­­

Dennis Shasha

Upstart Puzzles

**Privacy-Preserving Polling**

**(figure caption) “**We understand that your choice of candidate may be something you want to keep private. At the end of this process, only you will know for sure whether the choice you mention is your real choice or not.”

[ **figure design concept** Pollster and potential voter talking over the telephone. Pollster bubble: “Who are you going to vote for?” Voter: “Let me flip a coin and then I’ll tell you.”]

When people are asked whom they will vote for, they might not want to say. After all, other people might judge them, ask for contributions, or publish the answer. Suppose there are two candidates, randomly called B and T. Suppose, again for the sake of this hypothetical, that there is a slight stigma against people who support T.

The pollster says to them: "Please flip a coin. If the coin comes up tails, please tell us whom you like best. If it comes up heads, then always say T." That way, even if a person states an intention to vote for T, nobody knows for sure.

Warm-Up: Suppose the true probabilities are 60% for B and 40% for T. Suppose we poll 200 people. How many will say T in response to the poll with the coin flip rule and how many will say B?

Solution to Warm-Up: Approximately half the people, that is 100, will flip heads and will say T, regardless of their preferences. Of the other half, 60 will say B and 40 will say T. So 140 will say T and 60 will say B.

Warm-Up 2: Suppose 70% want T and 30% want B. We poll 200 people again. How many will say T in response to the poll with the coin flip rule and how many will say B?

Solution to Warm-Up2: 170 for T and 30 for B. So to find the true support for T and B, simply subtract from the T score half of the total number of people polled. Leave the B score alone.

But now suppose that a country is so divided that, depending on whom you talk to, there might be a stigma to vote for either candidate. Can the pollsters still do their job?

Question: Can you think of a protocol that will protect privacy for supporters both of B and T?

Solution: Here is one possibility. Tell the pollees (the people asked): "Please flip a coin twice. If it comes up heads both times, then please say T. If it comes up tails both times, please say B. With any other combination, please tell us the truth." Suppose again for the purposes of example 60% want B and 40% want T. If we poll 200 people, B will get 60 true answers and 50 because of double tails. The remaining 90 will go to T. Thus we subtract a quarter of the total number of people polled (50 in this example) from B (yielding 110-50 = 60), a quarter from T (yielding 90-50=40) and we get the correct answer.

The only trouble with this privacy-preserving approach is that it requires doubling the number of people polled to get the same effective sample size. In terms of our example, we need to sample 200 people to get an effective sample size of 100.

Question: Suppose that T has the support of approximately 60% of the people and B has the support of 40%. Suppose the stigma is against only B supporters. Assuming that the people who are polled a deck of cards, can you make it so that if a person responds B, then that person has roughly a 50% chance of actually supporting B, and can achieve an effective sample size of 100 by polling only 140 people.

Solution: Suppose that with probability 2/7 a person will say B regardless of his or her view. Then if 60% want T and 40% want B, B will receive (2/7)\*140 = 40 votes because of this 2/7 probability and another 40 from the committed B supporters. Such a scheme would the goal of having anyone who responds with B to the pollster to have a 50% chance of supporting B. To achieve this, the pollster says to the pollee: "Please take 7 cards including an ace, a 2, a 3, a 4, a 5, a 6, and a 7, regarless of suit. Shuffle the seven cards. Now turn over one card. If it's an ace or a 7, say B. Otherwise, please tell me what you really think." In this way, we have to interview only 40 extra people to get an effective sample size of 100 or in general 2/7 extra people.

Upstart: Generalize the above solution to k candidates all with approximately equal support. Each person should have a probability of no more than p of actually supporting the candidate he or she mentions. You can assume for this purpose that the pollee has access to a trusted random number generator that will give a number between 0 and 1 with uniform probability. It should be enough to use this random number generator just once per pollee.

All are invited to submit their solutions to [upstartpuzzles@cacm.acm.org](http://delivery.acm.org/10.1145/3050000/3040969/mailto:upstartpuzzles@cacm.acm.org); solutions to upstarts and discussion will be posted at <http://cs.nyu.edu/cs/faculty/shasha/papers/cacmpuzzles.html>

**Dennis Shasha** (dennisshasha@yahoo.com) is a professor of computer science in the Computer Science Department of the Courant Institute at New York University, New York, USA, as well as the chronicler of his good friend the omniheurist Dr. Ecco.

**pull quote**

When people are asked whom they will vote for, they might not want to say.