Subject Pool Description

All experiments for this project were conducted on Mechanical Turk (MTurk). Participants were selected according to the following criteria:

1. We restricted our sample to workers who had their location set to the US on the MTurk platform.
2. We restricted any individuals who reported their age as under 18 years from participating.
3. We restricted any individuals who reported the country in which they were currently living to be outside of the US from participating.
4. We ran the following set of bot screening questions and restricted participants who did not answer all the questions correctly from participating in the experiment.
    a. Father is to son as mother is to?
    b. Brother is to Father as Sister is to?
    c. Mother is to wife as father is to?
5. Participants were restricted to participating in only one of our experimental sessions.

Experimental protocol

1. Participants who accepted the Human Intelligence Task (HIT) on MTurk were provided with a link to the experiment on the Qualtrics platform.
2. Participants provided their country of residence, their nationality, their age and their MTurk Worker ID.
3. Participants were asked to answer 3 bot questions.
4. Participants were shown a Plain Language Statement about the experiment.
5. Participants were shown the consent form.
6. Participants were shown the instructions for the experiment:
    a. For Experiment 1, our instructions consisted of the following two pages:

For this experiment, on each page you will be presented a **bolded statement**.

For example: **Toronto is the capital of Canada.**

You will then be asked the following questions about each statement:
- (a) Is this statement more likely to be false or true?
- (b) What percentage of other people do you think thought the bolded statement was true?
- (c) What is the probability that the bolded statement is true?
- (d) What do you think is the average probability estimated by other people?

Note that:

If you select "true" for question (a), you will need to enter a probability between 50 and 100 for question (c).

If you select "false" for question (a), you will need to enter a probability between 0 and 50 for question (c).

>>

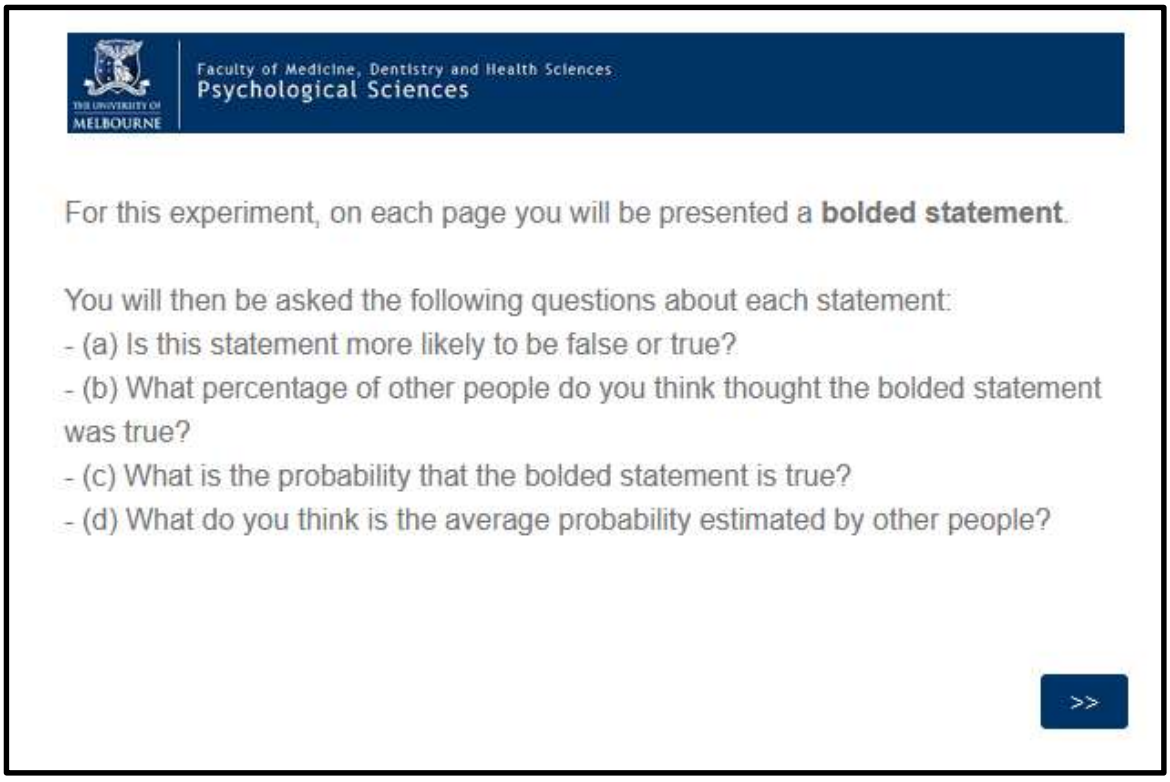
Followed by:

Please do not look up the answers to the following questions on the internet or otherwise. It is OK to make errors. We are interested in what you think the correct answer is, not your ability to look up the correct answer. Please answer each question as honestly as you can. **You will still be paid regardless of your performance, and we will provide you with all the answers at the end of the experiment.**

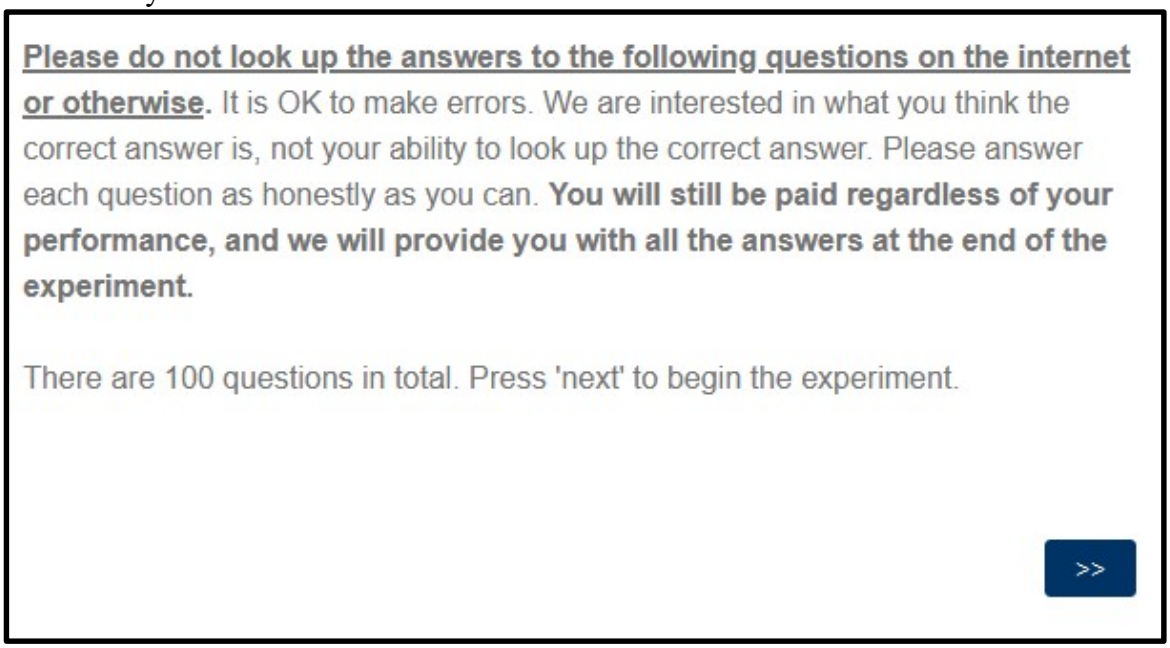
There are 50 questions in total. Press 'next' to begin the experiment.

>>

a. **For Experiment 2, our instructions consisted of the following two pages:**



Followed by:



7. Participants then completed the experiment.
8. Participants were asked whether they cheated or looked up any of the answers.
9. Participants were then provided with a completion code, which they then submitted via MTurk for payment.

Pilots

In addition to the reported results, we ran an initial pilot of our States dataset where we did not enforce consistency between answering true and false and the probabilistic meta-predictions. We had planned to drop inconsistent forecasters in this dataset. However, there were a moderate number of inconsistent reports and this led to a decrease in performance in the original SP algorithm but not the new SC algorithm. To ensure that the original algorithm was not disadvantaged, we reran the states experiment with an additional consistency check that did not allow participants to progress if they (i) reported that the answer was true but gave a probability of below 50% or (ii) reported that the answer was false but gave a probability of above 50%.