

How Helpful is a Coin Toss? Evaluations of Predictions at Chance Accuracy

Jay Naborn & Hannah Perfecto

Washington University in St. Louis

Contact: j.naborn@wustl.edu

Abstract

- Predictions at chance accuracy are uninformative, and they become informative as they depart from chance.
- We find people instead treat accuracy ratings as monotonically increasing in value from 0%.
- Across several binary prediction contexts and even when incentivized for accuracy, participants regarded predictions of forecasters performing at chance, vs. below-chance, accuracy as more helpful for decisions.
- We find evidence that overreliance on intuition drives this error and that further deliberation can correct it.

Background

- Decision makers look to forecasts to reduce uncertainty.
- To use forecasts effectively, they must assess which are worth using and which are not.
- Although considerable prior research has examined how people evaluate forecasts and forecasters (e.g., Bagchi & Ince, 2015; Dietvorst & Bharti, 2020), accuracy rates have so far been ignored.
- We compare evaluations of predictions at chance accuracy (50%, for binary outcomes) to those at below-chance accuracy. The former are wholly uninformative, whereas the latter may be informative, in that decision makers can expect the opposite of what they state.

Main Design

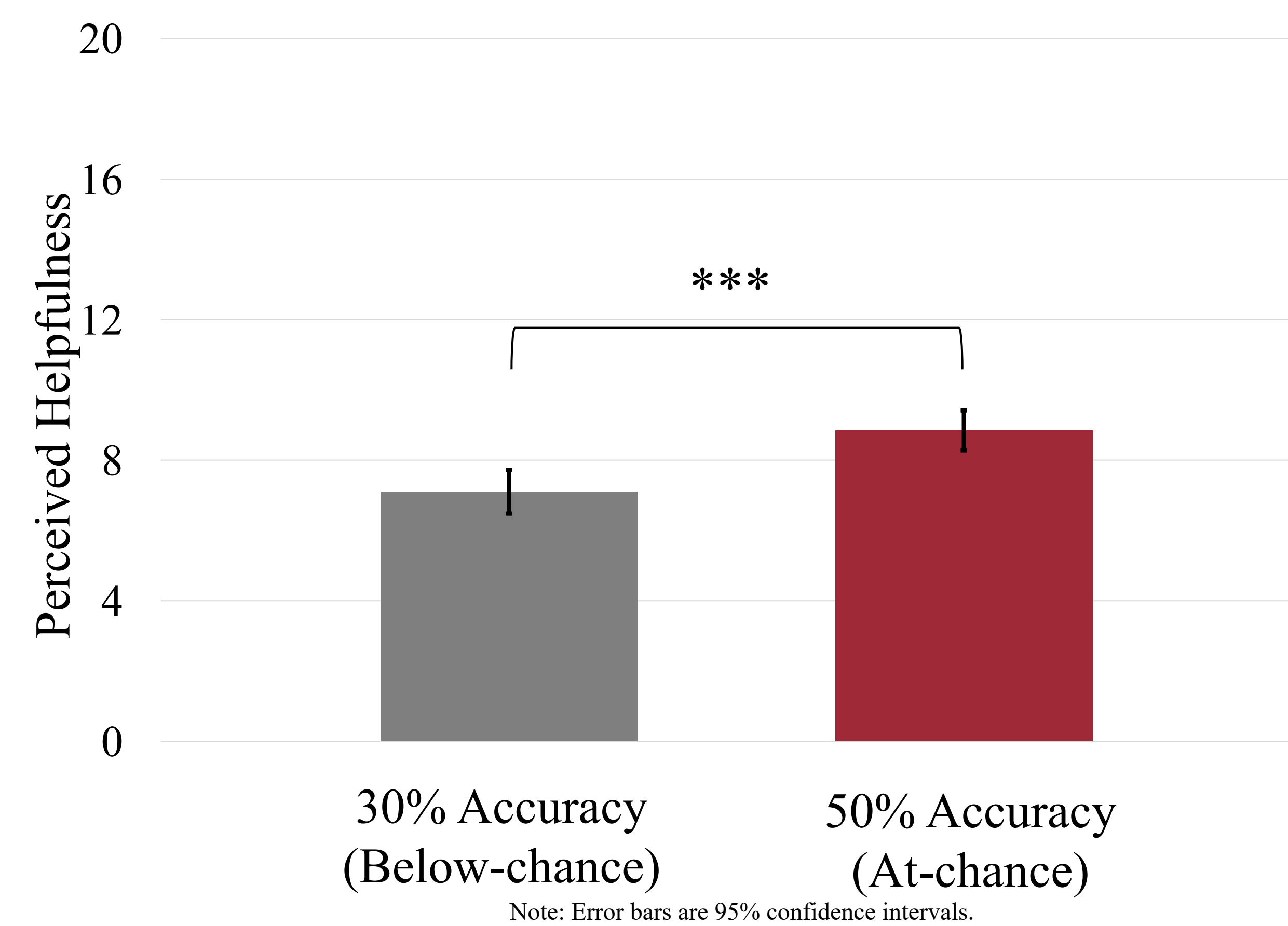
2-cell, between-subjects design (Accuracy: 30%, 50%)

N = 289 (MTurk) https://aspredicted.org/MP7_8WH

- **Scenario:** Travel website recommendations to either buy a ticket or wait, from algorithm predictions of upcoming price increases.
 - IV: Algorithm accuracy (30%, 50%)
 - DV: Helpfulness of knowing the website's prediction
 - 20-point slider scale (1 = "It would definitely not help save me money", 20 = "It would definitely help save me money.")
- **Result:** Predictions from the chance (50%) accuracy algorithm were seen as more helpful (M = 8.85, SD = 3.49) than those from the below-chance (30%) accuracy algorithm (M = 7.10, SD = 3.80), $t(287) = 4.10, p < .001$.

Main Result

- **Summary:** Predictions at chance accuracy, which are uninformative, were seen as more helpful than those at below-chance accuracy, which can indicate what to bet against.



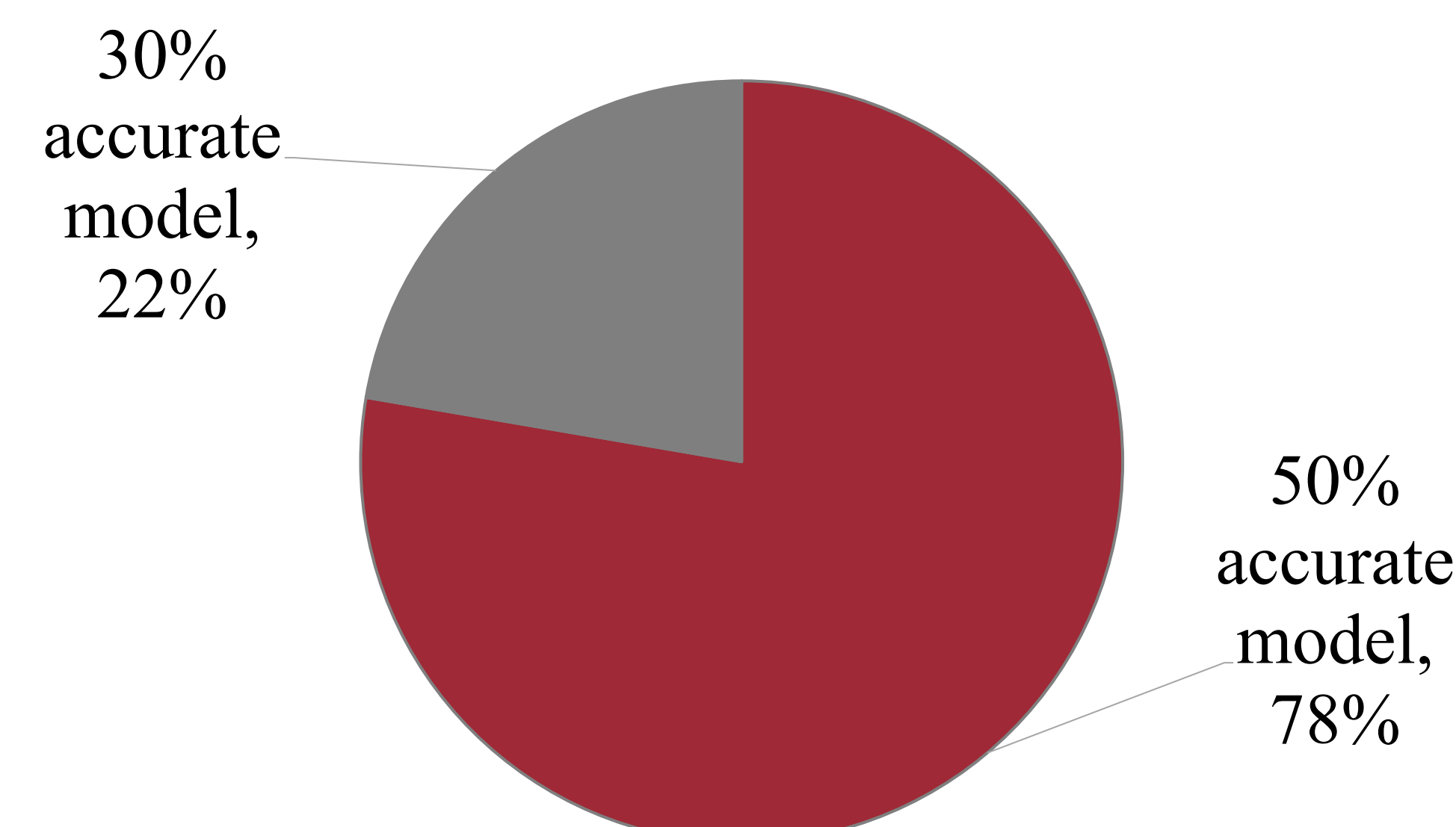
- This pattern also replicated in 3 other contexts (Total N = 743).

Incentivized Choice

Choice paradigm (30% Accuracy & 50% Accuracy)

N = 278 (MTurk) https://aspredicted.org/Q3X_VT1

- **Scenario:** Predicting the winner of a real Major League Baseball game. Correct predictions earned a \$0.50 bonus.
 - Before predicting, participants chose between seeing the prediction of a model at chance (50%) or at below-chance (30%) accuracy.
- **Result:** Most participants opted for the coin-toss model: 77.7% ($n = 216$) vs. 22.3% ($n = 62$), $\chi^2(1) = 85.31, p < .001$.



- **Summary:** Participants strongly preferred the (uninformative) chance accuracy model to the below-chance model.

Debiasing by Deliberation

Choice paradigm + Deliberation

N = 749 (MTurk)

https://aspredicted.org/C16_6R9

- **Hypothesis:** Overreliance on intuition causes the preference for chance over below-chance accuracy predictions.
- **Scenario:** Participants predicted the winner of an MLB game. Before predicting, they chose to see the prediction of a model of 30% or 50% accuracy.
 - After choosing, participants were asked to pause and think deeply about the task.
 - Finally, they again chose which prediction they wanted to see (30% accurate model or 50% accurate model).
- **Result:** These more deliberative, second choices were much less likely to be biased than were first choices, as indicated by McNemar's test, $X^2(1, N = 749) = 67.213, p < .001$.

Model Choices

Post-Deliberation Choice

Initial Choice	N = 749	50% (chance)	30% (below-chance)
	50% (chance)	572	73
30% (below-chance)	2	102	

- **Summary:** Consistent with our overreliance on intuition account, 97.3% of switches from first choices to second choices were switches from the below-chance model to the chance model.

Discussion

- We observe an intuitive preference for higher (versus lower) accuracy, even when the higher values correspond to objectively uninformative predictions (those equivalent to coin tosses).
- This bias raises concerns about people's evaluations of predictions, which are often used as decision inputs.
- We find that further deliberation can debias responses, providing evidence that (at least some) people can correct this bias and that overreliance on intuition drives the effect.

References

- Bagchi, R., & Ince, E. C. (2015). Is a 70% Forecast More Accurate than a 30% Forecast? How Level of a Forecast affects Inferences about Forecasts and Forecasters. *Journal of Marketing Research*, 53(1), 31–45. <https://doi.org/10.1509/jmr.12.0526>
- Dietvorst, B. J., & Bharti, S. (2020). People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error. *Psychological Science*, 31(10), 1302–1314. <https://doi.org/10.1177/0956797620948841>