**Title**
The Wisdom of the Crowd and Framing Effects in Spatial Knowledge



**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

**Authors**
Montgomery, Lauren E
Lee, Michael

**Publication Date**
2022

Peer reviewed

# The Wisdom of the Crowd and Framing Effects in Spatial Knowledge

**Lauren E. Montgomery (lmontgo1@uci.edu)**
Department of Cognitive Sciences, 2201 Social
Behavioral Sciences Gateway
Irvine, CA 92697 USA

**Michael D. Lee (mdlee@uci.edu)**
Department of Cognitive Sciences, 2201 Social
Behavioral Sciences Gateway
Irvine, CA 92697 USA

## Abstract

We study the wisdom of the crowd in the context of spatial knowledge, asking participants to identify US states and African countries on unlabeled tile maps. We use two question frames, asking participants to select where the target is present or eliminate where it is absent. Participants generally display overconfidence, often selecting small regions that do not include the target. We find strong wisdom of the crowd effects by aggregating participants' responses, especially by weighting the individual responses according to the size of their selection. The weighted crowd outperforms all but a few participants for the US states and all participants for the African countries. We also find wisdom of the crowd within effects, by aggregating the present and absent frames for the same participant. We discuss the implications of our findings for understanding how people express uncertain spatial knowledge and the potential use of crowd aggregation in real-world applications.

**Keywords:** wisdom of crowds; spatial knowledge; wisdom of the crowds within; framing effects

## Introduction

The wisdom of the crowd is the finding that a crowd's aggregate judgment is more accurate than the judgment of a randomly sampled individual in the crowd (Galton, 1907; Davis-Stober, Budescu, Dana, & Broomell, 2014; Surowiecki, 2004). Crowd superiority has been demonstrated in a range of contexts. The most common context is general knowledge, which examines the accuracy of answers to factual questions about geography, society, culture, entertainment, and other topics (Bennett, Benjamin, Mistry, & Steyvers, 2018; Lee, Steyvers, & Miller, 2014; Prelec, Seung, & McCoy, 2017). Another context involves forecasting and predictions about political, social, sporting, and other events (Armstrong, 2001; Boon, 2012; Da & Huang, 2020; Klugman, 1947; Lee, Danileiko, & Vi, 2018; Miller, Wang, Kulkarni, Poor, & Osherson, 2012; Page & Clemen, 2013). A third context involves group settings in which individuals interact or compete with each other to generate judgments or estimates about stimuli (Atanasov et al., 2017; Christiansen, 2007; Lee, Zhang, & Shi, 2011; Lyon & Pacuit, 2013; Ray, 2006). In all of these contexts, the required judgments can take different forms, including scalar estimates (Jenness, 1932; Farnsworth & Williams, 1936), discrete choice (Lee et al., 2018; Prelec et al., 2017), rank orderings (Bruce, 1935; Gordon, 1924; Knight, 1921; Lee et al., 2014; Miller et al., 2012), or sequential decisions (Thomas, Coon, Westfall, & Lee, 2021; Zhang & Lee, 2010).

In this study, we explore the wisdom of the crowd in the context of spatial knowledge by asking people to identify US states or African countries on unlabeled tile maps. Some previous research on spatial or geographical knowledge has focused on scalar estimates ("what is the height of Mount Everest?"), discrete choices ("is Reno east or west of San Diego?"), or rankings ("order the following US states from west to east") rather than direct spatial judgments. Other previous research has presented spatial targets and then required direct spatial judgments (Juni & Eckstein, 2017), although this type of task involves immediate perceptual rather than longer-term memory-based knowledge. The most relevant previous work studies how accurately people can identify locations on a map (Fu, Lee, & Danescu-Niculescu-Mizil, 2017; Fu, Wang, & Danescu-Niculescu-Mizil, 2020; Mayer & Heck, 2022). Our task involves people's memory for spatial knowledge and requires them to express that knowledge in a direct and detailed way by selecting a spatial region.

An interesting feature of our task is that it allows the same question to be framed in different ways. People are asked to identify a target US state or African country by selecting as many states or countries they need to be confident that the target is included in their set. We call this the present framing. They are also asked to identify a target state or country by indicating a set of states or countries that are *not* the target. We call this the absent framing. Being able to collect both of these judgments raises the issue of framing effects (Levin, Schneider, & Gaeth, 1998; Tversky & Kahneman, 1981) and, in particular, whether the inherent uncertainty in forming regions is managed differently between the frames. Previous research on elimination and inclusion, the same dichotomy that we use, suggests that using these frames will produce some non-complementarity in the generated choice sets (Shafir, 1993; Yaniv & Schul, 1997).

Asking multiple questions also allows us to consider the phenomenon known as the wisdom of the crowd within, in which multiple judgments from the same individual are aggregated. A basic challenge for the wisdom of the crowd within is that using only judgments from one individual results in correlated judgments, which limits the improvement in the aggregate. Accordingly, an effort is made to make the judgments as independent as possible. This has been achieved by increasing the time interval between estimates (Vul & Pashler, 2008) or having participants use various question



In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.* 

framing strategies, such as consider-the-opposite (Herzog & Hertwig, 2009; Lord, Lepper, & Preston, 1984), starting from scratch (Herzog & Hertwig, 2014), or having the individual combine their previous estimates in some way (Herzog & Hertwig, 2009; Larrick & Soll, 2006). These question framing strategies work because the participant has to consider additional information or approach the question differently. In our spatial knowledge context, being able to ask about the location of targets in terms of presence and absence provides two natural contexts for asking the same individual about the same information.

The remainder of this paper is organized as follows. We first describe the experimental design and the framing effects on participants' judgments and how participants manage the uncertainty inherent in the task. To test for the wisdom of the crowd, we develop two approaches for aggregating crowd judgments and compare their performance to individual judgments. To test for the wisdom of the crowd within effects, we examine improvements in individual judgments resulting from aggregating their two judgments. Finally, we examine how the crowd aggregate improves as a function of the number of individuals in the crowd. We close with a discussion of our main results and directions for future research.

# Experiment

## Participants

50 participants were recruited using Prolific (Prolific, 2022) for each of the US states and African countries conditions in a between-participants design. All participants were current US residents and provided basic demographic information including their age, whether they attended high school in the US, and their familiarity with each of the US states or African countries.

## Stimuli

Figure 1 shows the tile maps presented to participants in each trial. These are standard configurations used in data journalism.[1] The US states map was restricted to the 48 continental US states, and the African countries map was restricted to 51 of the 54 countries by excluding Comoros, Mauritius, and the Seychelles.

The tile maps make responding to the task simple and responses easy to visualize. They also introduce some irreducible uncertainty because even participants with perfect geographical knowledge will still be uncertain about the exact translation between the true geography and the tile layout. For example, South Africa could reasonably be any of the three tiles at the bottom of the African map. Thus, when responding to the questions, participants need to consider both the uncertainty in their spatial knowledge and the uncertainty that the tile layout introduces.

[1] See, for example, https://blog.apps.npr.org/2015/05/11/hex-tile-maps.html and https://public.tableau.com/app/profile/neil.richards/viz/Malaria_14/Dashboard1.

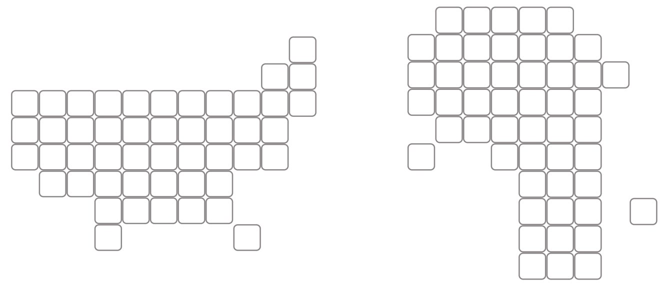

Figure 1: Tile map stimuli for the US states (left) and African countries (right) conditions.

## Method

Every participant was given every state or country as a target on a trial in both the present and absent framings. The two framings were blocked so that all of the targets were presented in one frame before changing to the other. The order of the framings was randomized, as was the order of the targets.

In the present framing, participants were asked "Where is X located? Select as few states/countries as possible, but be sure X IS in the states/countries you select." In the absent framing, participants were asked "Where is X NOT located? Select as many states/countries as possible, but be sure X IS NOT in the states/countries you select." Each question was answered sequentially with participants being asked not to look up any information but rely instead on their general knowledge and memory. Participants were not allowed to return to or view previous responses, and they did not receive any feedback. At the completion of all of the target questions in both frames, participants were asked for their demographic information.

# Framing Effects and Managing Uncertainty

To analyze framing effects, we looked at how complementary the participants' responses were. Complementary means that a participant's response contained the same information in both frames. Figure 2 shows participant-level responses for both the present and absent frames for four illustrative cases. In each panel, the tile for the target state or country is outlined in black. The participant's selections made only in the present frame are in blue, and their selections made only in the absent frame are in yellow. Tiles for states or countries selected in both frames are a blended blue-yellow color, and tiles selected in neither frame are white. This means that the extent of blue versus yellow regions indicates the confidence of the knowledge expressed by the participant. For example, the participant in panel A is very confident in locating California, the participant in panel C is less confident (and wrong) in locating the Democratic Republic of the Congo, and the participant in panel D has low confidence in locating Uganda.

The presence of blue-yellow and white tiles indicates that the participant's responses across the two frames are not perfectly complementary. In panel A, the participant made logically complementary selections for California, while in panel

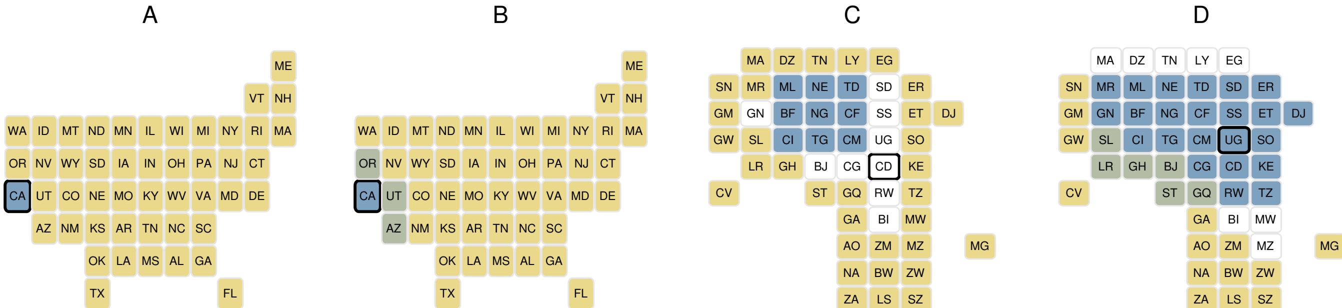


Figure 2: Four illustrative individual participant responses to particular target states and countries in both frames. States and countries selected in only the present frame are colored yellow, selected in only the absent frame are colored blue, selected in both frames are colored blue-yellow, and selected in neither frame are colored white.

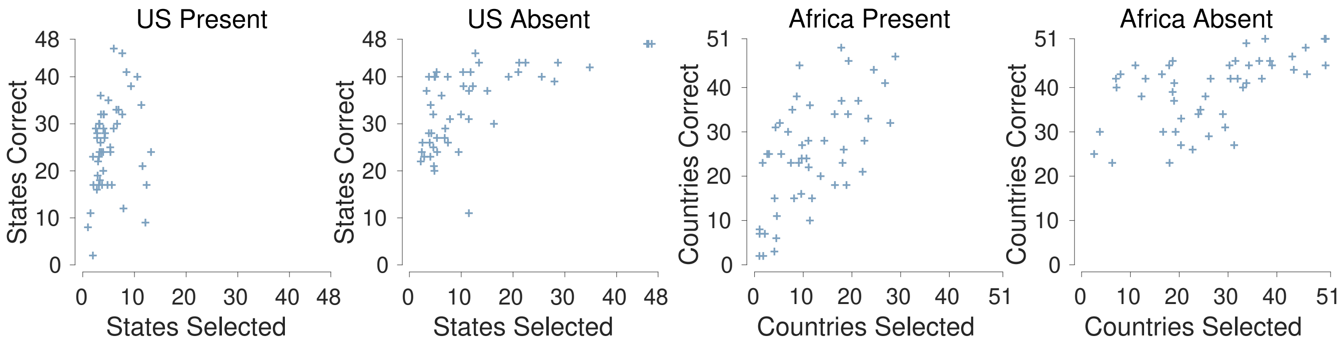


Figure 3: Individual participant performance in both conditions and both frames. Each blue cross corresponds to a participant, showing the average number of selections they made and the numbers of states or countries correctly included in their selections.

B the participant selected some states neighboring California in both the present and absent frames. This suggests the participant in panel B was less confident in the present than the absent frame. In contrast, the participant in panel C is more confident in the present frame and less confident in the absent frame. The participant in panel D is hard to characterize, since their present and absent frame responses are quite inconsistent, with some countries selected in both framings and others selected in neither. Participants rarely provided strictly complementary responses. On average, participants provided 5.7 complementary responses in the US states condition and 2.0 per country in the African countries condition. They had some overlap in 19.3 and 24.3 states and countries, respectively. They selected some tiles in neither frame in 38.4 and 45.9 states and countries, respectively.

Consistent with the task instructions, we measure a response as accurate if the target is included in the participant's selections in the present frame or not included in their selection in the absent frame, regardless of the size of the regions they selected. Figure 3 shows the relationship between the number of states or countries selected and this measure of participant accuracy. The four panels correspond to the US states and African countries conditions and the present and absent frames. To allow direct comparisons between the two frames, participant responses in the absent framing have been inverted so that they indicate the states or countries the participant selected as including the target. This means that less confident behavior now consistently corresponds to higher numbers of selections and more confident behavior corresponds to lower numbers of selections.

The striking feature of Figure 3 is that very few participants achieve high levels of accuracy. This likely reflects both a lack of perfect knowledge and a failure to compensate by selecting enough states or countries. In the present frame, participants selected an average of 5.2 states and 11.8 countries, correctly including an average of 25.5 states and 25.0 countries. More selections are made in the absent frame, especially for US states. The average numbers selected are 12.4 states and 26.8 countries. These expanded selections lead to greater average accuracies of 33.8 states and 39.6 countries.

There is no reason, however, participants cannot achieve perfect accuracy in both frames. In fact, this is what the task instructions require. A participant who has little relevant geographical knowledge should select many of the states or countries in the present frame and few in the absent frame. No participants were completely accurate in the US states condition. The four participants who achieved complete accuracy in the African countries condition did so in the absent frame by eliminating very few countries. The fact that most participants achieve modest accuracy suggests that they are

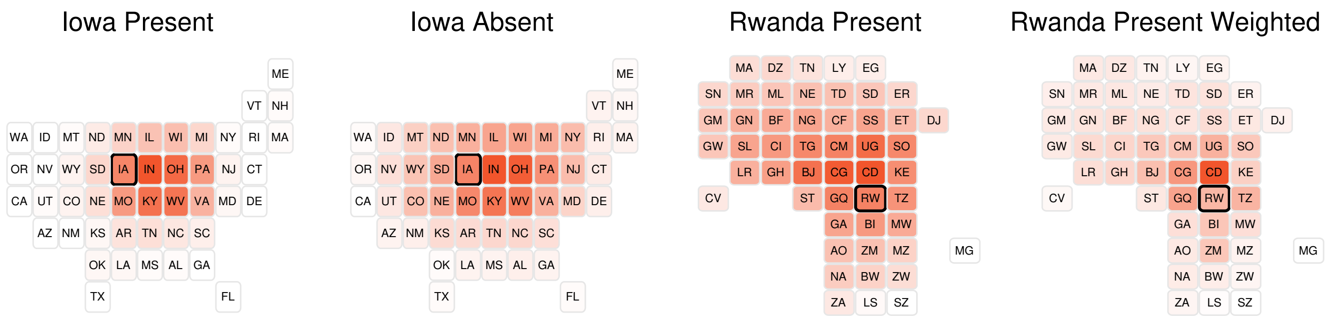


Figure 4: Examples of aggregate crowd responses. The two left-most panels show the unweighted proportion of participants who selected each state while targeting Iowa in the present and absent frames, with darker red colors indicating greater proportions. The two right-most panels show the unweighted and confidence-weighted proportion of participants who selected each country while targeting Rwanda in the present frame.

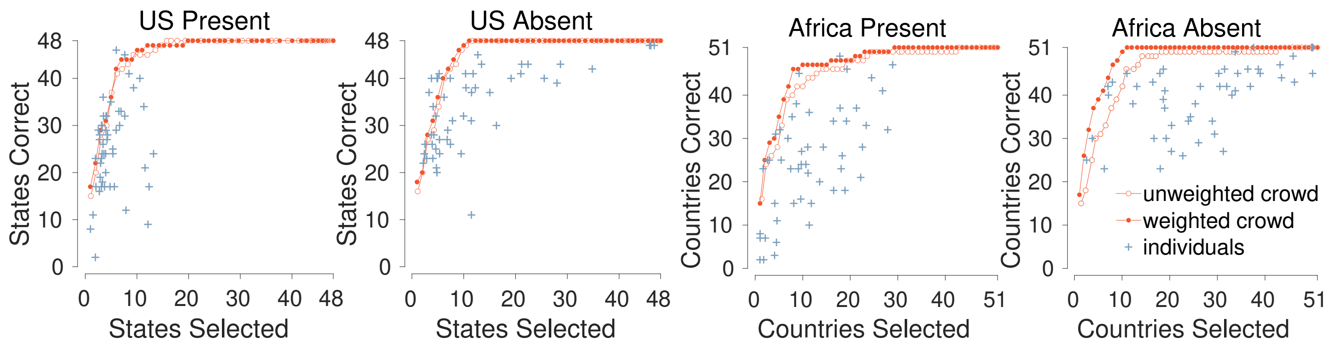


Figure 5: Individual participant and crowd performance in both conditions and both frames. Blue crosses correspond to participant performance. Red curves correspond to unweighted and weighted crowd performance, showing the average number of selections made and the numbers of states or countries correctly included.

overconfident in their selections. The explanation cannot be as simple as wanting to avoid effort, since the way to achieve high accuracy in the absent frame is the least effortful. Most participants provide effortful responses in the absent frame that still exhibit overconfidence.

## The Wisdom of the Crowd

The simplest way to form an aggregate crowd judgment is to count the proportion of times each state or country is selected by a participant. A more complicated method weighs the individual selections according to their confidence. A natural measure of confidence is the number of states or countries selected: that is, the number selected in the present frame and the number not selected in the absent frame. For example, if a participant selects 10 states, each of their selections will have 1/10th the value of a participant who just selected one state. Weighting individual judgments in this way implements the idea that more confident participants should have more influence on the crowd judgment (Lyon & Pacuit, 2013).

Figure 4 demonstrates these two approaches to crowd aggregation using heat map visualizations. The states and countries are shaded according to the aggregated group proportions. The left-most panels show the present and absent frames for the target state Iowa. It is clear that the crowd selection is more concentrated (less disperse) in the present frame, consistent with individuals making relatively fewer selections. The right-most panels show the unweighted and confidence-weighted crowd judgments for the target country Rwanda. The confidence-weighted aggregate is much more concentrated than the unweighted aggregate. This is a natural consequence of giving less weight to each selection made by participants who made many selections overall.

Crowd judgments are inherently graded and give a probability that each state or country is the target, unlike individual judgments in which every state or country is either selected or not selected. Accordingly, there is no natural single measure of crowd accuracy. Instead, there is a set of measures, depending on where the graded responses are thresholded. A simple way to set these thresholds is by ranking the probabilities and setting a threshold $k$ so that only states or countries in the top-$k$ are considered to be selected by the crowd. For example, if $k = 1$, the crowd response is the modal (most likely) state or country. In all four of the illustrative examples in Figure 4, this response would be incorrect. As the threshold is increased, to allow the top-two or top-three or more possibilities, the crowd will become more accurate at the expense of

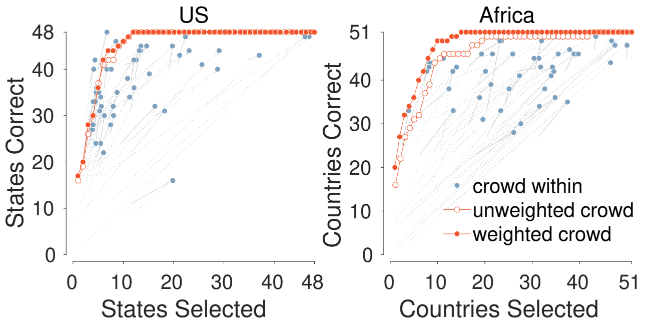


Figure 6: Wisdom of the crowd within performance for both conditions. Blue markers show aggregate performance across both frames for individual participants with lines connecting to their performance in each frame. The red curves show the the unweighted and confidence-weighted crowd performance.

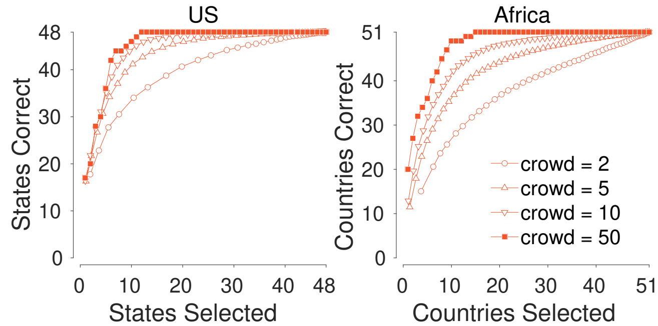


Figure 7: Performance of the crowd based on different numbers of individuals for both conditions. Each curve corresponds to the performance of the confidence-weighted crowd, including responses for both frames for crowds ranging from 2 to 50 individuals.

making more selections.

Figure 5 superimposes crowd performance on the individual performance shown in Figure 3. The red curves correspond to crowd performance, starting with the modal response and ranging to increased numbers of selections and accuracy (the non-integer values for selections are the result of ties in probabilities). These curves are shown for both the unweighted and confidence-weighted crowds. Better performance corresponds to small numbers of selections with high accuracy. The unweighted and confidence-weighted curves are very similar in the US states condition but the weighted crowd clearly performs better in the African countries condition, especially for the absent frame.

Comparing crowd and individual performance depends on how the goals of the task are interpreted. A strict literal interpretation of the task is that perfect accuracy is required using as few selections as possible. By this measure, the crowd outperforms every individual because it is capable of perfect accuracy. Almost every participant in both frames fails to achieve this. The unweighted crowd reaches perfect accuracy with 15.8, 11.7, 42.7, and 43.2 selections for US present, US absent, Africa present, and Africa absent cases, respectively, while the confidence-weighted crowd needs 19.9, 11.1, 29.3. and 11.1 selections. It is clear that the weighted crowd outperforms the unweighted crowds in the relatively low-knowledge African countries condition.

A less strict assessment of individual and crowd performance allows for less than perfect accuracy while still requiring relatively few selections. Visually, this corresponds to being at the top-left of the graphs shown in Figure 5. In the present frame of the US states condition, there are two participants whose performance is above and to the left of the crowd curve, and another three or four who are close. A similar result holds for the absent frame. In the African countries condition, there is one participant who meets this criterion in the present frame and none in the absent frame. A reasonable conclusion is that the crowd aggregate is superior to at least 90% of participants in the US states condition and essentially all participants in the African countries condition. For the vast majority of participants in all conditions and frames, the crowd's performance is both ordinally better and quantitatively much better.

## The Wisdom of the Crowd Within

To examine the wisdom of the crowd within, we combined the selections made in the present and absent frames by the same participant for the same target. We also created crowd aggregate responses by combining the selections made by all of the participants in both frames. Figure 6 shows the results of these analyses. The blue dots correspond to individual participants, showing the average of the number of states or countries they selected over both framings, and the accuracy of their crowd-within aggregate. Accuracy is measured in terms of whether the correct state or country was selected in either the present or absent framing. The blue lines connect the aggregate individual performance to performance for just the present and absent frames separately (i.e., to the performance measures shown in Figures 3 and 5). These wisdom of the crowd within aggregates allow us to evaluate how perceptually similar participants treat the two structurally identical tasks as complementary responses would be exactly overlaid in Figure 6.

By its construction, the crowd-within aggregate always involves as many or more states or countries being selected as in the separate frames. Our interest is whether this increase significantly improves accuracy. Visually, this corresponds to crowd-within performance that shifts significantly upward without shifting far to the right. Figure 6 makes clear that, for most of the participants in both conditions, the crowd-within aggregate leads to an increase in accuracy. The mean increase in accuracy is 11.5 states and 17.9 countries. Much of this improvement comes from the absent frame selections broadening the selections to include the target as shown by the crowd-within aggregates moving diagonally toward the upper right in Figure 6. There are also cases in which two relatively narrow selections in the frame are combined to form an improved selection and where the crowd-within aggregates mainly shift upward with little movement to the right. For example, for the best performing individual in the US states condition, the crowd-within aggregate has perfect accuracy based on an average of 6.8 states being selected. This individual's crowd-within aggregate combined their present frame accuracy of 46 states, based on 6.0 selections, with their absent frame accuracy of 40 states, based on 5.0 selections. The crowd aggregation over both frames shown by the red curves continues to be well performed.

## Crowd Size

Given the clear wisdom of the crowd effect, an interesting follow-up question is how many individuals are needed for effective crowd performance. Figure 7 shows the confidence-weighted crowd-within responses averaged over many subsets of 2, 5, or 10 randomly selected participants and the full crowd of 50 participants. The full crowd is the one considered in Figure 6, which uses all of the participant and frame information about each target. Both conditions show the same expected pattern of improved performance as the crowd size increases. There is an especially large improvement as the crowd increases from the smallest possible size of 2 to the still small size of 5. This pattern of initial quick improvement as the crowd size first grows followed by a long period of more gradual improvement is consistent with previous findings (Han & Budescu, 2019; Steegen, Dewitte, Tuerlinckx, & Vanpaemel, 2014; Vul & Pashler, 2008). For the US states condition, a crowd size of 10 is almost as well performed as the full crowd. For the African countries condition, the full crowd is clearly better performed than the smaller crowds. We interpret this result as showing that the more difficult African countries condition, about which participants had less knowledge, benefits more from incorporating more participants to capture the more sparsely distributed knowledge.

## Discussion

We studied spatial knowledge in an experiment that asked participants to select regions on unlabeled tile maps to identify target US states or African countries. We asked for the knowledge to be expressed in two different ways, by framing the question in terms of identifying regions in which the target was present or eliminating regions from which the target was absent. Our first interest was in how people manage their uncertainty about the spatial location of the target, and whether this is affected by the different frames. Our second interest was whether wisdom of the crowd effects, including wisdom of the crowd within, are present for spatial knowledge.

We found that participants were consistently overconfident in their management of uncertainty, often to a very large degree. Many participants selected regions in the present frame that were too narrow and failed to include more than half of the targets. They were also overconfident in the absent frame, although to a lesser degree. The consistent pattern of overconfidence in both frames eliminates simple explanations in terms of minimizing effort and suggests that people are overconfident in their spatial knowledge. This sort of overconfidence is consistent with classic findings from the judgment and decision-making literature (Lichtenstein, Fischhoff, & Phillips, 1982; Paese & Sniezek, 1991; Russo & Schoemaker, 1992; Welsh & Begg, 2018).

We also found strong wisdom of the crowd effects. Both unweighted and confidence-weighted aggregate crowd judgments outperformed the vast majority of individual participants. This was especially true for the more difficult African countries condition, suggesting most individuals have significant gaps in their knowledge but that collectively a crowd can perform well. At the individual level, we found that combining judgments from the same participants across both present and absent frames improved performance. A crowd aggregate that combined all participants and both frames achieved very good performance in both conditions. For the US states domain, a crowd of around 10 people proved enough to exhibit good performance, but the lower-knowledge African countries domain benefited from larger crowds.

Our results have implications both for understanding human cognition and practical applications. It is important to understand why people are overconfident in the regions they select, how robust this behavior is, and whether it can be mitigated. Future experiments should consider other spatial knowledge domains and other methods for expressing spatial knowledge, such as point estimates of locations or free-form selections of regions rather than discrete choices on tile maps. It is also important to understand how framing effects interact with the management of uncertainty. Our results suggest that the absent frame reduces overconfidence, but this could arise from the nature of the task design, and more robust replication is needed. In terms of practical applications, the demonstration of strong wisdom of the crowd effects holds promise for real-world problems like search and rescue (Breivik, Allen, Maisondieu, & Olagnon, 2013; Lin, Huynh, Barrington, & Lanckriet, 2013), military targeting (Council, 2013; Qing & Fang, 2021), and other problems where a spatial region needs to be identified based on human knowledge (e.g. Drew et al., 2013; Fu et al., 2017, 2020; Krupinski, 2010; Lin, Huynh, Lanckriet, & Barrington, 2014).

Finally, future work should apply cognitive modeling methods to understand people's behavior and potentially improve the wisdom of the crowd. This approach has proved fruitful in other cognitive domains including probability estimation, category learning, and sequential decision making (Danileiko & Lee, 2017; Lee & Danileiko, 2014; Thomas et al., 2021). Modeling how people select states and countries based on their knowledge should allow inferences about parameters that correspond to their uncertainty and decision-making strategies. A model-based approach to crowd aggregation may outperform the simple statistical methods on which our wisdom of the crowd results are based.

## Acknowledgments

We are thankful for the helpful comments and suggestions from David Budescu and two anonymous reviewers. This research was supported by the U. S. Air Force Research Laboratory's Continuous Learning Branch. LEM's collaboration was enabled through an appointment to the Oak Ridge Institute for Science and Education (ORISE) Summer Research Internship Program. MDL's collaboration was enabled through an appointment to the ORISE Faculty Research Program. The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

## References

Armstrong, J. S. (2001). Combining forecasts. In A. J. S (Ed.), *Principles of forecasting: A handbook for researchers and practitioners.* Springer.

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., … Mellers, B. (2017). Distilling the wisdom of crowds: Prediction markets vs. prediction polls. *Management Science*, *63*(3), 691–706.

Bennett, S. T., Benjamin, A. S., Mistry, P. K., & Steyvers, M. (2018). Making a wiser crowd: Benefits of individual metacognitive control on crowd performance. *Computational Brain & Behavior*, *1*, 90–99.

Boon, M. (2012). Predicting elections: A 'wisdom of crowds' approach. *International Journal of Market Research*, *54*(4), 465–483.

Breivik, Ø., Allen, A. A., Maisondieu, C., & Olagnon, M. (2013). *Advances in search and rescue at sea* (Vol. 63). Springer.

Bruce, R. S. (1935). Group judgments in the fields of lifted weights and visual discrimination. *The Journal of Psychology*, *1*(1), 117–121.

Christiansen, J. D. (2007). Prediction markets: Practical experiments in small markets and behaviours observed. *The Journal of Prediction Markets*, *1*(1), 17–41.

Council, N. R. (2013). *Future U.S. workforce for geospatial intelligence*. Washington, DC: National Academies Press.

Da, Z., & Huang, X. (2020). Harnessing the wisdom of crowds. *Management Science*, *66*(5), 1847–1867.

Danileiko, I., & Lee, M. D. (2017). A model-based approach to the wisdom of the crowd in category learning. *Cognitive Science*, *42*(S3), 861–883.

Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision*, *1*(2), 79–101.

Drew, T., Vo, M. L., Olwal, A., Jacobson, F., Seltzer, S. E., & Wolfe, J. M. (2013). Scanners and drillers: Characterizing expert visual search through volumetric images. *Journal of Vision*, *13*(10), 1–13.

Farnsworth, P. R., & Williams, M. F. (1936). The accuracy of the median and mean of a group of judgments. *The Journal of Social Psychology*, *7*(2), 237–239.

Fu, L., Lee, L., & Danescu-Niculescu-Mizil, C. (2017). When confidence and competence collide: Effects on online decision-making discussions. In *Proceedings of the 26th international conference on world wide web* (pp. 1381–1390).

Fu, L., Wang, A. Z., & Danescu-Niculescu-Mizil, C. (2020). Confidence boost in dyadic online teamwork: An individual-focused perspective. In *Proceedings of the international aaai conference on web and social media* (Vol. 14, pp. 197–208).

Galton, F. (1907). Vox populi. *Nature*, *75*, 450–451.

Gordon, K. (1924). Group judgments in the field of lifted weights. *Journal of Experimental Psychology*, *7*(5), 398–400.

Han, Y., & Budescu, D. (2019). A universal method for evaluating the quality of aggregators. *Judgment and Decision Making*, *14*(4), 395–411.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*(2), 231–237.

Herzog, S. M., & Hertwig, R. (2014). Think twice and then: Combining or choosing in dialectical bootstrapping? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 218–232.

Jenness, A. (1932). The role of discussion in changing opinion regarding a matter of fact. *The Journal of Abnormal and Social Psychology*, *27*(3), 279–296.

Juni, M. Z., & Eckstein, M. P. (2017). The wisdom of crowds for visual search. *Proceedings of the National Academy of Sciences*, *114*(21), E4306–E4315.

Klugman, S. F. (1947). Group and individual judgments for anticipated events. *The Journal of Social Psychology*, *26*(1), 21–28.

Knight, H. C. (1921). *A comparison of the reliability of group and individual judgments*. Unpublished master's thesis, Columbia University (1921).

Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics*, *72*(5), 1205–1217.

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, *52*(1), 111–127.

Lee, M. D., & Danileiko, I. (2014). Using cognitive models to combine probability estimates. *Judgment and Decision Making*, *9*(3), 259–273.

Lee, M. D., Danileiko, I., & Vi, J. (2018). Testing the ability of the surprisingly popular method to predict nfl games. *Judgment and Decision Making*, *13*(4), 322–333.

Lee, M. D., Steyvers, M., & Miller, B. (2014). A cognitive model for aggregating people's rankings. *PLoS ONE*, *9*(5), 1–9.

Lee, M. D., Zhang, S., & Shi, J. (2011). The wisdom of the crowd playing the price is right. *Memory & Cognition*, *39*, 914–923.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188. doi: 10.1006/obhd.1998.2804

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (chap. 22). Cambridge University Press.

Lin, A. Y., Huynh, A., Barrington, L., & Lanckriet, G. (2013). Search and discovery through human computation. In P. Michelucci (Ed.), *Handbook of human computation.* Springer New York.

Lin, A. Y., Huynh, A., Lanckriet, G., & Barrington, L. (2014). Crowdsourcing the unknown: The satellite search for genghis khan. *PloS ONE*, *9*(12), 1–17.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*(6), 1231–1243.

Lyon, A., & Pacuit, E. (2013). The wisdom of crowds: Methods of human judgement aggregation. In P. Michelucci (Ed.), *Handbook of human computation.* Springer.

Mayer, M., & Heck, D. W. (2022). *Cultural consensus theory for two-dimensional data: Expertise-weighted aggregation of location judgments.* PsyArXiv. Retrieved from psyarxiv.com/unhvc doi: 10.31234/osf.io/unhvc

Miller, M. K., Wang, G., Kulkarni, S. R., Poor, H. V., & Osherson, D. N. (2012). Citizen forecasts of the 2008 US presidential election. *Politics & Policy*, *40*(6), 1019–1052.

Paese, P. W., & Sniezek, J. A. (1991). Influences on the appropriateness of confidence in judgment: Practice, effort, information, and decision-making. *Organizational Behavior and Human Decision Processes*, *48*(1), 100–130.

Page, L., & Clemen, R. T. (2013). Do prediction markets produce well-calibrated probability forecasts? *The Economic Journal*, *123*(568), 491–513.

Prelec, D., Seung, H. S., & McCoy, J. (2017). A solution to the single-question crowd wisdom problem. *Nature*, *541*(7638), 532–535.

Prolific. (2022). *Online participant recruitment for surveys and market research.* Retrieved from https://www.prolific.co

Qing, S., & Fang, L. (2021). Research on the intelligent combat decision-making under the simulation and deduction system. In *2021 international conference on big data and intelligent decision making (bdidm)* (pp. 206–209). Guilin, China.

Ray, R. (2006). Prediction markets and the financial "wisdom of crowds". *The Journal of Behavioral Finance*, *7*(1), 2–4.

Russo, J. E., & Schoemaker, P. J. H. (1992). Managing overconfidence. *Sloan Management Review*, *33*(2), 7–17.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition*, *21*(4), 546–556.

Steegen, S., Dewitte, L., Tuerlinckx, F., & Vanpaemel, W. (2014). Measuring the crowd within again: A pre-registered replication study. *Frontiers in Psychology*, *5*(786), 1–8.

Surowiecki, J. (2004). *The wisdom of crowds*. New York: Random House.

Thomas, B., Coon, J., Westfall, H. A., & Lee, M. D. (2021). Model-based wisdom of the crowd for sequential decision-making tasks. *Cognitive Science*, *45*(7).

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*(7), 645–647.

Welsh, M. B., & Begg, S. H. (2018). More-or-less elicitation (mole): reducing bias in range estimation and forecasting. *EURO Journal on Decision Processes*, *6*(1-2), 171–212.

Yaniv, I., & Schul, Y. (1997). Elimination and inclusion procedures in judgment. *Journal of Behavioral Decision Making*, *10*(3), 211–220.

Zhang, S., & Lee, M. D. (2010). Cognitive models and the wisdom of crowds: A case study using the bandit problem. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 32, p. 1118-1123).