IT'S A PUZZLEMENT By JON EVANS

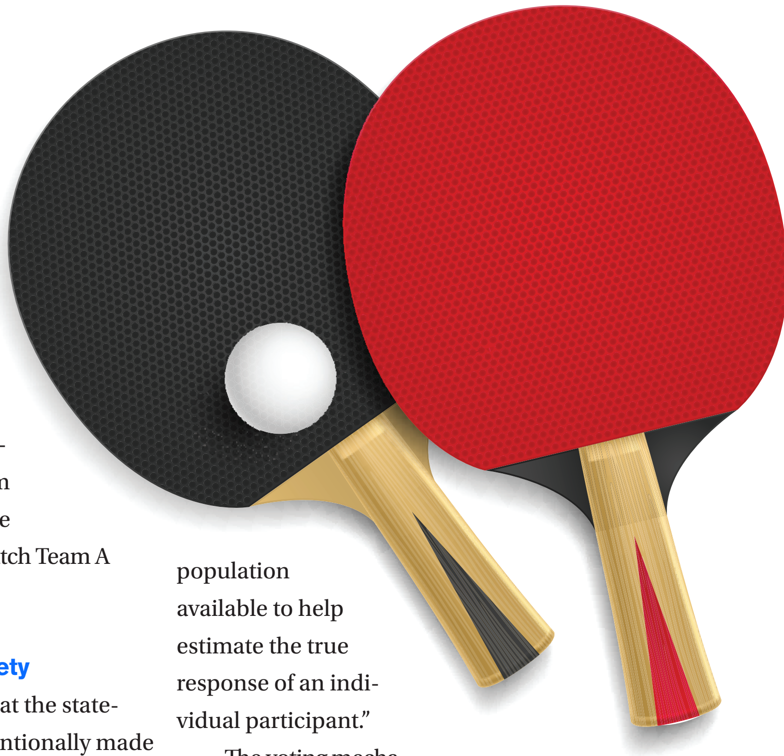
Ping-Pong Team Strategy

Two teams of ping-pong players, Teams A and B, face off in a game under special rules. Two players, one from each team, face off in each match. When the first point is scored, the match ends and the losing player is eliminated from further play. The first team to run out of players loses.

Individual players are rated for strength *S*, measured in the average seconds of playing time until that player gives up a point to the opposing player. For example, if Player X is rated at 20 seconds, then at each instant of playing time Player X gives up an average of 0.05 points per second. The probability of giving up a point is the same at any given instant, independent of the opposing player, and "memoryless," meaning that the probability at any instant is completely independent of whatever happened in earlier instants. Here is the strength table of all the starting players in the teams.

Team A		Team B	
Player	Strength (sec.)	Player	Strength (sec.)
A1	40	B1	90
A2	30	B2	20
A3	25	B3	15
A4	20	B4	10
A5	15	B5	5

Suppose before each match Team B first selects one of its remaining players to play and then Team A can select with this information. What is the best possible strategy for Team B to select players and what is the expected probability Team B will win under this strategy? What is the worst possible strategy for Team B to select players and what is the expected probability Team B will win under this strategy? What are the strategies and probabilities, best and worst for Team A? Now, answer all the same questions if before each match Team A has to select a player first.



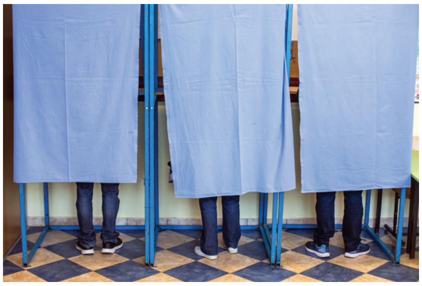
Polling privacy and safety

Apologies to readers, in that the statement of this puzzle unintentionally made the solution much more ambigu-ous than intended. The wording of the puzzle was slightly flawed or incom-plete, defining the 60/40 standard only as "even if a participant's identity and reported response are disclosed, the true intended response of the participant could only be determined with 60% probability of being correct." The problem is that this definition may be impossible if there is information about the population as a whole prior to the survey (e.g., 90% of the population intends to vote for Candidate 1); in some cases, Bayesian estimates then may always allow a higher than 60% probability that the true intended response for a given participant can be determined from the reported response. For a meaningful solution, we will include the condition that "There is no other information about the candidate preference proportions of the population available to help estimate the true response of an individual participant."

The voting mechanism software could employ a random generator (triggered when a survey button is pressed) that reports the true response with probability *p*, but otherwise reports the result of a random 50/50 coin-flip for the two candidates. Then the probability that a voter's true response is reported is $p + (1\text{-}p)/2 = (1 + p)/2$. For the 60/40 standard, *p* should be set to 20%. For N total surveyed voters with $N1$ being the number of true responses for Candidate 1 and $M1$ being the reported responses for Candidate 1, $E[M1] = p\,N1 + (N\text{-}N1)(1\text{-}p)\,/2$. So $N1_{hat} = \frac{(2M1-N+Np)}{2p}$ is an unbiased estimator that can be used to determine the outcome of the survey.

Let q = the true proportion of the population that would respond for Candidate 1 if surveyed, then for N=1:

$\text{Var}[M1] = \text{Var}[E[N1|\text{ respondents true preference}]]$



$+ E[\mathrm{Var}[N1|$ respondents true preference$]]$

$= (\frac{1+p}{2})^2 q + (\frac{1-p}{2})^2 (1-q) - (\frac{1+p}{2} q + \frac{1-p}{2}(1-q))^2 + (\frac{1+p}{2})(\frac{1-p}{2})$.

In general,

$\mathrm{Var}[M1] = N((\frac{1+p}{2})^2 q + (\frac{1-p}{2})^2 (1-q) - (\frac{1+p}{2} q + \frac{1-p}{2} (1-q))^2 + (\frac{1+p}{2})(\frac{(1-p)}{2}))$.

$\mathrm{Var}[N1_{hat}] = \frac{Var[M1]}{p^2}$.

$$\mathrm{Var}[\frac{N1_{hat}}{N}] = \frac{(\frac{1+p}{2})^2 q + (\frac{1-p}{2})^2 (1-q) - (\frac{1+p}{2} q + \frac{1-p}{2} (1-q))^2 + (\frac{1+p}{2})(\frac{1-p}{2})}{N p^2}$$

For the 60/40 standard, since $p = 20\% = 1/5$, after some algebra

$$Var[\frac{N1_{hat}}{N}] = \frac{6+q-q^2}{N}.$$

On the other hand, for no privacy standard $p = 100\% = 1$:

$$Var[\frac{N1_{hat}}{N}] = \frac{Var[M1]}{N} = \frac{Var[N1]}{N} = \frac{q-q^2}{N}.$$

So, introducing the privacy standard increases the variance of the estimate by a factor of

$$\frac{6+q-q^2}{q-q^2} = 1 + \frac{6}{q-q^2}.$$

Unfortunately, this factor goes to $+\infty$ as q goes to either 0 or 1. In general, there is not necessarily a possible proportional increase in the sample size N that would keep the standard deviation of $\frac{N1_{hat}}{N}$ to no more than 3%. As a practical matter, if we make a reasonable assumption that q is no more extreme than 10% or 90%, then

$$\mathrm{Var}[\frac{N1_{hat}}{N}] \le (1 + \frac{6}{0.09})\, \mathrm{Var}[\frac{N1_{hat}}{N}] = \frac{203}{3}\mathrm{Var}[\frac{N1_{hat}}{N}]$$

Thus, $\mathrm{St.Dev}[\frac{N1_{hat}}{N}] \le \mathrm{Sqrt}[\frac{203}{3}]\ \mathrm{St.Dev.}[\frac{N1}{N}]$. If the sample size N is increased by a factor s to $N' = sN$, then $\mathrm{St.Dev}[\frac{N1_{hat}}{N'}] \le \mathrm{Sqrt}[\frac{203}{3}]\ \mathrm{St.Dev.}[\frac{N1}{N}] / \mathrm{Sqrt}[s]$. Since $\mathrm{St.Dev.}[\frac{N1}{N}] = 2\%$, to keep $\mathrm{St.Dev}[\frac{N1_{hat}}{N'}] \le 3\%$ requires that $Sqrt[\frac{203}{3}] / \mathrm{Sqrt}[s] \le 3/2$ or that $s \ge \mathrm{Sqrt}[\frac{203}{3}](2/3) \approx 5.484$. As a practical matter, a bit less than increasing the sample size by a factor 5.5 should reasonably insure that the "sampling error" (after the 60/40 standard mechanism is applied) will still be 3% or less.

Know the answer? Send your solution to ar@casact.org.