



# Harnessing the wisdom of the inner crowd by exploiting the confidence in your decisions

Aleksandra Litvinova, Stefan M. Herzog, Ralph Hertwig

www.mpib-berlin.mpg.de/arc

177

## 1. ABSTRACT

The “wisdom of crowds” effect describes the phenomenon whereby aggregating judgments of multiple individuals can lead to a more accurate judgment than the judgment of the average individual. In the absence of other people’s judgments, could one individual harness the wisdom of her “inner crowd”? We investigated whether individuals can improve the quality of confidence judgments in two-alternative forced choice tasks by either (a) **averaging** confidence judgments or (b) by selecting the judgment with the higher confidence (i.e. **maximizing**). We tested the performance of both strategies by means of a simulation study and three empirical studies. Our results show that (1) individuals can harness the wisdom of their inner crowd by exploiting confidence judgments and that (2) averaging is the safer and more robust strategy.

## 2. THEORETICAL BACKGROUND

When an individual holds two conflicting beliefs about a given issue but needs to make a final decision with an associated final confidence, there are at least two strategies that one can apply based on two confidence judgments:

1. **Maximizing:** Select the decision with the higher confidence (Koriat, 2012), and report the maximum confidence as the final confidence.
2. **Averaging:** Aggregate both confidence judgments by averaging them and report the averaged confidence as the final confidence (Ariely et al., 2000).

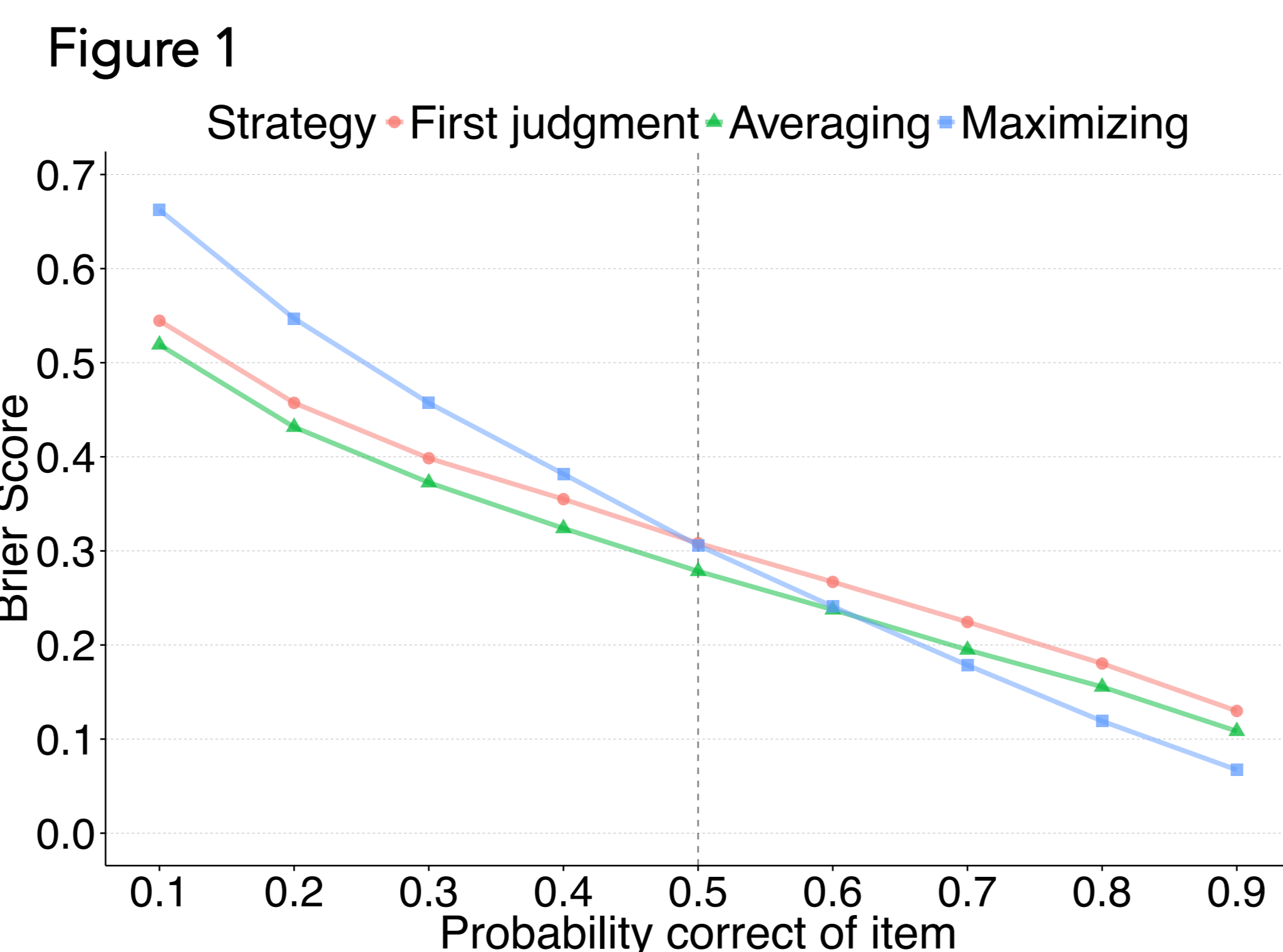
To assess the quality of confidence judgments we calculated the **Brier score**, which is the mean squared deviation of one’s confidence judgment ( $c$ ) from the actual outcome ( $o$ ).

$$BS = \frac{1}{N} \sum_{i=1}^N (c_i - o_i)^2$$

Smaller scores indicate higher quality. In our simulation study we investigated how the wickedness of the environment influences the quality of averaged or maximum confidence judgments.

Environments can be relatively *kind*, where the majority of people agrees on the correct answer (probability correct > 0.5) or relatively *wicked*, where the majority of people agrees on the wrong answer (probability correct < 0.5).

## 3. SIMULATION RESULTS



Based on uncorrelated confidence judgments our simulation study shows that:

1. **Averaging is a safe and robust strategy, because it consistently improves the quality of confidence judgments irrespective of the environment**
2. **Maximizing is a risky strategy, because it can have strong effects, yet they depend on the wickedness of the environment.**

## 4. EMPIRICAL STUDIES

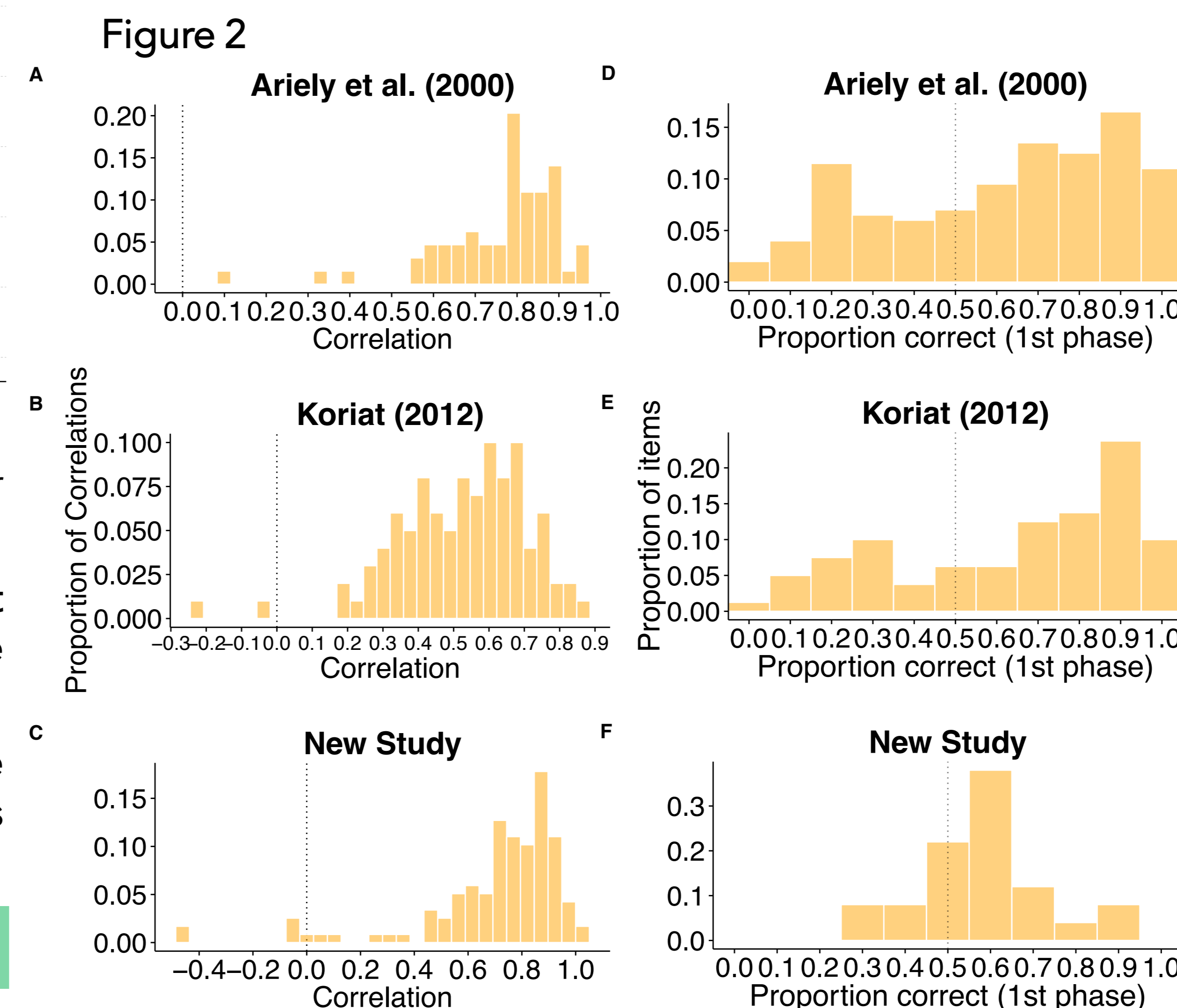
How correlated are an individual’s actual confidence judgments? And can averaging or maximizing boost the quality of confidence judgments in empirical studies, where individuals have no metacognitive insight about the wickedness of the environment? We investigated both questions in two previously published datasets and one new experiment.

**Predictions:** Based on the simulation results, we predicted that averaging will consistently improve the quality of confidence judgments and therefore is the more robust strategy when individuals have no metacognitive insight about the wickedness of the environment.

## 5. METHODS

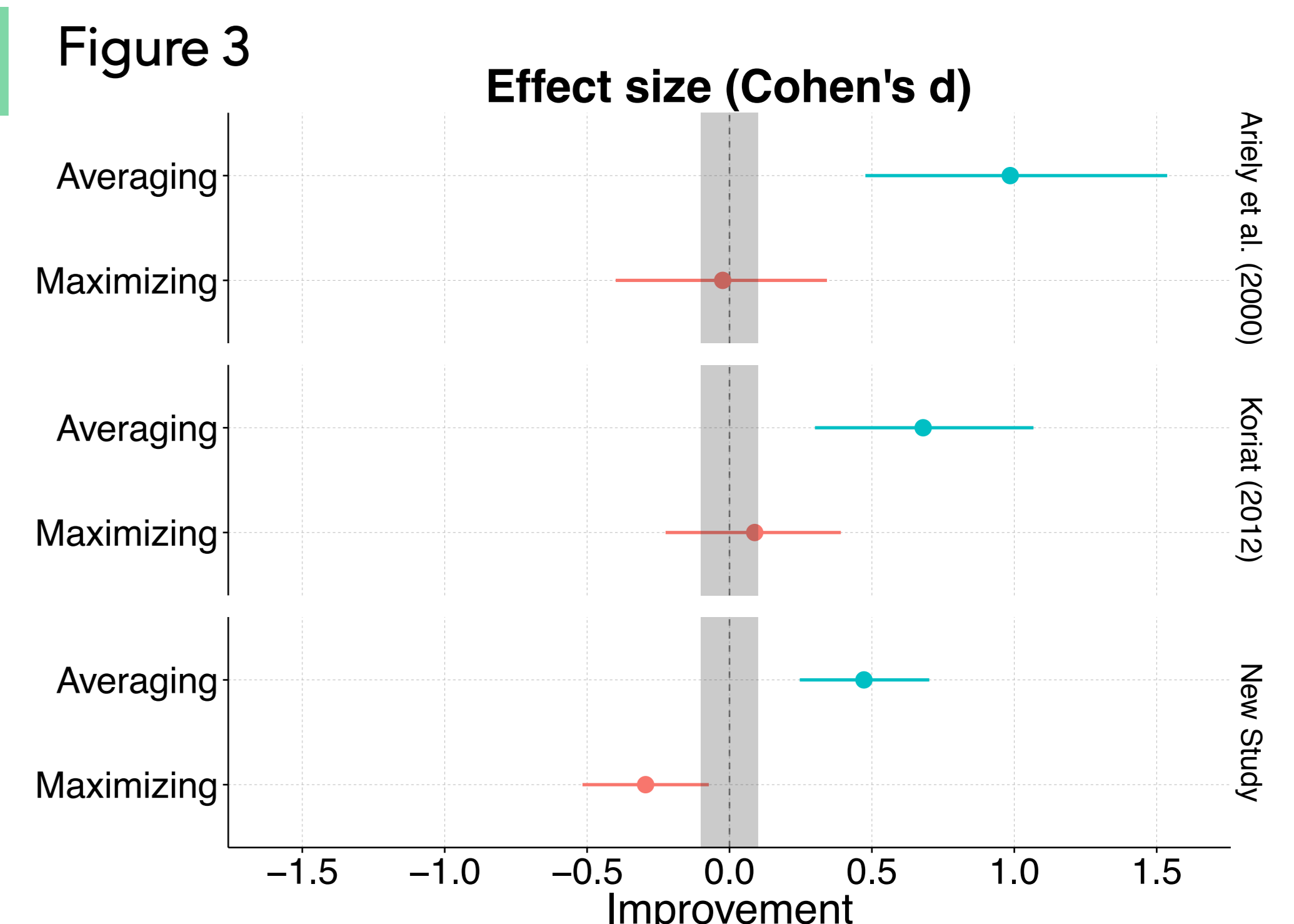
Ariely et al. (2000)	Koriat (2012)	New Study
<ul style="list-style-type: none"> <li>• 64 participants answered 200 questions about relative city populations twice within one session</li> <li>• Repeated questions were maximally separated in time</li> </ul>	<ul style="list-style-type: none"> <li>• 50 participants compared the length or surface area of 80 pairs of figures</li> <li>• With a one week interval</li> </ul>	<ul style="list-style-type: none"> <li>• 309 participants answered 25 general knowledge questions twice within one session</li> <li>• Repeated questions were maximally separated in time</li> </ul>
“In 1992 the population of Cleveland exceeded that of Albuquerque”	“Which object has the larger surface area?”	“Who was born first?”
True or false?	A or B?	Buddha or Aristotle?

## 6. RESULTS



**Figure 2 (A – C):** Individual’s first and second confidence judgments are not perfectly correlated, thereby providing room for both strategies to exert an effect on the quality of confidence judgments. **Figure 2 (D – F):** Most items in the empirical studies are kind, that is, for most items the majority of people agreed on the correct answer. However, there is a considerable proportion of wicked items, which poses a risk for the maximizing strategy to have an overall negative effect on the quality of confidence judgments.

**Figure 3** shows the effect sizes of differences in Brier scores when averaging or maximizing confidence judgments relative to first judgments. Values above 0 indicate an improvement, values below 0 indicate harmed scores.



In all three studies, averaging boosts the overall quality of confidence judgments (relative to first judgments), whereas maximizing sometimes tends to improve and sometimes harms the quality of confidence judgments.

## 7. CONCLUSION

The presented results have two implications:

1. **People can harness the wisdom of their inner crowd by exploiting confidence judgments.**
2. **When in doubt about the environment, it is safer to consistently average than to consistently maximize confidence judgments.**

### References

Ariely, D., Tung Au, W., Bender, R. H., Budescu, D. V.,... & Zauberman, G. (2000). The effects of averaging subjective probability estimates between and within judges. *Journal of Experimental Psychology: Applied*, 6(2), 130-147.  
Koriat, A. (2012). When are two heads better than one and why? *Science*, 336(6079), 360-362.