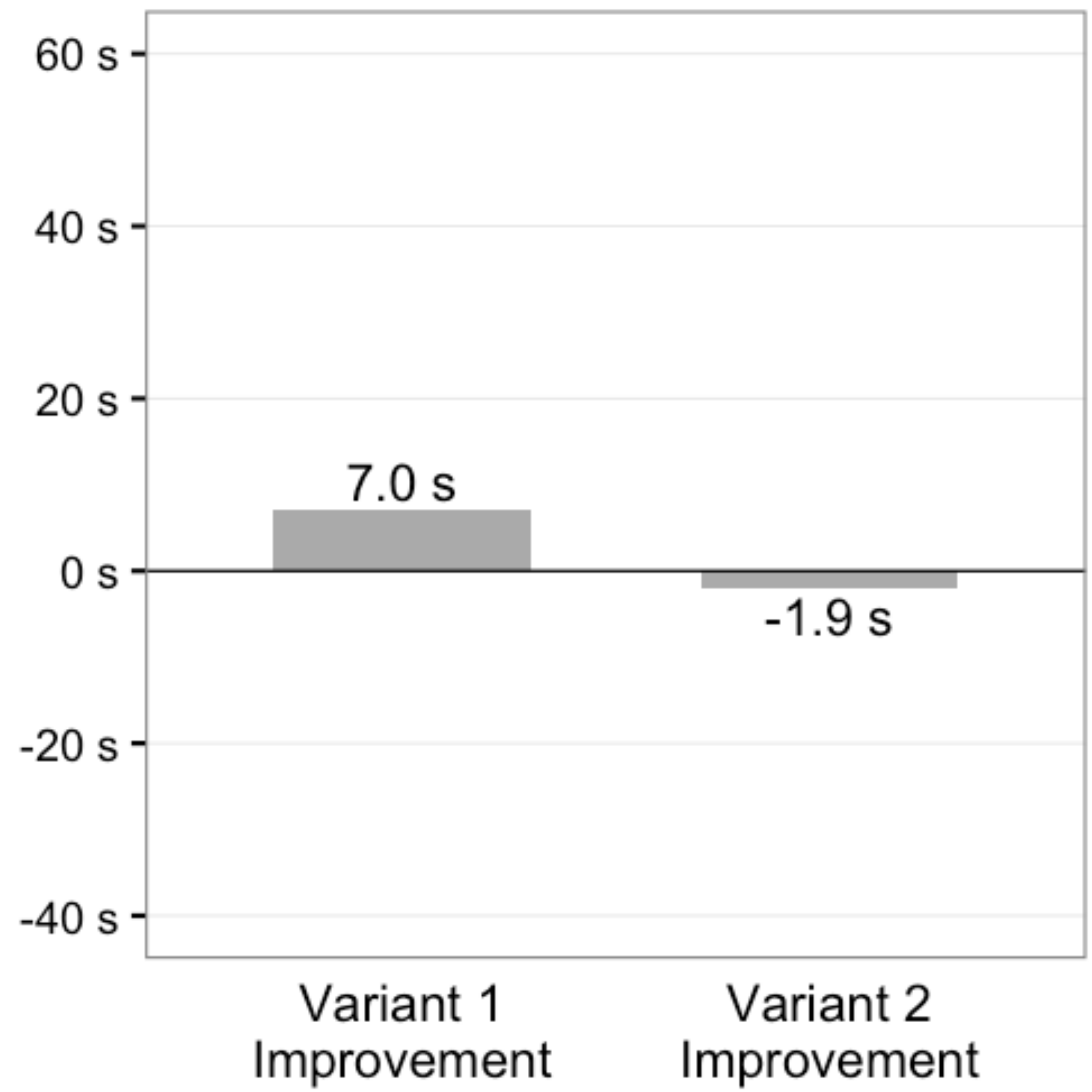
Dance of the sample means



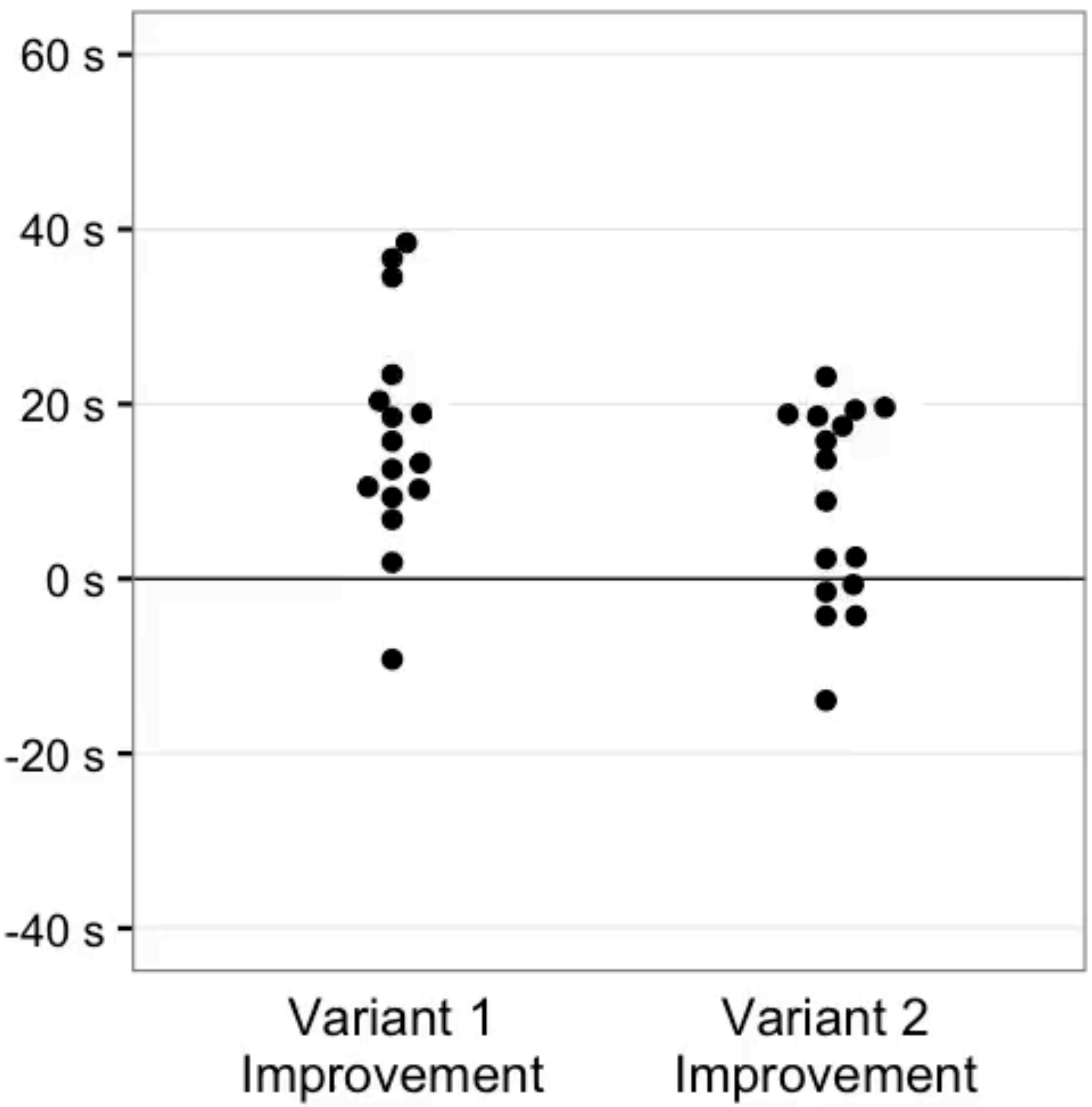


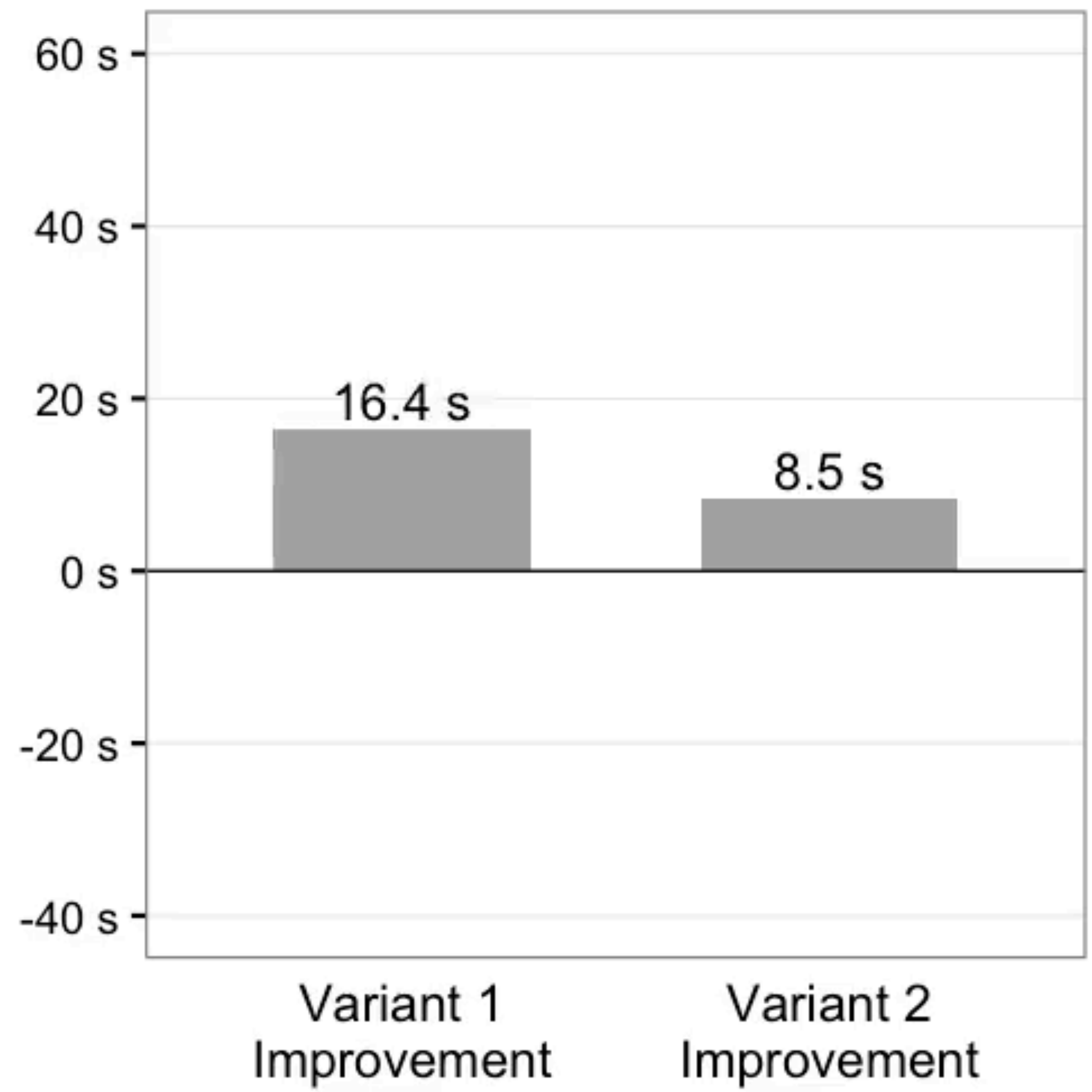
Dance of the sample means



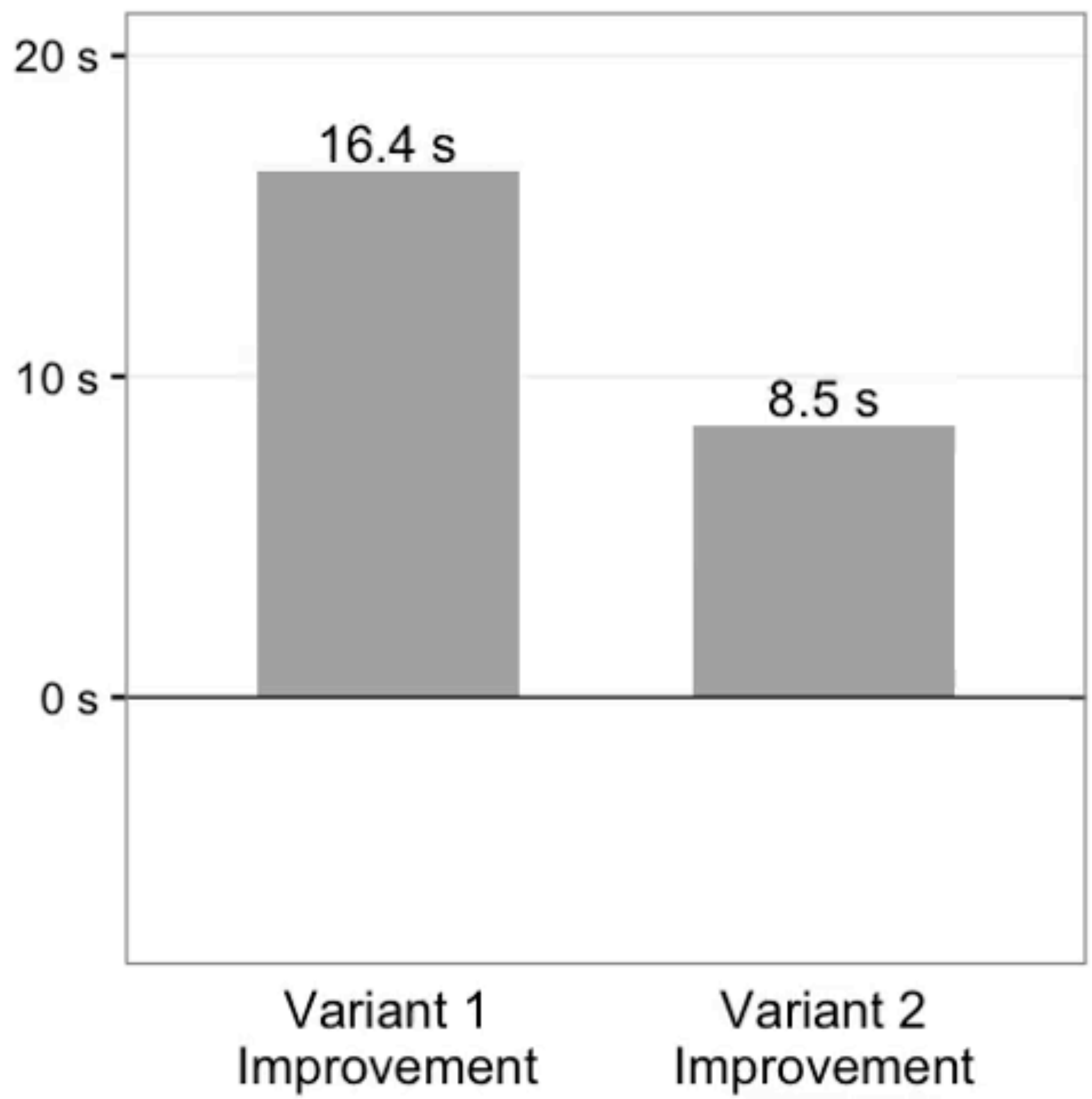


Dance of the sample means

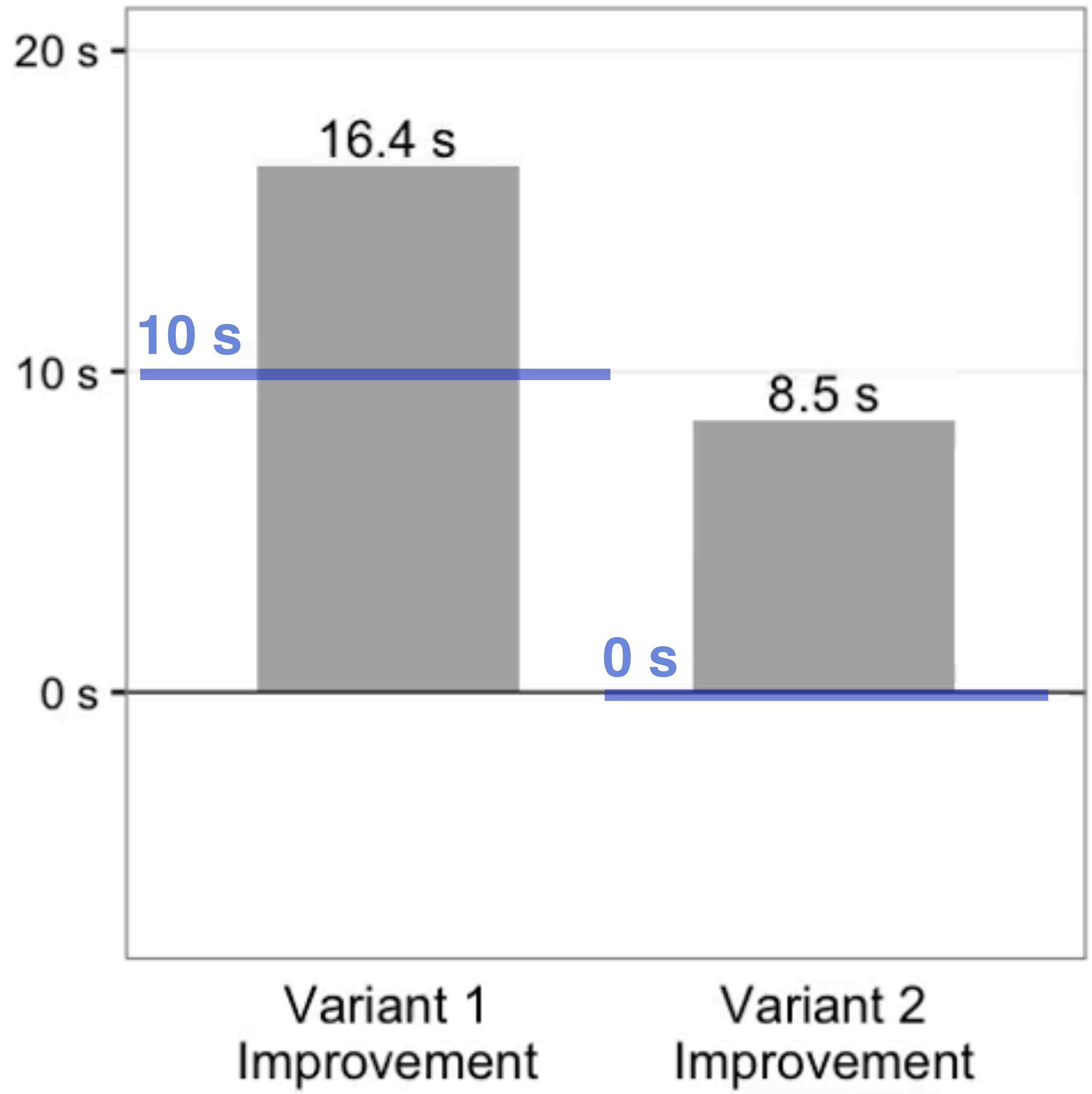




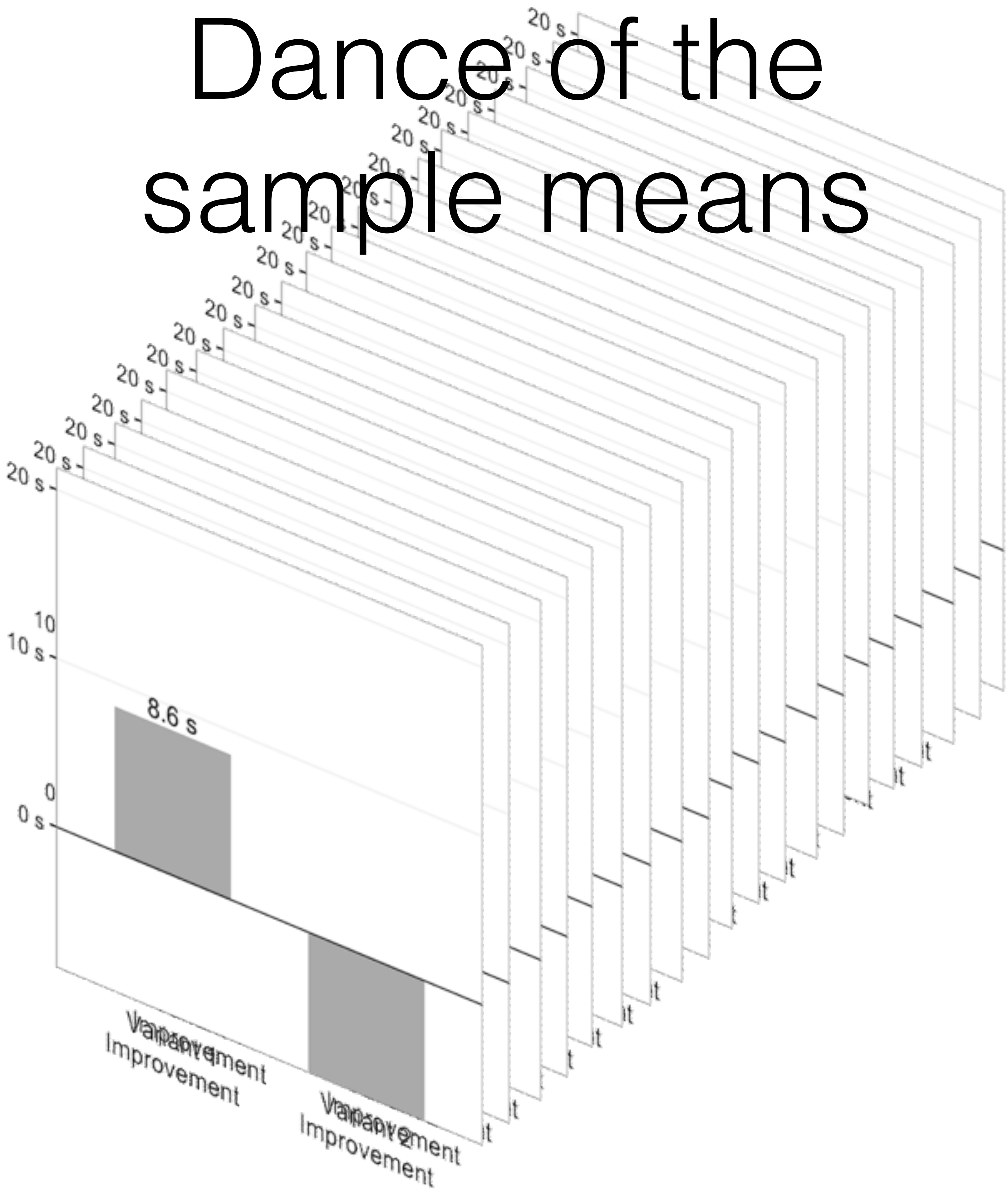
Dance of the sample means



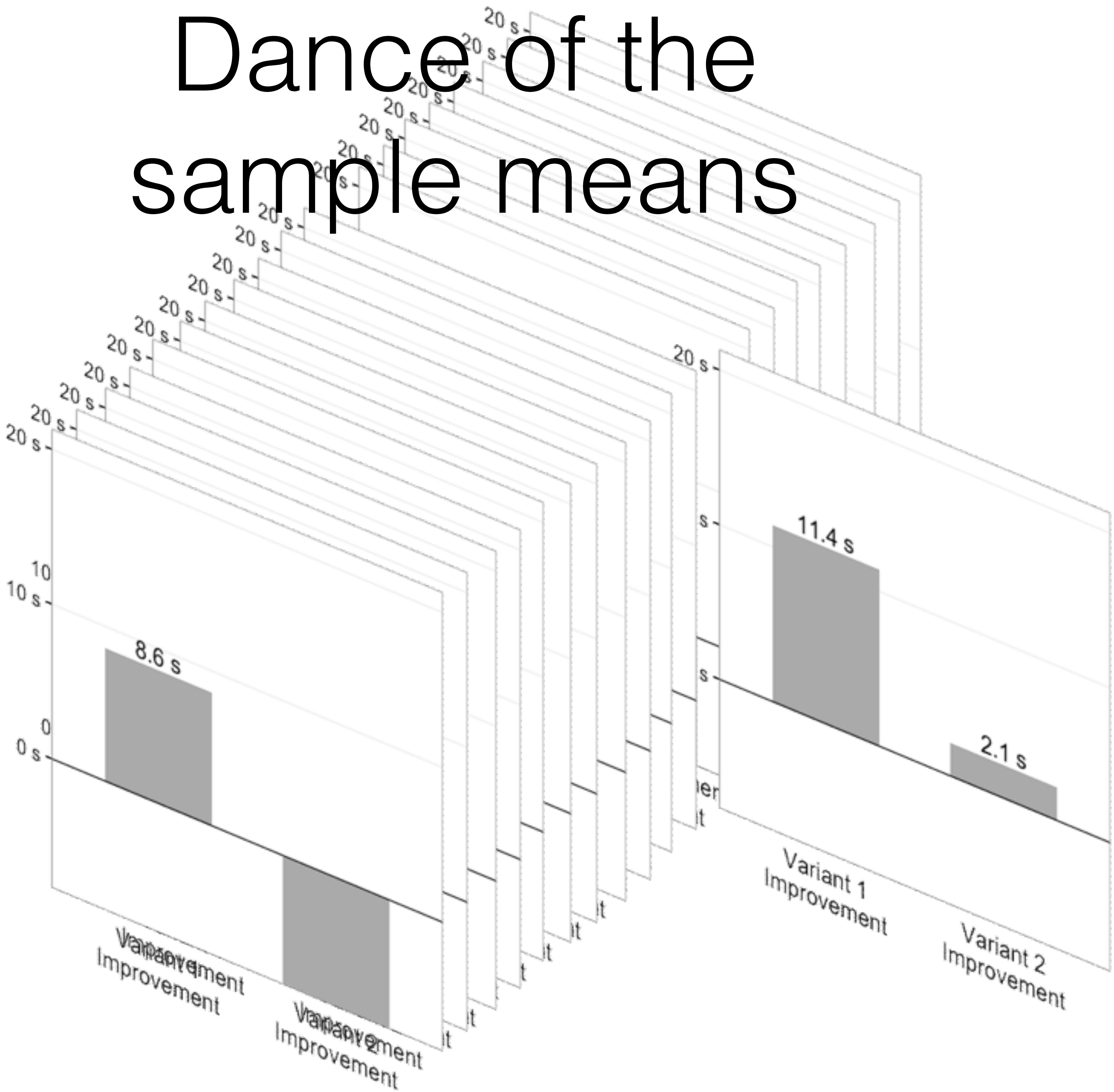
Dance of the sample means



Dance of the sample means



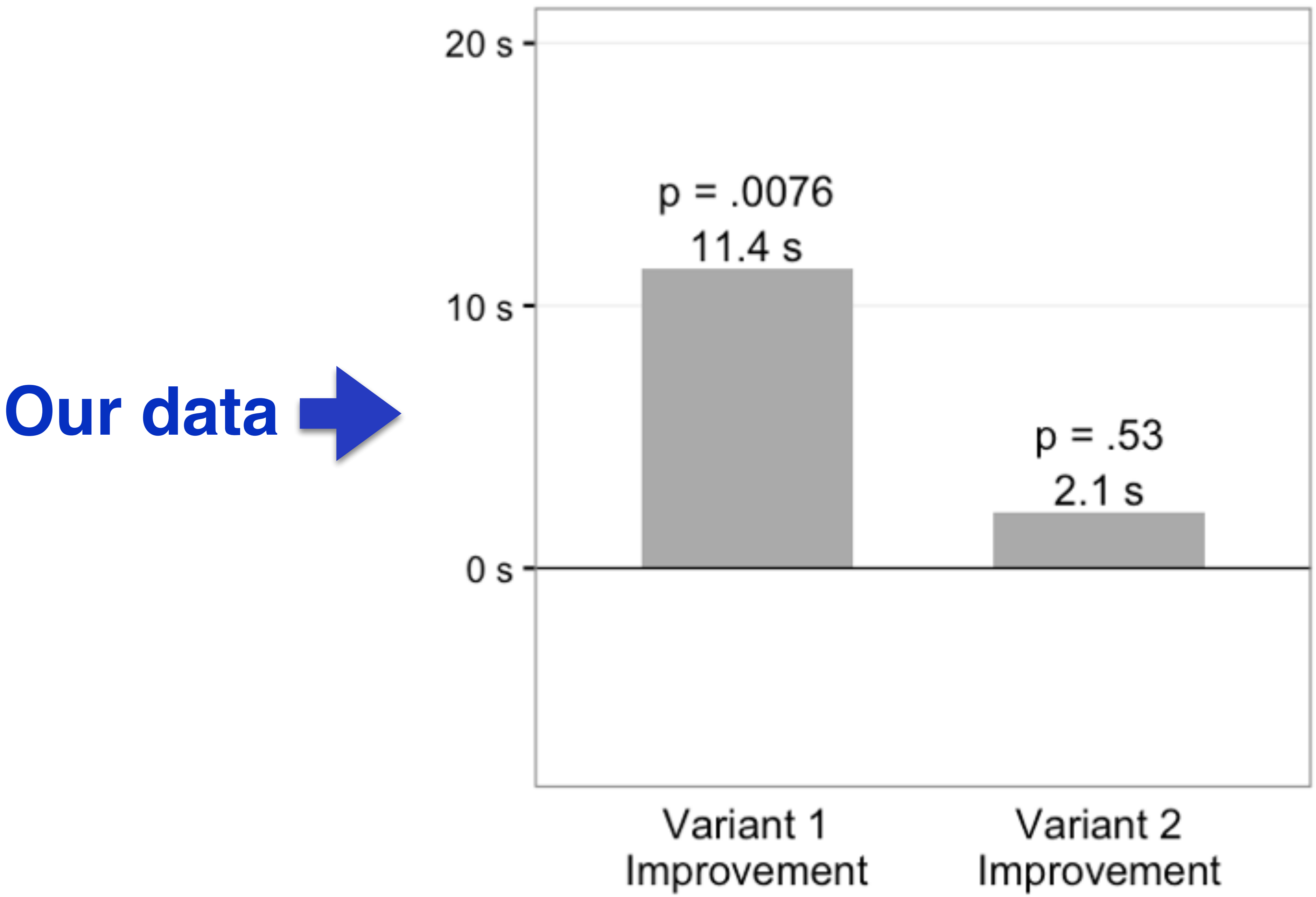
Dance of the sample means



Dance of the sample means

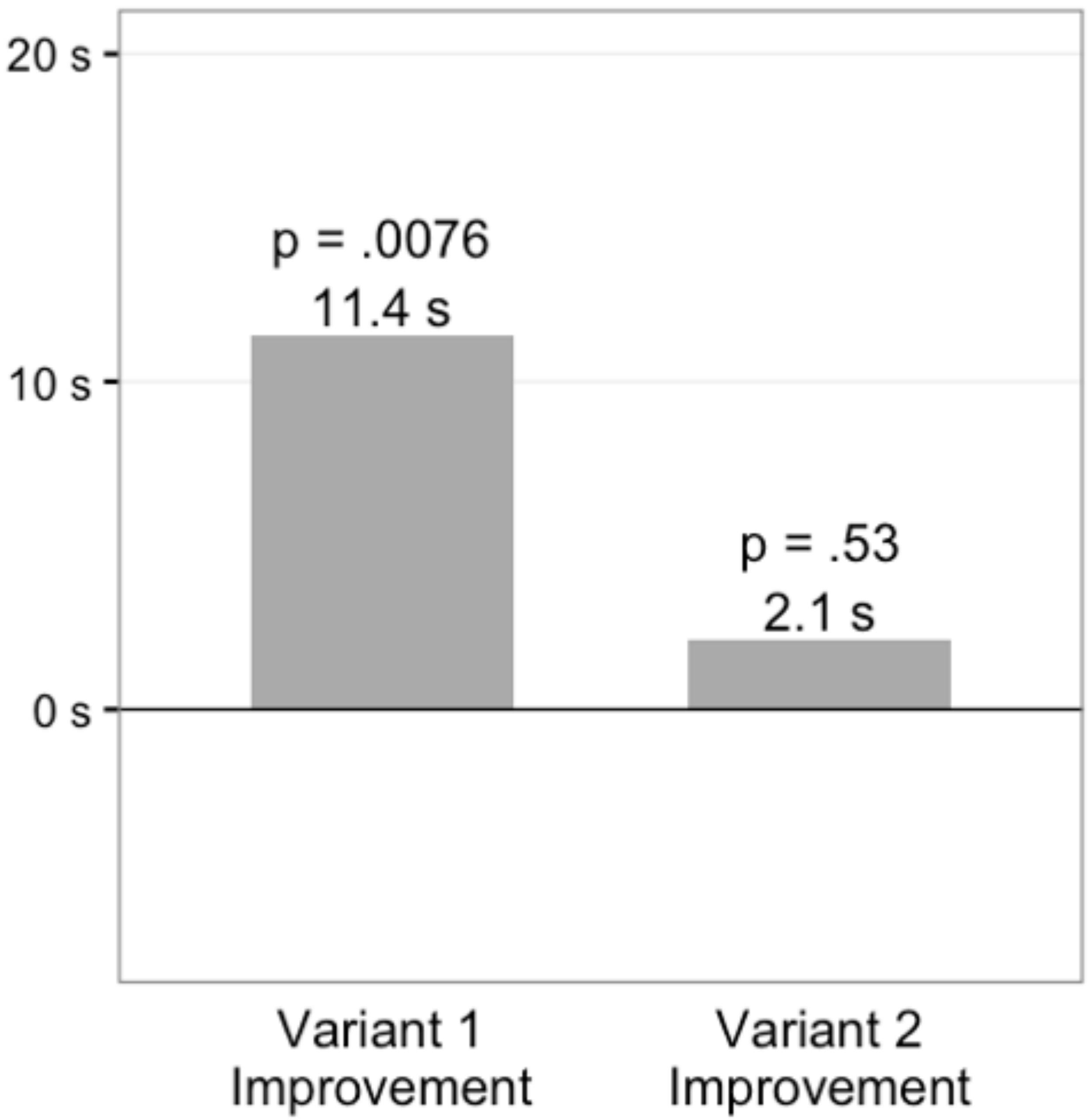
Our data

p-values

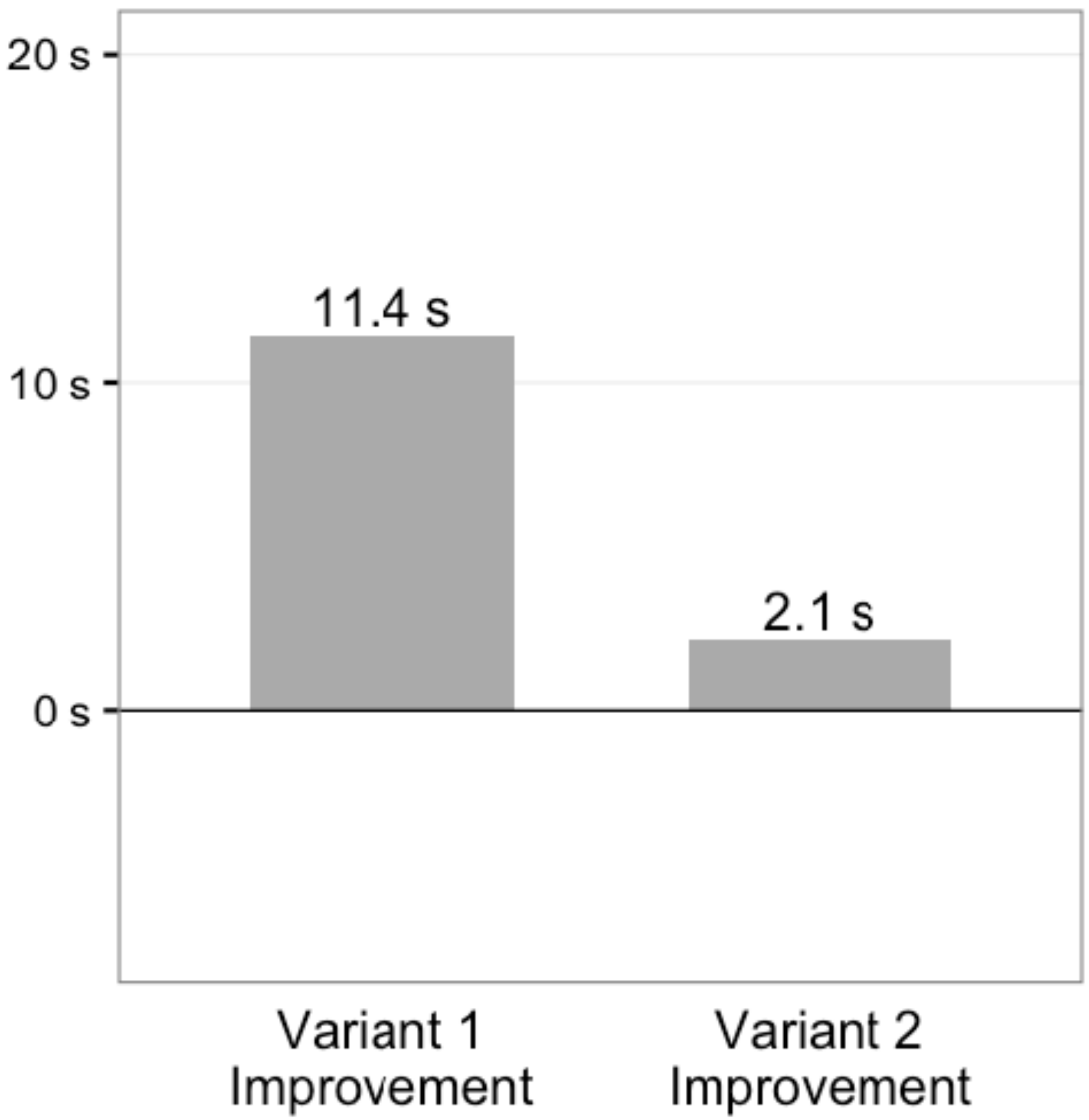


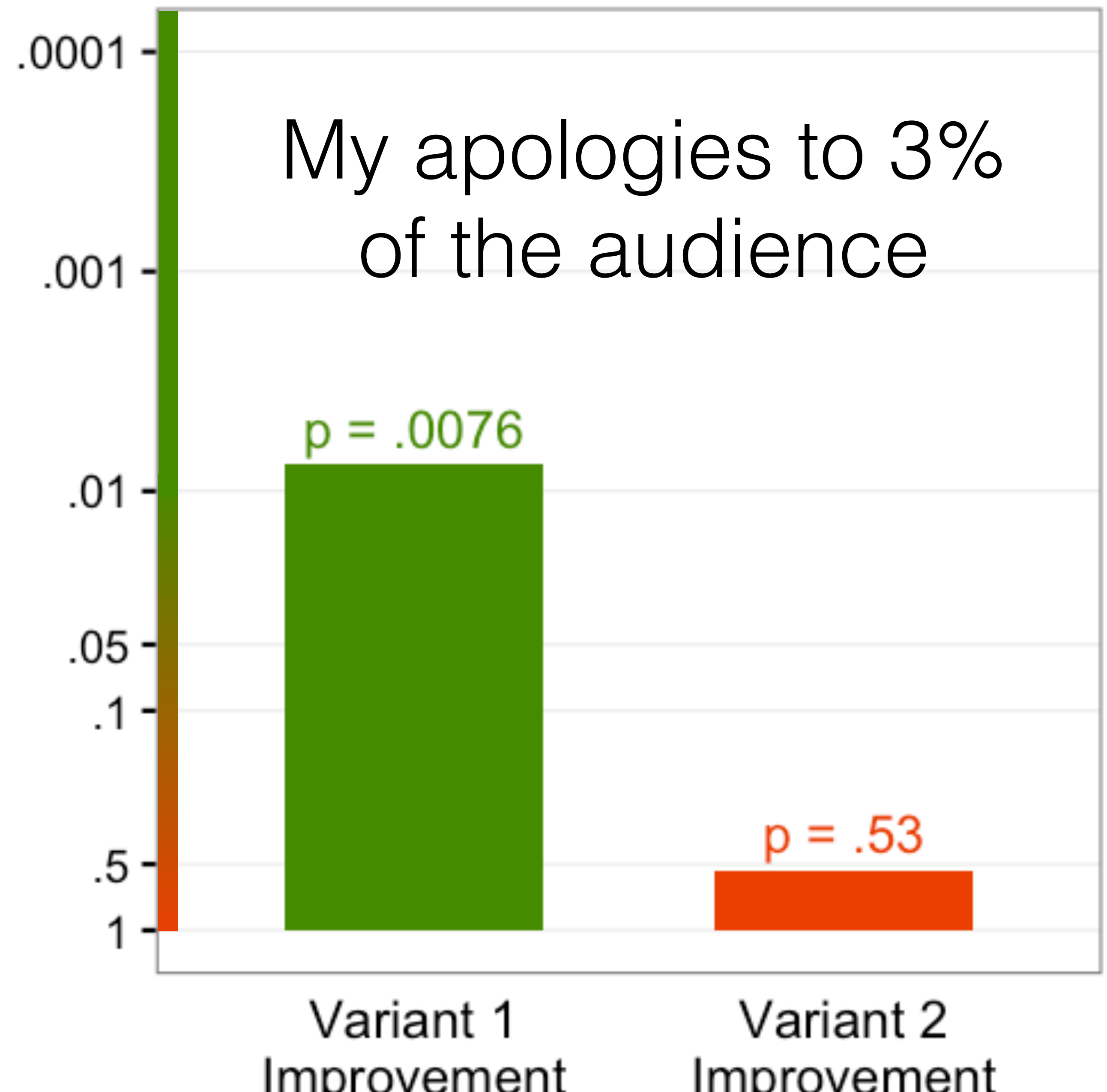
Our data
20 s
10 s
0 s
p = .0076
11.4 s
p = .53
2.1 s
Variant 1 Improvement
Variant 2 Improvement

Dance of the p-values

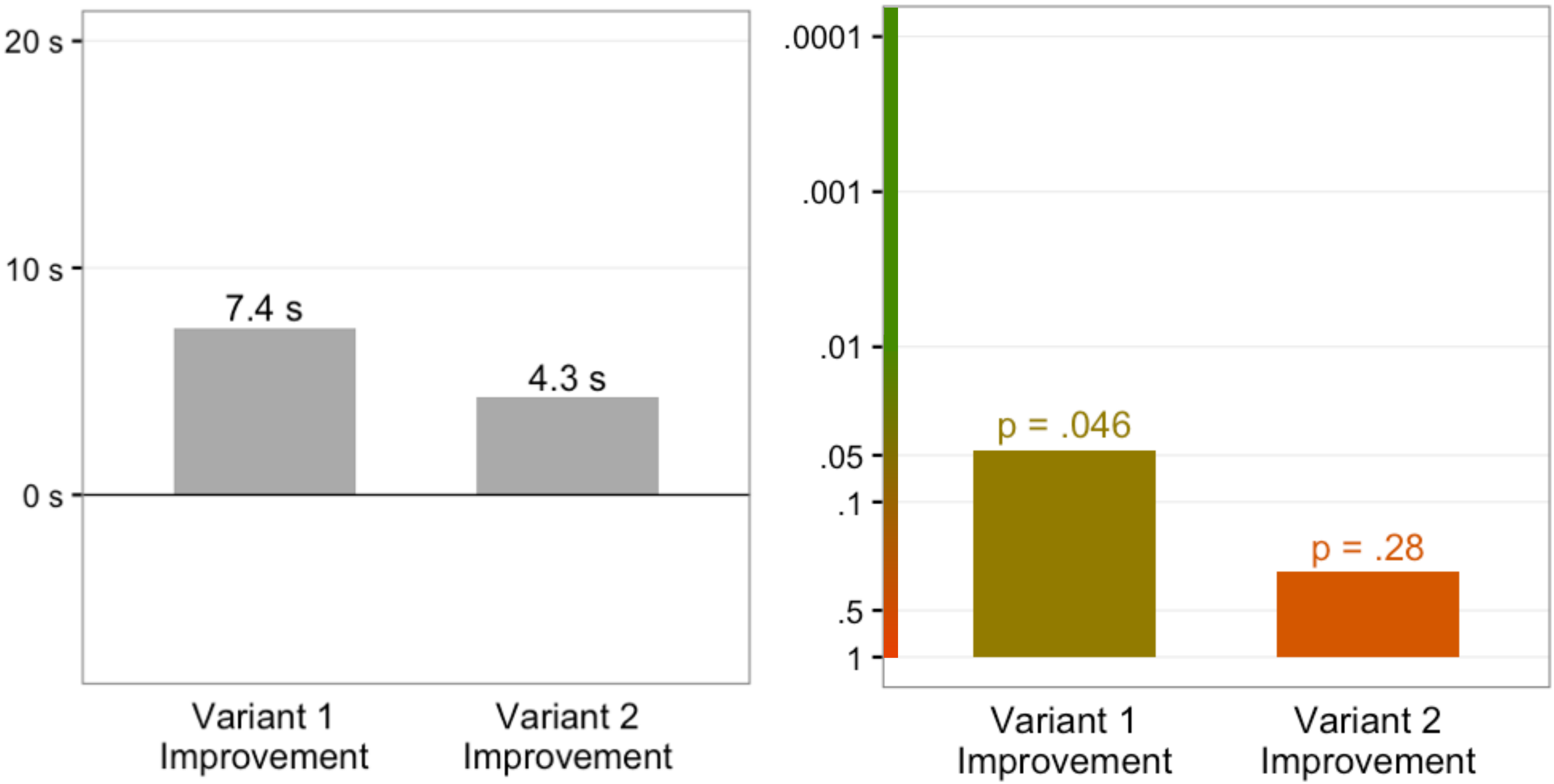


Dance of the p-values

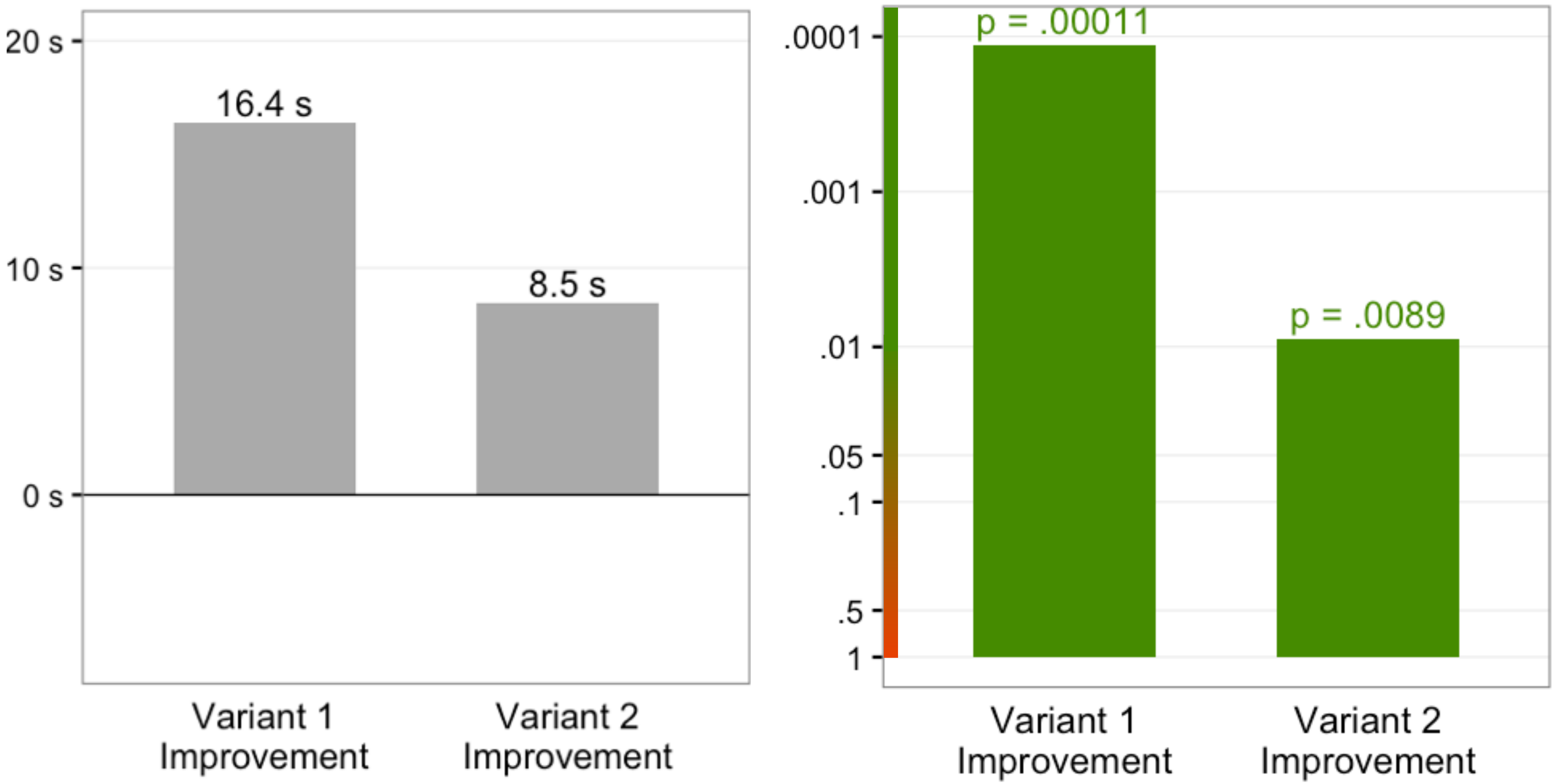




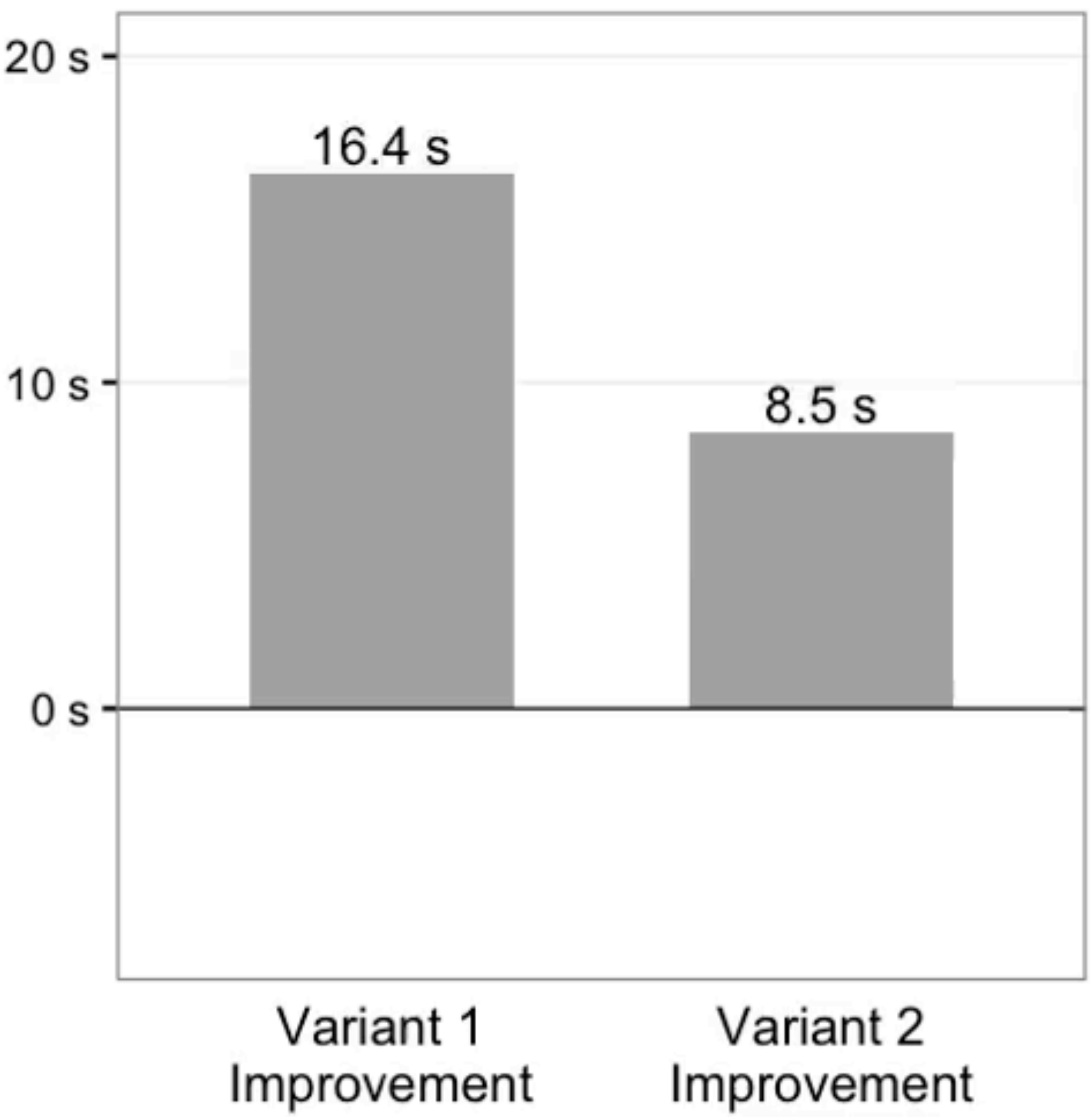
Dance of the p-values

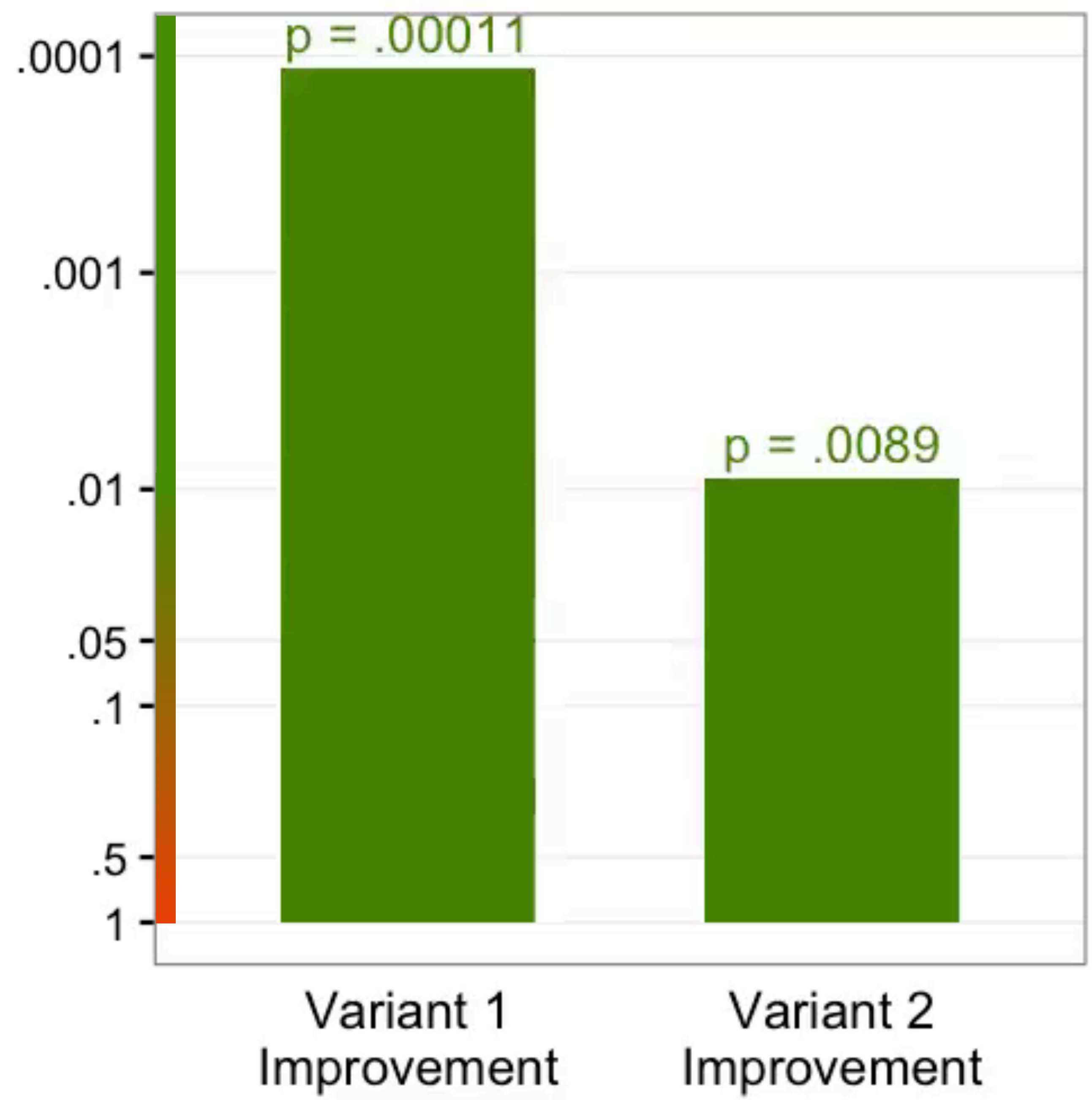


Dance of the p-values

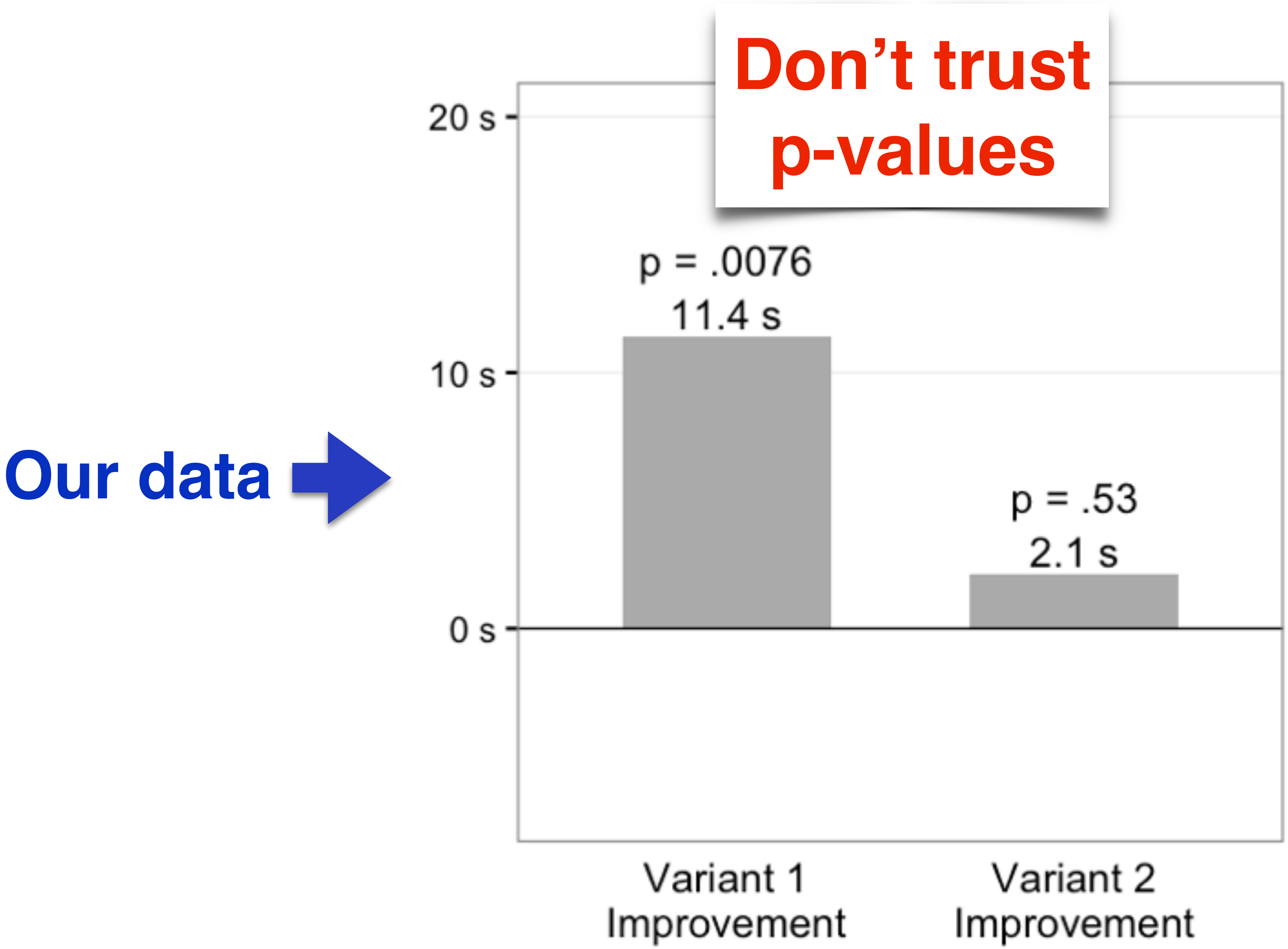


Dance of the p-values





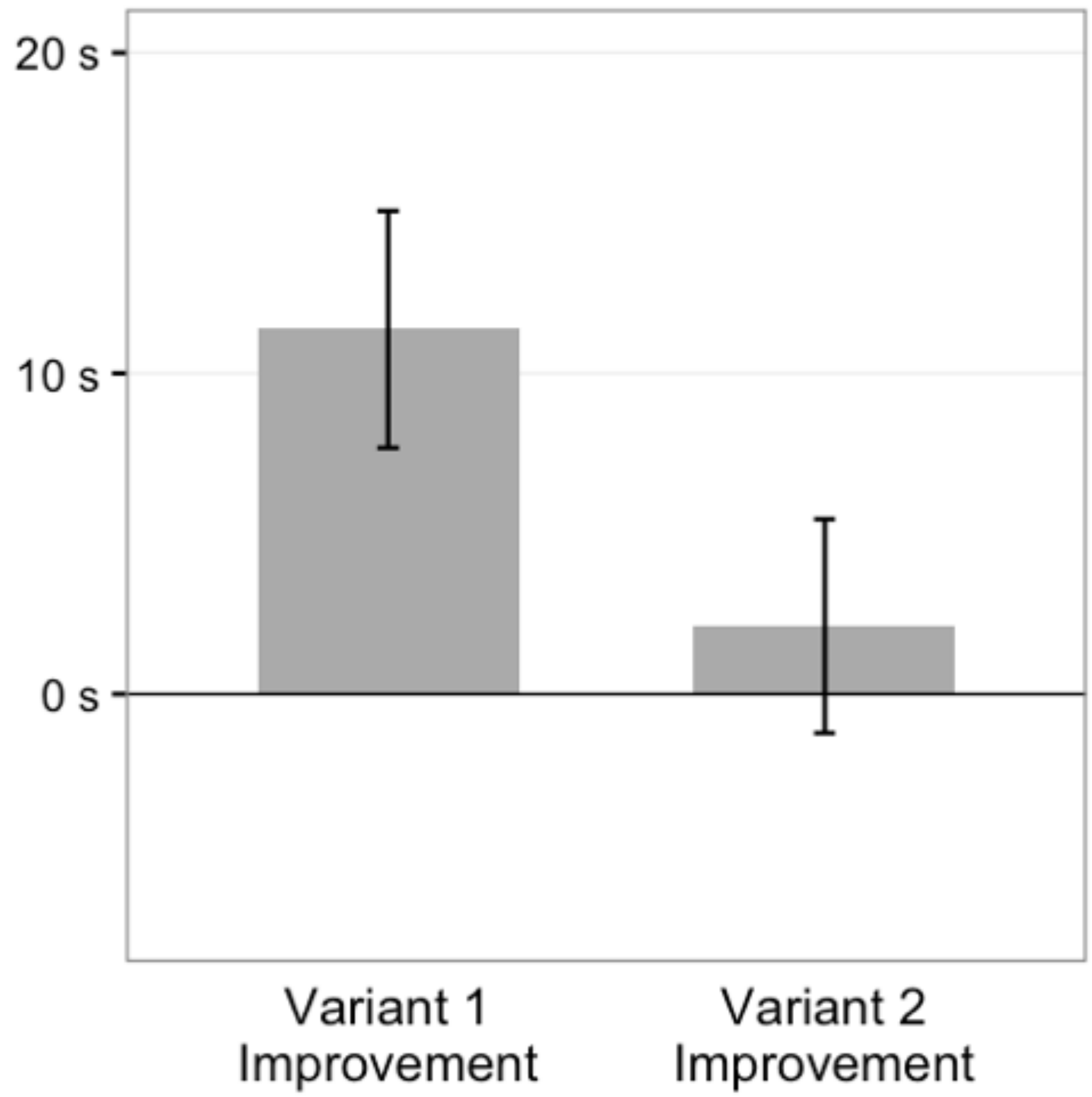
p-values



Don't trust p-values
Our data
20 s
10 s
0 s
p = .0076
11.4 s
p = .53
2.1 s
Variant 1 Improvement
Variant 2 Improvement

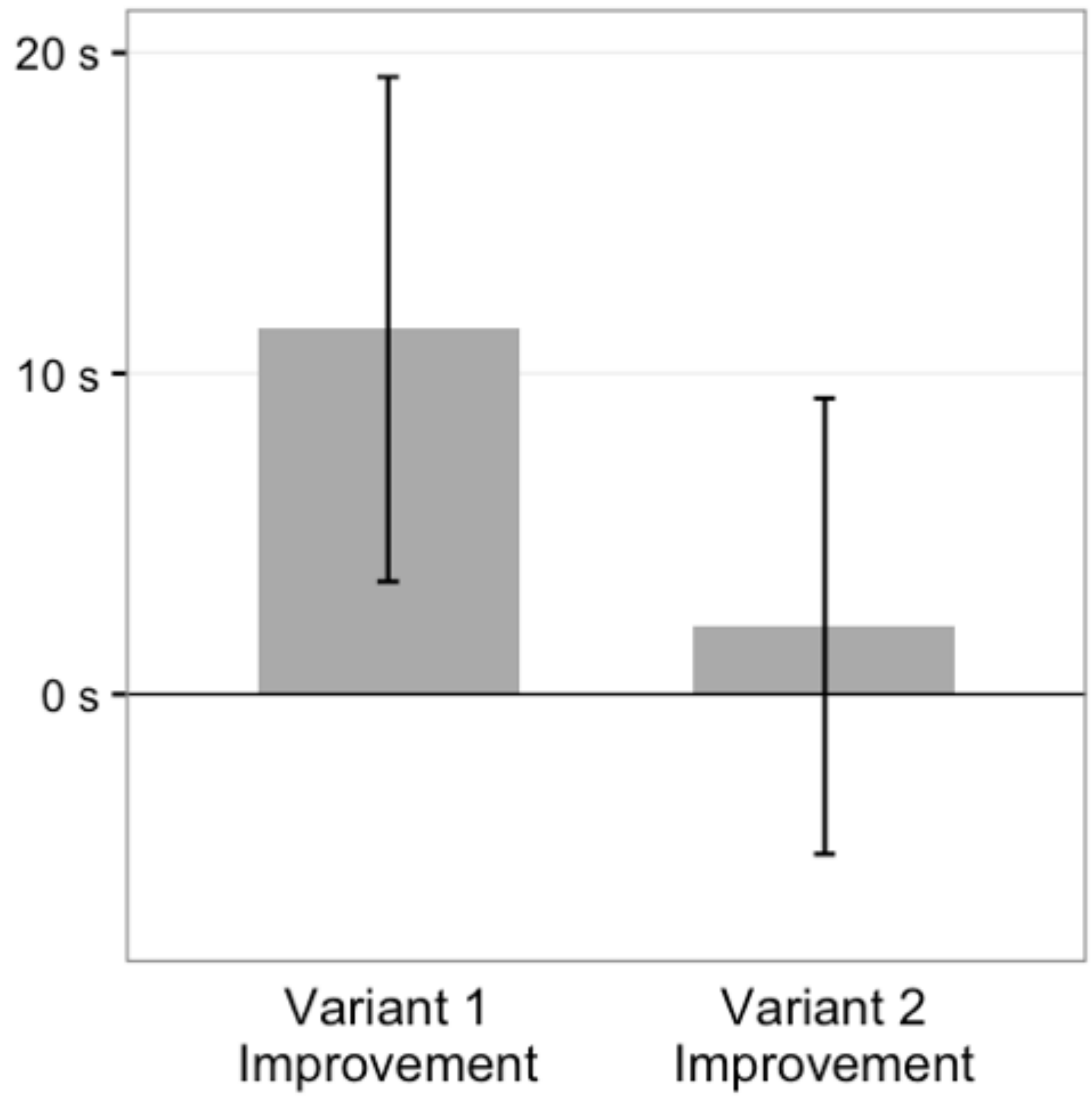
Standard errors

Our data



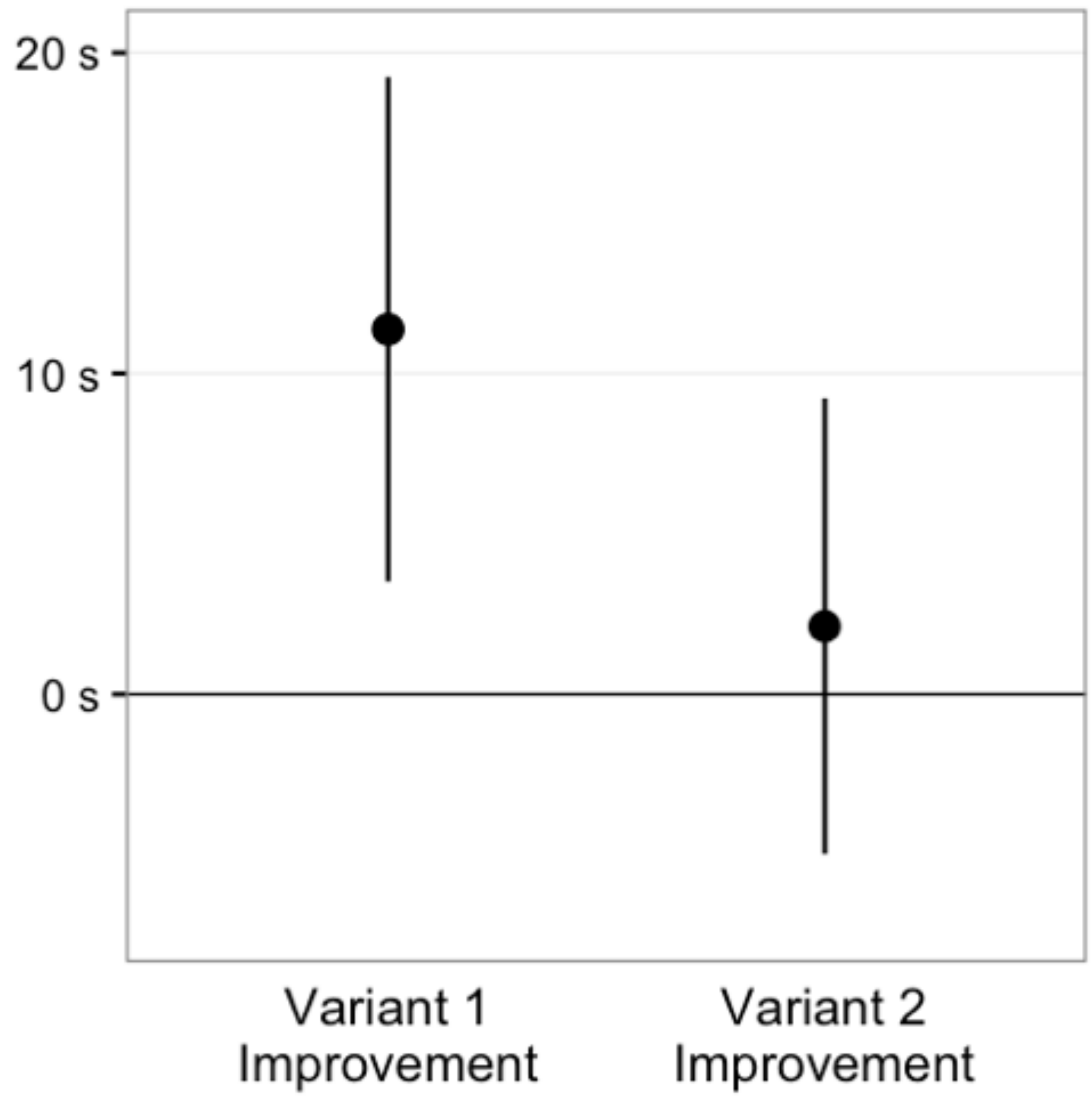
95% confidence intervals

Our data

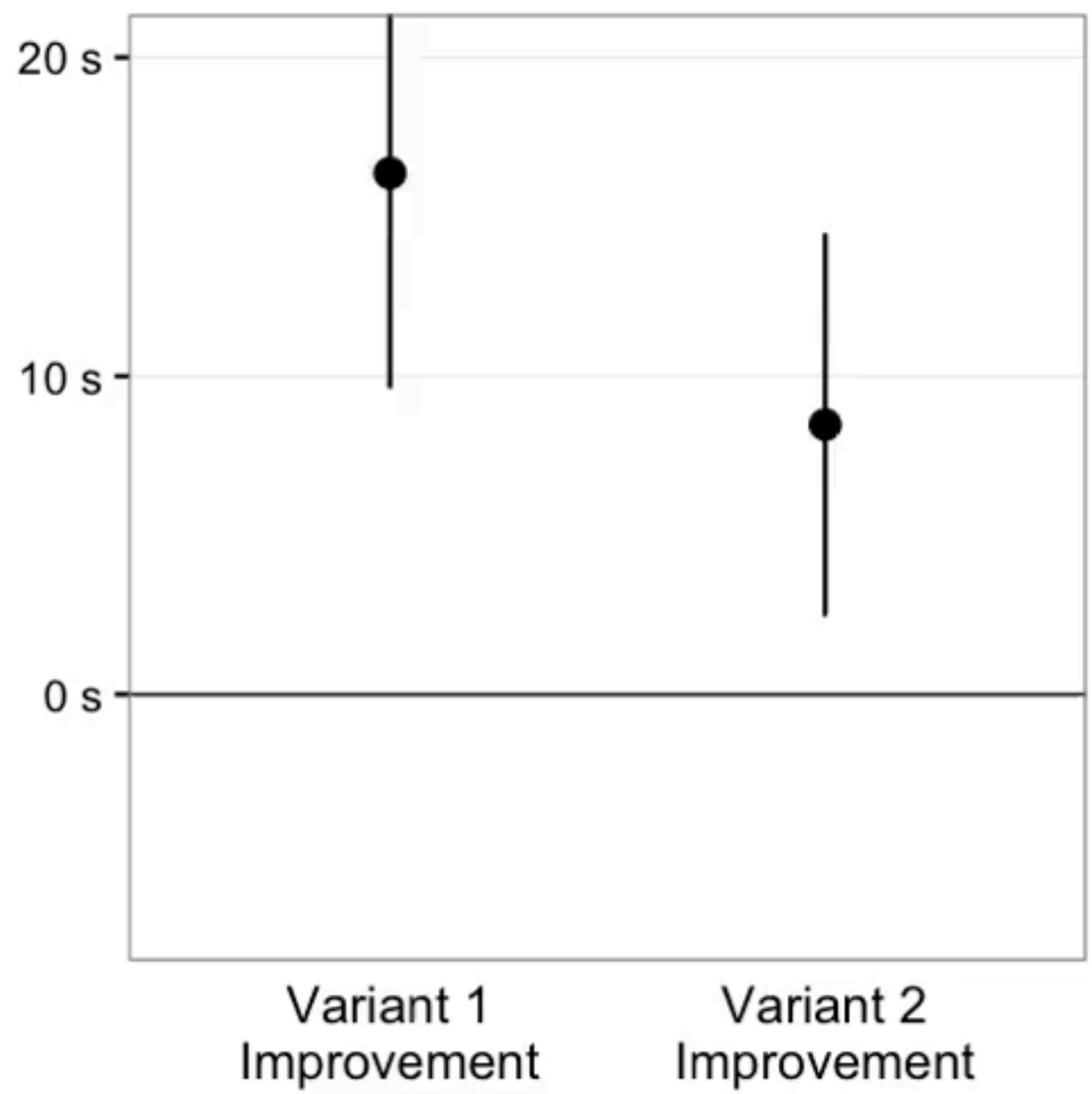


95% confidence intervals

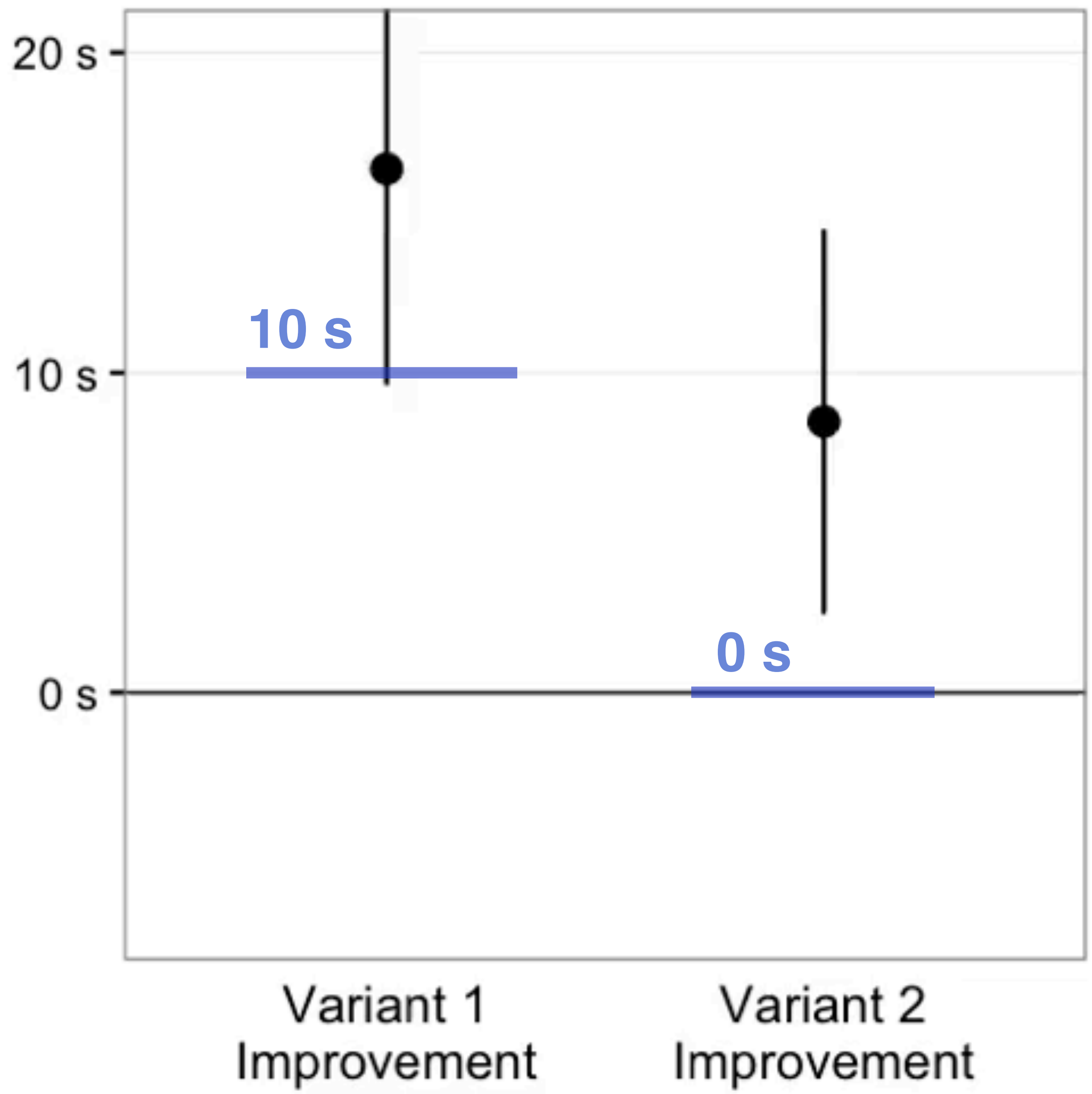
Our data



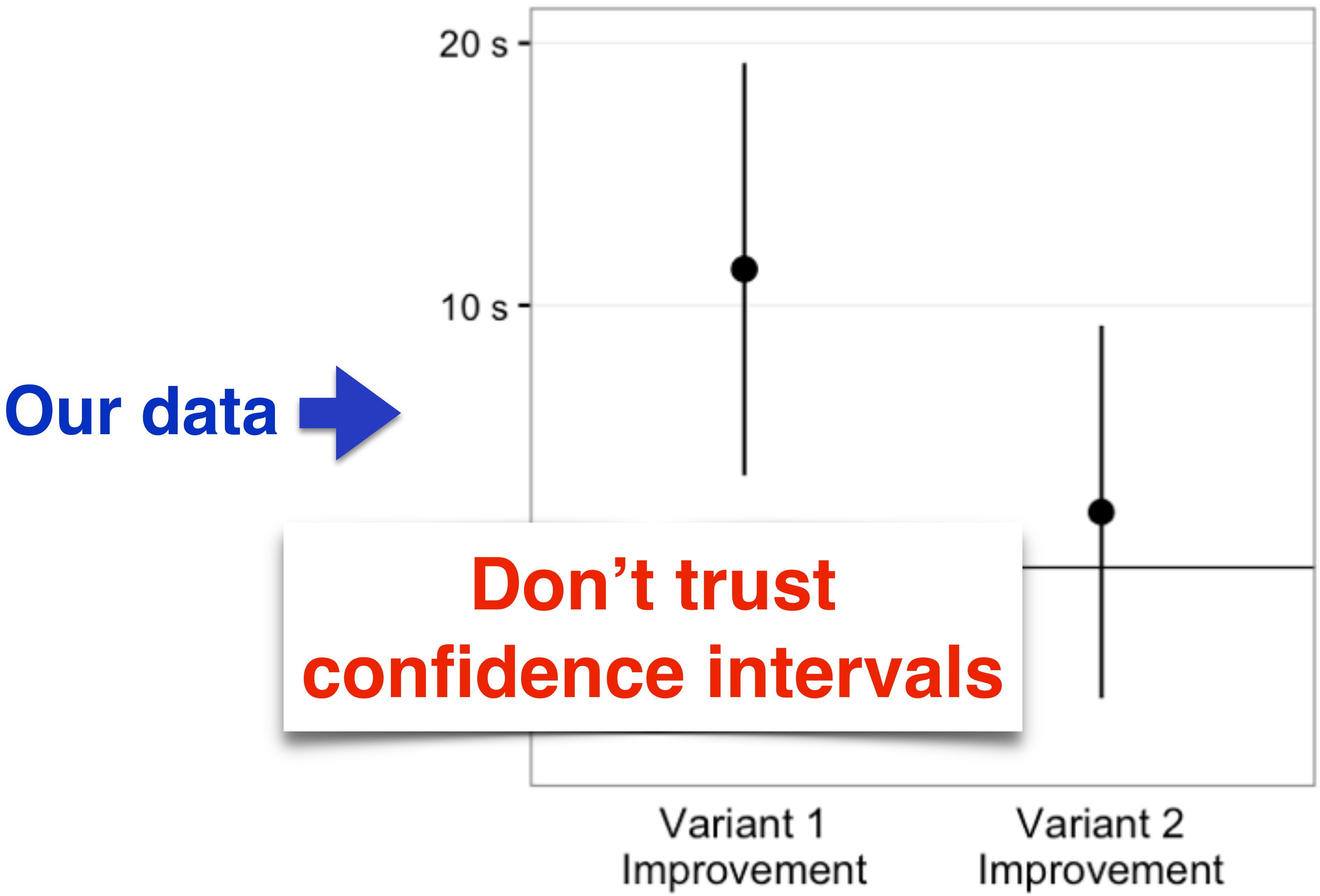
Dance of the confidence intervals



Dance of the confidence intervals

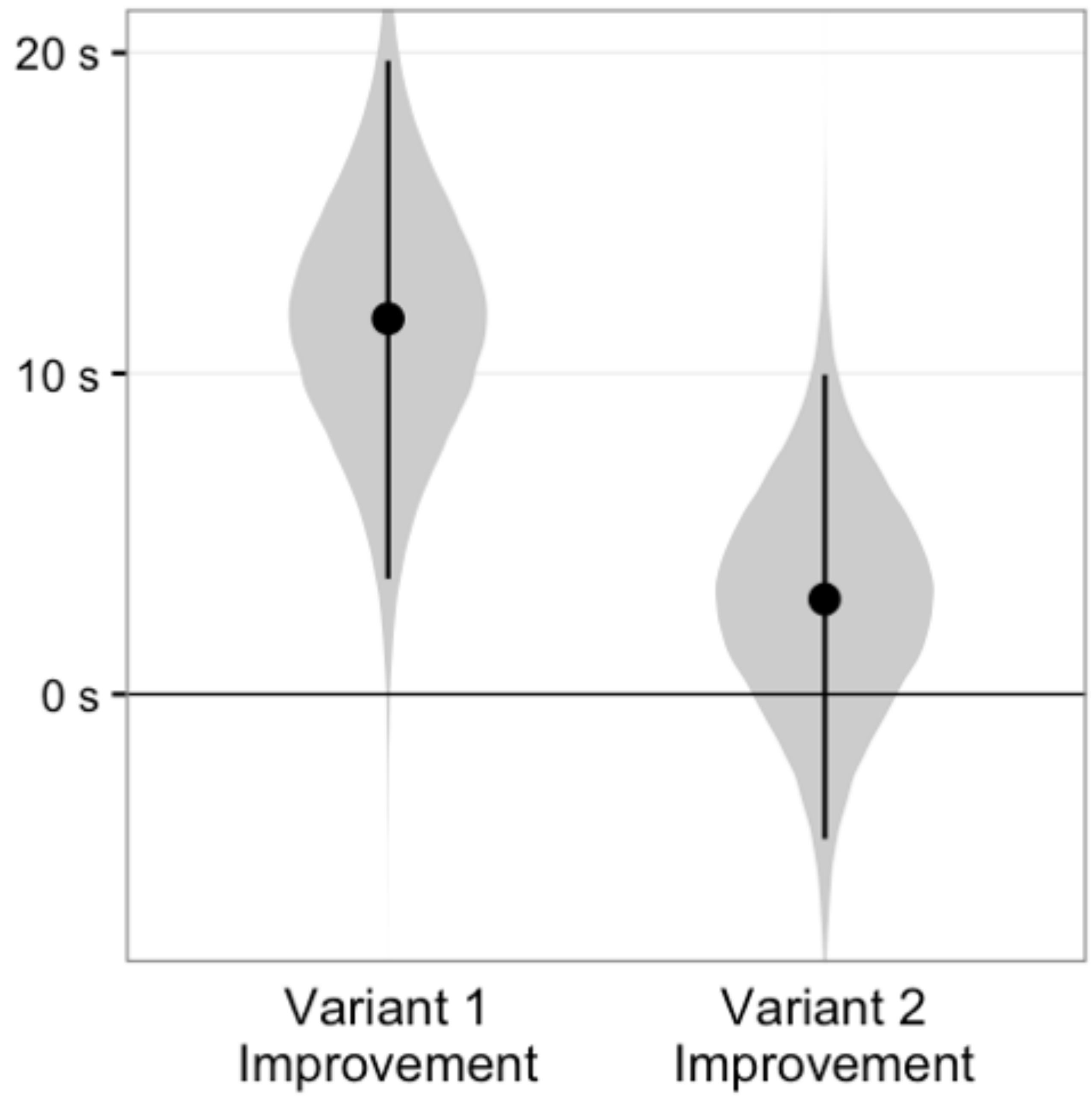


95% confidence intervals



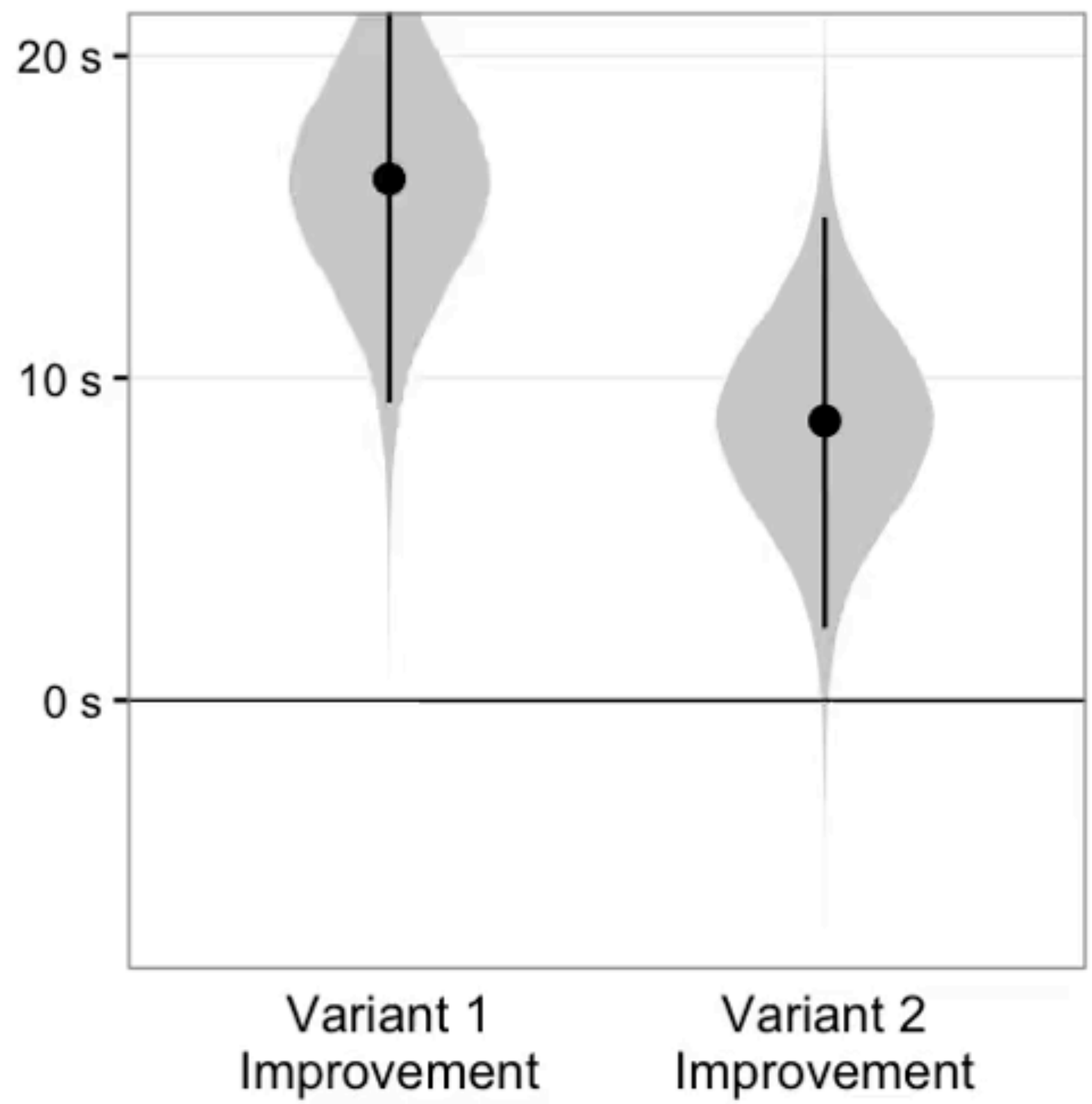
Bayesian credible intervals and posterior distributions

Our data



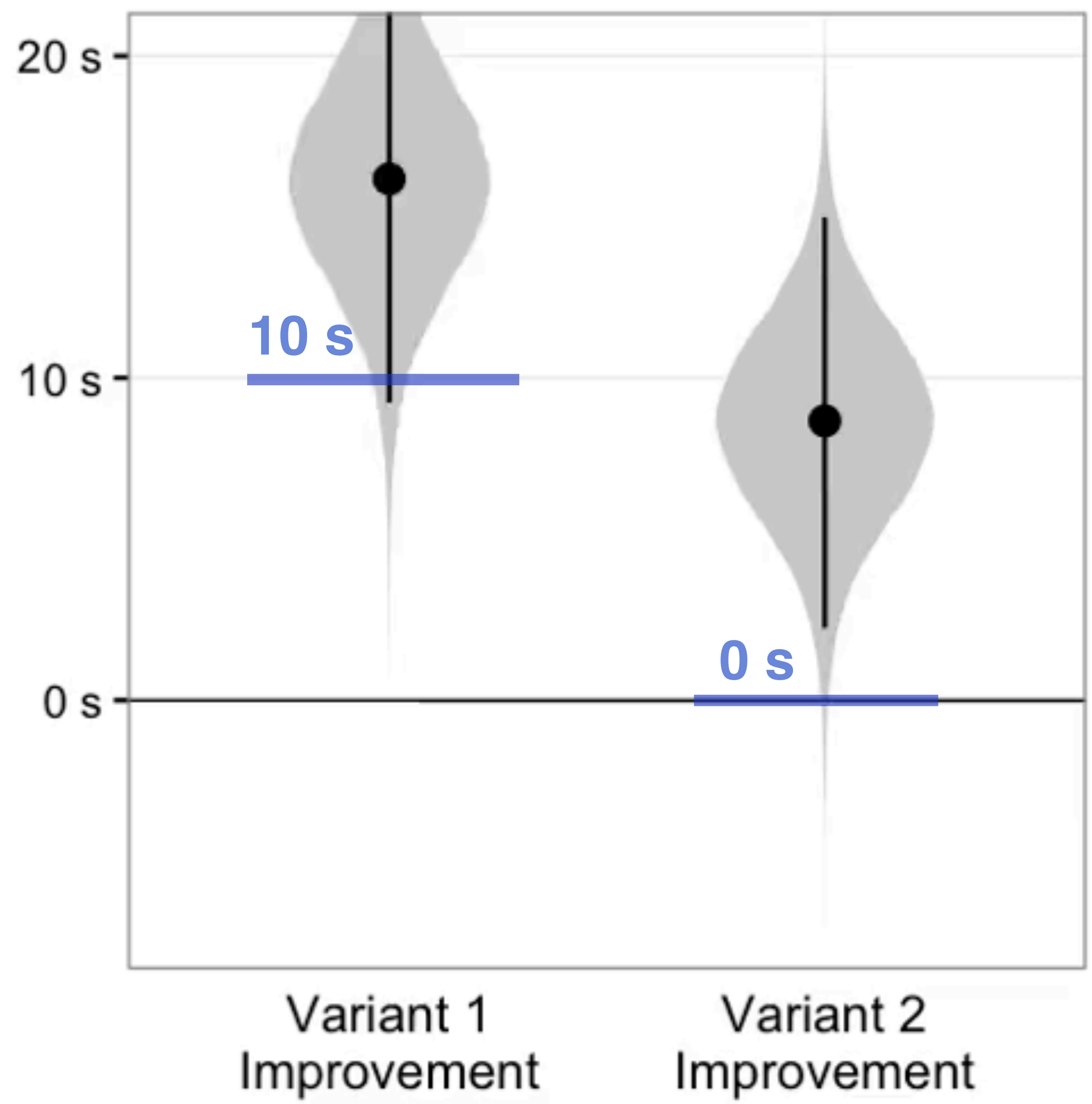
“Eye plots” made using (Kay, 2016)

Bayesian dance



"Eye plots" made using (Kay, 2016)

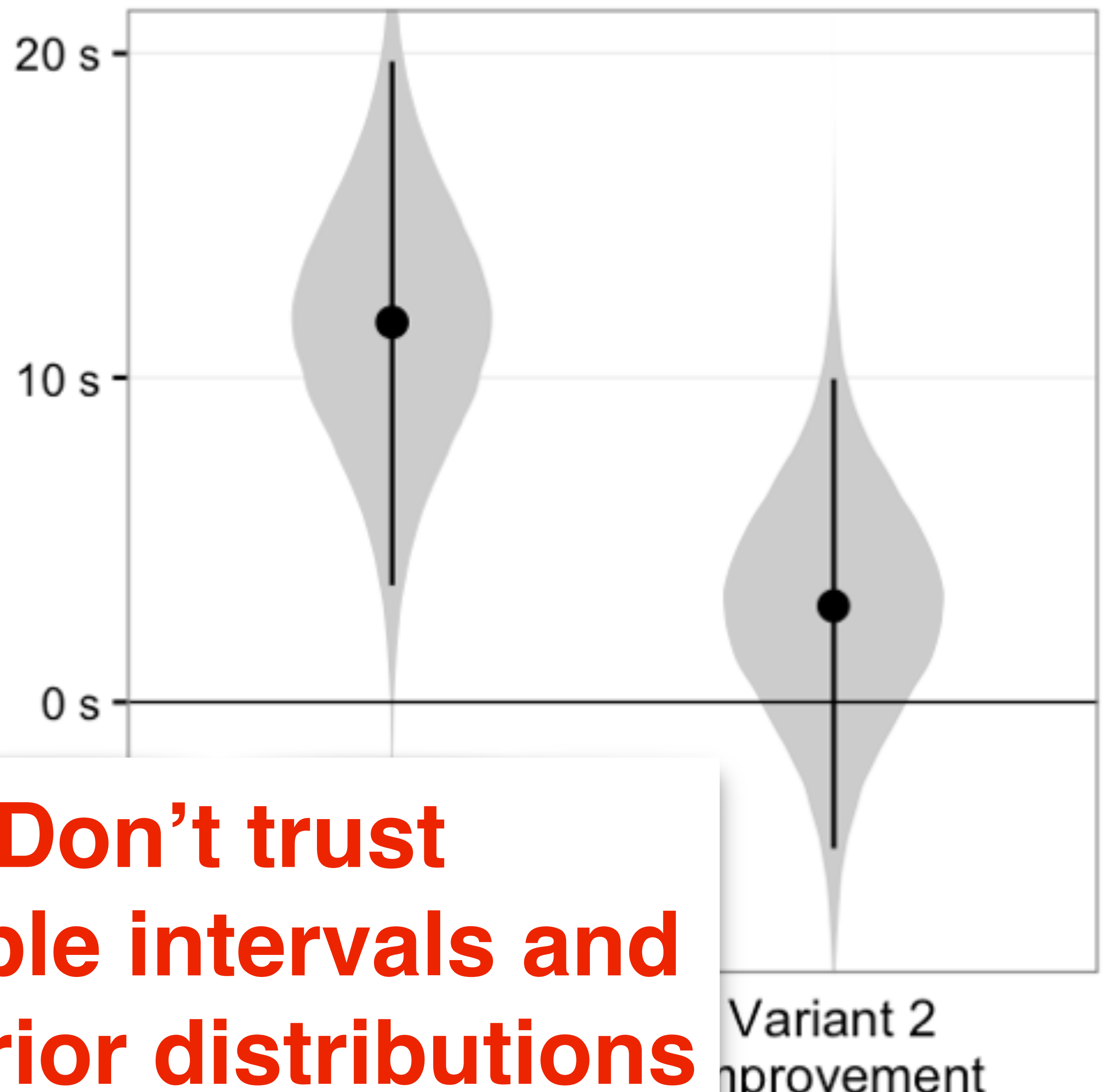
Bayesian dance



“Eye plots” made using (Kay, 2016)

Bayesian credible intervals and posterior distributions

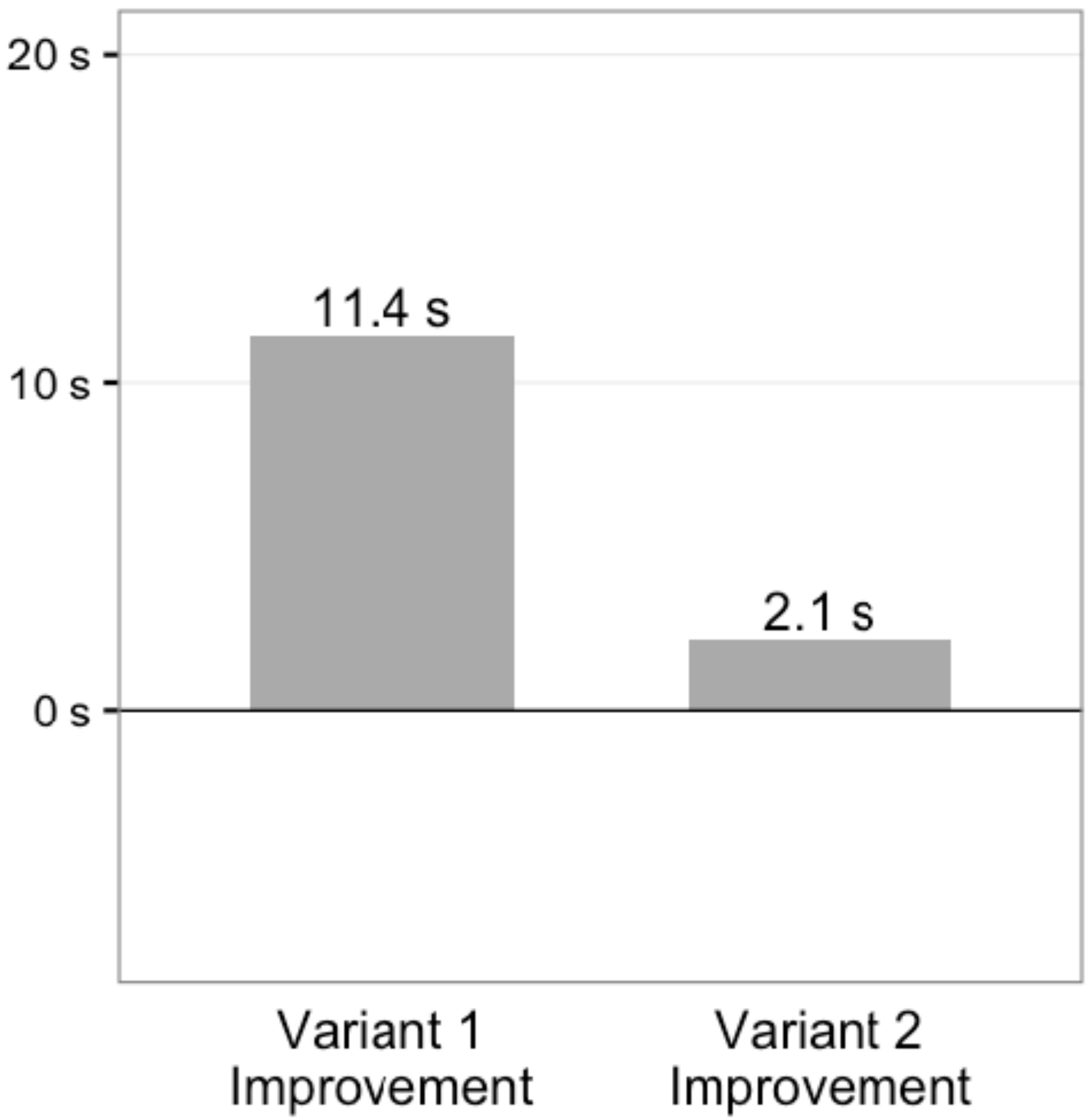
Our data

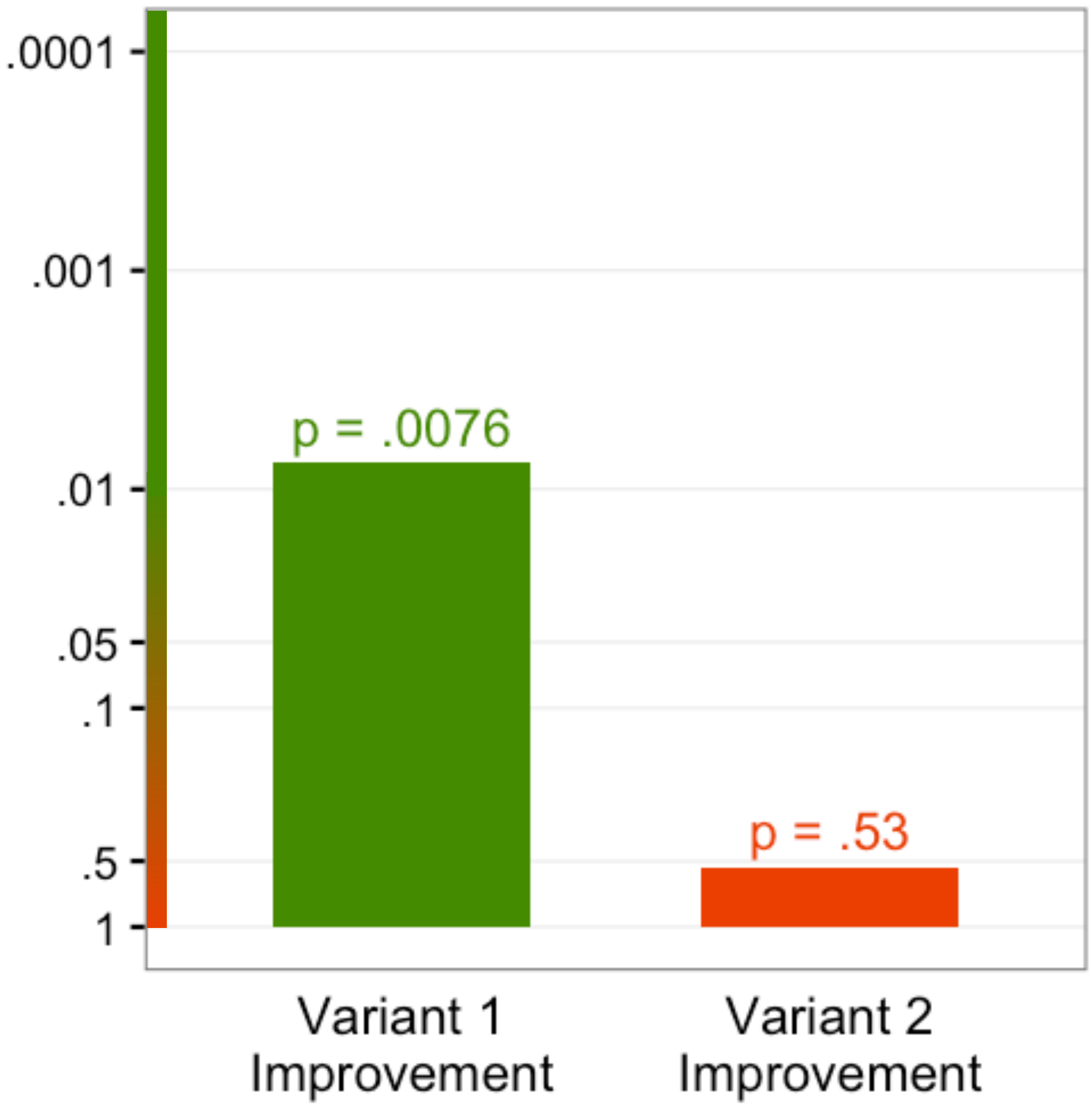


Don’t trust credible intervals and posterior distributions

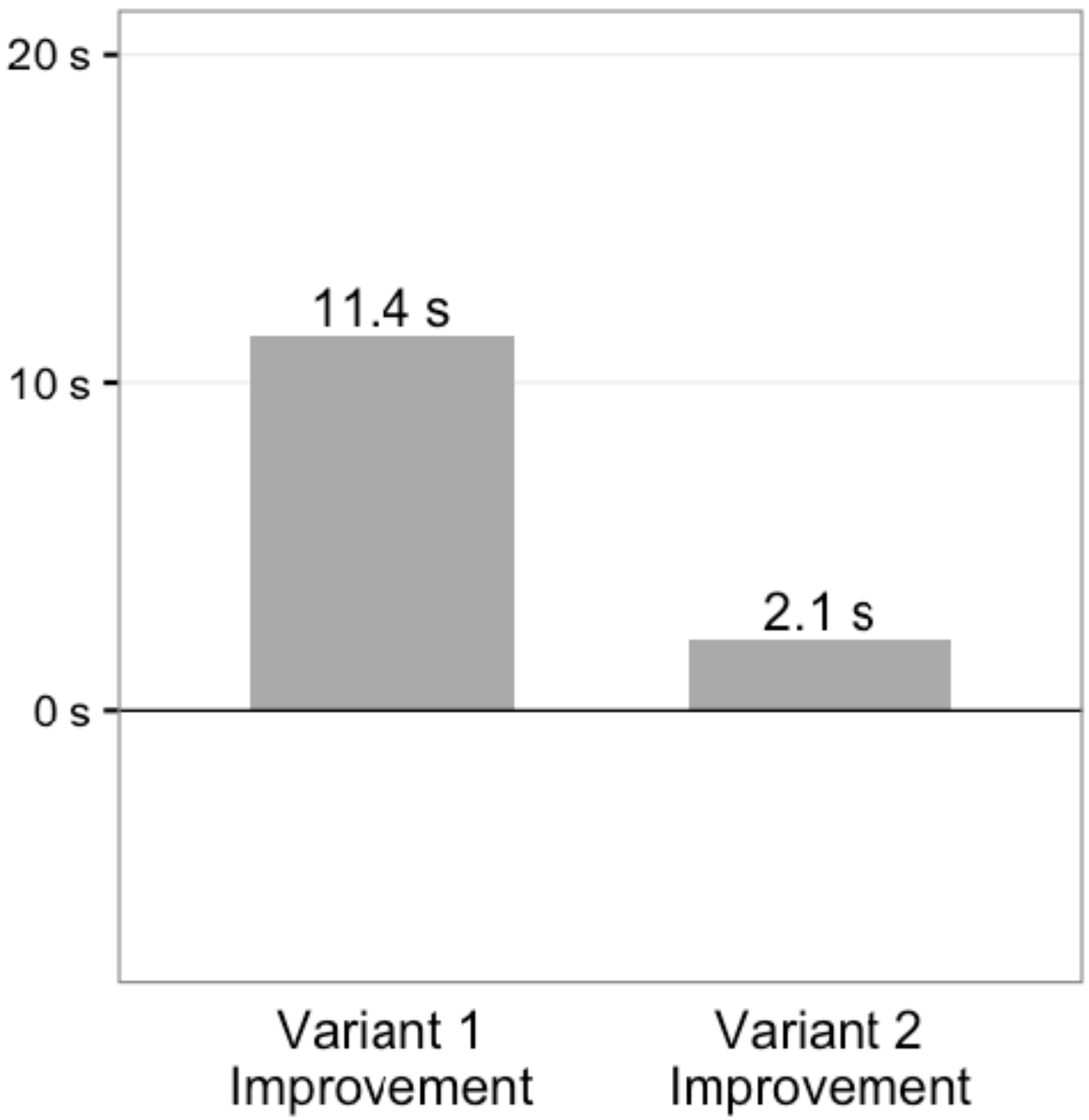
“Eye plots” made using (Kay, 2016)

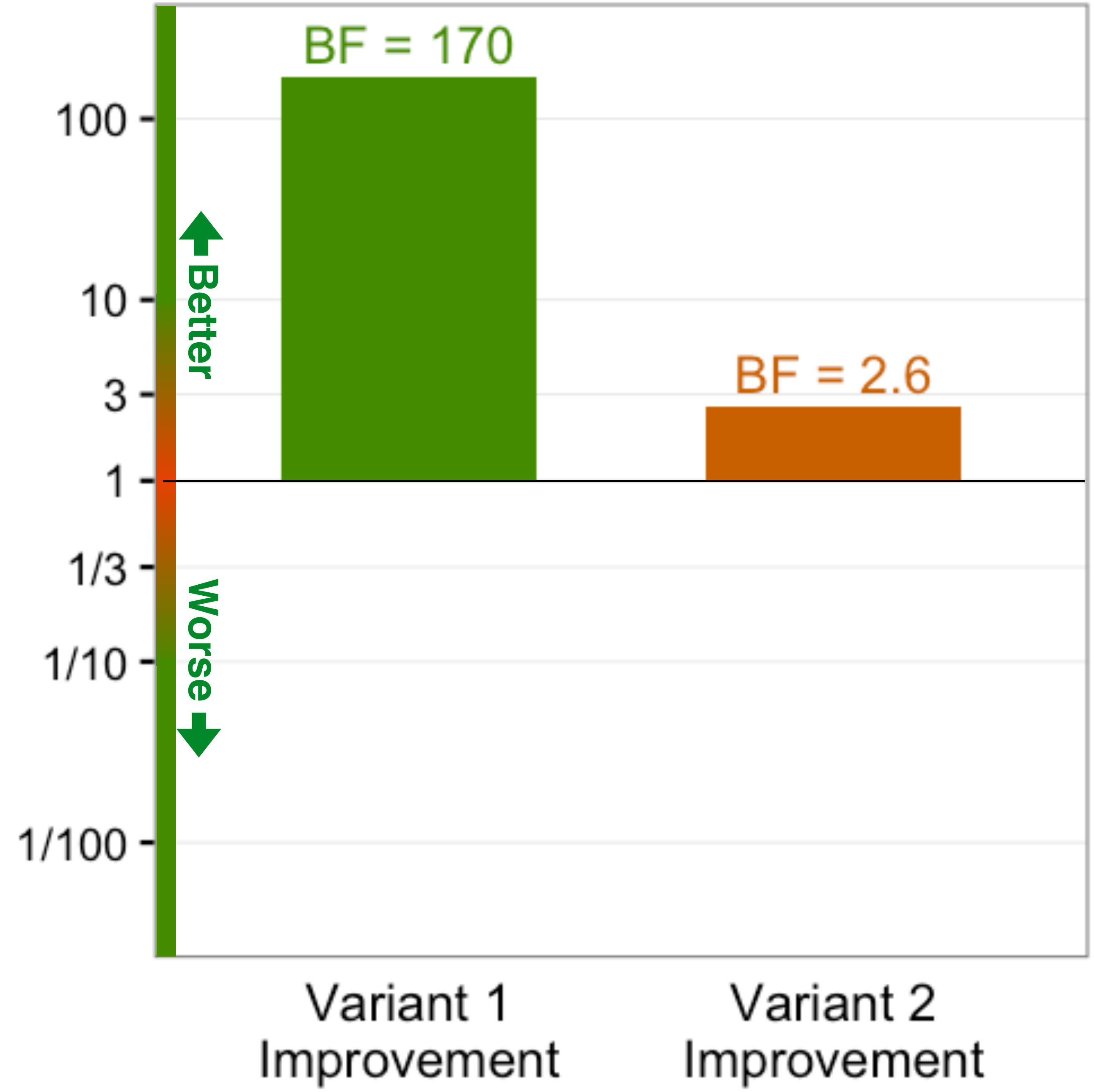
p-values


20 s
10 s
0 s
11.4 s
2.1 s
Variant 1
Improvement
Variant 2
Improvement

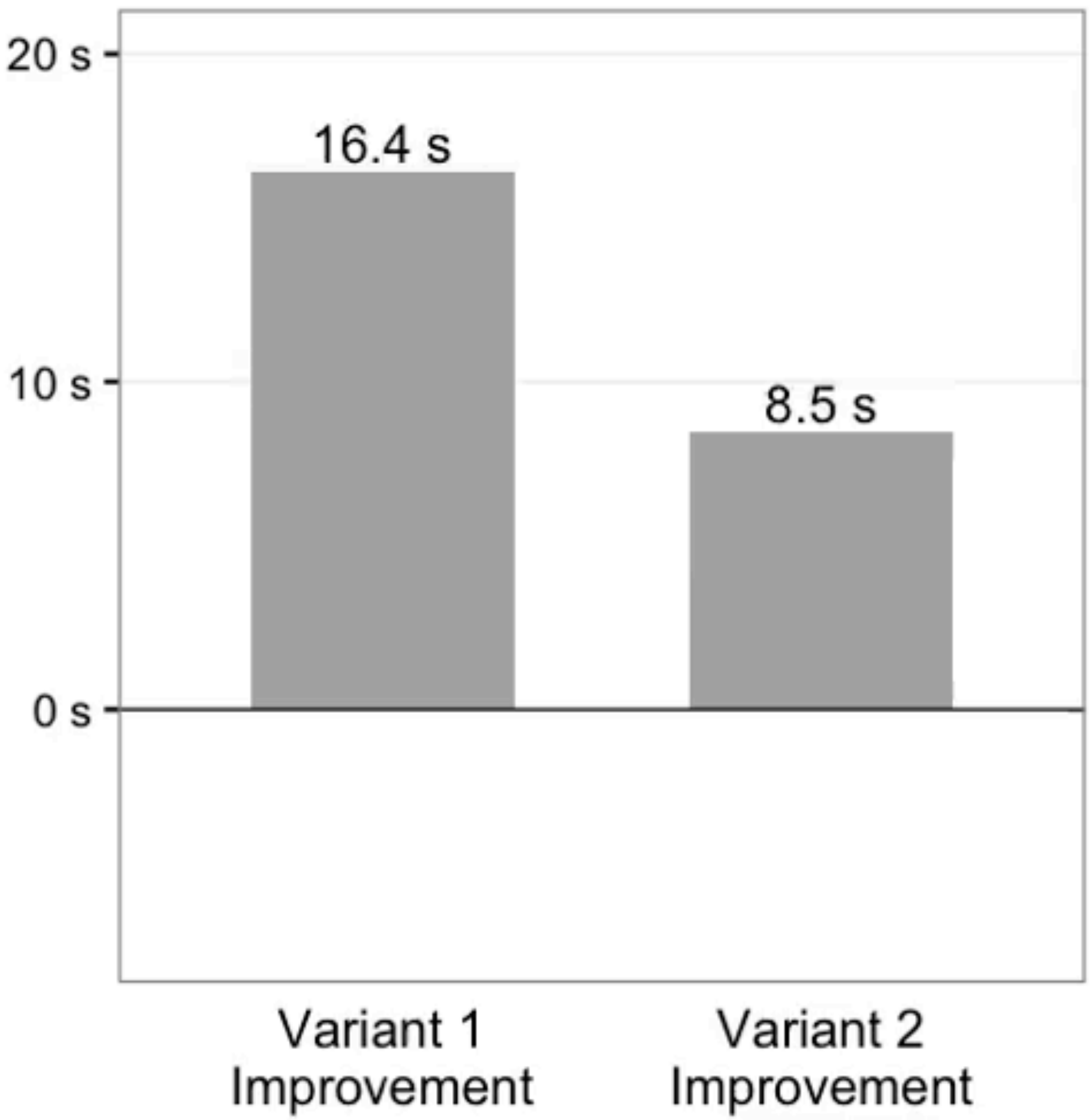

.0001
.001
.01
.05
.1
.5
1
p = .0076
p = .53
Variant 1
Improvement
Variant 2
Improvement

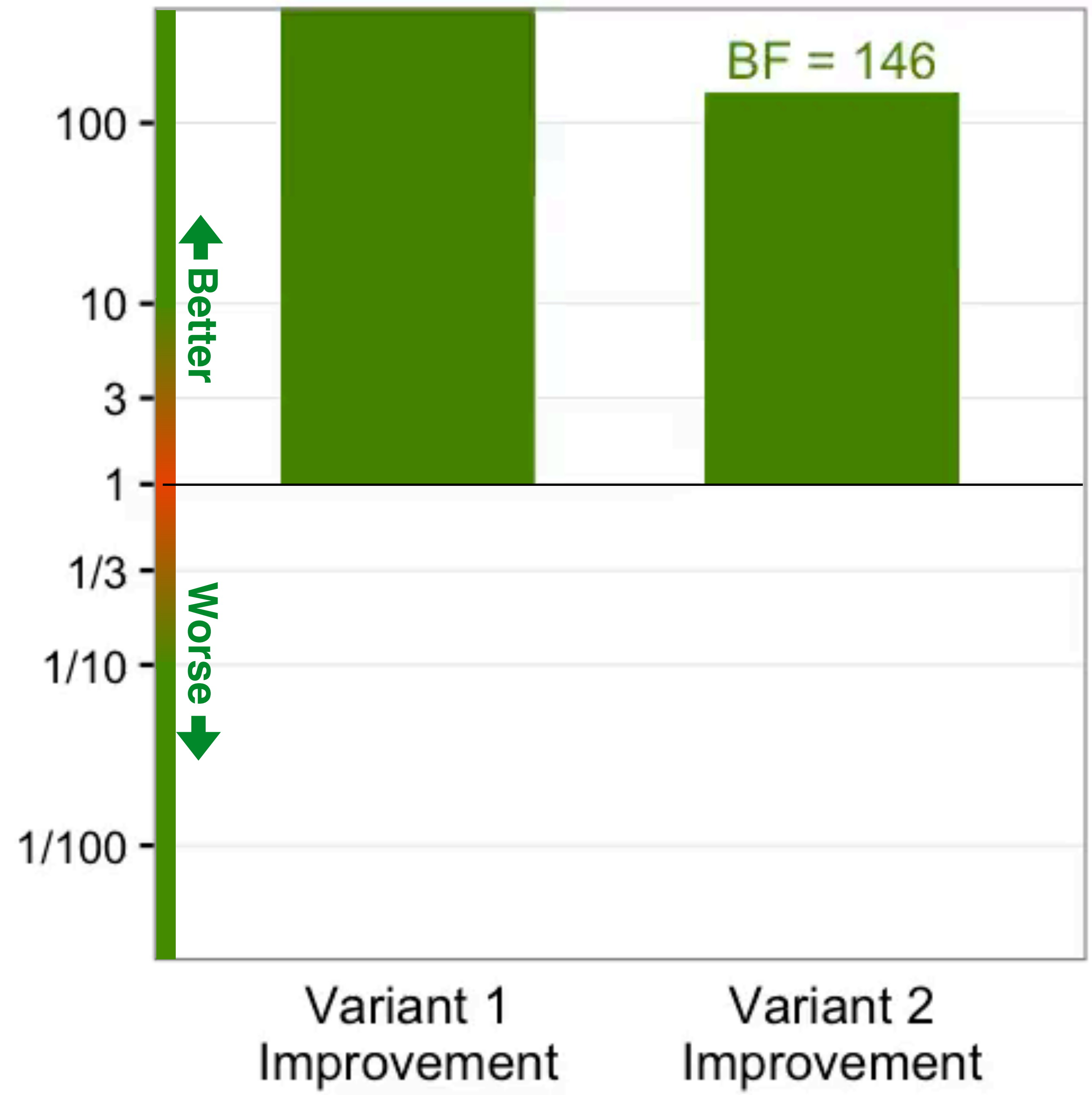
Bayes Factors



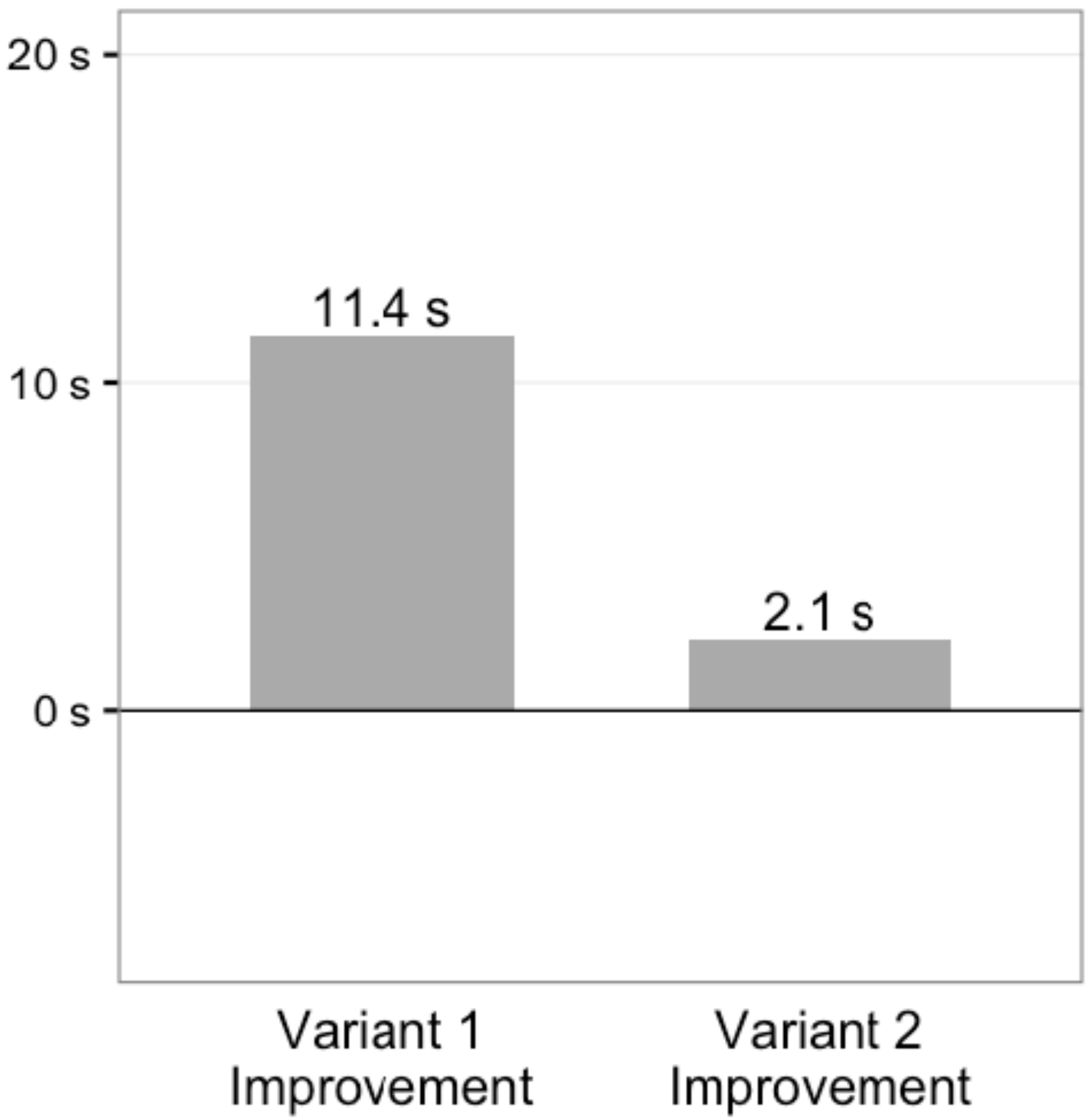


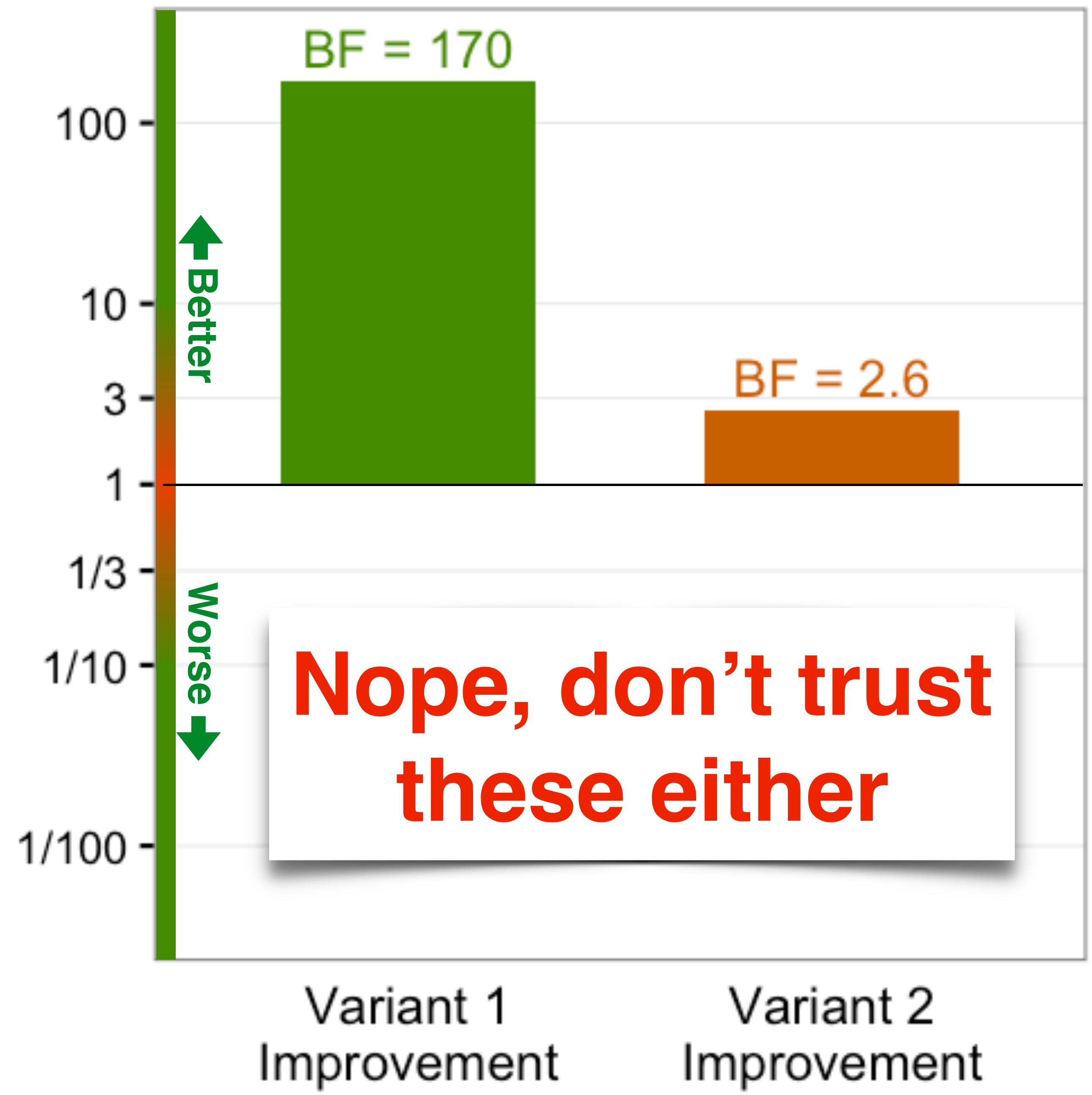
Dance of the Bayes Factors





Bayes Factors





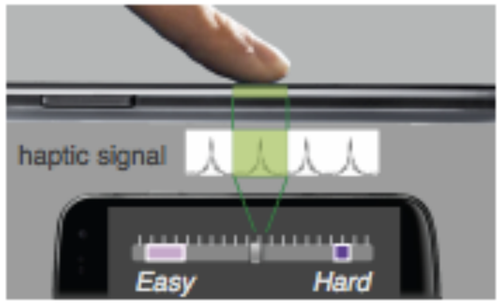


Figure 1: Example of *Easy* (left) and *Hard* (right) targets on the touch slider. In the haptic condition (*HS*), 20 evenly-spaced detents were simulated with vibrotactile feedback.

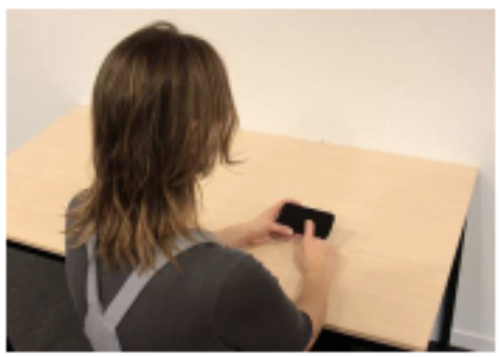


Figure 2: A participant completing our study.

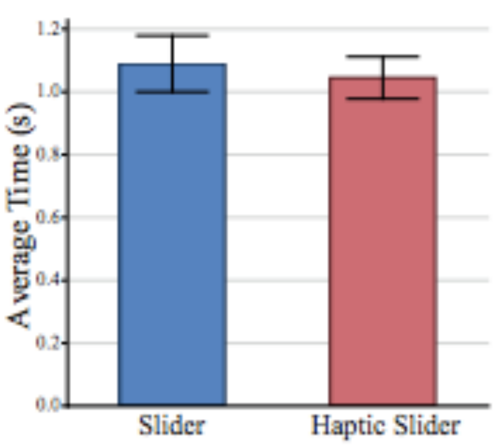


Figure 3: Time by Technique.

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 × 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 highly significant effect of Technique but no significant effect of Difficulty and no Technique×Difficulty interaction (see Table 1).

Table 1: Anova table.

Source	df	F	Sig.
Technique	1,11	12.7336	0.0044**
Difficulty	1,11	2.7084	0.1281
Technique×Difficulty	1,11	4.0402	0.0696

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)

Universe 2

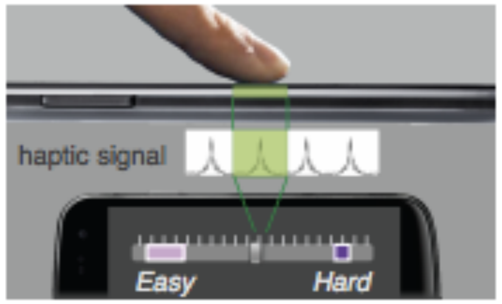


Figure 1: Example of *Easy* (left) and *Hard* (right) targets on the touch slider. In the haptic condition (*HS*), 20 evenly-spaced detents were simulated with vibrotactile feedback.

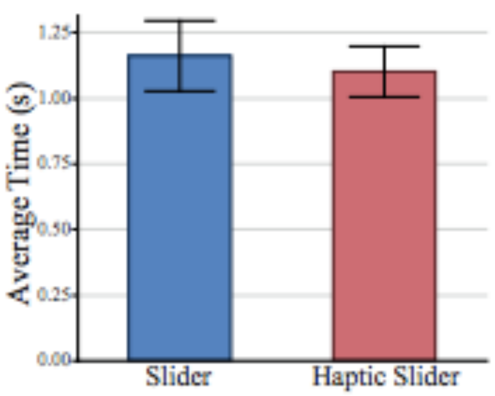


Figure 2: Time by Technique.

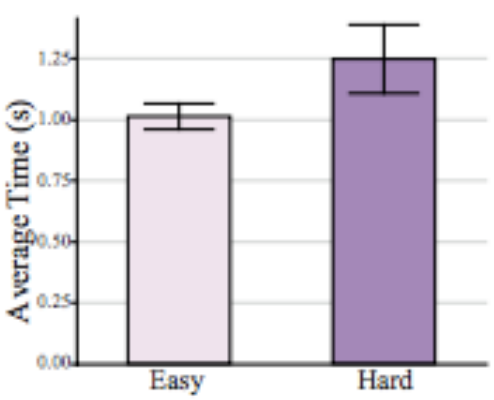


Figure 3: Time by Difficulty.

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 × 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 both Technique and Difficulty, but no significant Technique×Difficulty interaction effect (see Table 1).

Table 1: Anova table.

Source	df	F	Sig.
Technique	1,11	5.1139	0.0450*
Difficulty	1,11	6.2892	0.0291*
Technique×Difficulty	1,11	1.3669	0.2671

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

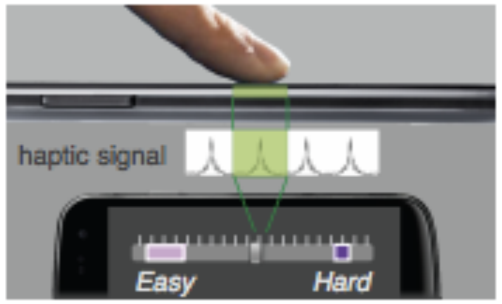


Figure 1: Example of *Easy* (left) and *Hard* (right) targets on the touch slider. In the haptic condition (*HS*), 20 evenly-spaced detents were simulated with vibrotactile feedback.

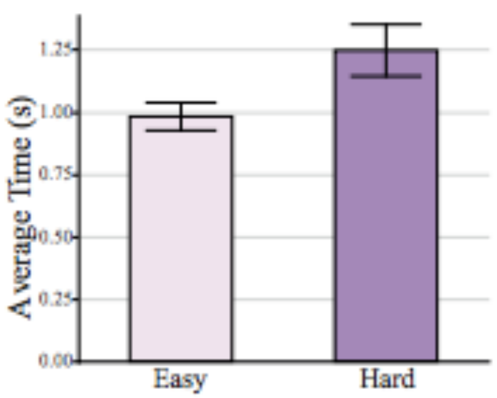


Figure 2: Time by Difficulty.

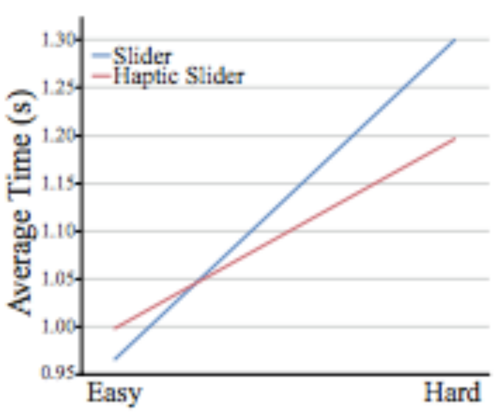


Figure 3: Time by Difficulty and Technique.

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).

Table 1: Anova table.

Source	df	F	Sig.
Technique	1,11	3.2748	0.0977
Difficulty	1,11	14.2324	0.0031**
Technique×Difficulty	1,11	14.9541	0.0026**

Our analysis confirms the effect of difficulty (avg. Times: *Easy*=0.98s, *Hard*=1.25s, see Figure 2). Student's t-tests reveal no significant difference between techniques for *Easy* (avg. Times: *S*=0.96s, *HS*=1.00s, $p = 0.1416$), and a highly significant difference between techniques for *Hard*, with a 9.2% increase in speed with *HS* (avg. Times: *S*=1.30s, *HS*=1.19s, $p = 0.0069$) (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 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

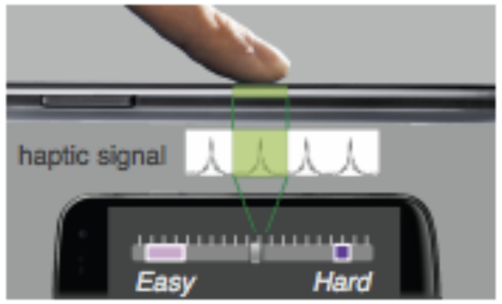


Figure 1: Example of *Easy* (left) and *Hard* (right) targets on the touch slider. In the haptic condition (*HS*), 20 evenly-spaced detents were simulated with vibrotactile feedback.

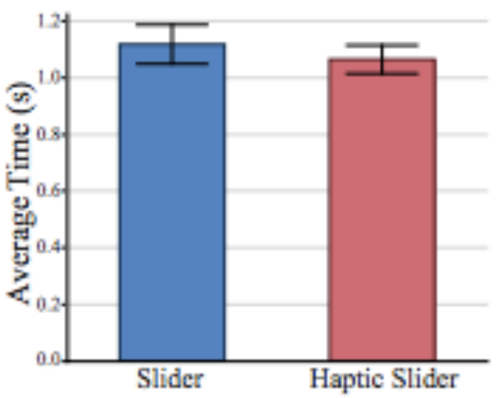


Figure 2: Time by Technique.

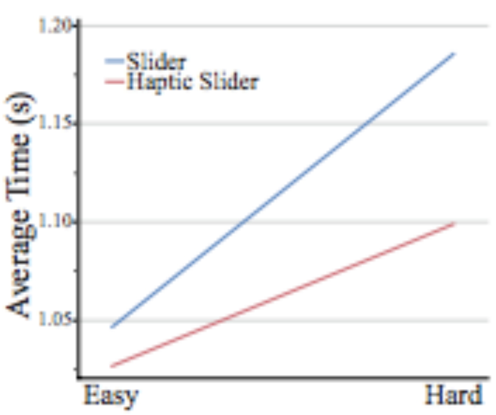


Figure 3: Time by Difficulty and Technique.

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 1: Anova table.

Source	df	F	Sig.
Technique	1,11	7.2144	0.0212*
Difficulty	1,11	4.1479	0.0665
Technique×Difficulty	1,11	5.5941	0.0375*

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 a 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

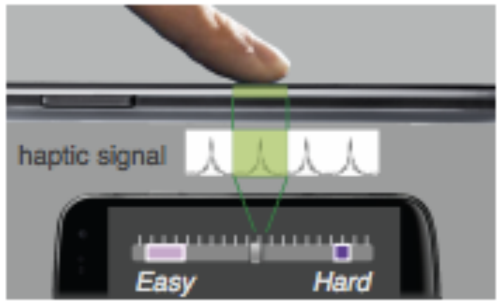


Figure 1: Example of *Easy* (left) and *Hard* (right) targets on the touch slider. In the haptic condition (*HS*), 20 evenly-spaced detents were simulated with vibrotactile feedback.

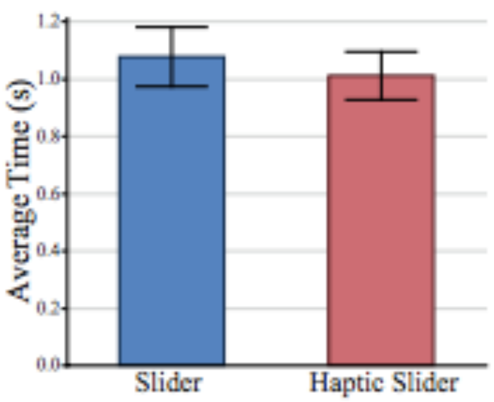


Figure 2: Time by Technique.

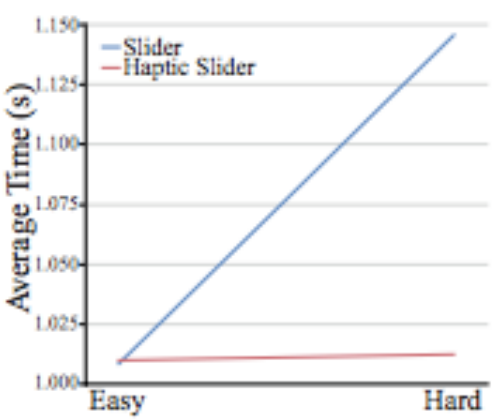


Figure 3: Time by Difficulty and Technique.

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 × 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 1: Anova table.

Source	df	F	Sig.
Technique	1,11	6.0536	0.0317*
Difficulty	1,11	1.0392	0.3299
Technique×Difficulty	1,11	9.4480	0.0106*

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)

Universe 6

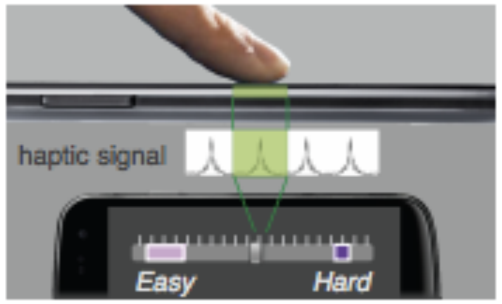


Figure 1: Example of *Easy* (left) and *Hard* (right) targets on the touch slider. In the haptic condition (*HS*), 20 evenly-spaced detents were simulated with vibrotactile feedback.

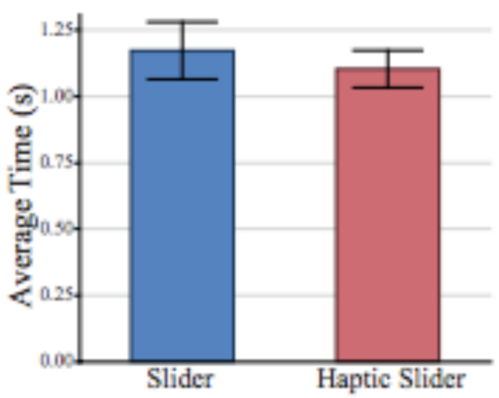


Figure 2: Time by Technique.

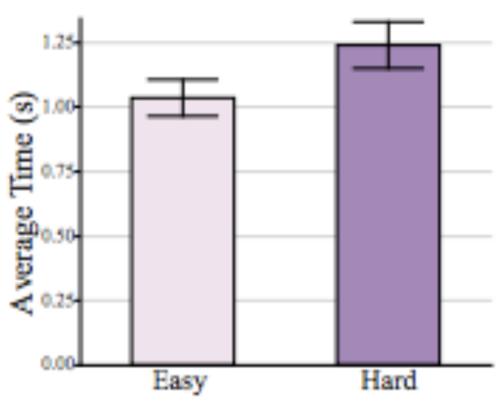


Figure 3: Time by Difficulty and Technique.

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 × 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 highly significant effect of Technique, and a very highly significant effect of Difficulty, and no Technique×Difficulty interaction (see Table 1).

Table 1: Anova table.

Source	df	F	Sig.
Technique	1,11	13.1323	0.0040**
Difficulty	1,11	21.9758	0.0007***
Technique×Difficulty	1,11	3.9159	0.0734

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

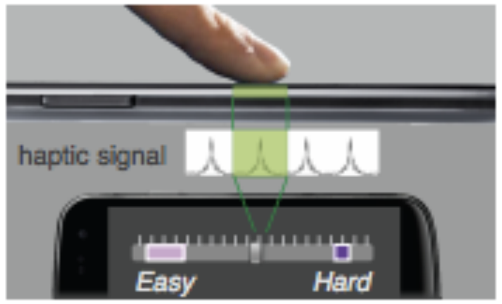


Figure 1: Example of *Easy* (left) and *Hard* (right) targets on the touch slider. In the haptic condition (*HS*), 20 evenly-spaced detents were simulated with vibrotactile feedback.

Figure 2: A participant completing our study.

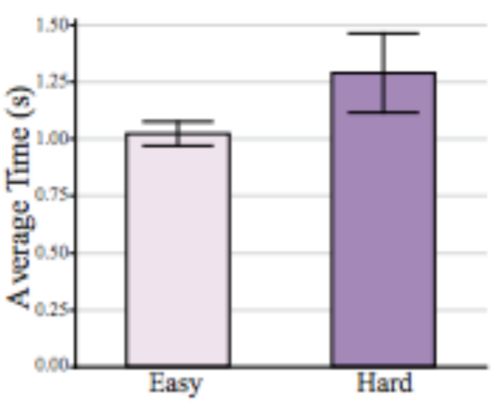


Figure 3: TIME by DIFFICULTY and TECHNIQUE.

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 1: ANOVA table.

Source	df	F	Sig.
Technique	1,11	4.6215	0.0547
Difficulty	1,11	4.8698	0.0495*
Technique×Difficulty	1,11	1.8322	0.2030

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 hedonistic 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

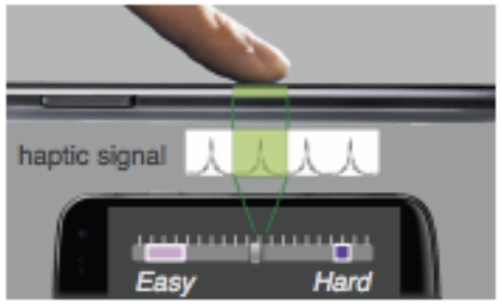


Figure 1: Example of *Easy* (left) and *Hard* (right) targets on the touch slider. In the haptic condition (*HS*), 20 evenly-spaced detents were simulated with vibrotactile feedback.

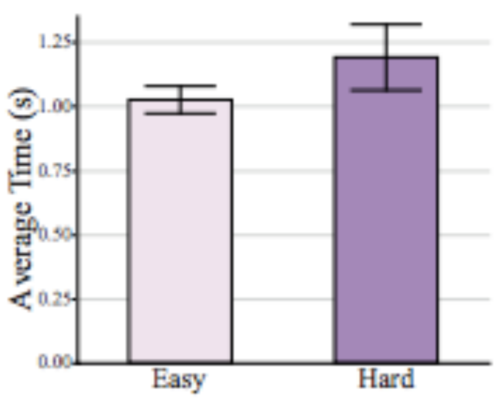


Figure 2: TIME by TECHNIQUE.

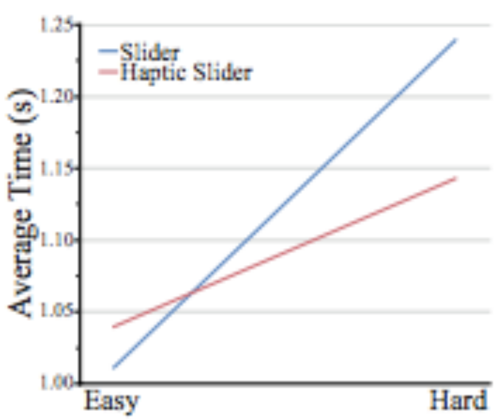


Figure 3: TIME by DIFFICULTY and TECHNIQUE.

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 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 *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 significant effect of DIFFICULTY with also a very highly significant TECHNIQUE×DIFFICULTY interaction effect (see Table1).

Table 1: ANOVA table.

Source	df	F	Sig.
Technique	1,11	2.1350	0.1719
Difficulty	1,11	5.1621	0.0442*
Technique×Difficulty	1,11	22.6791	0.0006***

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 a 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 haptic detents on touch sliders for tasks that require fine control, but not for tasks where coarse control is sufficient.

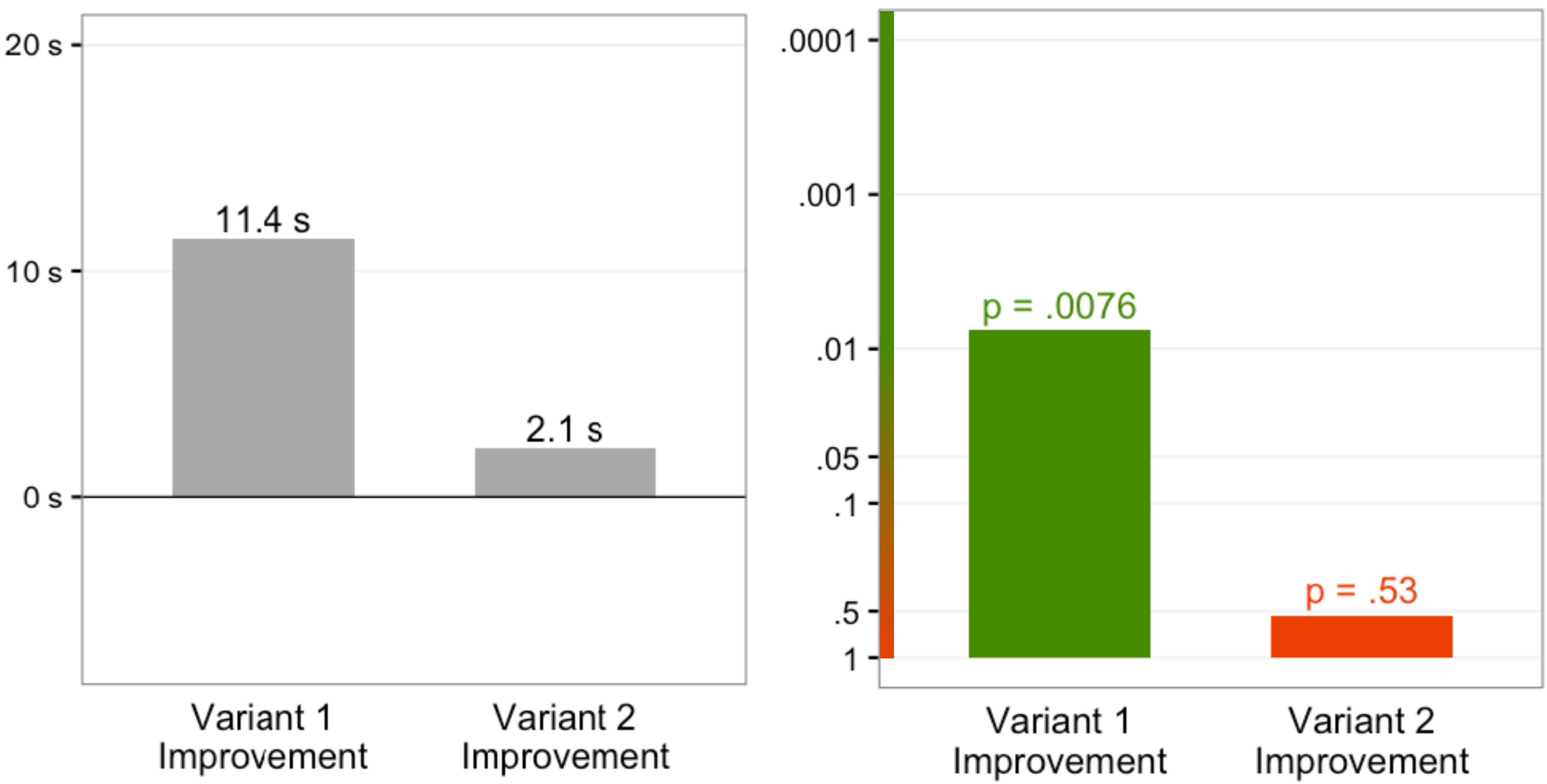
(Dragicevic, Chevalier and Huot, 2014)

Lessons learned so far

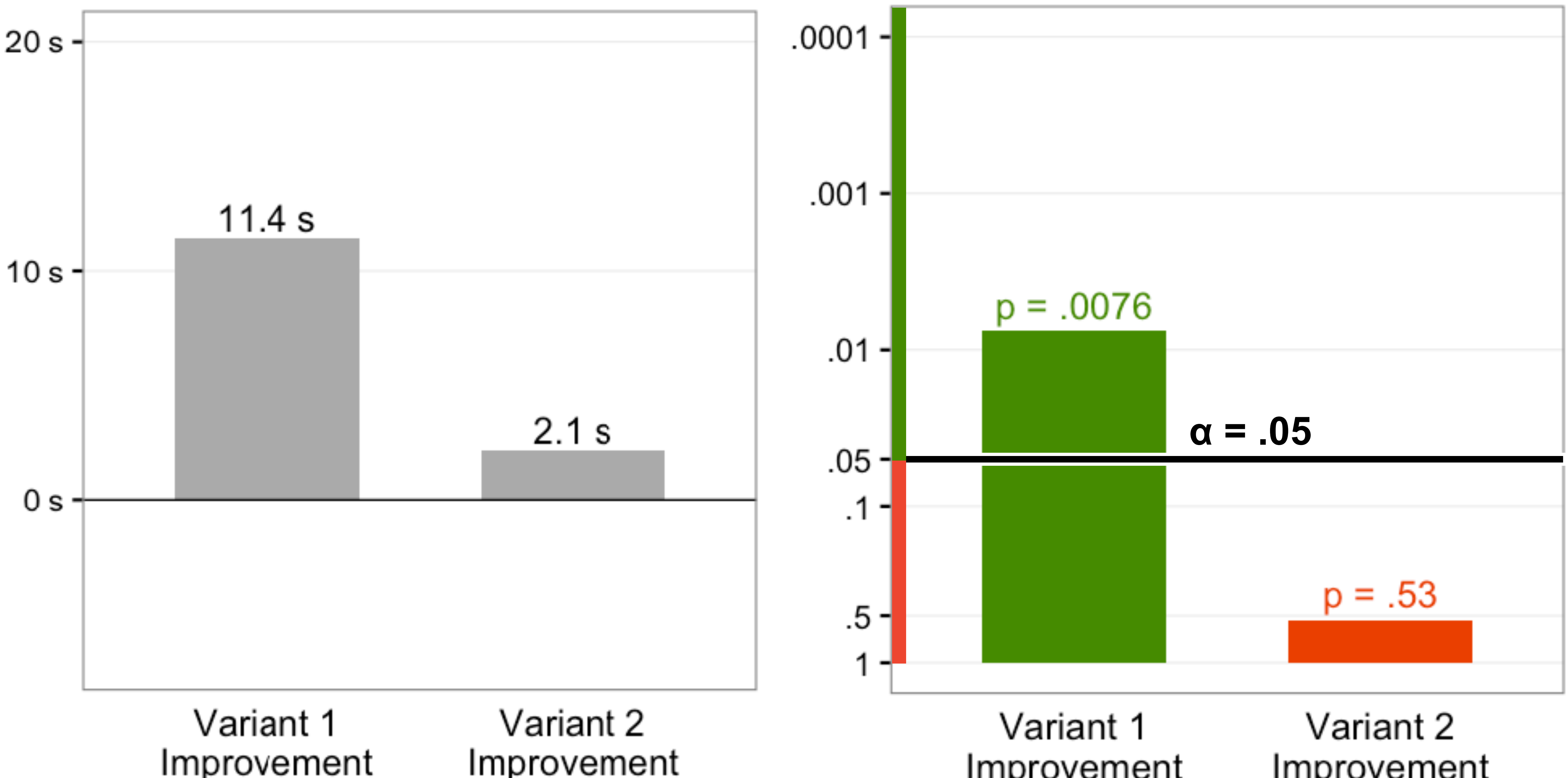
- Everything dances
    - Descriptive and inferential statistics alike
    - Dances propagate along the analysis pipeline
- There is no way around it
- But there must be ways!

False solution:
discretize / dichotomize

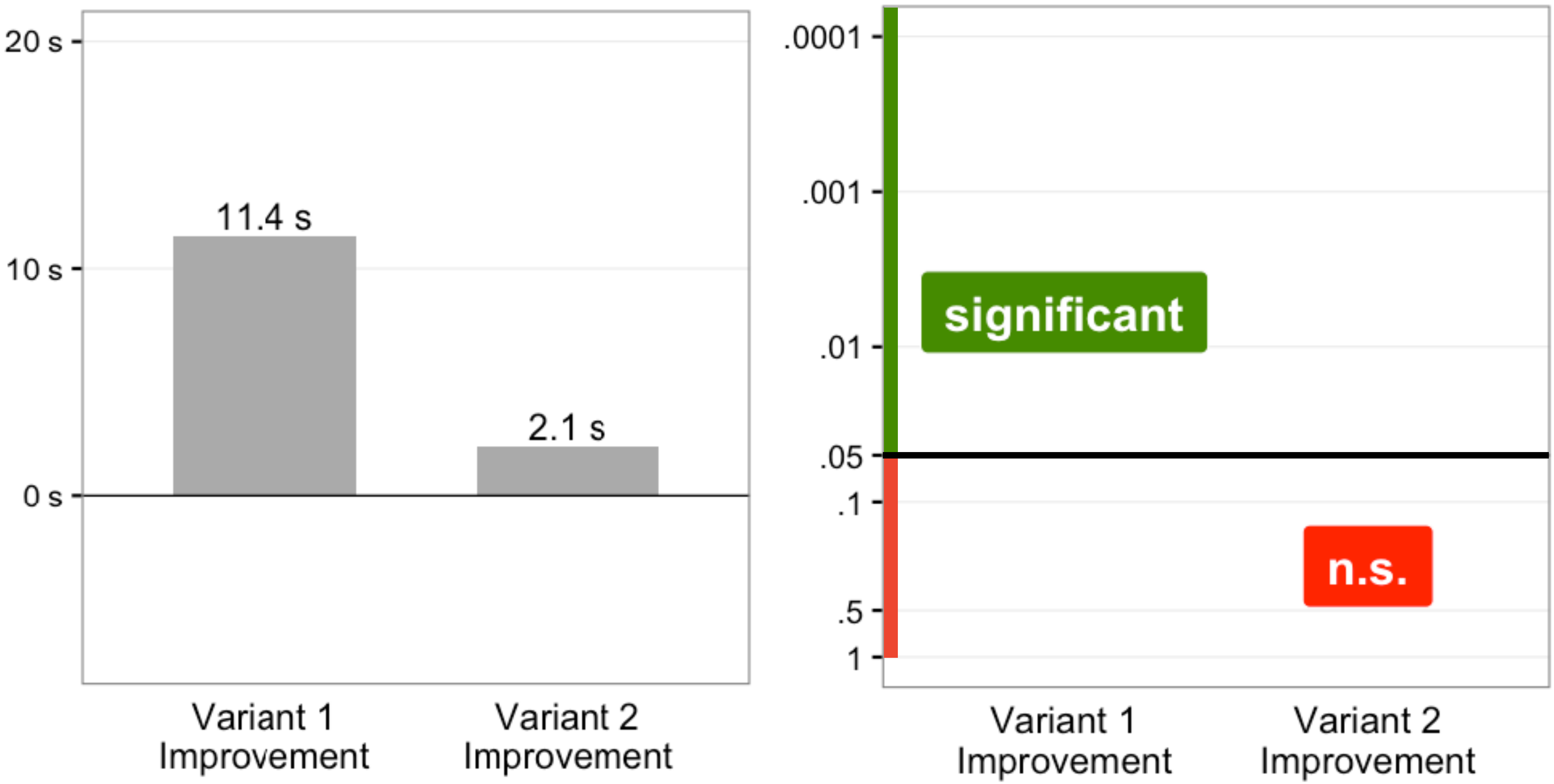
False solution: discretize / dichotomize



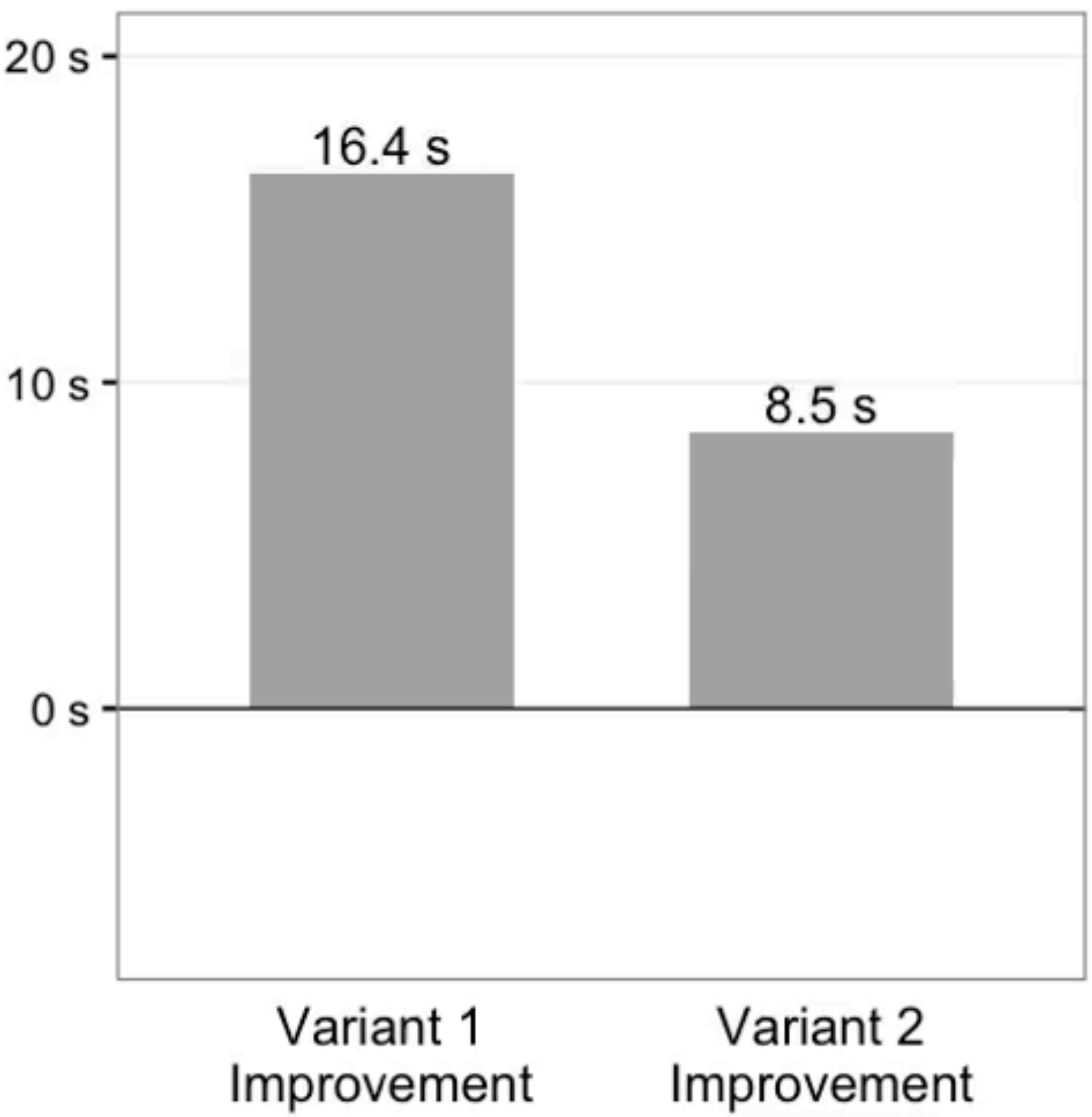
False solution: discretize / dichotomize

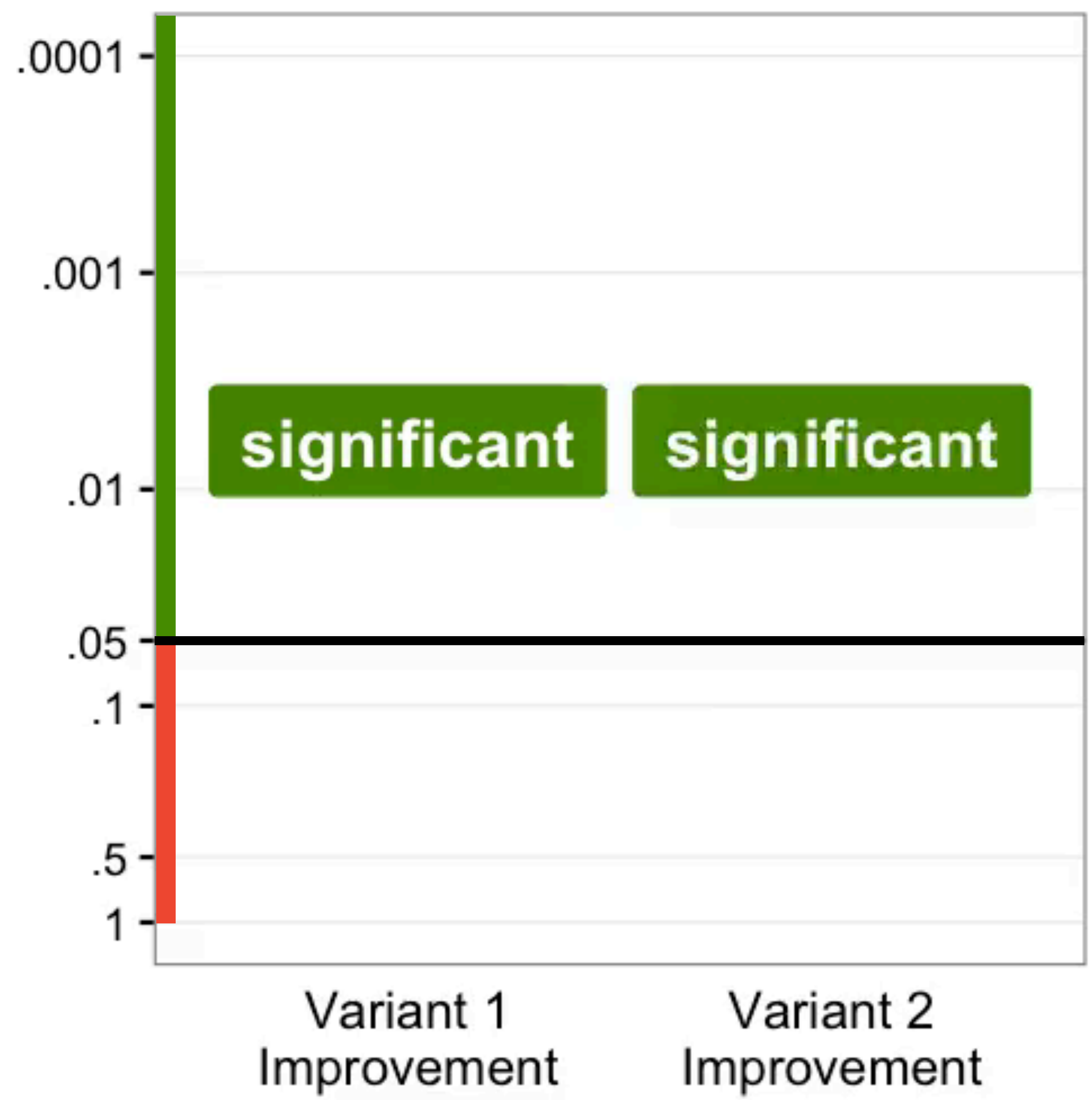


False solution: discretize / dichotomize

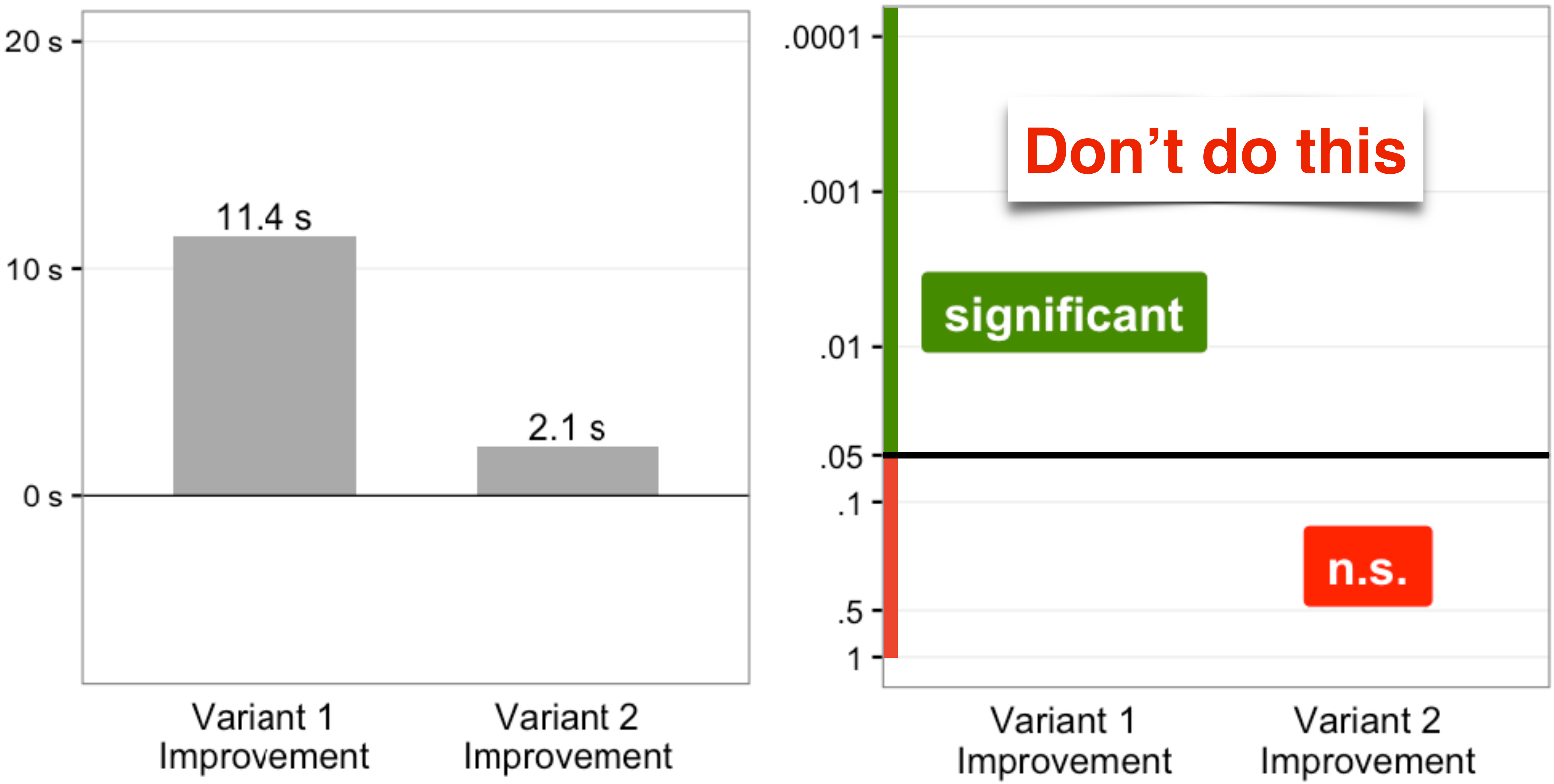


False solution: discretize / dichotomize

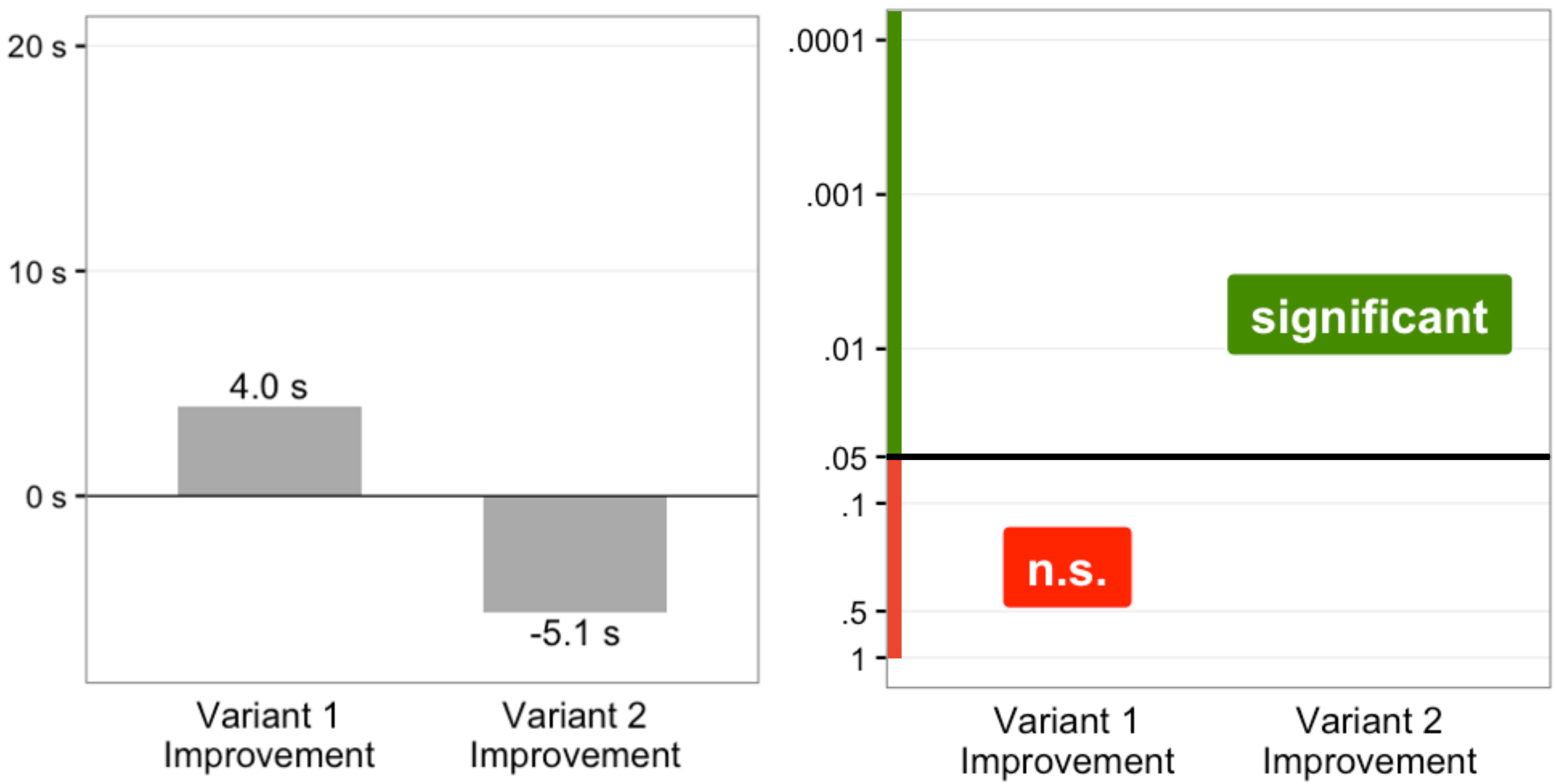




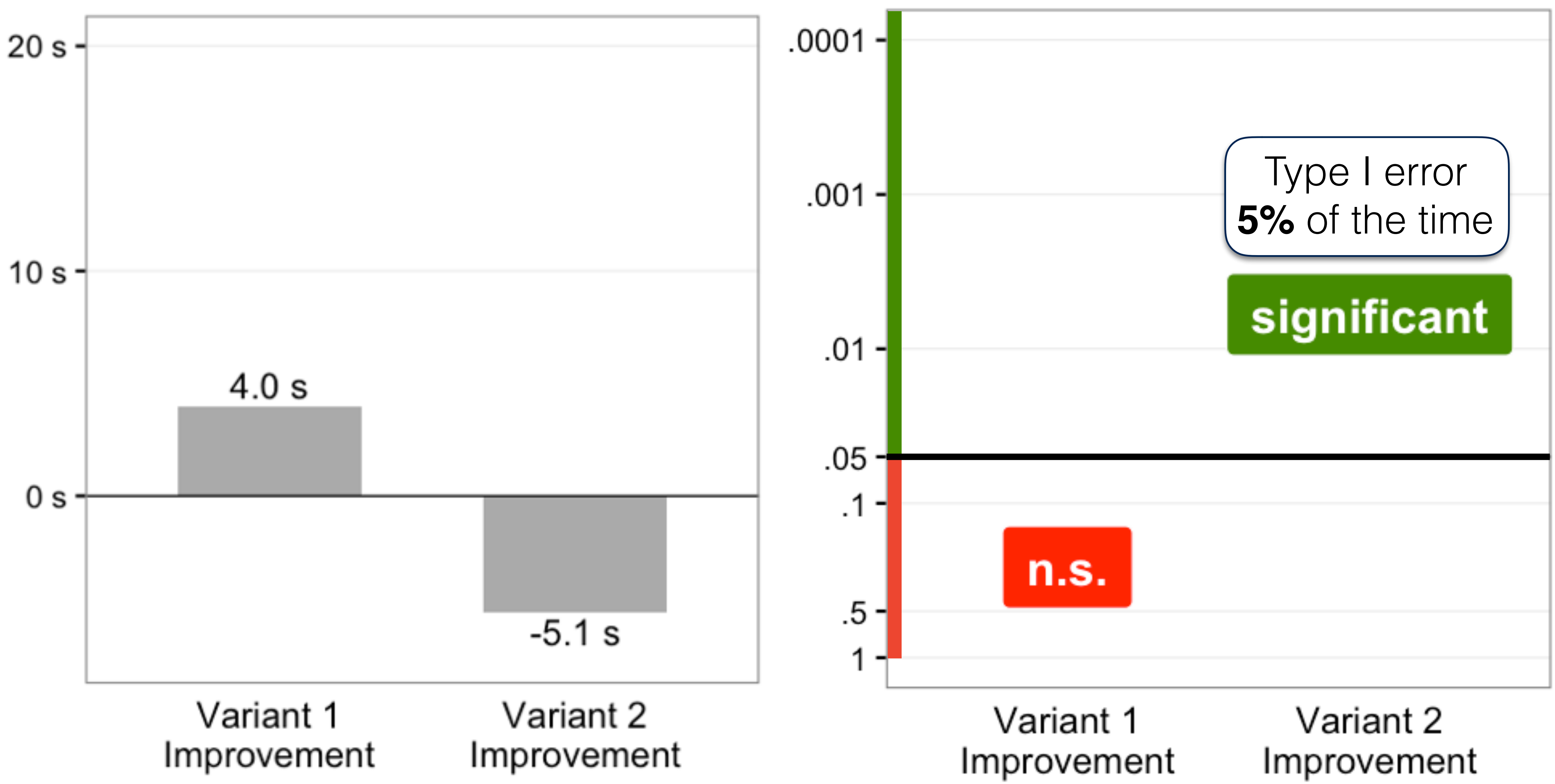
False solution: discretize / dichotomize



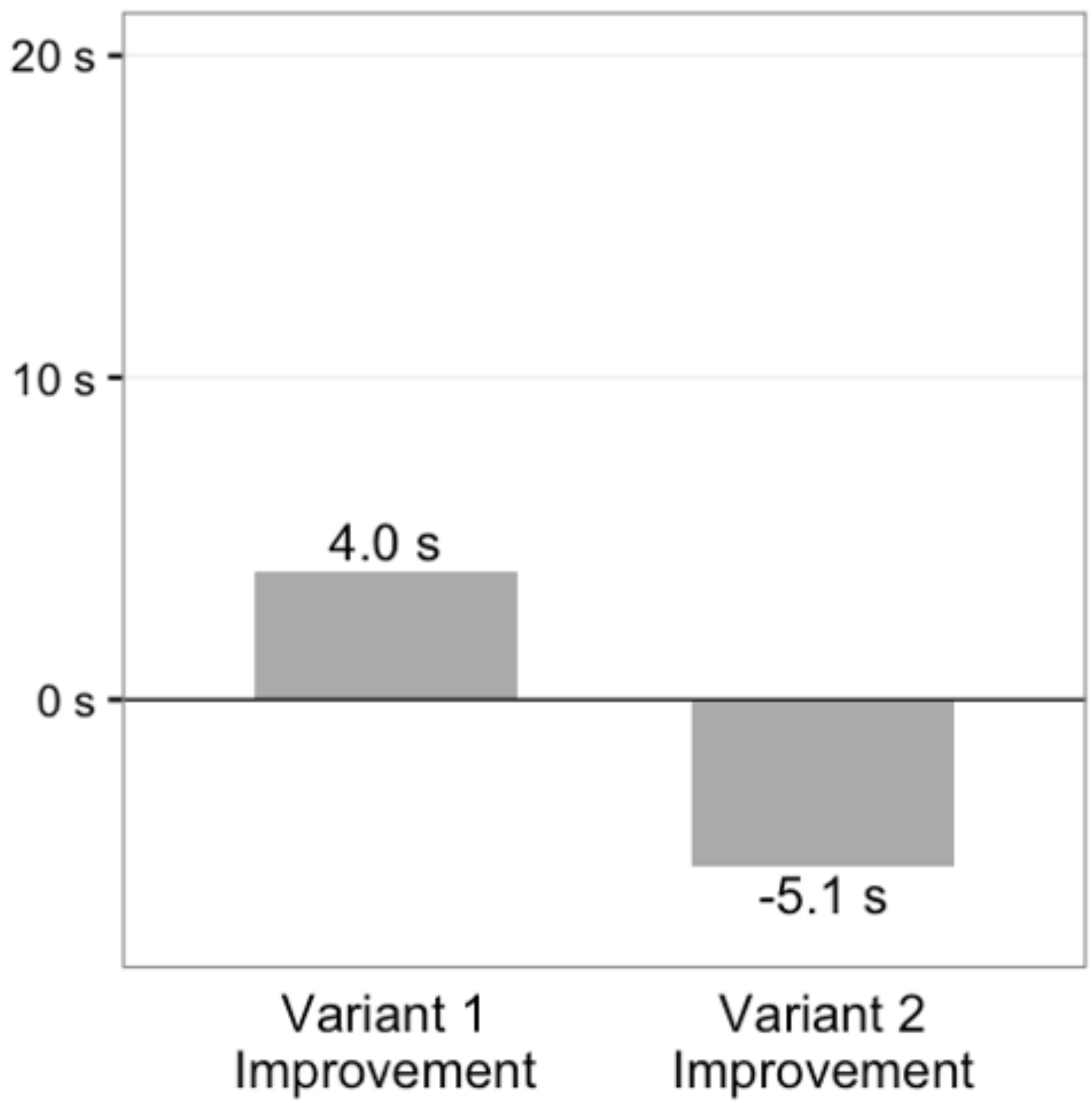
False solution: discretize / dichotomize

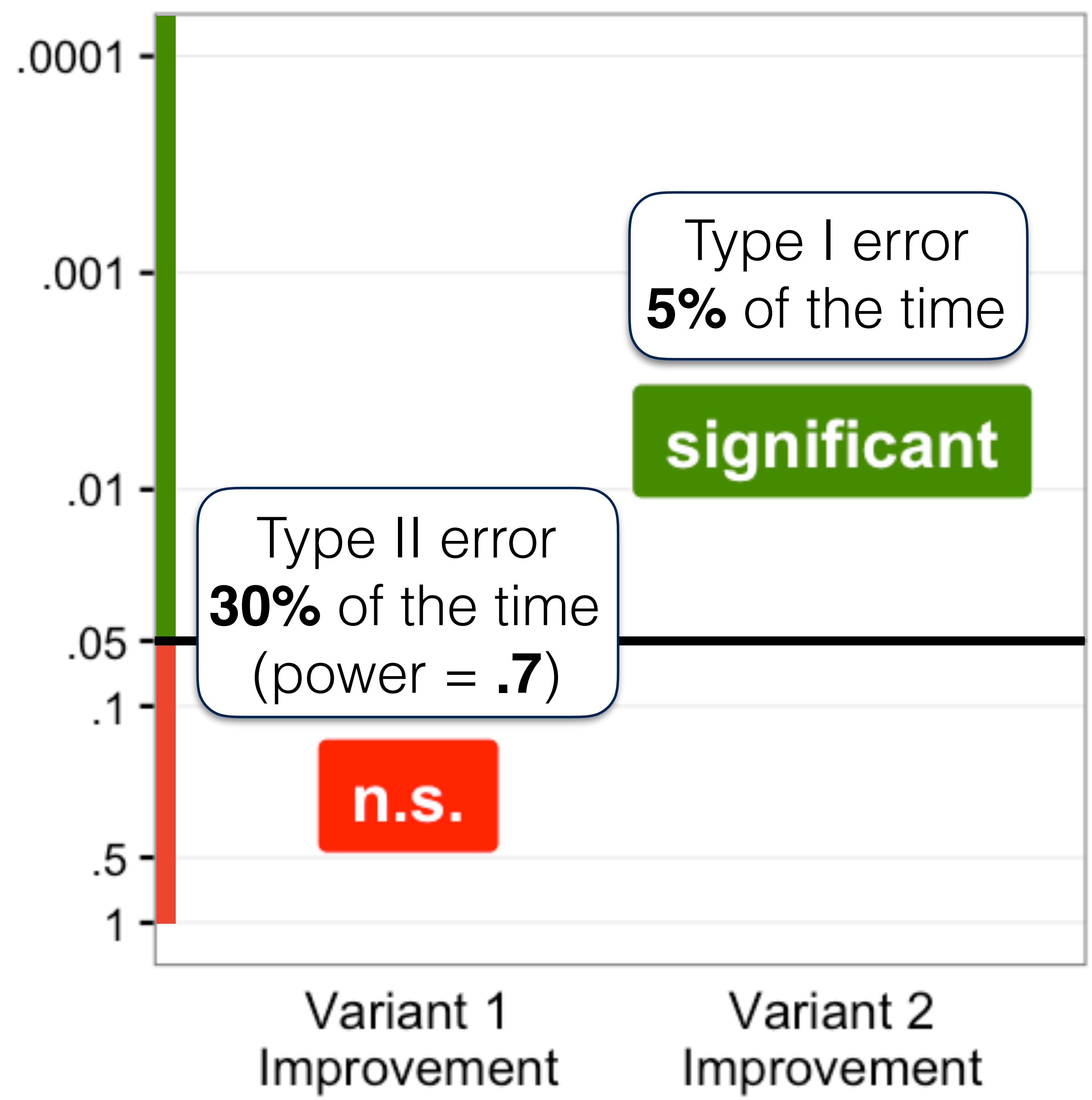


False solution: discretize / dichotomize

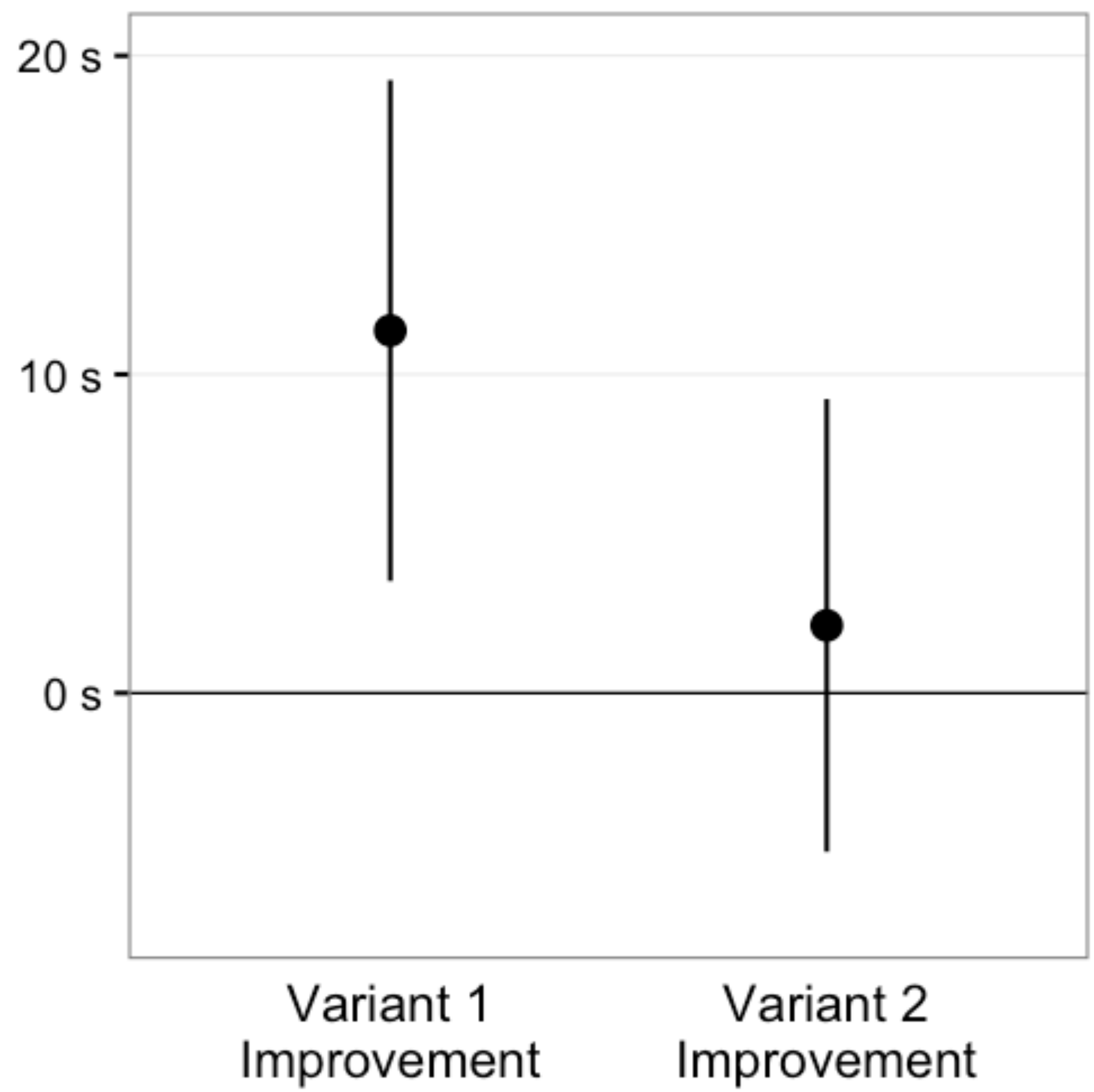


False solution: discretize / dichotomize

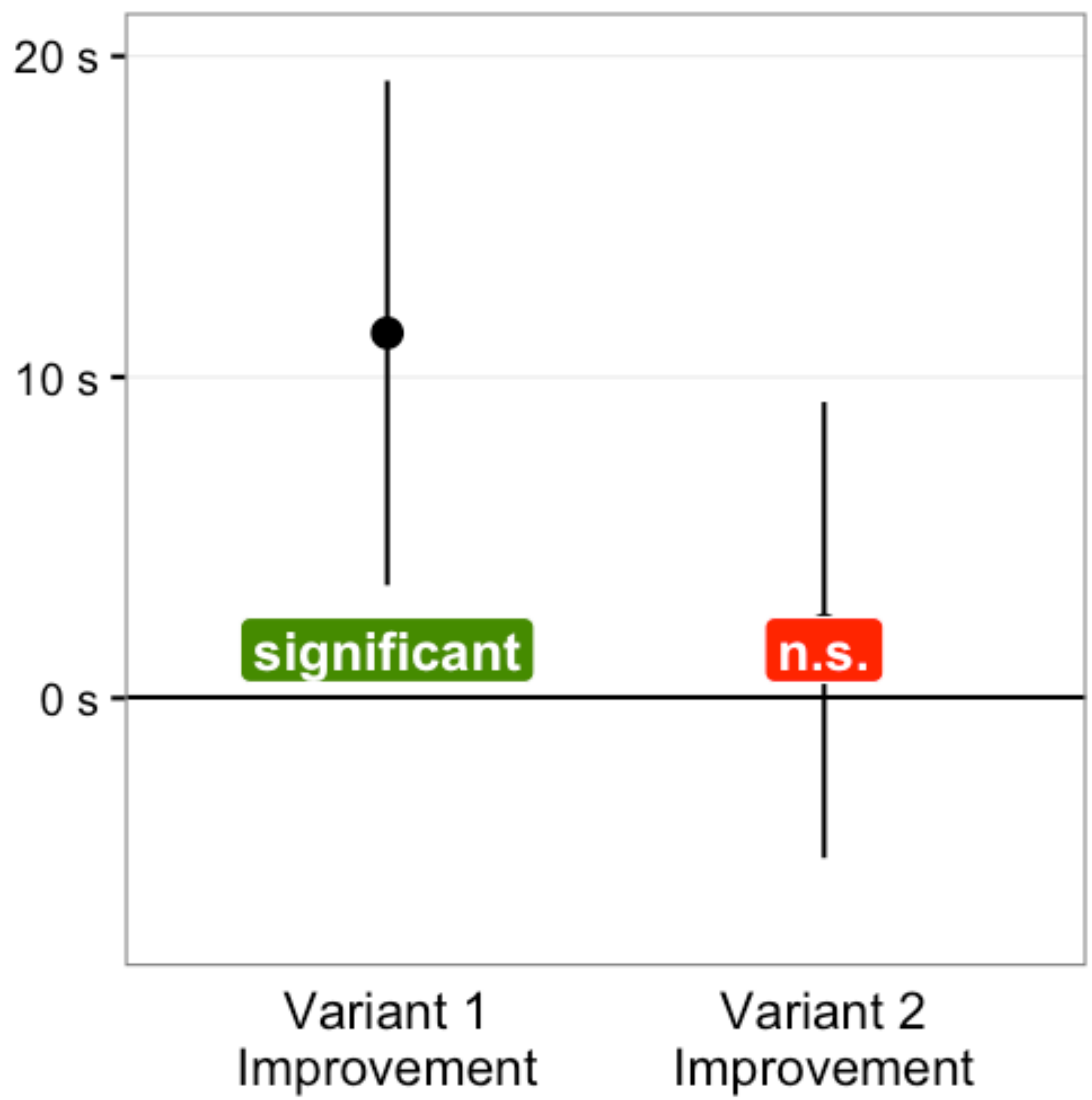




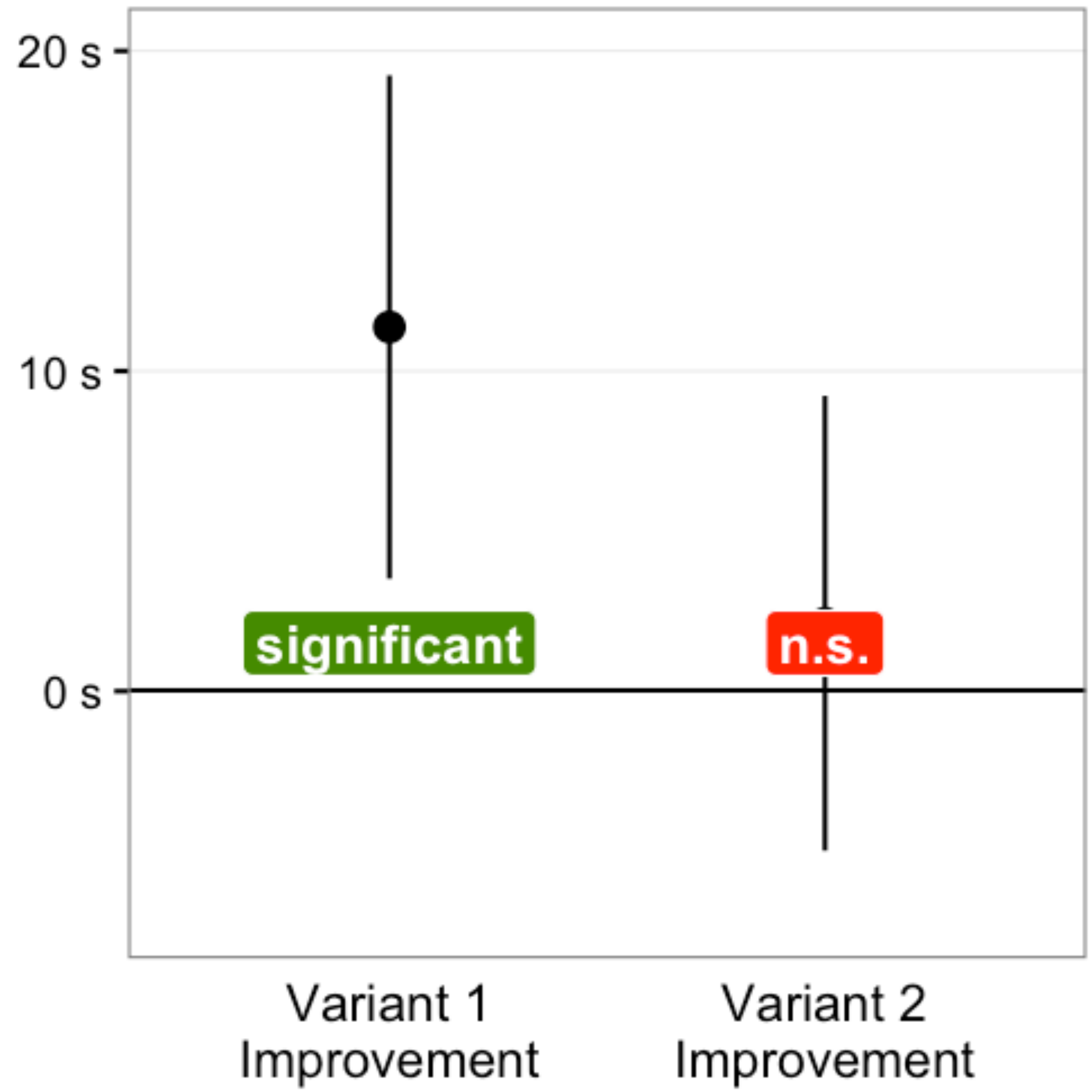
False solution: discretize / dichotomize



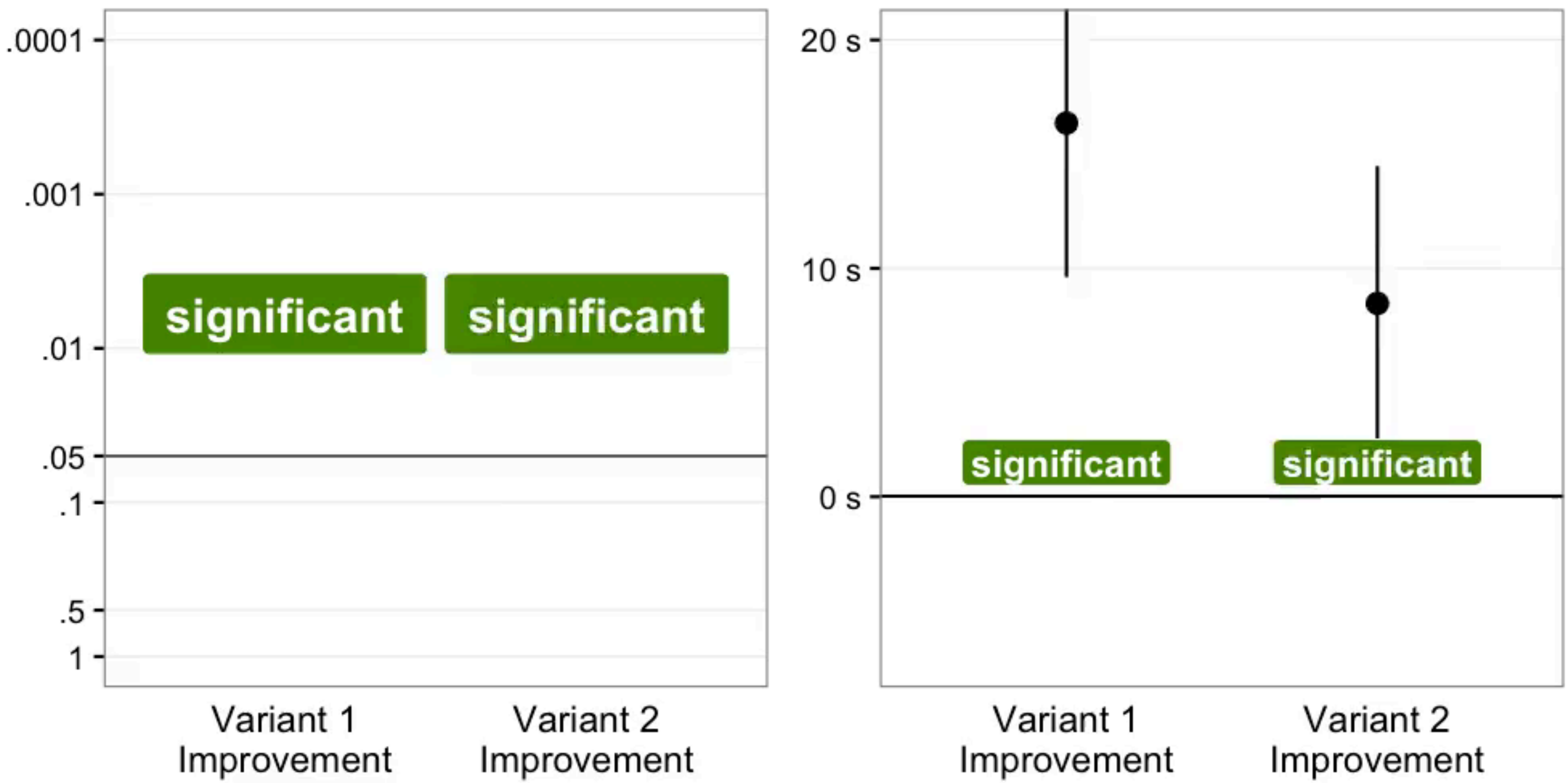
False solution: discretize / dichotomize



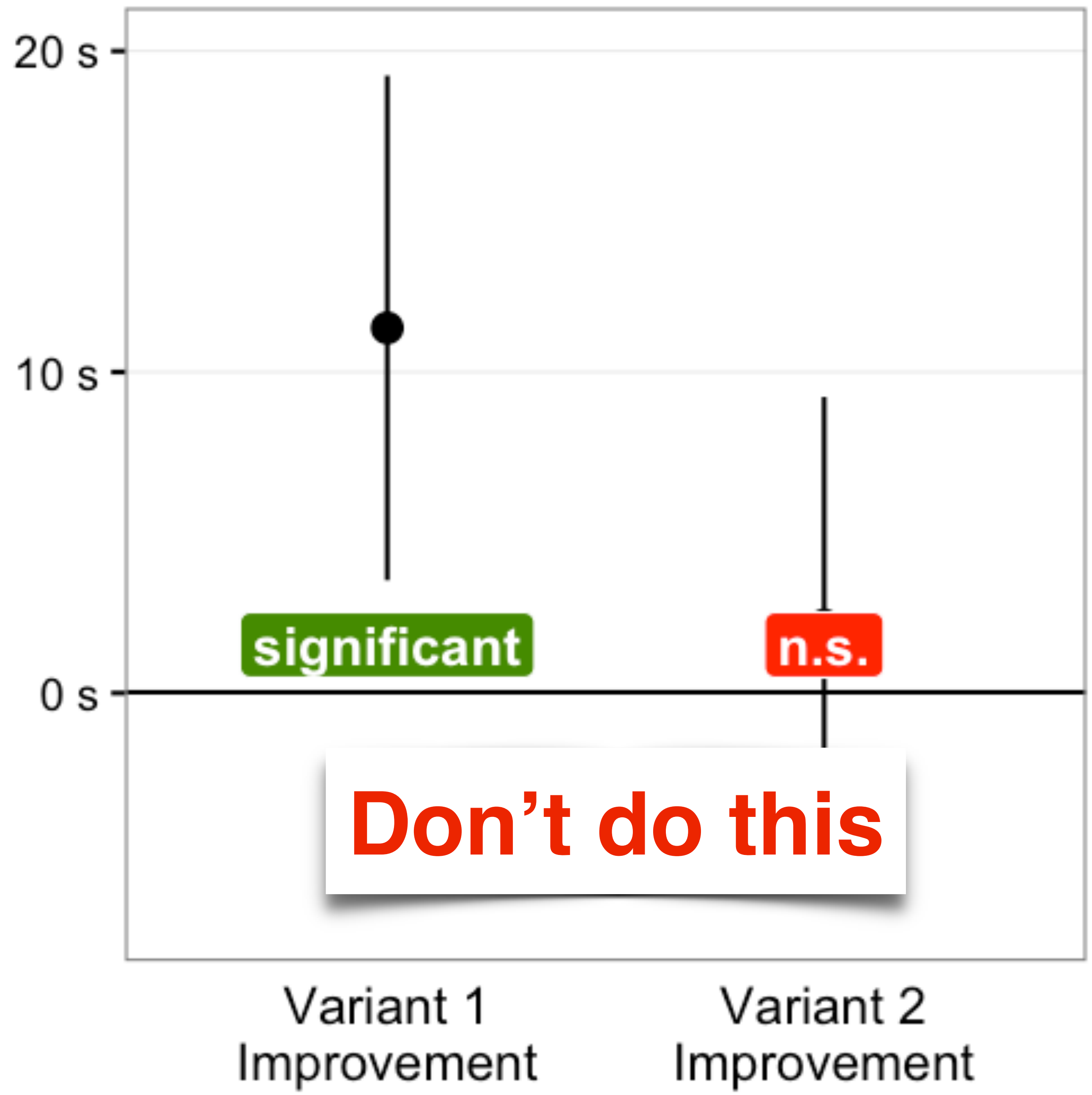
False solution: discretize / dichotomize



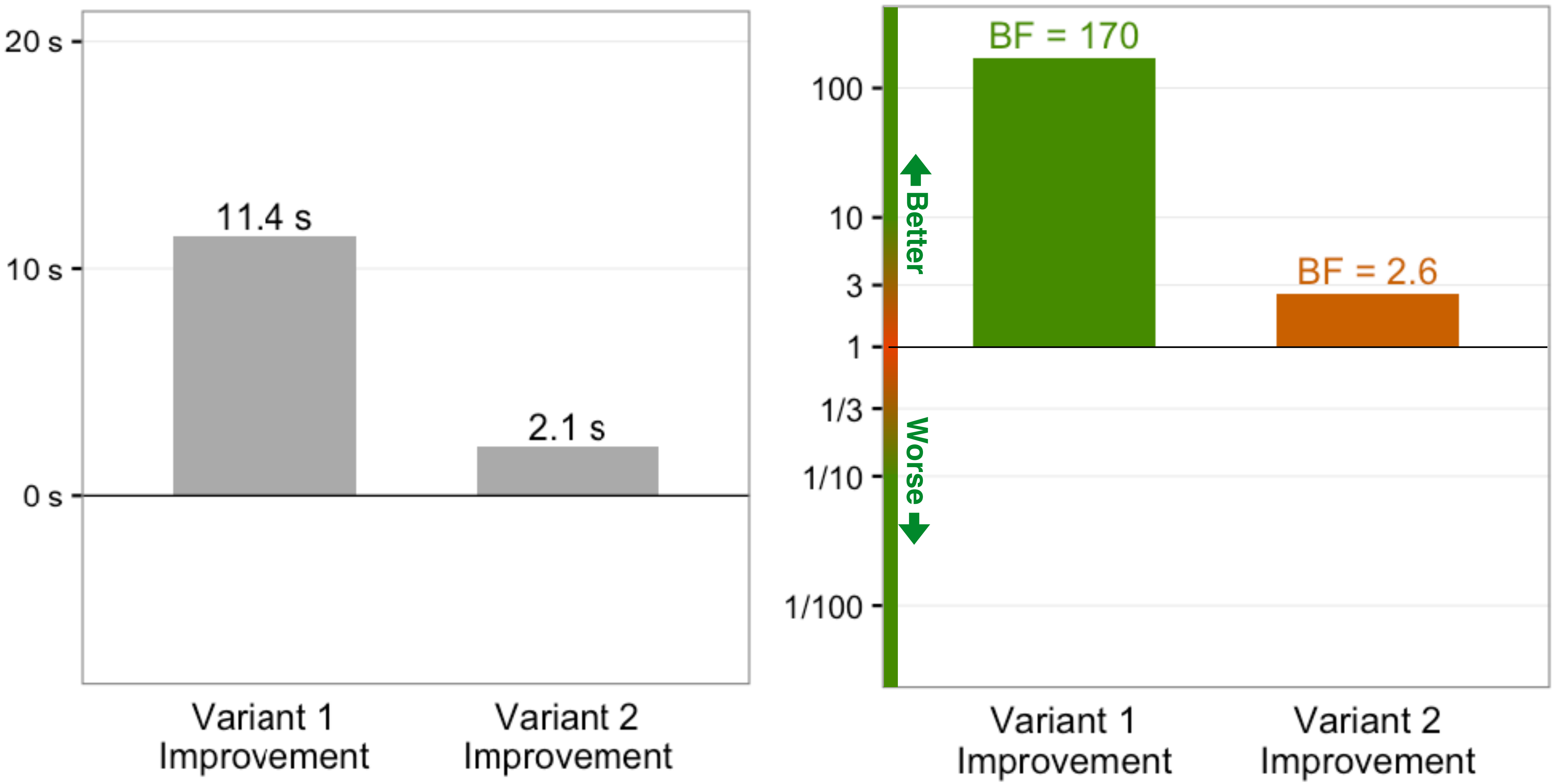
False solution: discretize / dichotomize



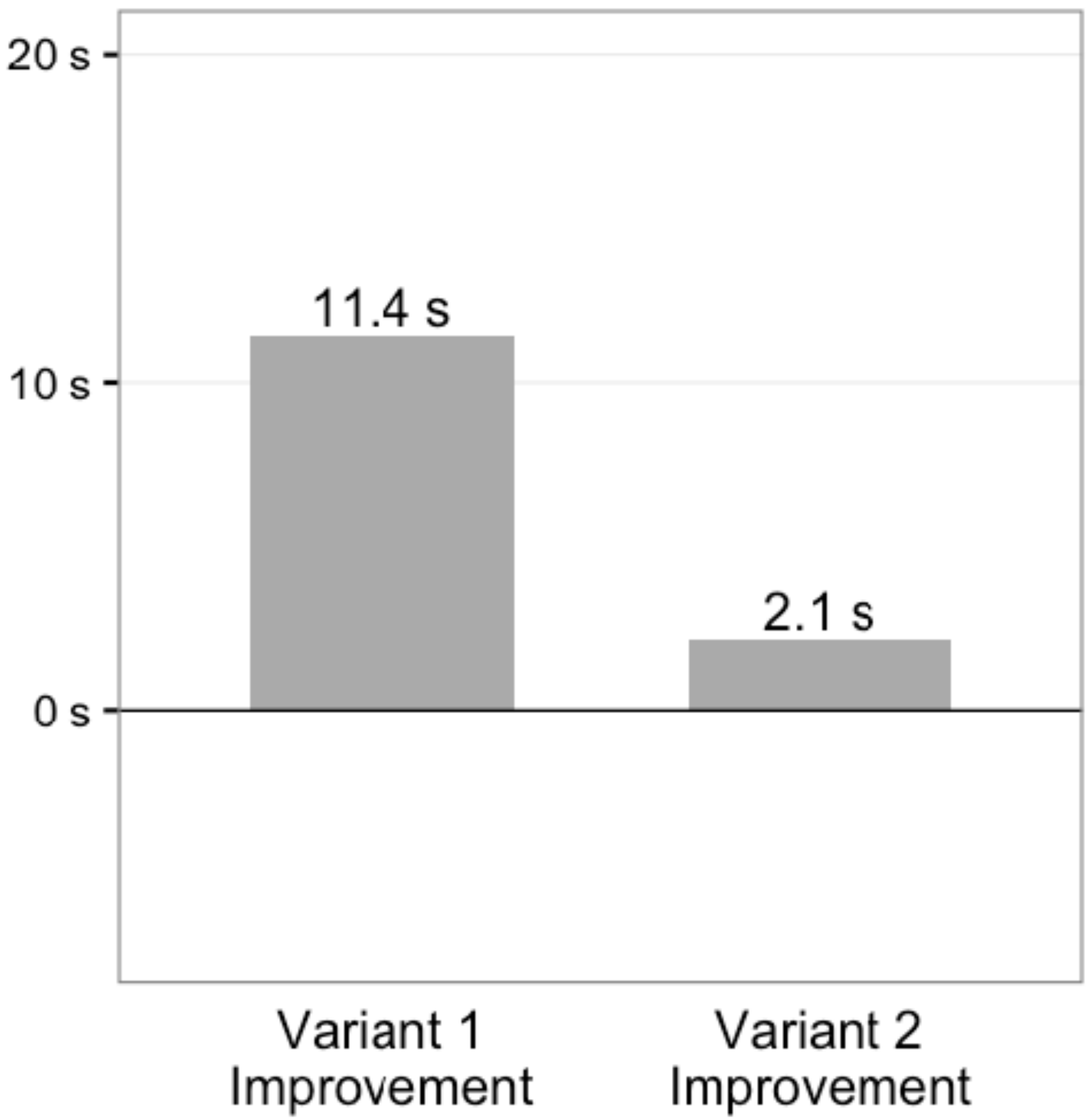
False solution: discretize / dichotomize

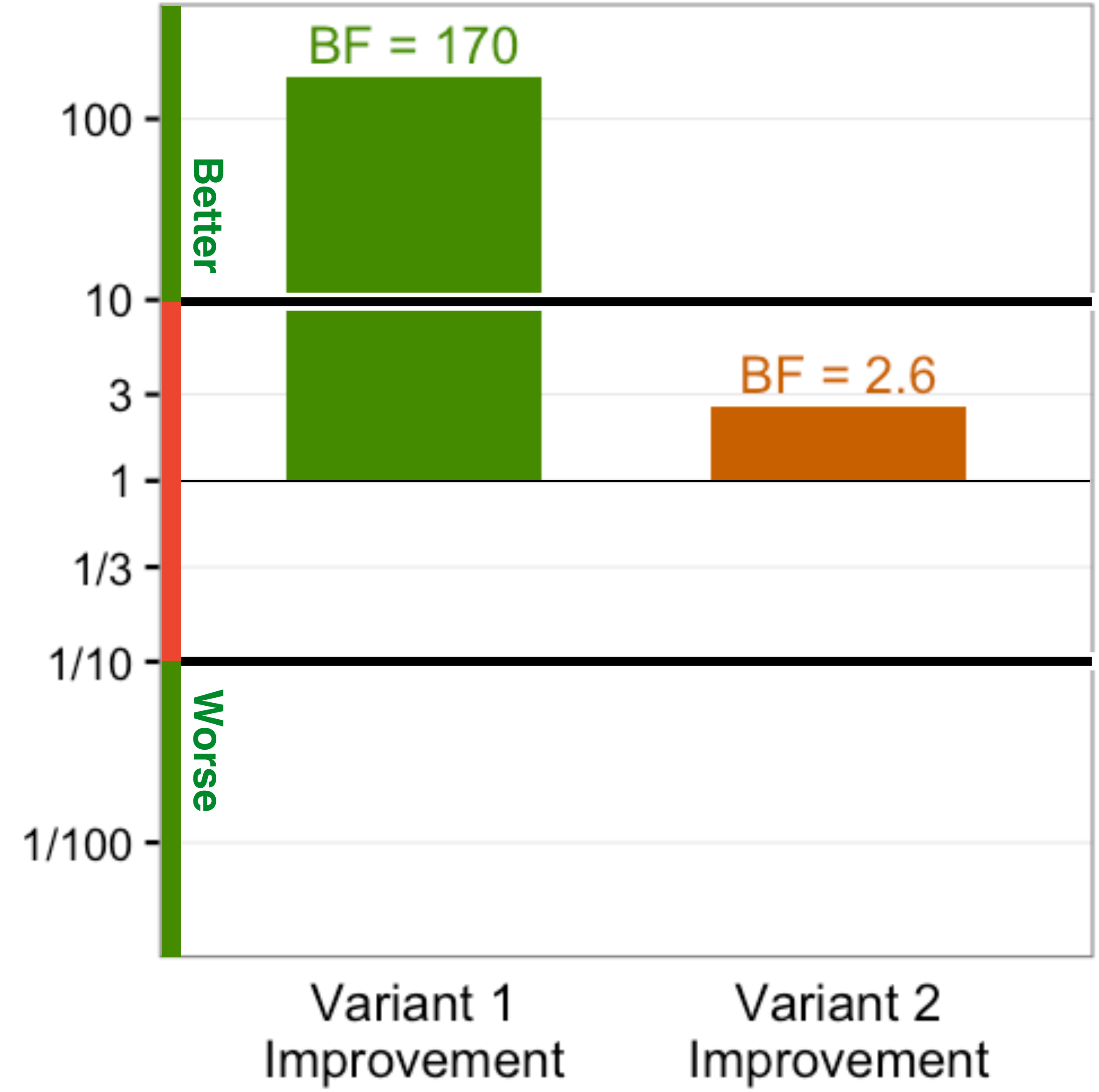


False solution: discretize / dichotomize

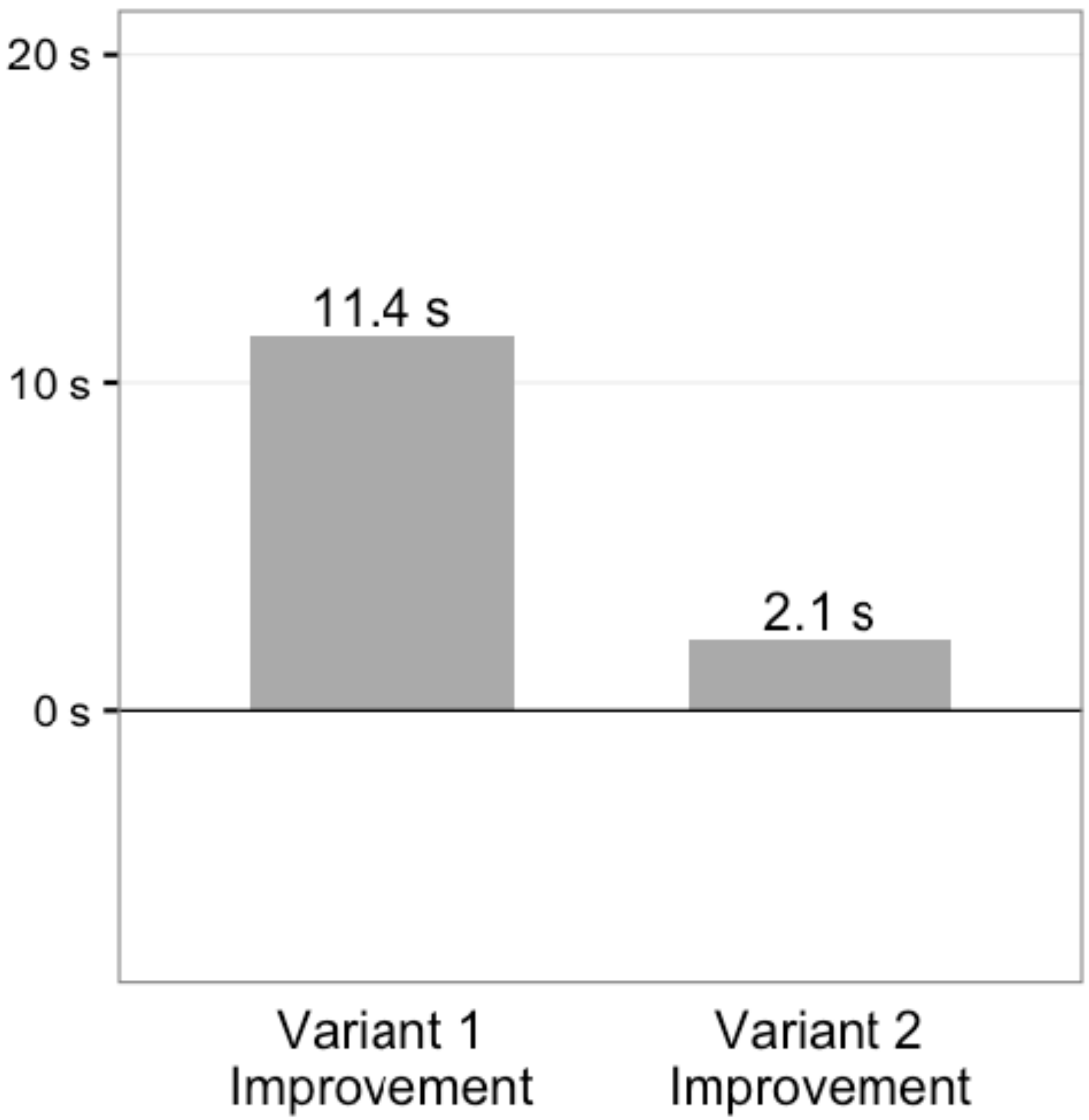


False solution: discretize / dichotomize

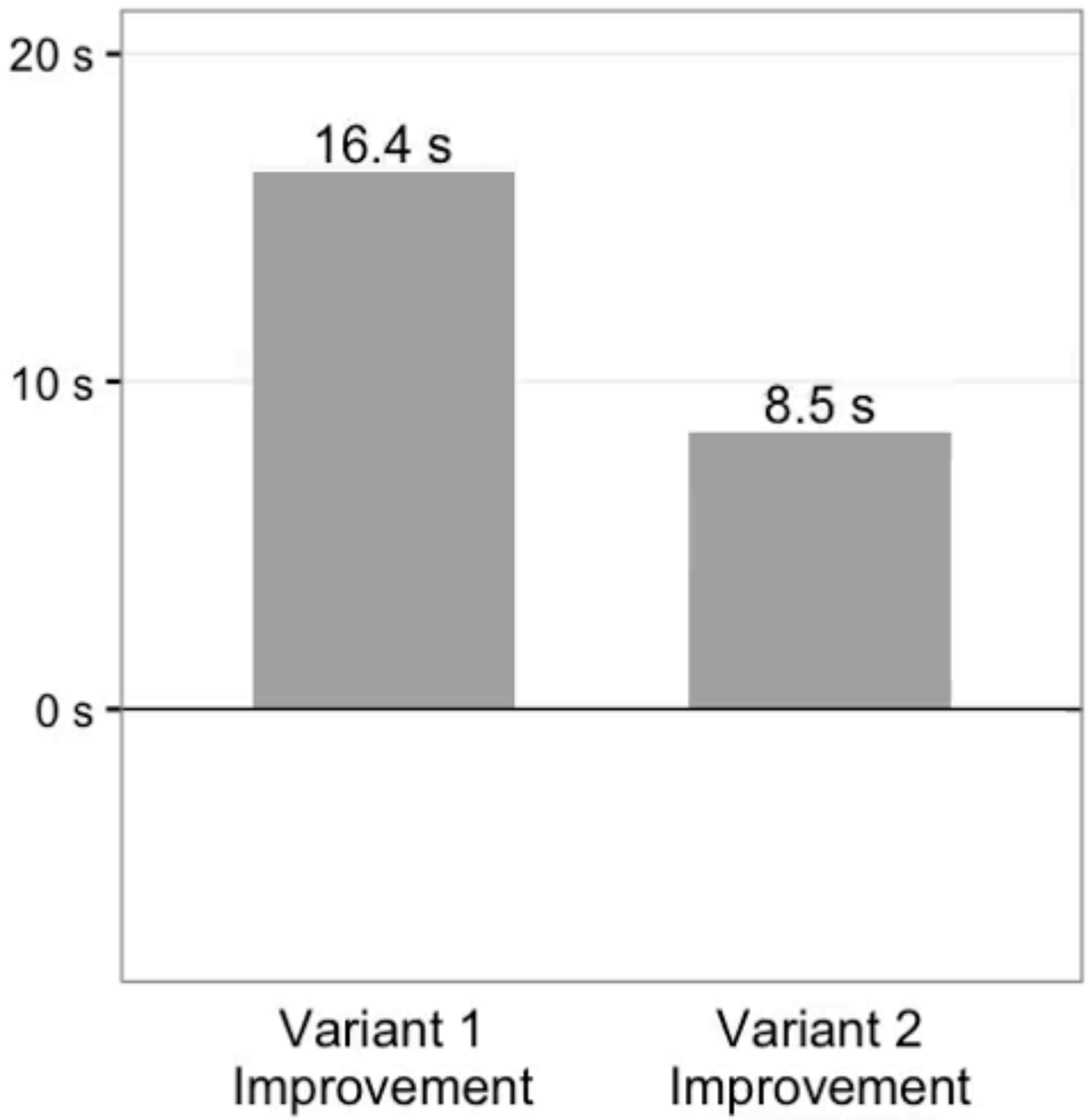


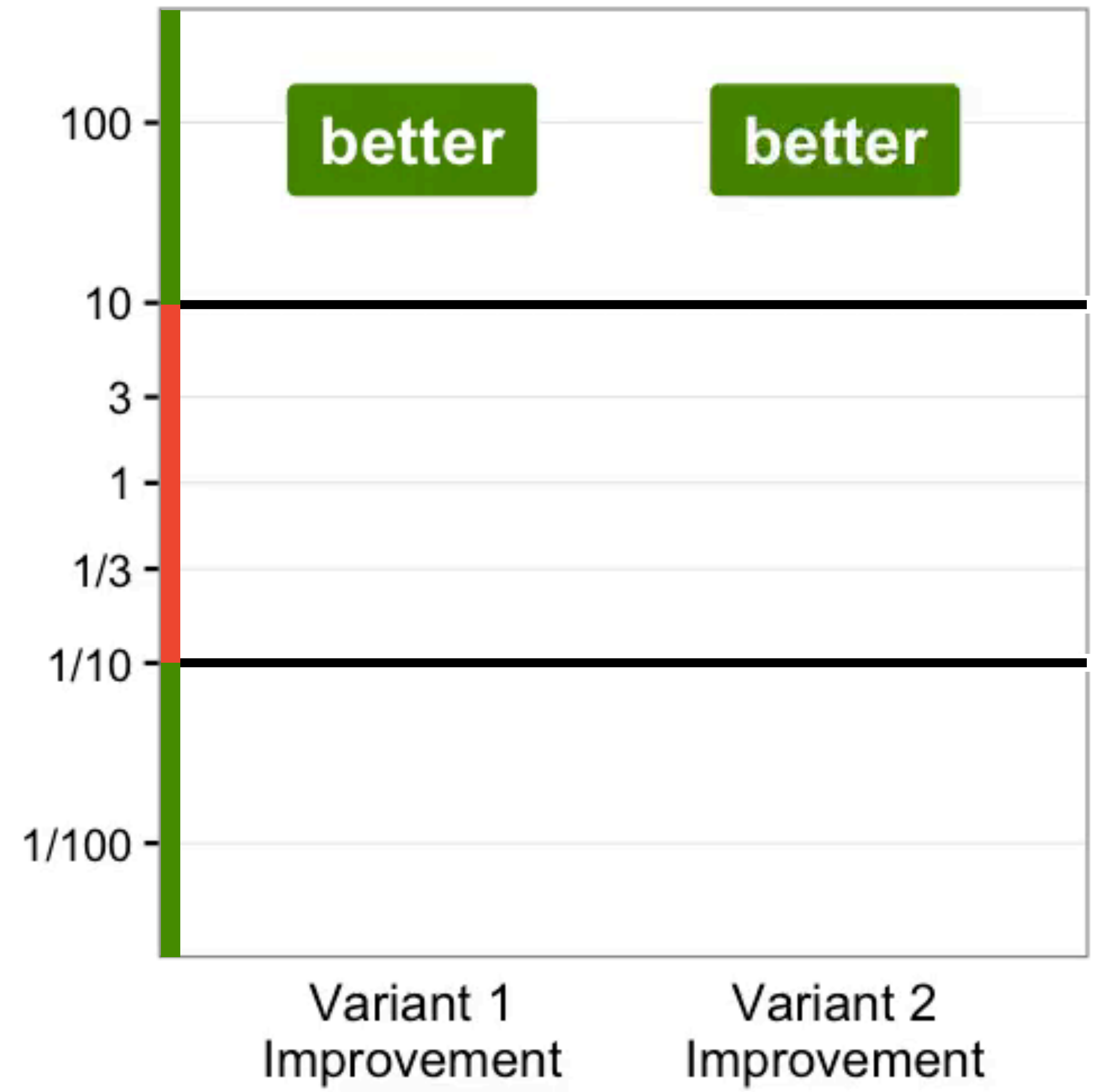


False solution: discretize / dichotomize

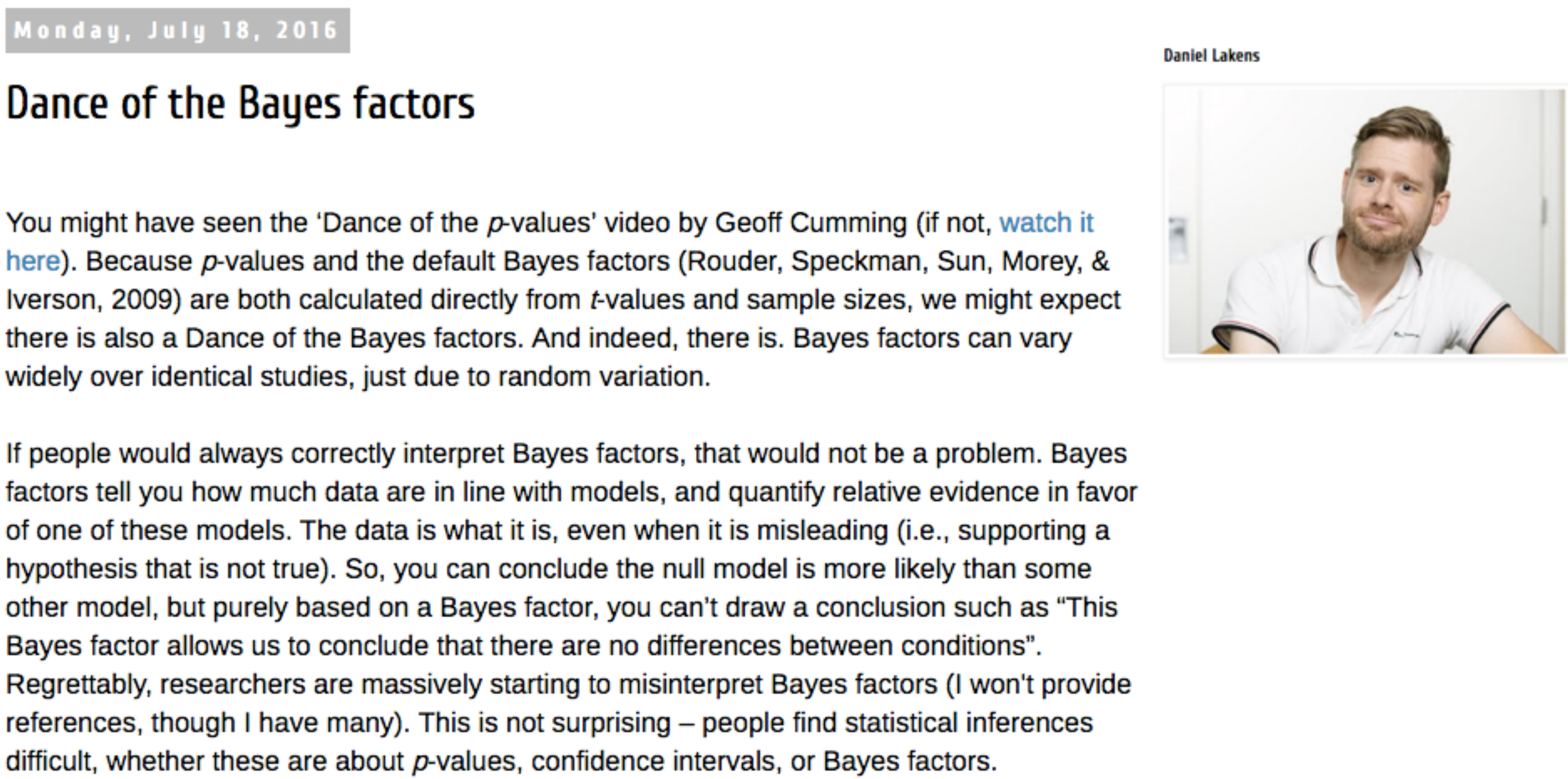


False solution: discretize / dichotomize



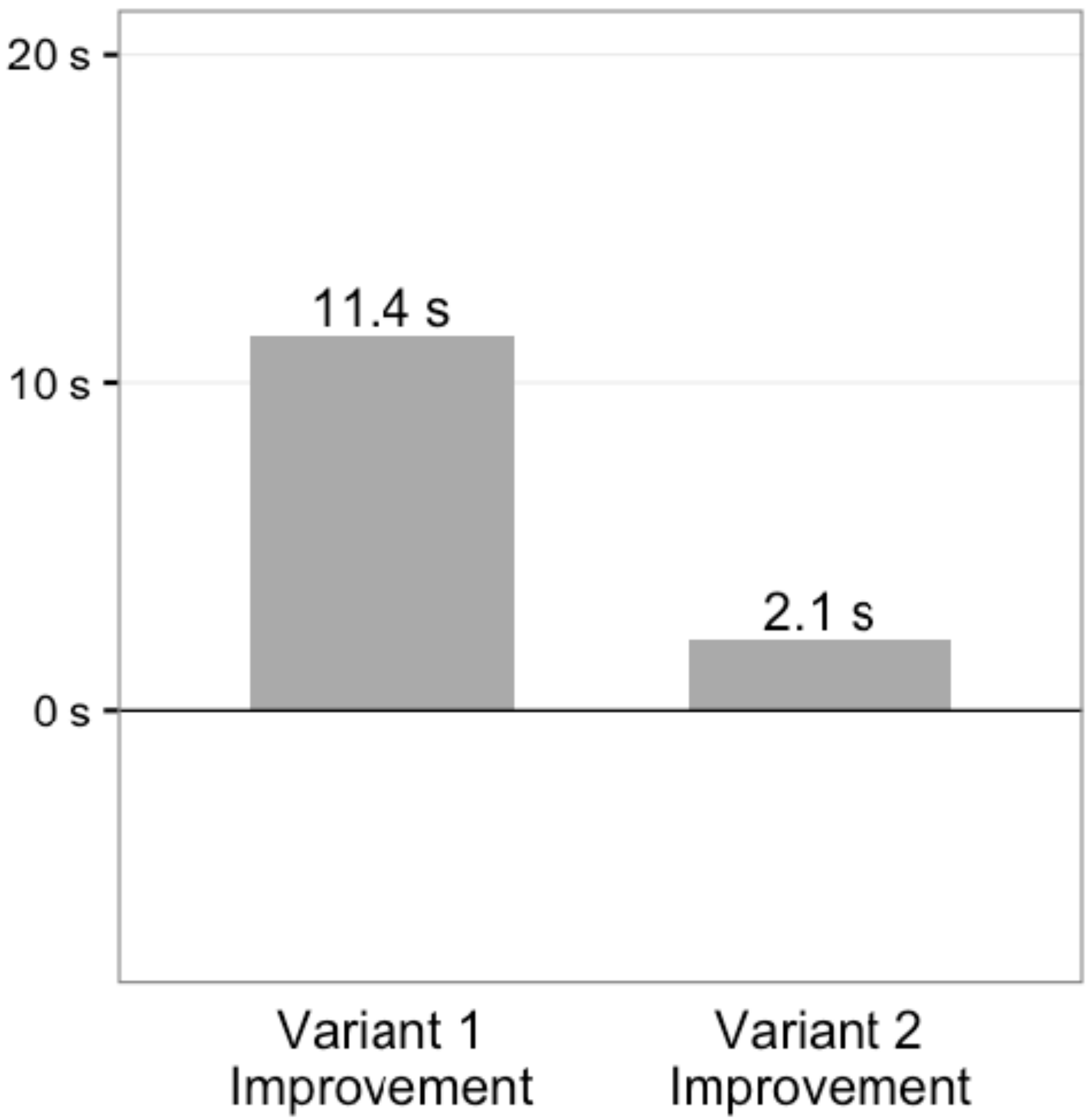


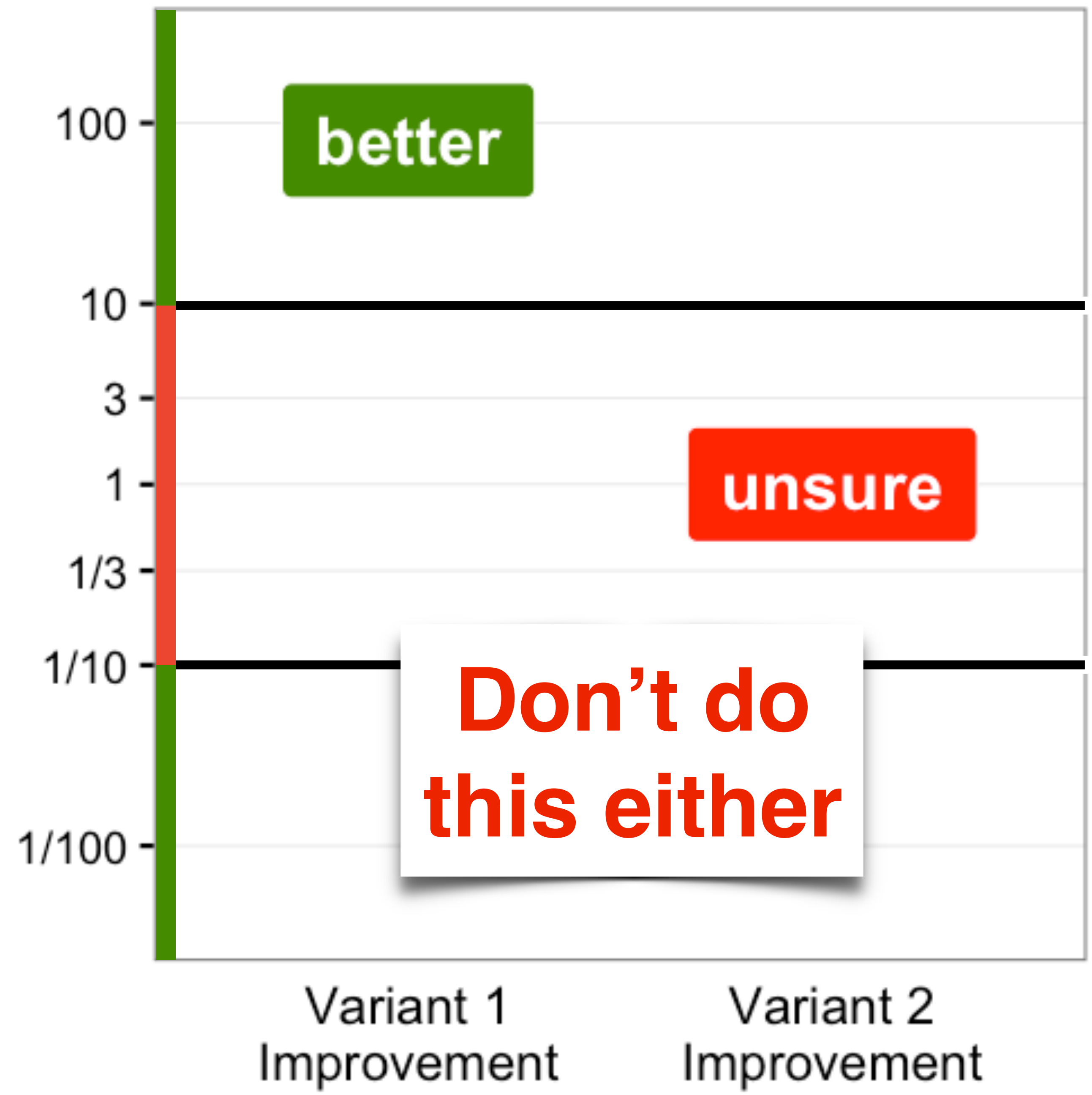
False solution: discretize / dichotomize



(Lakens, 2016)

False solution: discretize / dichotomize

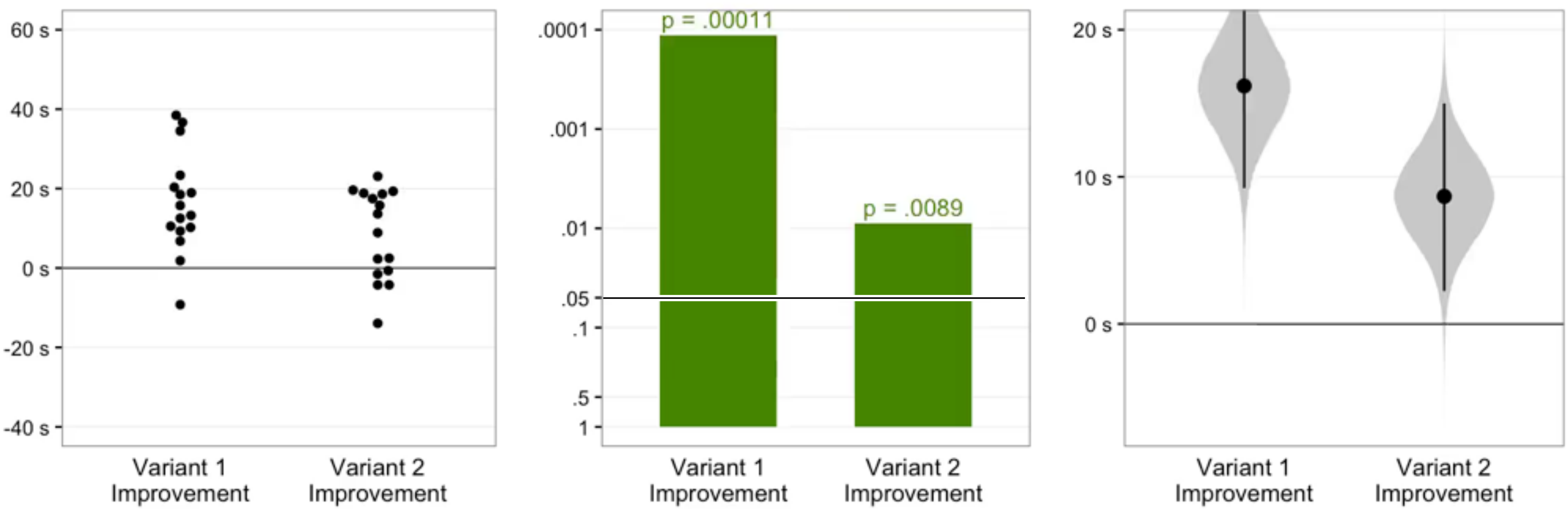




Better solution:
use large samples

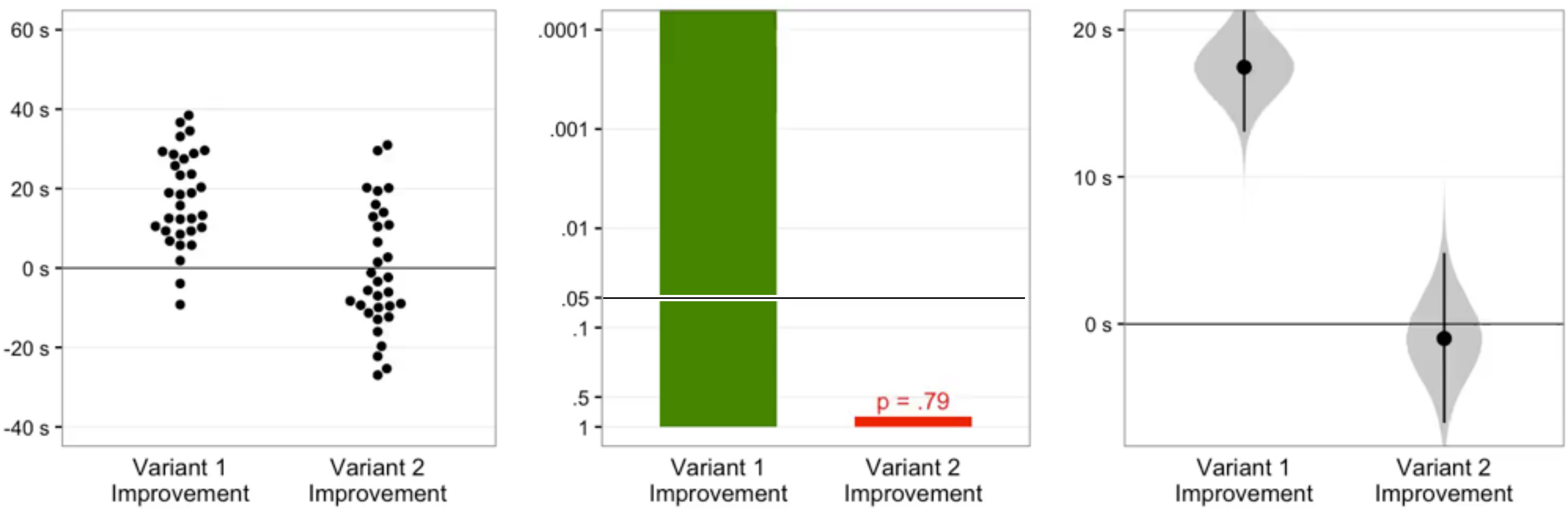
Better solution:
use large samples

n = 16



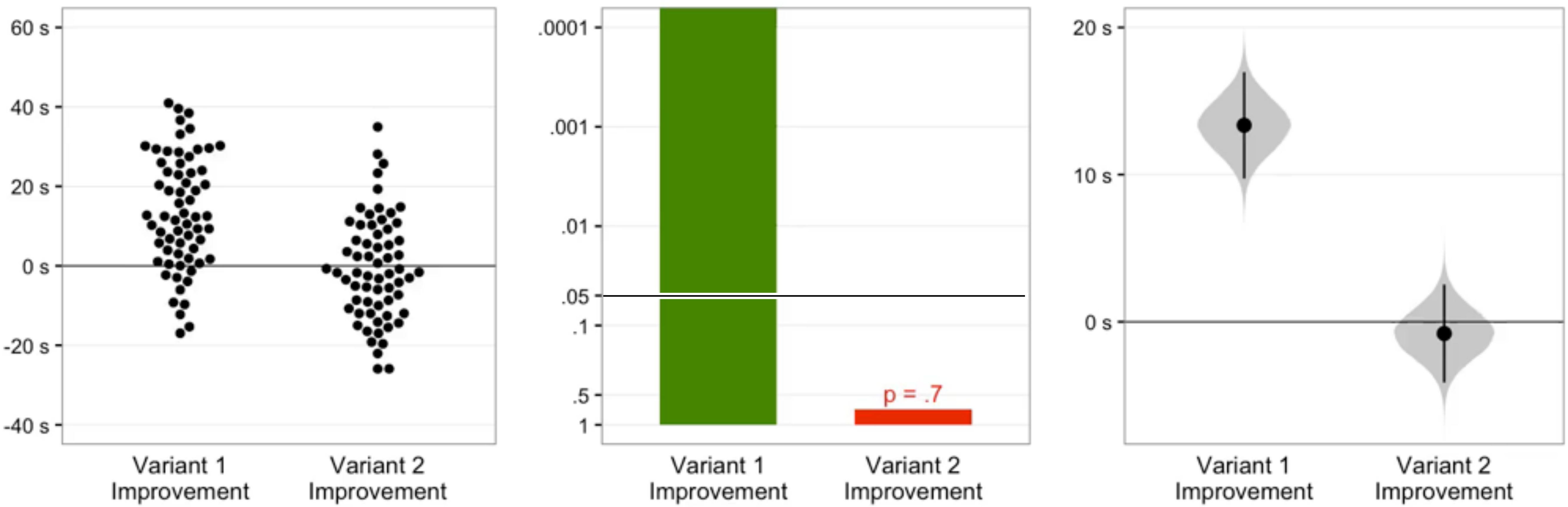
Better solution: use large samples

n = 32



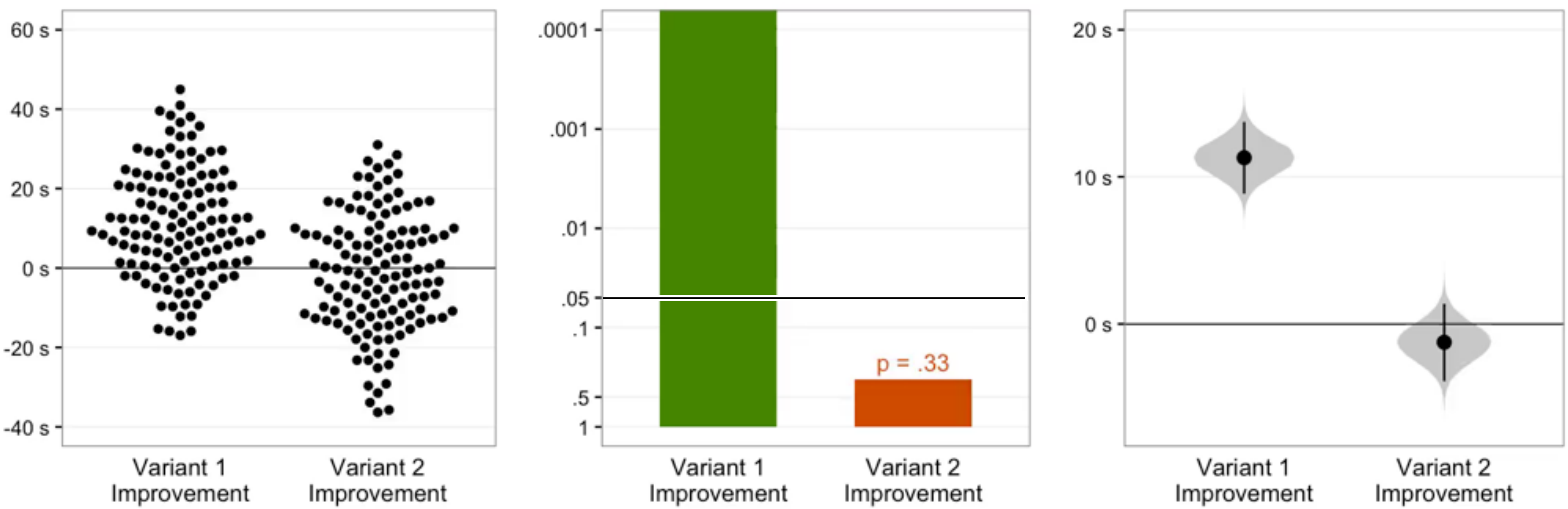
Better solution: use large samples

$n = 64$



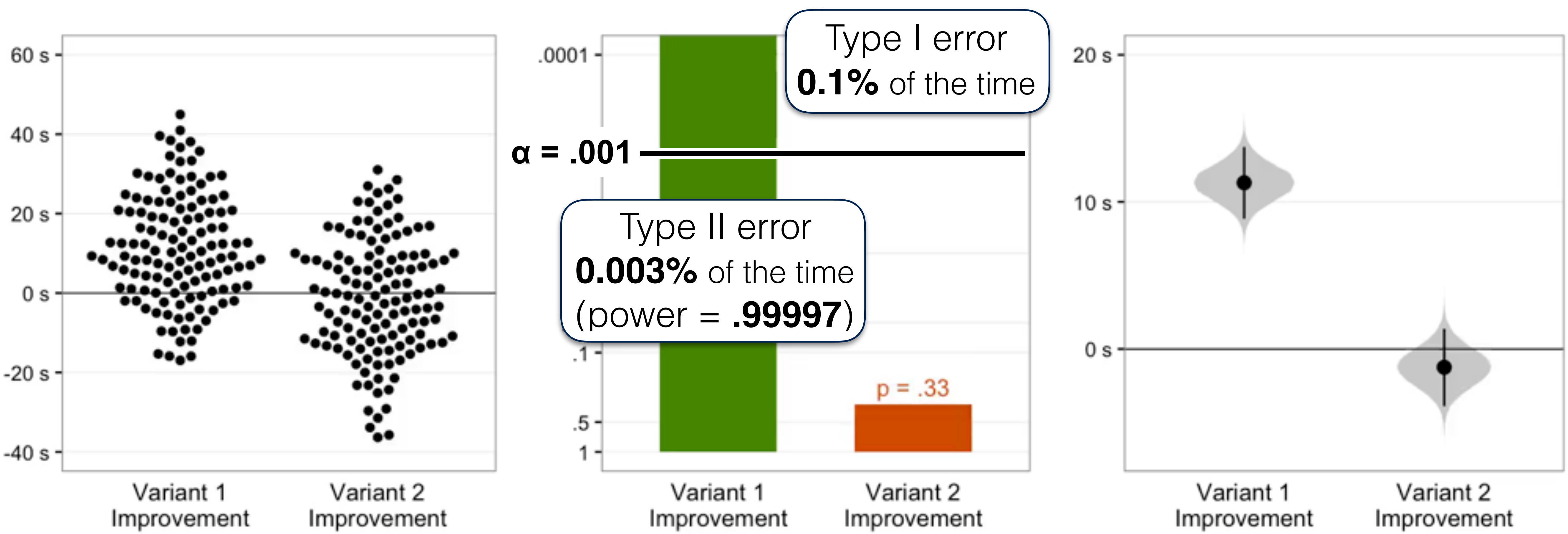
Better solution: use large samples

n = 128



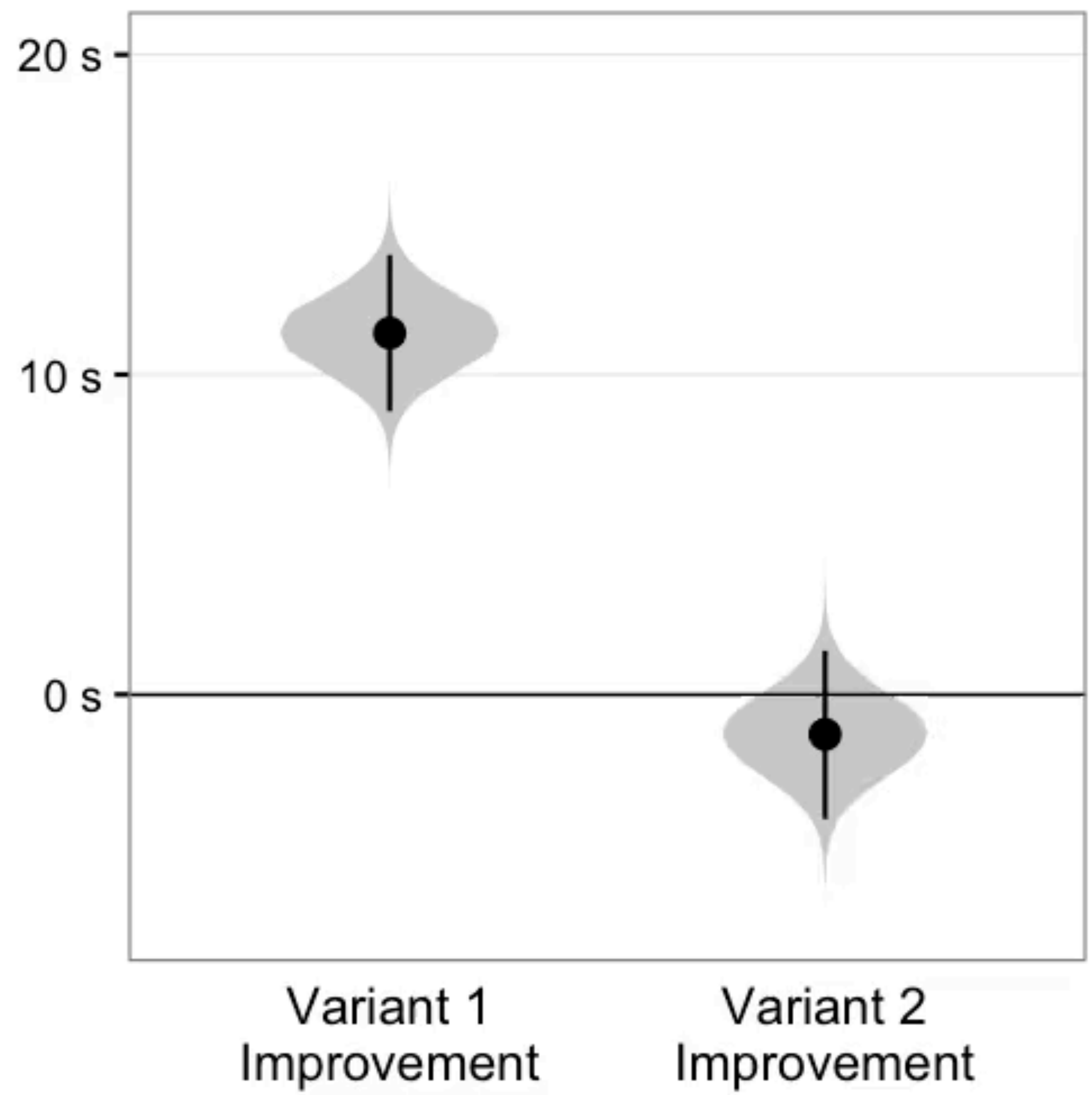
Better solution: use large samples

$n = 128$



Better solution: use large samples

n = 128

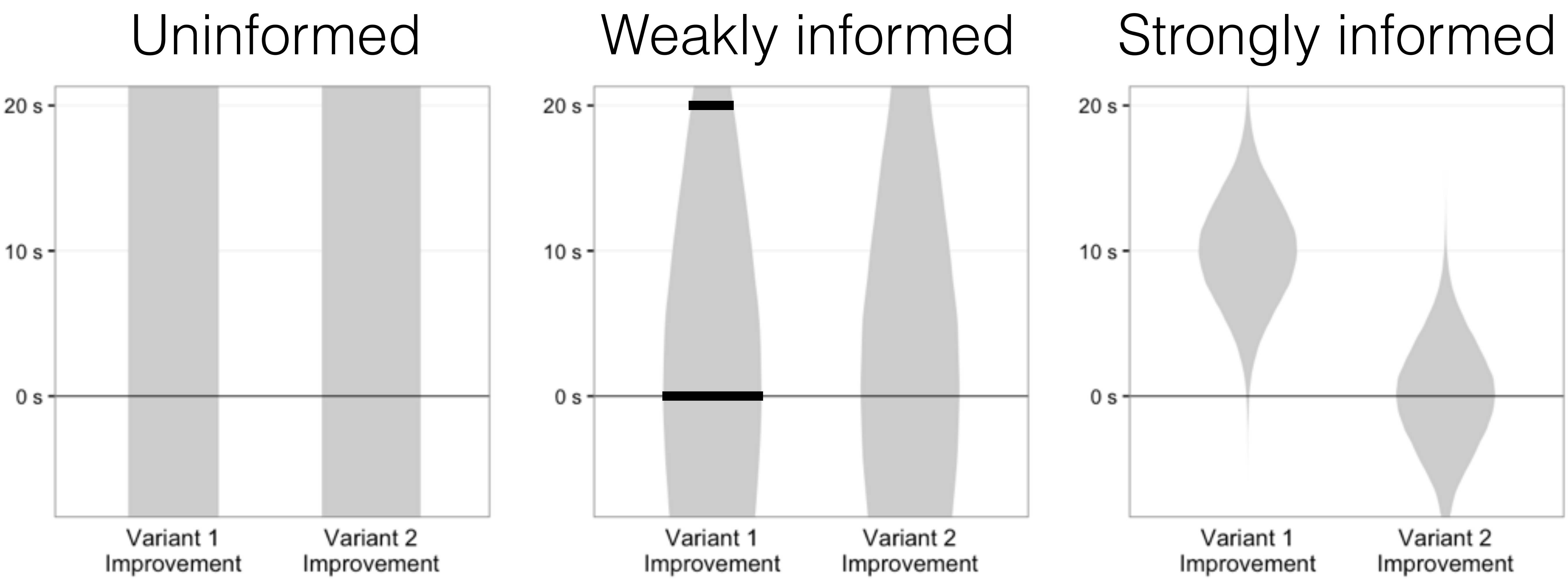


Limits of large samples

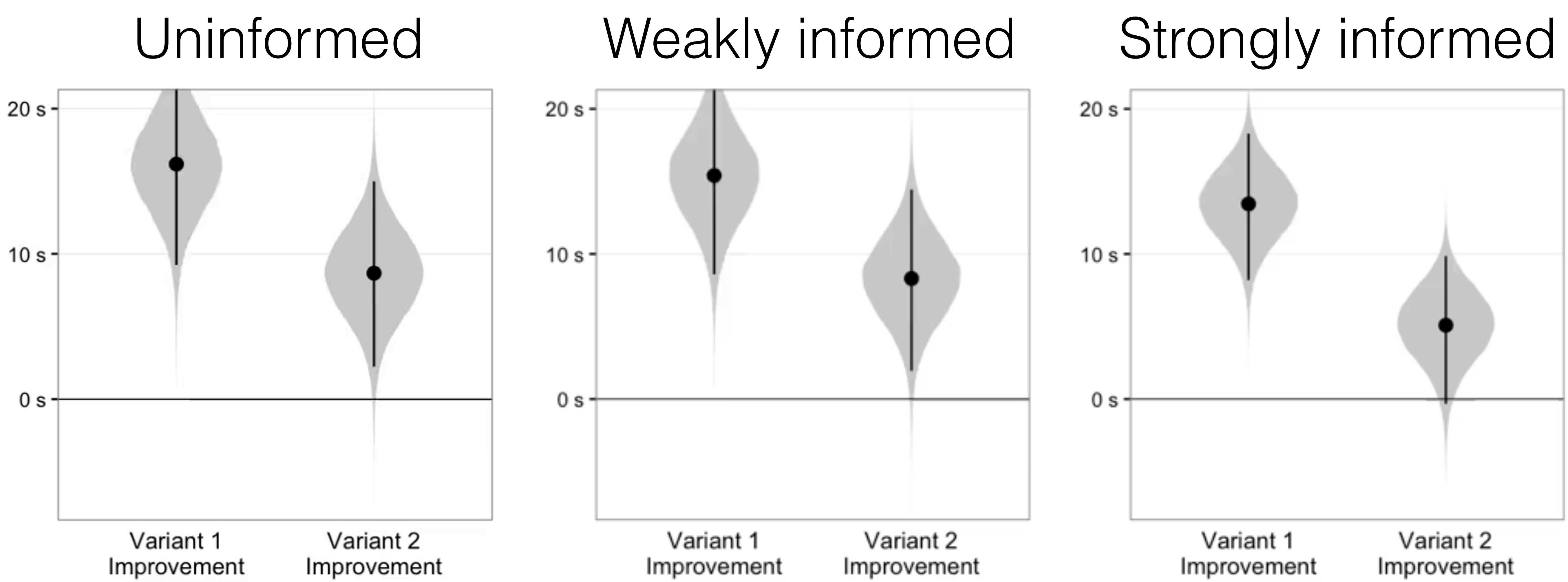
- More participants stabilize dances only slowly
- Running participants can be costly
- Power can be increased by better measurement
- Researchers would still report secondary findings with lower power
- Still need ways of conveying uncertain results

Another solution:
use informed priors

Another solution: use informed priors



Another solution: use informed priors



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 willingless, new principles

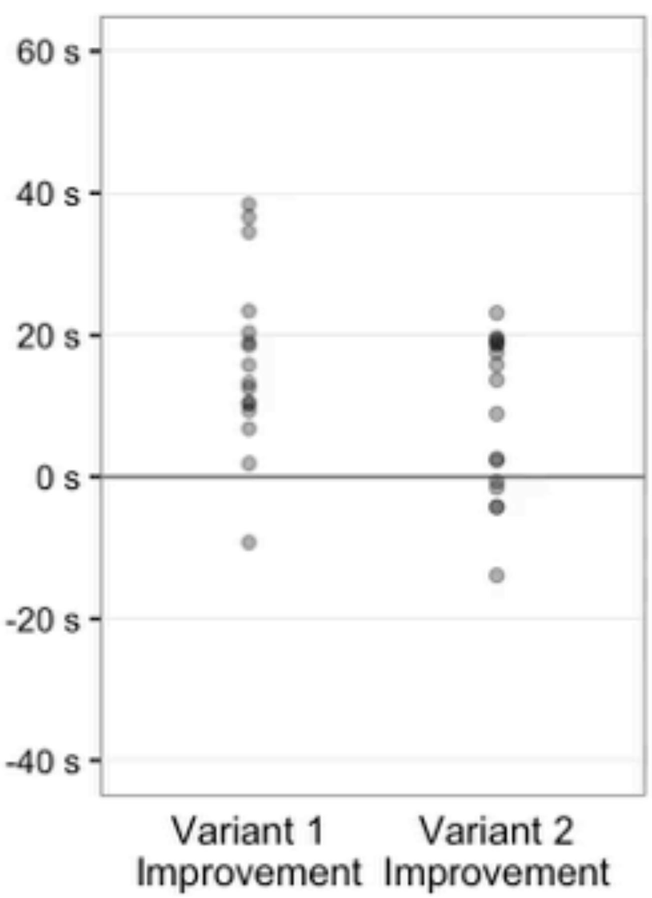
Robustness principle

- 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

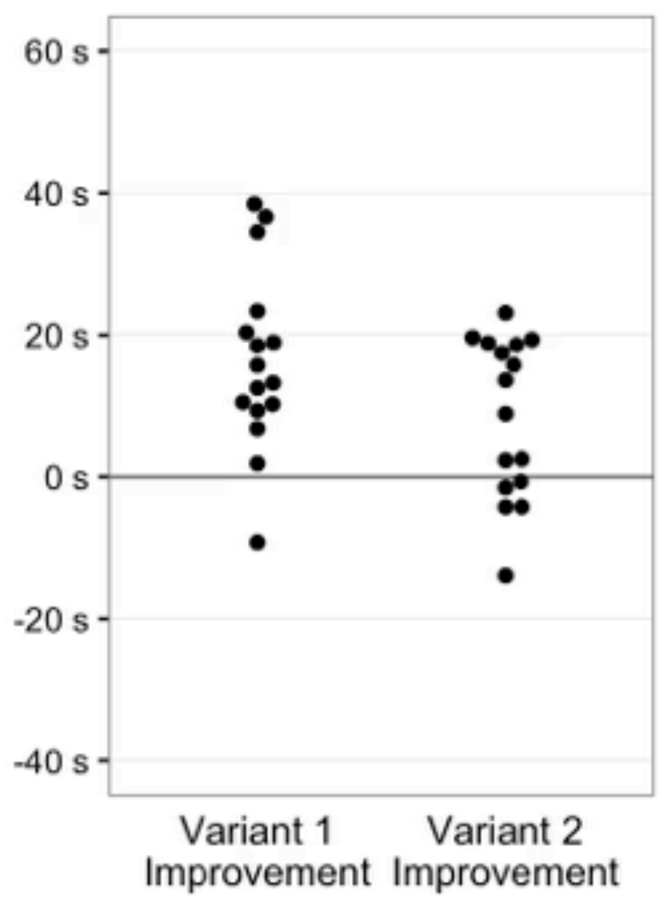
Examples

Plotting distributions

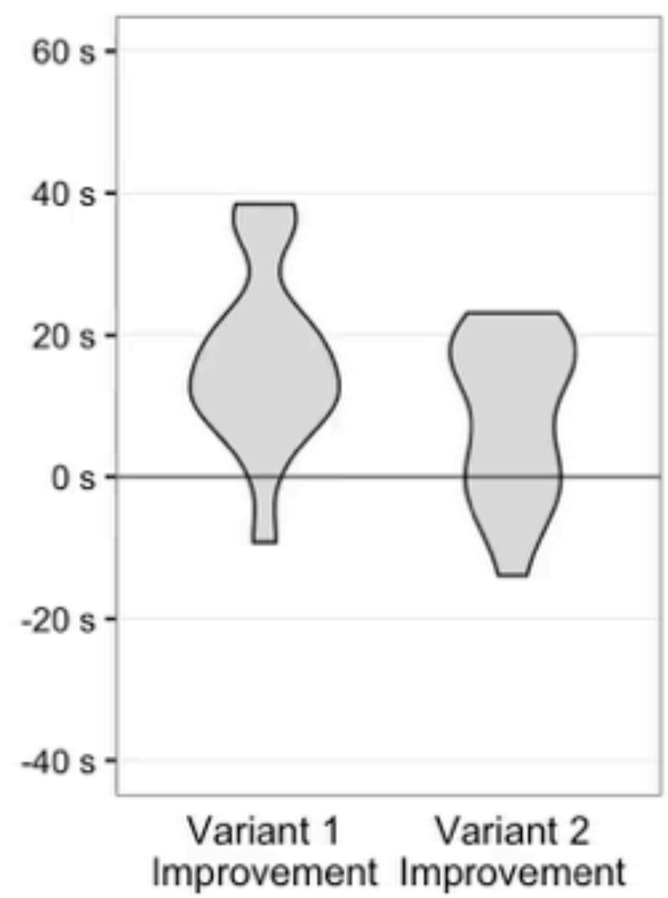
Dot plot



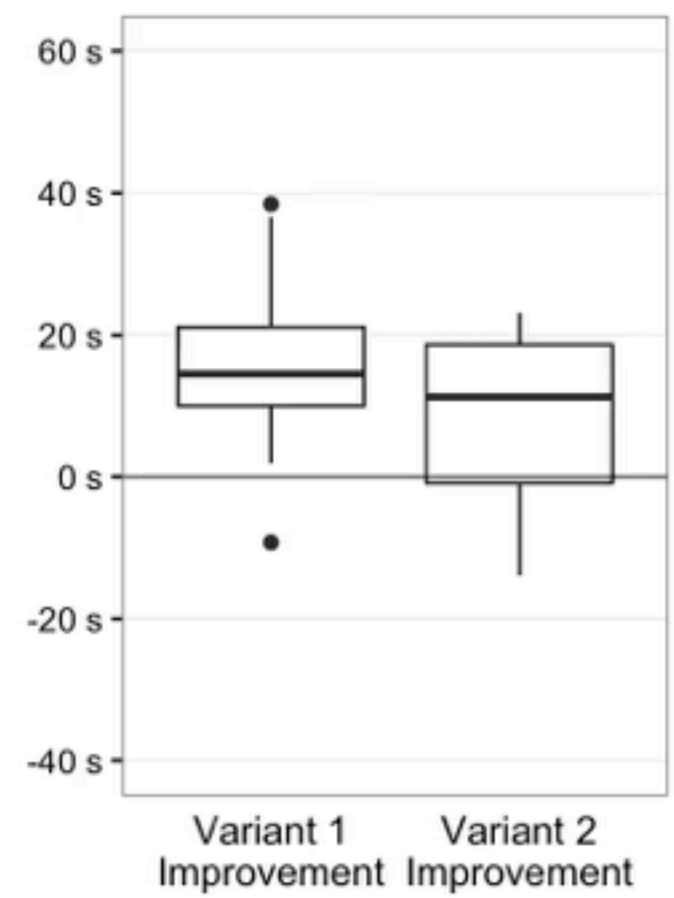
Bee swarm



Violin plot



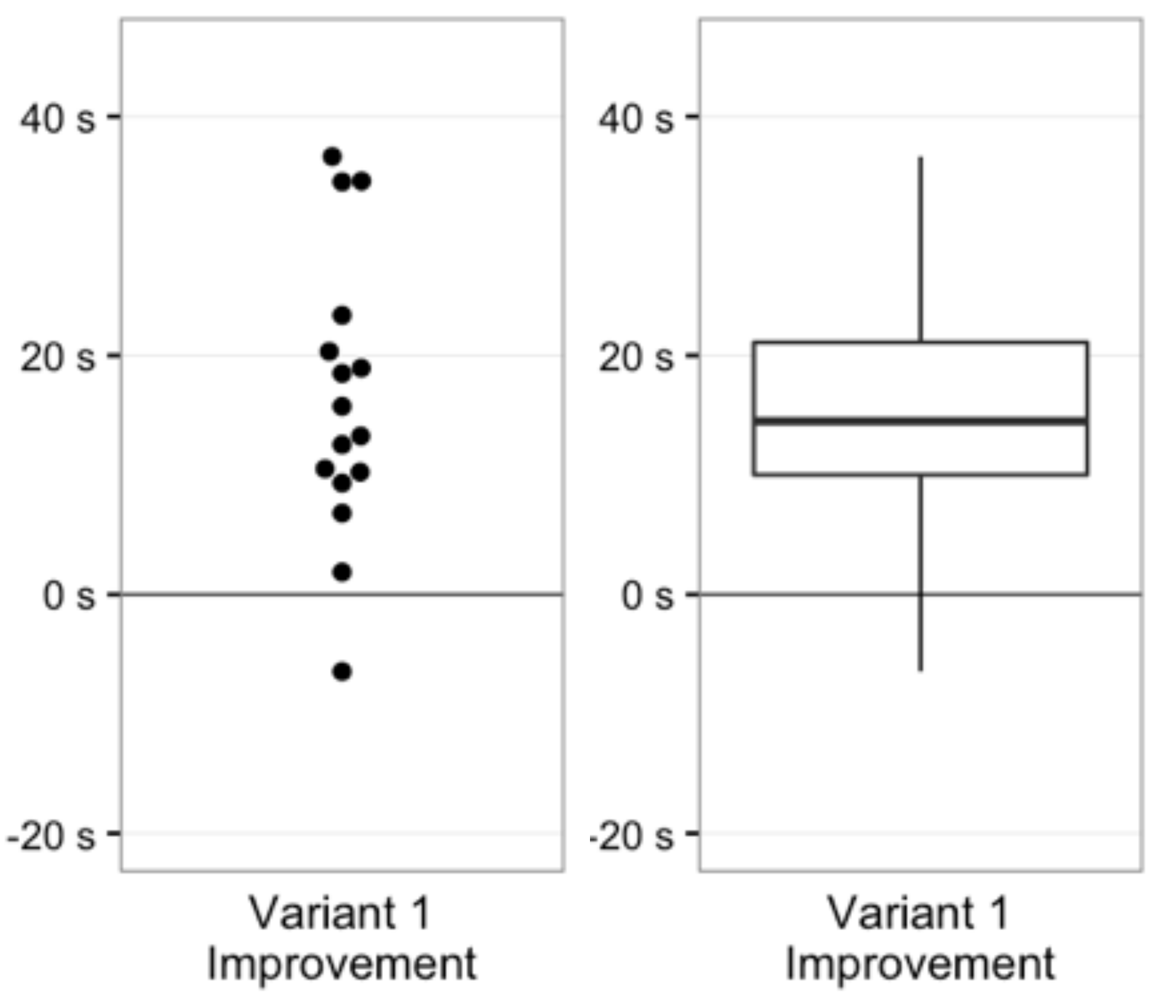
Box plot



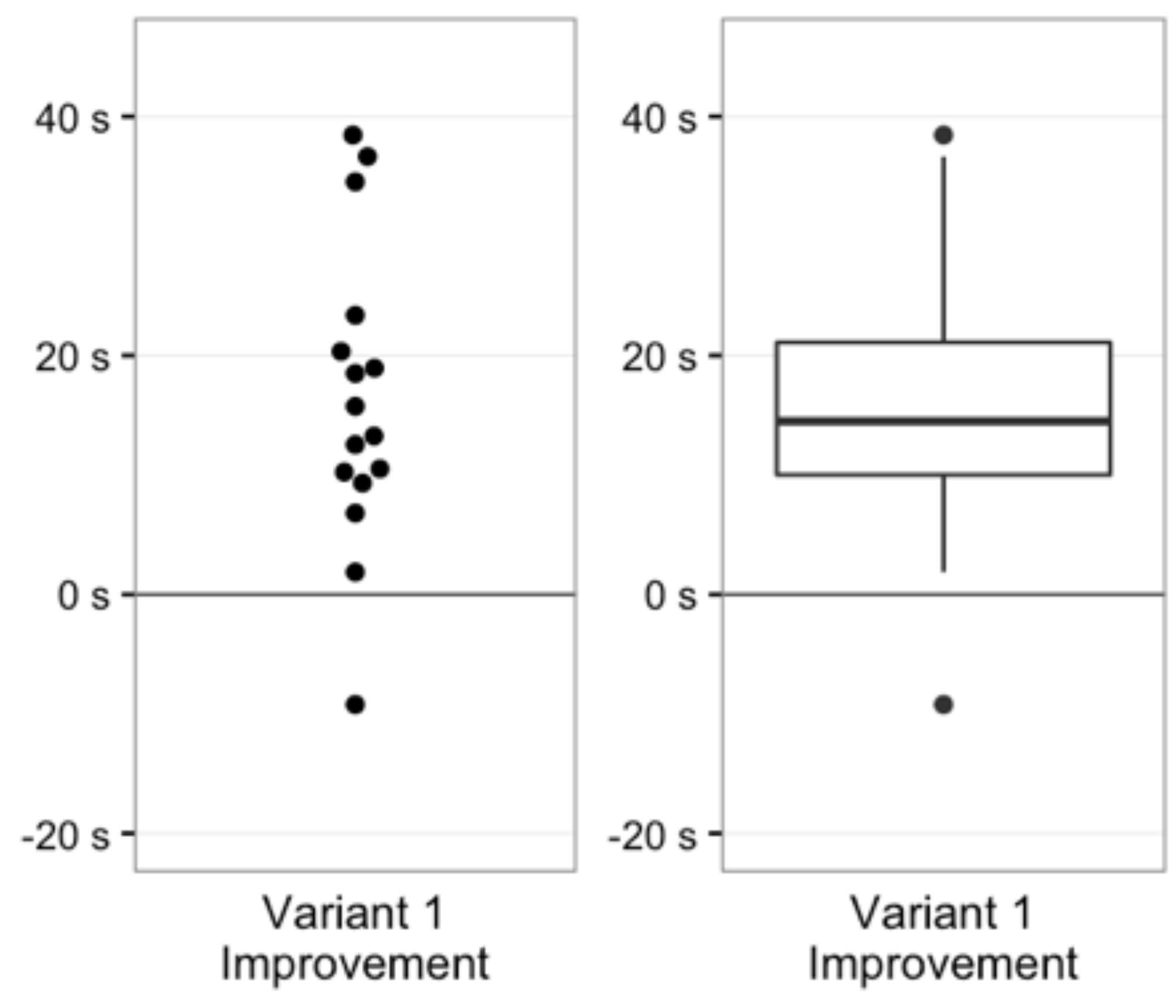
Examples

Plotting distributions

Dataset 1

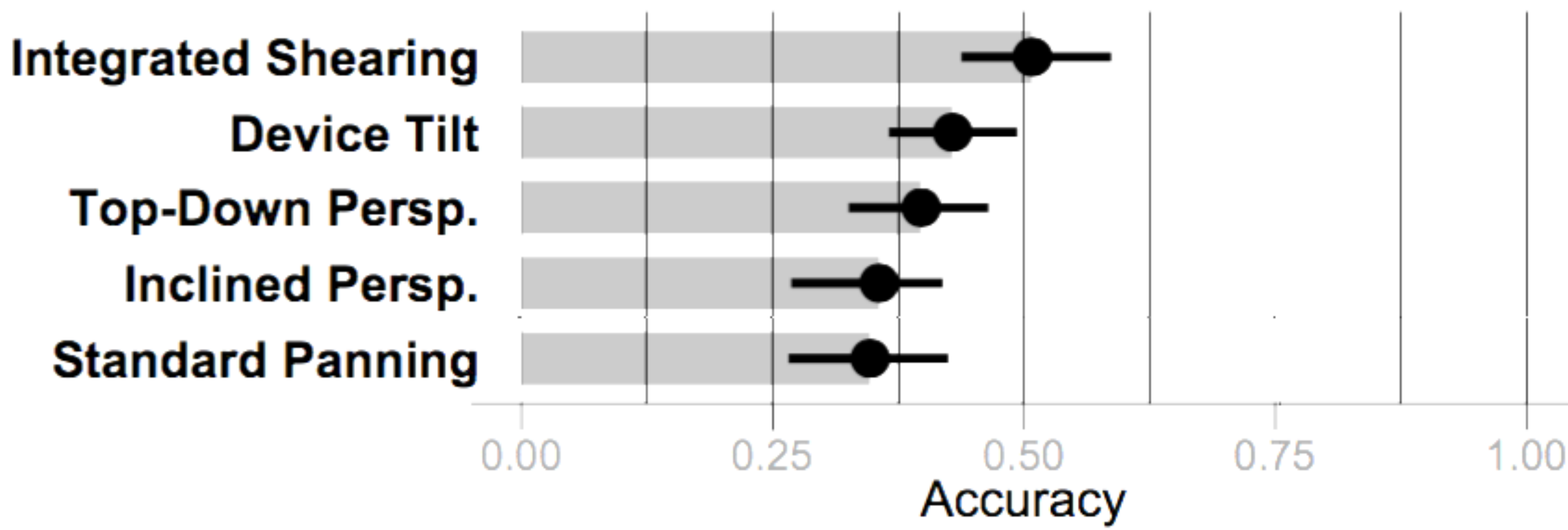


Dataset 2



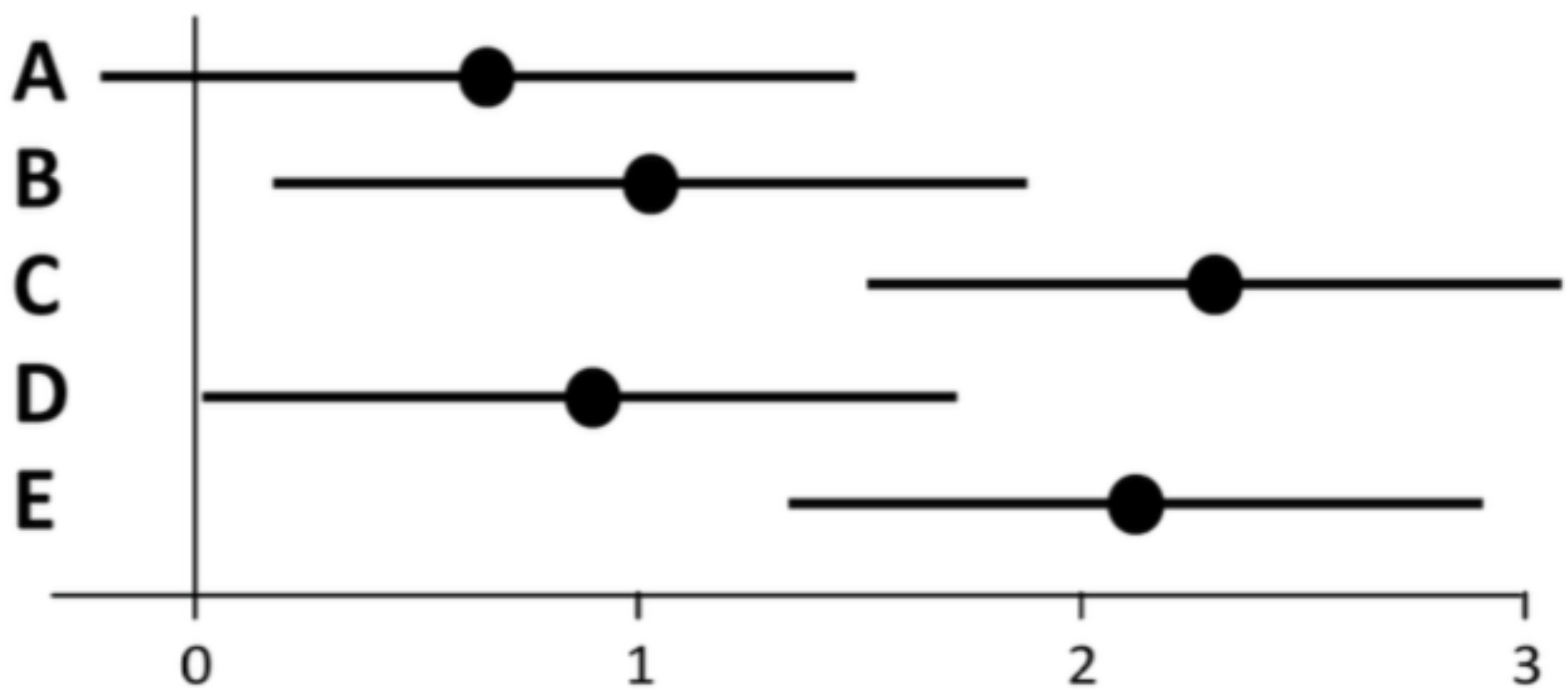
Examples

Sorting by effect size



Examples

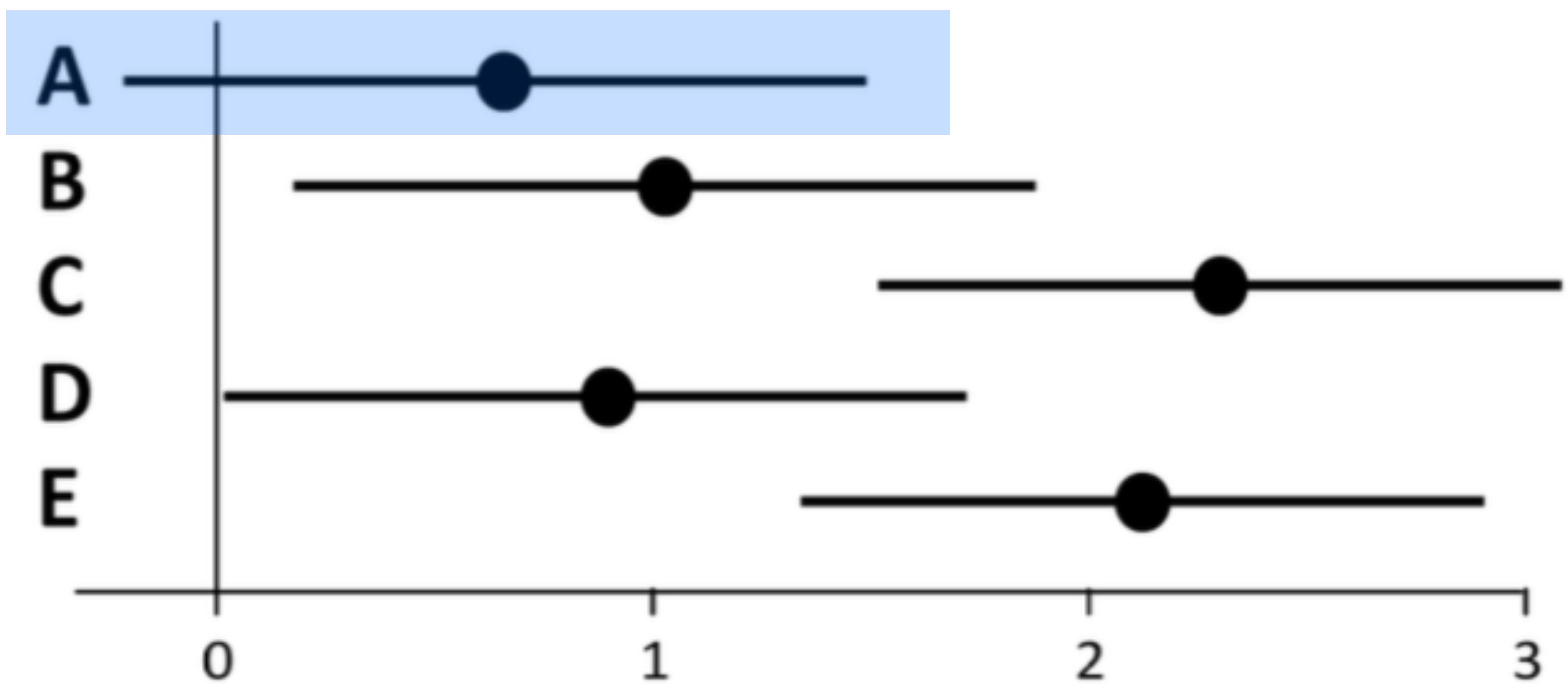
Interpreting multiple CIs



(Dragicevic, 2016)

Examples

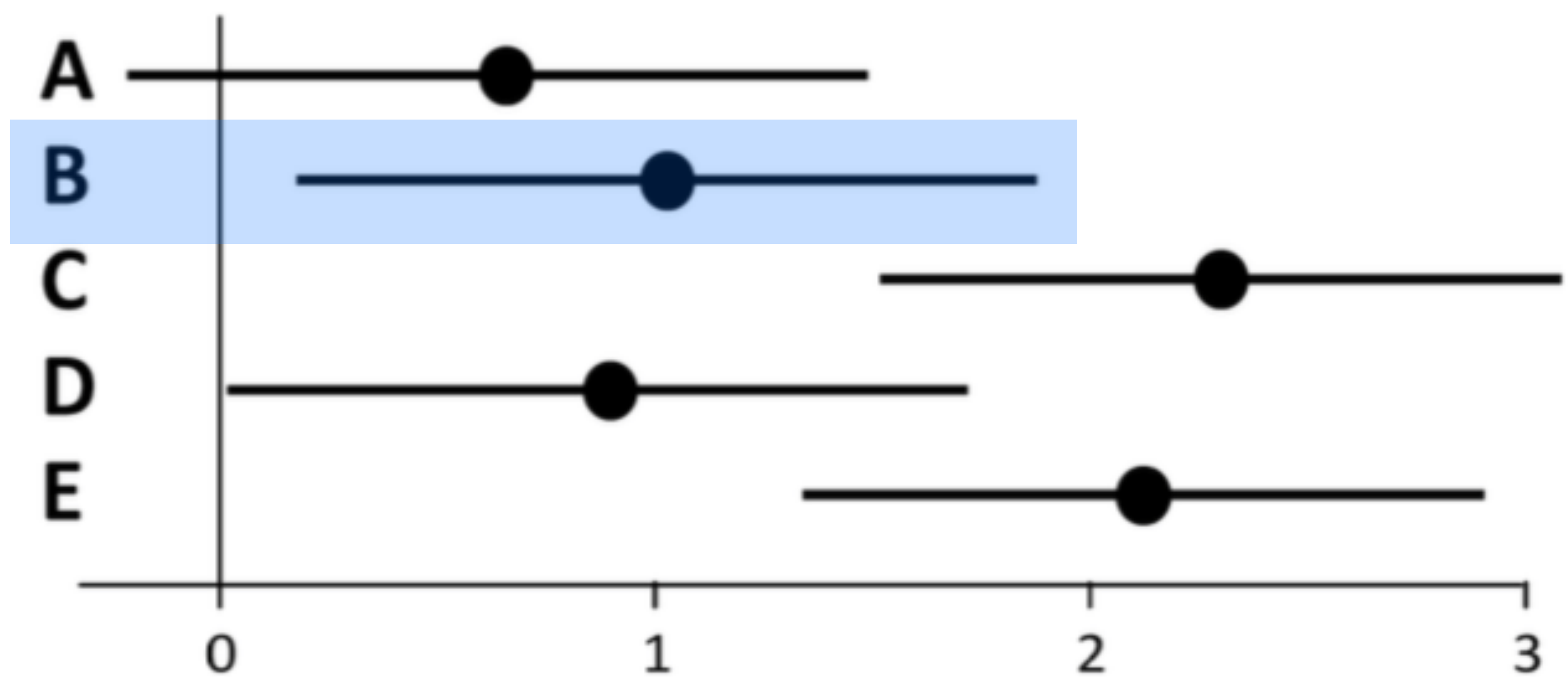
Interpreting multiple CIs



(Dragicevic, 2016)

Examples

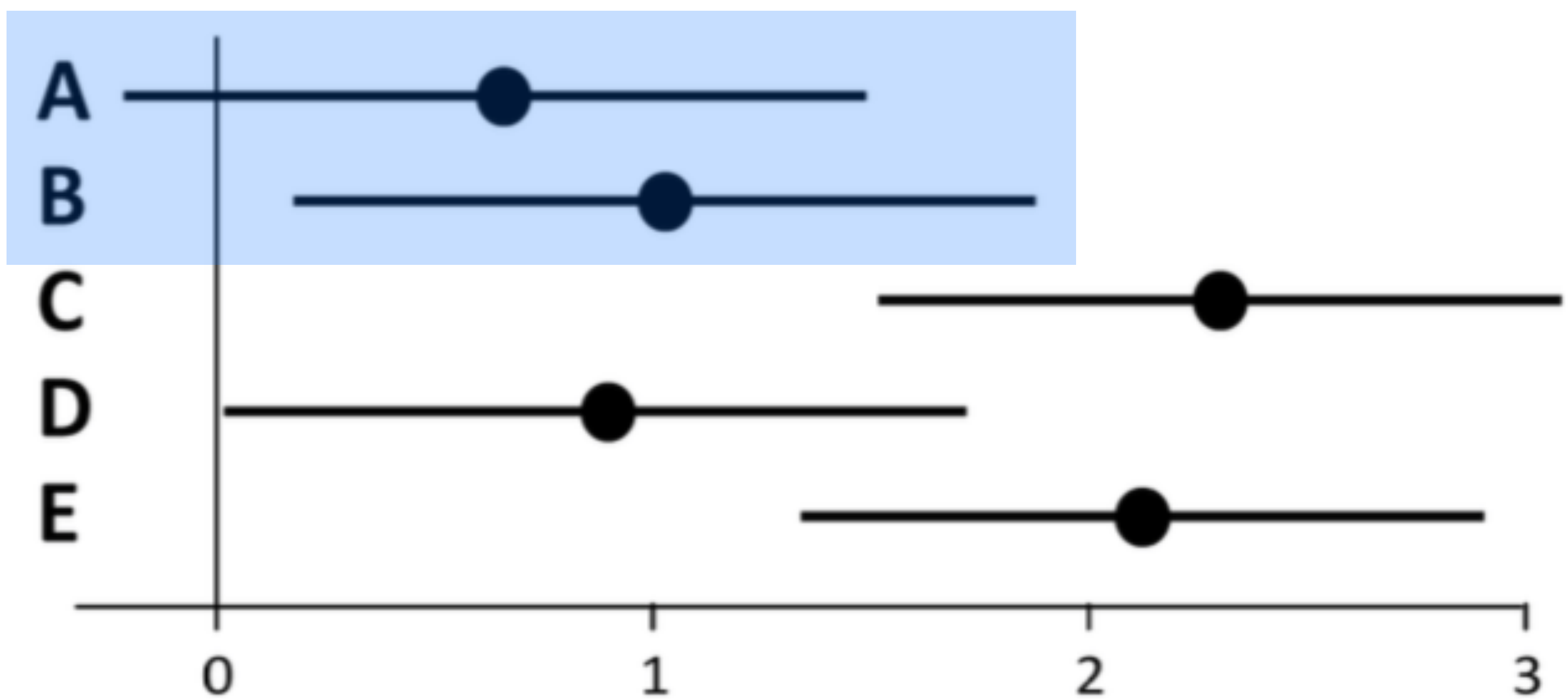
Interpreting multiple CIs



(Dragicevic, 2016)

Examples

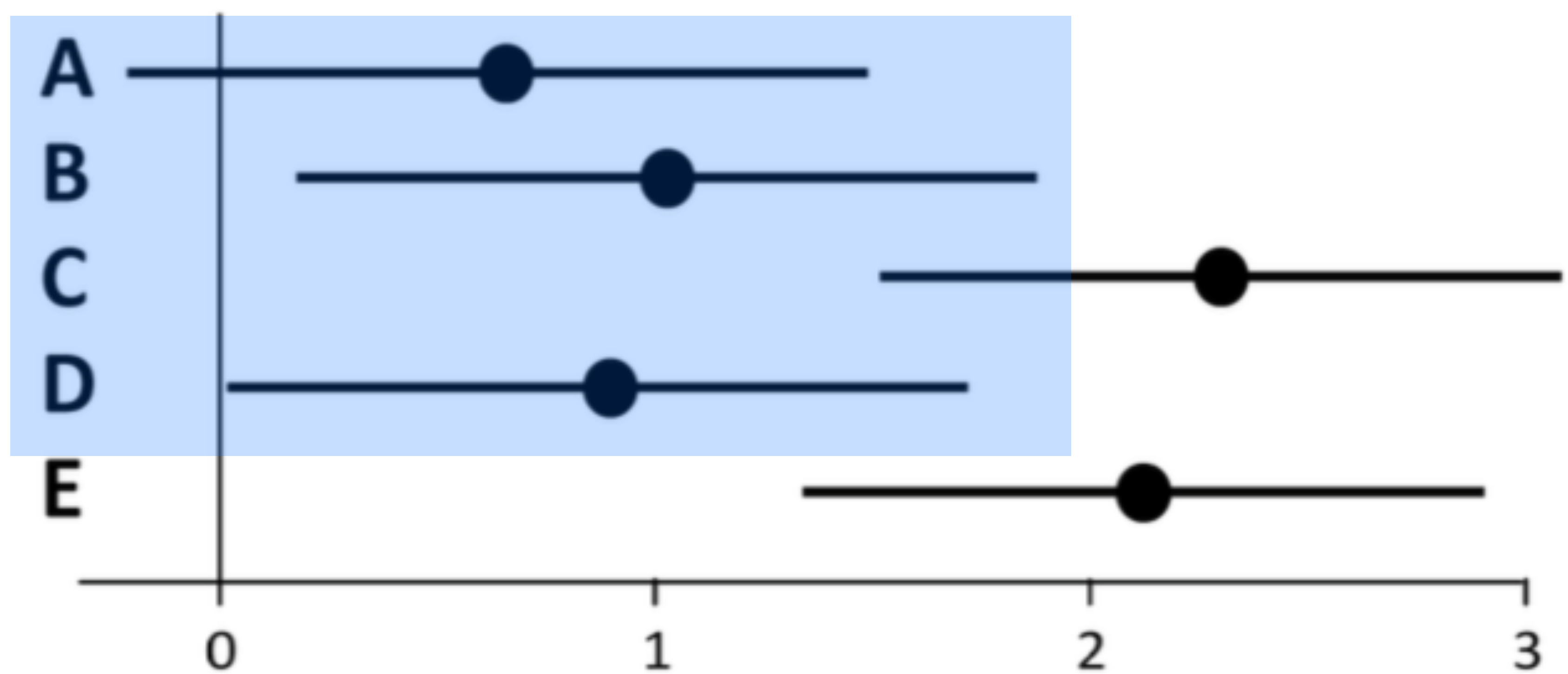
Interpreting multiple CIs



(Dragicevic, 2016)

Examples

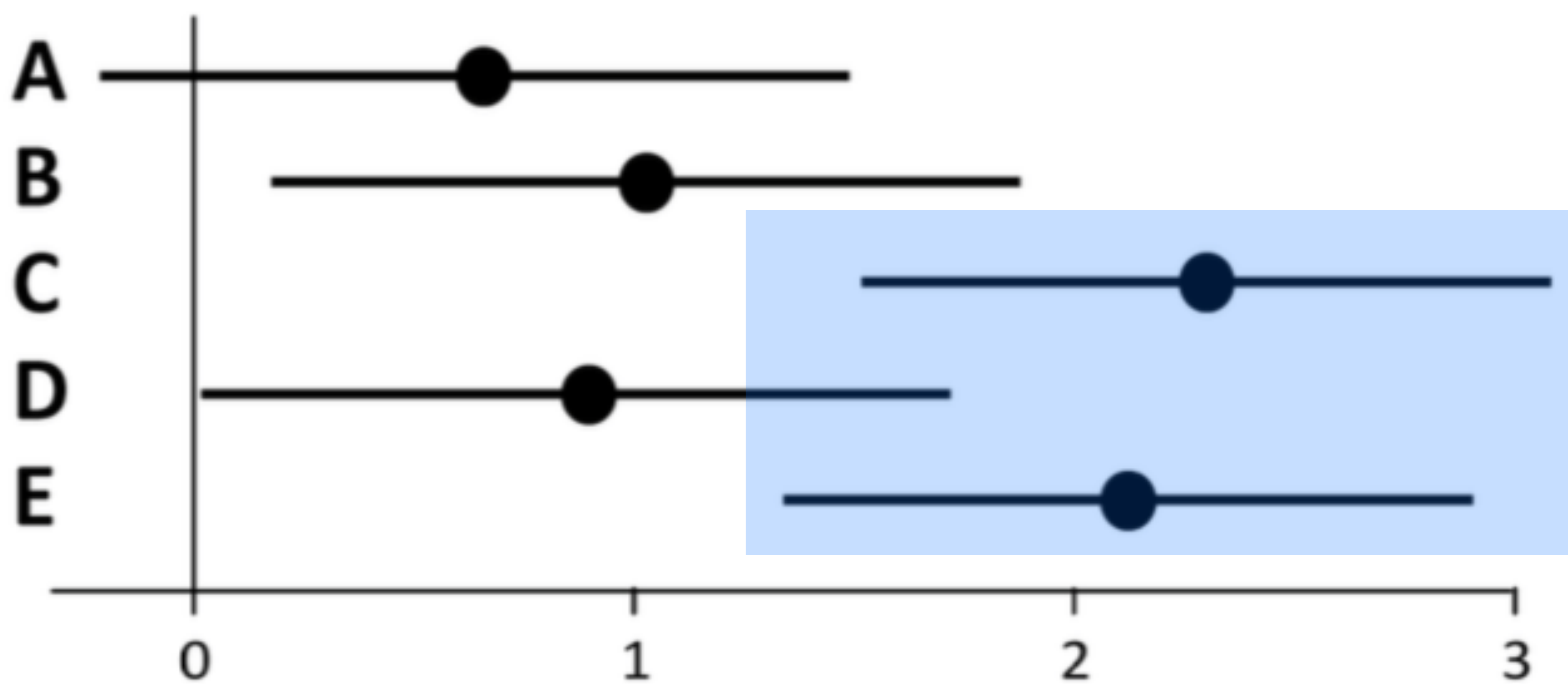
Interpreting multiple CIs



(Dragicevic, 2016)

Examples

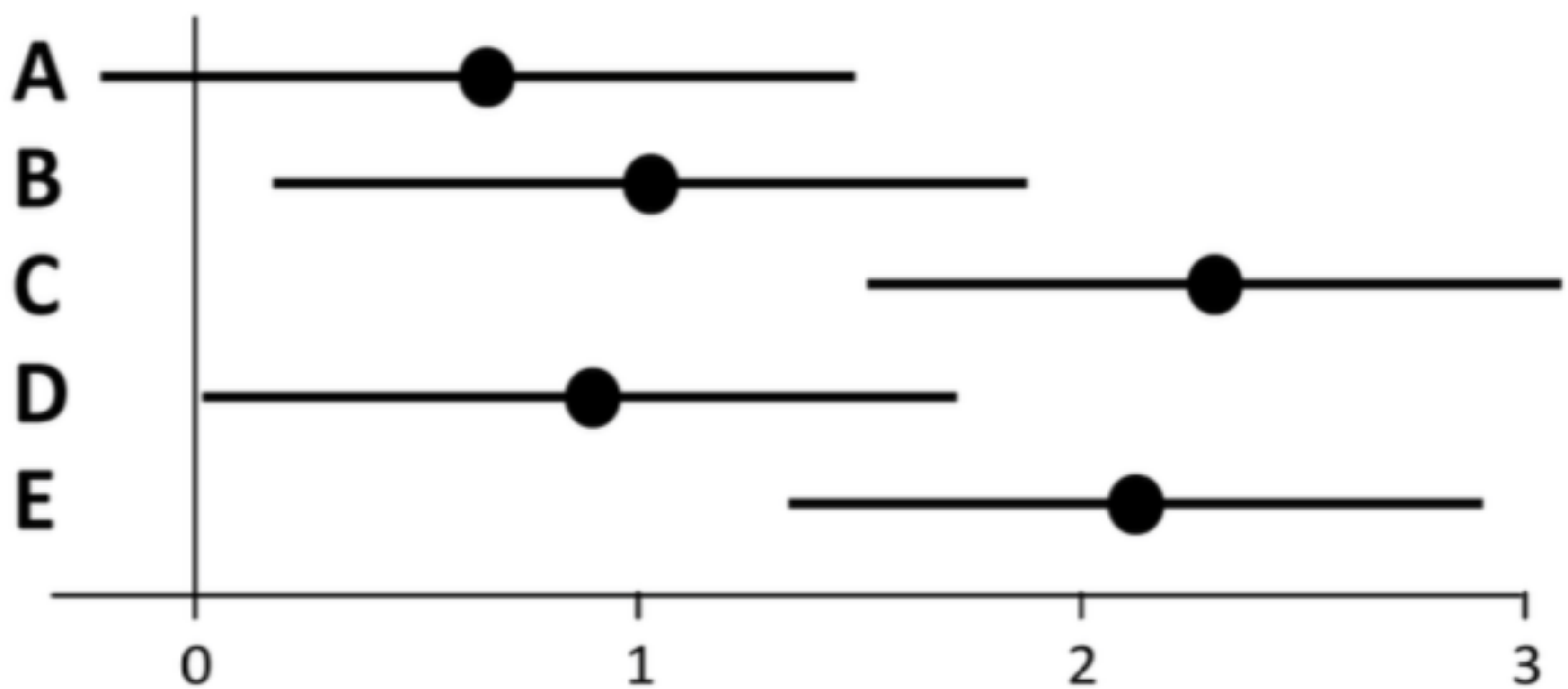
Interpreting multiple CIs



(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

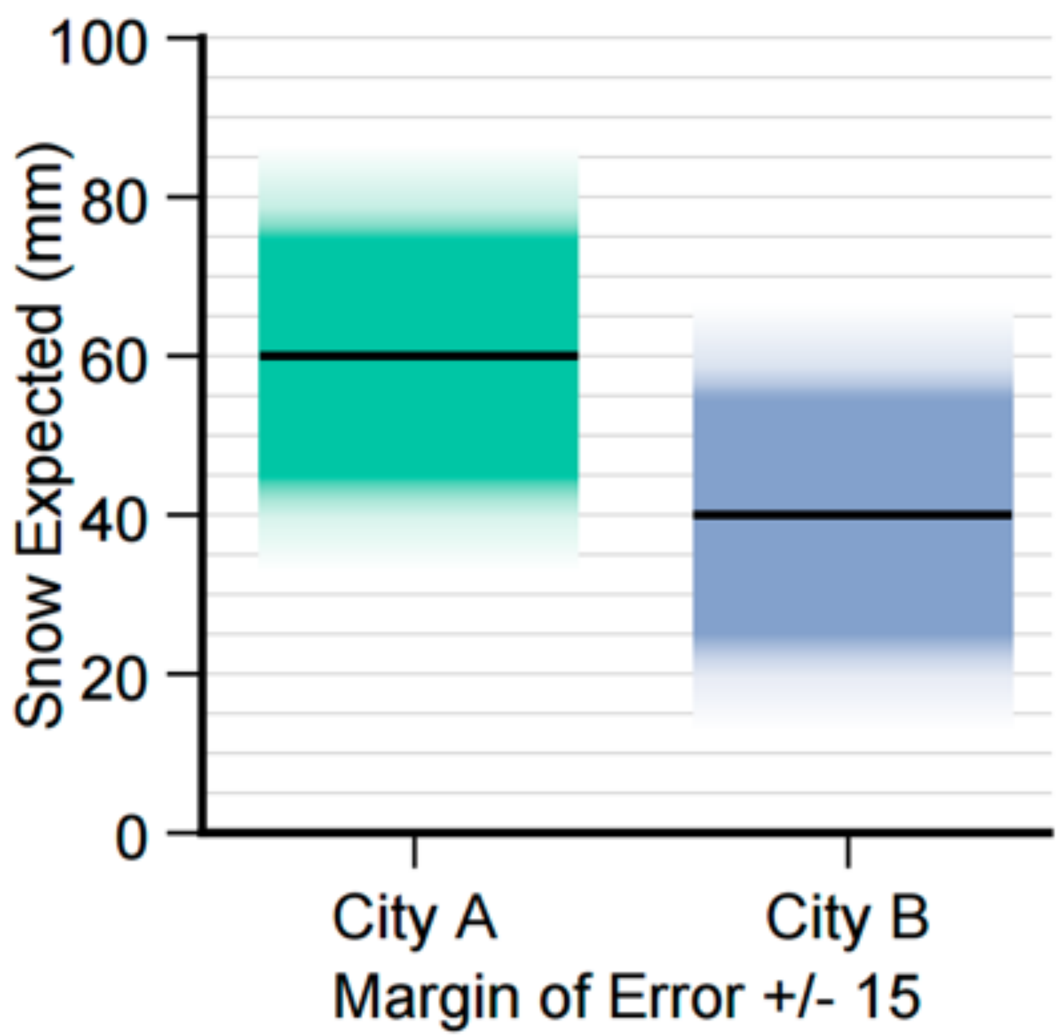
Conclusions

- We need more research on this!

Conclusions

- We need more research on this!

Gradient plots

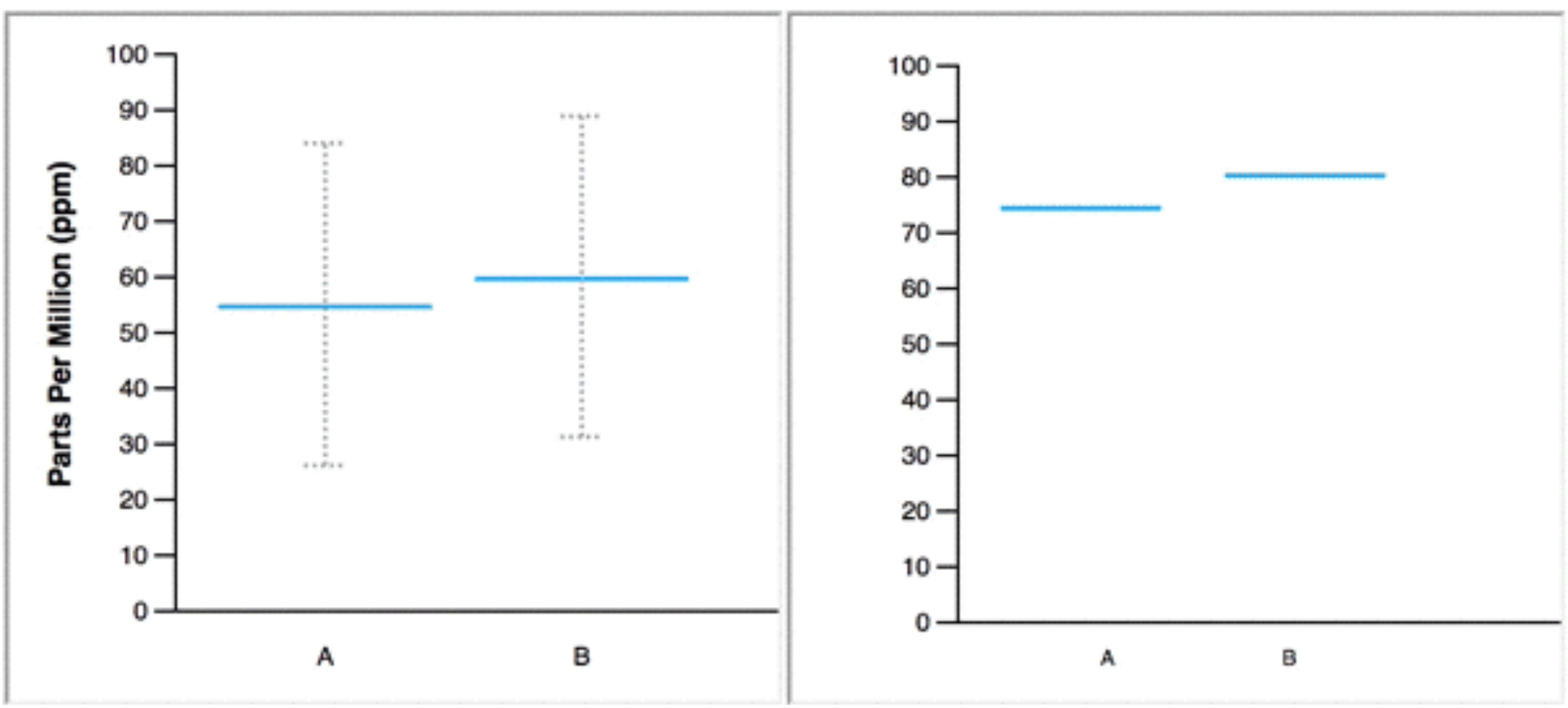


(Correll and Gleicher, 2014)

Conclusions

- We need more research on this!

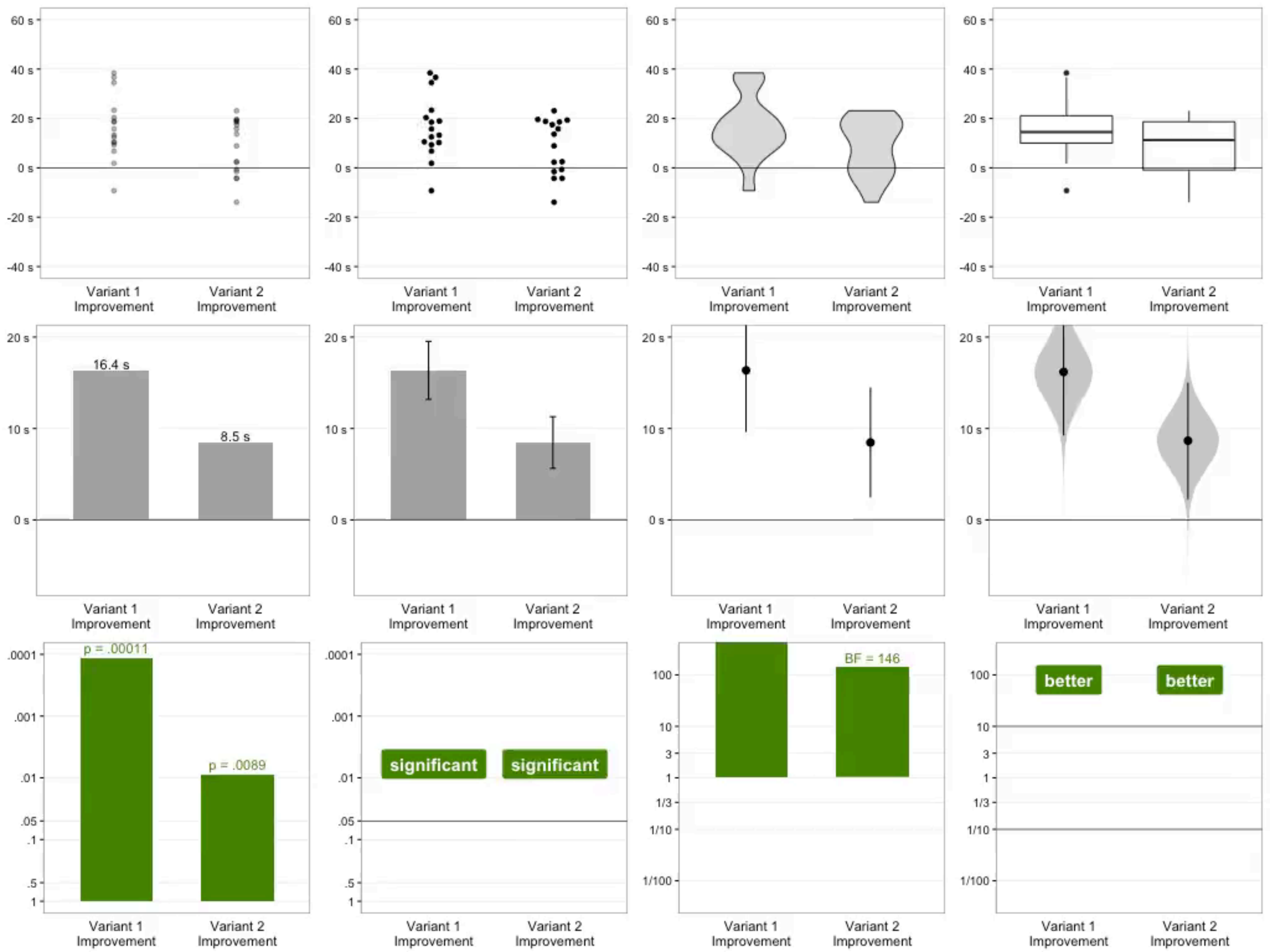
Hypothetical outcome plots



(Hullman, Resnick and Adar, 2015)

Conclusions

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



60 s
40 s
20 s
0 s
-20 s
-40 s
Variant 1 Improvement
Variant 2 Improvement
16.4 s
8.5 s
10 s
p = .00011
p = .0089
.0001
.001
.01
.05
.1
.5
1
significant
significant
BF = 146
100
10
3
1
1/3
1/10
1/100
better
better