Appendix A: Experiment 1 Design and Planned Analyses

Finalized: February 16, 2010

Qualtrics survey

1. Instructions.

   Tell them about the 1MM ten-flip sets. (Order 1: 2-3-4-5 and Order 2: 4-5-2-3 are randomized.)

2. **2^5**.

   They first get these 20 questions:

   a. 20 literally random questions where we show 5 of the 10 flips and ask about one of the others, e.g., Which is more frequent, xxHHHTTxHx or xxHHHTTxTx?

      i. Address the problem that since both are equally likely neither response could be wrong: Don't have equal payoffs, but offer (say 51 if they say H and they're right but they get 49 if they say T and they're right. These money amounts are randomly chosen from the set 55/45, 53/47, 51/49, 49/51, 47/53, and 45/55 and are never equal (.50).

3. **Head counts**. Blocks (a – e) are presented in random order. The questions within each block are also presented in random order.

   a. Estimate the proportion of the 10-flip sets that have 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 heads. (1 question with 11 options. We reward them using the squared distance of their answer from the truth for one Randomly Selected Interval., aka STASI. The summed probability across the 11 options must equal 100.)

      When we run STASI, we let them specify a distribution without constraining it to sum to 100. If it doesn't sum to 100, we normalize it to 100 and give them the question again asking them if it is acceptable. Also, instruct them to complete the different options of the distribution in a (randomly) pre-specified order. .)

   b. Estimate the proportion of 10-flip sets that have 0-3 heads, 4, 5, 6, and 7-10 heads. (1 question with 5 options. Probabilities 17.2, 20.5, 24.6, 20.5, and 17.2 respectively. Score with STASI.)

   c. Estimate the proportion of 10-flip sets that have 0-4, 5, and 6-10 heads. (1 question with 3 options. Probabilities 37.7, 24.6, and 37.7, respectively. Score with STASI.)

   d. 11 questions:

      i. What proportion of sets have a head as the first flip?

      ii. Among the sets where the 1st flip is a head, what fraction of draws have the 2nd flip being a head?

      iii. 3rd

iv. 4th
v. 5th
vi. 6th
vii. 7th
viii. 8th
ix. 9th
x. Among the sets where the first nine flips are heads, what fraction of draws have the 10th flip being a head?

- For each of these 11, randomly pick the first sequence as being either H or T.
- We reward them for accurately estimating these proportions according to STASI.

e. Estimate the proportion of the 10-flip sets that have 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 heads. (11 separate questions, not constrained to sum to 100%.)

4. Tell them about the 1MM thousand-flip sets.
5. They answer these questions in random order:
   a. Estimate the proportion of the 1000-flip sets that have 0 to 50, 51 to 150, 151 to 250, 251 to 350, 351 to 450, 451 to 550, 551 to 650, 651 to 750, 751 to 850, 851 to 950, 950 to 1000 heads/tails. (1 question with 11 options, whose summed probability must equal 100. Score with STASI.)
   b. Estimate the proportion of 1000-flip sets that have 0 to 480, 481 to 520, and 521 to 1000 heads/tails. (1 question with 3 options. Probabilities 31%, 38%, and 31% respectively. Score with STASI.)
   c. Estimate the proportion of 1000-flip sets that have 0 to 450, 451 to 55%, 550 to 1000 heads/tails. (STASI)
   d. Estimate the proportion of 1000-flip sets that have 0 to 50, 51 to 150, 151 to 250, 251 to 350, 351 to 480, 481 to 520, 521 to 650, 651 to 750, 751 to 850, 851 to 950, 950 to 1000 heads/tails (1 question with 11 options, whose summed probability must equal 100. Score with STASI)
6. Ask people how their statistics/math teacher would have answered:
   a. Among the sets where the first nine flips are heads, what fraction of draws have the 10th flip being a head?
   b. One of the random questions from 3d., 3c, 2
7. Everyone answers demographic questions (age, gender, income)
8. Everyone takes Dan's "2-min" math quiz

Debriefing

9. Email them the actual distribution we generated for them, along with feedback on their performance.

Compensation

10. Winners determined by lottery. They are contacted by email and asked to submit a signed receipt. Once we get the receipt, we mail them a check.

•

Specific Hypotheses: Comparisons with normative beliefs

1) Subjective sampling distributions will have "fat tails."
   a) Calculate sum of weights on 0, 1, 2, 3 and 7, 8, 9, 10 in ***3a***.
   b) Examine 0-3 and 7-10 in ***3b***.
   c) Calculate sum of weights on 0-5%, 5-15%, 15-25%, 25-35% and 65-75%, 75-85%, 85-95%, 95-100% in ***5a***.
2) Beliefs about sequences will exhibit GF (gambler's fallacy).
   a) The 11 Qs in ***3d*** will show GF.
   b) Every random question in ***3e*** that has a majority of heads will have people predicting a tail, and vice-versa.
3) Exact representativeness vs. 1/n
   a) Exact representativeness
      i) People will overestimate the probability of exactly 5/10 heads when asked 0-4, 5, 6-10 (***3c***).
      ii) People will overestimate the probability of 48-52/100 heads when asked 0-47 heads, 48-52, and 52-100 heads (***5b***) – or will this go the other way as "fat tails" predicts?
   b) 1/n
      i) The probability attached to 48-52% will be higher in ***5b*** than ***5d***
      ii) The probability attached to 45-55% will be higher in ***5c*** than ***5a***
4) Sample-size neglect: 10-flip distribution (***3a)*** and 100-flip distribution (***5a***) will be "constant in proportions" a la K&T.
   a) NOTE: These distributions are not strictly comparable, but this is nonetheless the comparison that K&T made.

Specific Hypotheses: Within-subject inconsistencies

5) Support theory / conjunction fallacy
   a) When asked the frequency of an outcome separately (***3e***), people will put higher weight than if asked to give a frequency distribution (***3a***). Consequently, the probabilities in ***3e*** will sum to more than 1.
   b) The sum of weights on 0, 1, 2, 3 in ***3a*** will exceed the weight on 0-3 in ***3b***. Similarly for 7-10.
   c) The weight on 48-52 heads in ***5b*** will exceed the weight on 45-55% in ***5a***. – this feels like the "Linda problem" / exact representativeness
   d) Partitioning up the outcome space will spread out participants' estimates of proportion. (5a will have a flatter distribution than 5c. 5d will be flatter than 5b.)
6) People will have subjective sampling distributions that are too flat and put too much weight in the tails, and yet believe in *negative* autocorrelation of signals.
   a) Compare the subjective sampling distribution in ***3a*** with GF in ***2***.

b) Compare the product of the probabilities in ***3d*** with the probability of 10 heads reported in ***3a*** and ***3e***.

Specific Hypotheses: Others

7) People will report more accurate beliefs when asked what their math/statistics teacher would say (***6ab***).
8) People who are better at math (***8***) will be better on ***6ab***, but still exhibit the other biases and inconsistencies.

Specific Hypotheses: Logistical

9) Fatigue: People will respond more noisily (but not systematically different in expectation) when asked the same question late in the survey rather than early.

10) Do the results depend on the order? Specifically test whether the order of 3a first or last matters.

Notes to remember for later:

- If the "statistics/math teacher would have answered" questions generates different answers than the beliefs questions, maybe do a condition where we ask the "teacher" question first.
- Do a lab follow-up study where we actually have subjects implement a distribution of 10-flip sequences at the end of the experiment.
- If we get crappy data, don't give up. Run a simpler study with more targeted hypotheses that is less likely to exhaust participants.

Stakes: $50 prize for every 10 subjects.

Tentative Schedule
Program web page: by Mar 1
Fix, debug, test: by mid-March
Collect data late March

Dan volunteers to take first crack at data analysis.

Appendix B: Experiment 2 Design and Planned Analyses (Preregistered)

Finalized: April 6, 2014

Subject sample: 300 UC-Berkeley undergrads

Incentives: For each question, the participant will earn some number of lottery tickets between 0 and 100. The number of lottery tickets is determined as follows:

- For betting questions, participants choose between (A) and (B) and get the number of lottery tickets specified for an option if the coin flips satisfy the condition described in the chosen option.
- For histogram questions, one bin in the histogram is selected at random, and the participant earns a number of lottery tickets that depends on the participant's reported frequency for that bin in the set of coin flips and the actual frequency for that bin according to: # lottery tickets earned = $100 \times [1 - (\text{participant's reported frequency} - \text{actual frequency})^2]$.
- For questions that directly elicit a frequency from the participant, the number of lottery tickets will again be determined as per the formula in the previous bullet point.

After the experiment is completed, the computer will randomly determine, for each question, whether the participant won a prize. In any given question, the probability of winning a prize will be (# lottery tickets won in that question) / 2,000. That is, the probability will be some percentage between 0% (which occurs if the participant earns 0 lottery tickets in the question, the minimum possible) and 5% (which occurs if the participant earns 100 lottery tickets in the question, the maximum possible).

If the participant wins n prizes, then the value of *each prize* will be $\$n$. In other words, the participant's incentive payment will be $\$n^2$. (This procedure of having the value of the prize for *each* question be increasing in the total number of prizes won sharpens the incentives and was pioneered by McKelvey and Page (1990).) In addition to this incentive payment, each participant will also be paid a flat show-up of $5.

The following calculation guided our choice of the size of our incentive payments. There are 74 questions in the experiment in which a participant could win a prize, so the maximum possible incentive payoff is $74^2 = \$5476$, but this is fantastically unlikely. To conservatively estimate the cost of the experiment, we assume that all participants answer all questions with maximal accuracy, making the probability of winning each prize 5% (the upper bound). In that case, the expected value of $\$n^2$ is $17.205. Adding the $5 show-up fee, the conservative estimate of the cost per participant is $22.205. In addition, the incentive for the math test is $.25 per item, and we'll expect to pay $1 per person, on average. Given

our intention to recruit 300 participants, our conservative estimate of the cost of the experiment is $23.205 × 300 = $6961.50.

Section order: We will keep together as blocks the questions regarding samples of size 10, 1000, and 1 million coin flips, since there are common instructions for each block, but we will randomize the order of the blocks. Within each block, we will group together each type of question, but also randomize the order of the groups within each block as well as the order of the questions within each group.

Note: In describing the questions below, we phrase them in terms of the number of heads. However, for each participant and each block of questions, the computer randomly varied whether the questions were framed in terms of heads or tails.

Note: In order to simplify the process of paying subjects, in places that called for selecting a random flip number or a random flip outcome, we pre-computed a large set of these random numbers and then selected at random from this set.

Note: If the journal in which this paper is eventually published is cooperative, we will ask researchers who want to access the data to provide a brief paragraph on their intended analyses. The description would not be vetted—we would give the data automatically—but this procedure will hopefully make transparent which hypotheses/research questions were of interest to them before they got a chance to conduct any analyses. We will reserve the right to post or share these analysis plans.

I. Questions of Primary Interest

I.A. 10-flip sets

I.A.1. Betting questions (10 Questions)

In question q = 1,…,10 (in a random order), flips 1, 2,…, q-1 are shown (and determined randomly), and the participant must make a bet about whether the qth flip is H or T. The participant is told that we will pick one sequence at random from the million sets (matching the sequence shown), and they will be paid if the sequence we draw matches what they picked. If the participant picks H, then the payoff is $x \in \{45, 47, 49, 51, 53, 55\}$ lottery tickets (the value of x is randomly chosen), and if the participant picks T, then the payoff is the symmetric (around 50) number of lottery tickets.

Planned analyses:

(a) Testing gambler's fallacy (GF): For a given *q*, how the frequency of predicting H varies with the prior number of heads.
   i. t-test of each frequency of prediction's difference from 50%.
   ii. Linear regression (linear probability model) of prediction on number of prior H's.

(b) How the frequency of predicting H varies with the payoff to H: Including payoffs as regressors in the model from (a.ii) above.

(c) Consistency of GF across questions: Comparing people's predictions to the Rabin-Vayanos model as fitted on data from I.A.3 (streak questions): Both with and without the payoff information included as an input to the model's predictions.

I.A.2. Histogram questions (14 Questions)

(A) 11-bin elicitation (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 heads) using SPIES (Subjective Probability Interval EstimateS), in which subjects are asked to estimate the probability of each of the entire set of mutually exclusive possible outcomes, and these estimates must sum to 100%.
(True probabilities 0.1%, 1.0%, 4.4%, 11.7%, 20.5%, 24.6%, 20.5%, 11.7%, 4.4%, 1.0%, and 0.1%, respectively.)

(B) 5-bin equalized elicitation (0-3, 4, 5, 6, 7-10 heads) using SPIES.
(True probabilities 17.2%, 20.5%, 24.6%, 20.5%, 17.2%, respectively.)

(C) 3-bin elicitation (0-4, 5, 6-10 heads) using SPIES.
(True probabilities 37.7%, 24.6%, 37.7%, respectively.)

(D) 11 separate-screen questions (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 heads), not constrained to sum to 100% (each counted as a <u>distinct</u> question).

Planned analyses:
(a) Test for bin effects.
   i. Compare the mean probability assigned to 0-3 in (B) with the sum of the mean probabilities assigned to 0, 1, 2, 3 in (A); similarly, compare 7-10 with the sum of 7, 8, 9, 10.

   ii. Compare the mean probability assigned to 0-4 in (C) with the sum of the mean probabilities assigned to 0, 1, 2, 3, 4 in (A); similarly, compare 6-10 with the sum of 6, 7, 8, 9, 10.

iii. Compare the mean probability assigned to 0-4 in (C) with the sum of the mean probabilities assigned to 0-3, 4 in (B); similarly, compare 6-10 with the sum of 6, 7-10.

iv. Compare the mean probabilities from (D) to those from (A).

(b) Test for NBLLN in N = 10.

i. Examine whether (B) generates approximately equal mean elicited probabilities across the five bins (which is close to normatively correct).

ii. Examine whether (C) generates overweighting of 5 heads (whereas NBLLN + bin effects would predict underweighting of 5 heads).

(c) Test for the presence of exact representativeness.

i. Examine whether the Rabin-Vayanos model estimated in I.A.1.d falls short of predicting respondents' predictions about 5 heads out of 10 flips.

I.A.3. Streak questions (10 Questions)

10 sequence questions in a random order:
"Some of the sets of 10 flips had Heads come up on the [first *j* flips]. For all of these, please estimate the percentage that also came up Heads on the next flip."

Planned analyses:

(a) Test for GF: How the predicted frequency of H varies with the number of prior H's.

i. t-test of each predicted frequency's difference from 50%.
ii. Linear regression of predicted frequencies on number of prior H's.

(b) Test for GF at the individual level: Fraction of individuals who always report 50%, always report weakly less than 50% (at least one frequency strictly), always report weakly less than 50% (at least one frequency strictly).

(c) Calibrate GF: Estimate the Rabin-Vayanos model by non-linear least squares on the first-difference formulation of the model.

I.A.4. Parameter-uncertainty questions (3 Questions)

Ask about the percentage of flips in a randomly-chosen location (flip number 1-10) that are H, given that flips at *M* randomly chosen other locations came up H. (Note that these randomly-chosen flip locations are left unspecified.) We will do this with *M* = 1, 2, and 5. We do not anticipate much different from 50% in most of these

answers, because we do not believe the parameter uncertainty is likely to be significant. GF does not predict very strong effects here; while the framing of the question might magnify a form of LSN if participants isolate the set in their thinking and produce answers below 50%, we conjecture participants won't see the question that way. .

Planned analyses:

(a) Test for parameter uncertainty.

i. Test on average whether the judged probability of one random flip being a head, conditional on other random flips being heads, is equal, greater than, or less than 50%.

Note: The premise of this test is the following: if we do not tell participants anything about the positions of our 1, 2, or 5 flips, then there is a logical connection between their answers here and their histogram answers. For intuition, suppose that participants' histograms imply that they have overly dispersed belief about distributions. Then, roughly speaking, telling them that some set of randomly chosen flips are heads should make them think heads is more likely for another randomly chosen flip. To be precise, the probability that one randomly-chosen flip will be a head given that M other randomly-chosen flips are heads must satisfy:

$$\begin{aligned}
\Pr\left(1_{1/(N-M)}|M_{M/N}\right) &= \frac{\Pr\left(1_{1/(N-M)} \cup M_{M/N}\right)}{\Pr\left(M_{M/N}\right)} \\
&= \frac{\sum_{k=0}^{N} \Pr\left(1_{1/(N-M)} \cup M_{M/N}|k \text{ heads out of } N\right)\Pr(k \text{ heads out of } N)}{\sum_{k=0}^{N} \Pr\left(M_{M/N}|k \text{ heads out of } N\right)\Pr(k \text{ heads out of } N)} \\
&= \frac{\sum_{k=0}^{N} \Pr\left((M+1)_{(M+1)/N}|k \text{ heads out of } N\right)\Pr(k \text{ heads out of } N)}{\sum_{k=0}^{N} \Pr\left(M_{M/N}|k \text{ heads out of } N\right)\Pr(k \text{ heads out of } N)} \\
&= \frac{\sum_{k=0}^{N} \left(\frac{k}{N}\right)\left(\frac{k-1}{N-1}\right)\cdots\left(\frac{k-M}{N-M}\right)\Pr(k \text{ heads out of } N)}{\sum_{k=0}^{N} \left(\frac{k}{N}\right)\left(\frac{k-1}{N-1}\right)\cdots\left(\frac{k-M+1}{N-M+1}\right)\Pr(k \text{ heads out of } N)},
\end{aligned}$$

where the notation $x_{y/z}$ signifies the event that x flips were heads out of y flip numbers that were randomly chosen from a set of z flip numbers, and $\Pr(k \text{ heads out of } N)$ is the probability that exactly k flips are heads in the total sample of N flips. From a participant's histogram beliefs, we would know $\Pr(k \text{ heads out of } N)$ and could calculate the above expression exactly.

In practice, because we won't have exact histograms (but only bins), we will need to calculate lower bounds for the above expression, based on "worst-case scenarios" for how our elicited bin beliefs are generated from (unobserved) full histogram beliefs. Based on our calculations using our earlier experiment, we believe that if (as anticipated) subjects report $\Pr\left(1_{1/(N-M)}|M_{M/N}\right)$ as 50% or less than 50%, then we will have strong power to reject the hypothesis that the full histogram beliefs come from "parameter uncertainty," or are consistent with *any* theory of what sequences might occur. We also intend to calculate the above

expression after fitting a beta distribution from elicited bin beliefs (this is a reasonable way of inferring a smooth full histogram, rather than a worst-case full histogram), as in (b) discussed next.

(b) Calibrate beta-binomial parameter uncertainty.
    i. Find the best-fitting parameters of the beta distribution for the 5-bin elicitation histogram (I.A.2.(B)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.A.4).

I.B. 1,000-flip sets

I.B.1. Histogram questions (8 Questions)

(A) 11-bin elicitation (0-50, 51-150, 151-250, 251-350, 351-450, 451-549, 550-649, 650-749, 750-849, 850-949, 950-1000 heads) using SPIES.
(True probabilities all basically 0%, except the middle three, which are: 0.1%, 99.8%, 0.1%, respectively.)

(B) 3-bin narrowed elicitation (0-480, 481-519, 520-1000 heads) using SPIES.
(True probabilities 11%, 78%, 11%, respectively.)

(C) 3-bin elicitation (0-450, 451-549, 550-1000 heads) using SPIES.
(True probabilities 0.1%, 99.8%, 0.1%, respectively.)

(D) 11-bin narrowed elicitation (0-50, 51-150, 151-250, 251-350, 351-480, 481-519, 520-649, 650-749, 750-849, 850-949, 950-1000 heads) using SPIES.
(True probabilities all basically 0%, except the middle three, which are: 11%, 78%, 11%, respectively.)

(E) 3-bin distorted-900 elicitation (0-909, 910, 911-1000 heads) using SPIES.
(True probabilities 100%, 0%, 0%, respectively.)

(F) 3-bin distorted-500 elicitation (0-499, 500, 501-1000 heads) using SPIES.
(True probabilities 48.7%, 2.5%, 48.7%, respectively.)

(G) 3-bin equalized elicitation (0-493, 494-506, 507-1000 heads) using SPIES.
(True probabilities 34.05%, 31.9%, 34.05%, respectively.)

(H) 5-bin equalized elicitation (0-487, 488-496, 497-503, 504-512, 513-1000 heads) using SPIES.
(True probabilities 21.46%, 19.78%, 17.52%, 19.78%, 21.46%, respectively.)

Planned analyses:

(a) Test for bin effects.

i. Compare the mean probability assigned to 0-450 in (C) with the sum of the mean probabilities assigned to 0-50, …, 351-450 in (A); similarly, compare 550-1000 with the sum of 550-649, …, 950-1000.

ii. Compare the probability assigned to 910 heads in (E) to the probability assigned to 850-949 heads in (A).

iii. Compare the probability assigned to 500 heads in (F) to the probability assigned to 451-549 heads in (A).

(b) Test for NBLLN in N = 1000.

i. Examine whether (G) generates approximately equal mean elicited probabilities across the three bins.

ii. Examine whether (H) generates approximately equal mean elicited probabilities across the five bins.

(c) Test for the presence of "exact representativeness."

i. Test whether the average probability assigned to 481-519 in (B) exceeds the average probability assigned to 451-549 in (C).

I.B.2. Streak questions (10 Questions)

10 sequence questions in a random order that start at a randomly chosen flip *j*th location between 1 and 990:
"Please estimate the percentage that had Heads come up on flip [*j*]."
"Some of the sets of 1000 flips had Heads come up on flip [*j*]. For all of these sets, please estimate the percentage that also came up Heads on the next flip."
"Some of the sets of 1000 flips had Heads come up on each of flip [*j* through *j*+1]. For all of these sets, please estimate the percentage that also came up Heads on the next flip."
And so on, up to *j*+9.

Planned analyses:

(a) Test for GF: How the predicted frequency of H varies with the number of prior H's.

i. t-test of each predicted frequency's difference from 50%.

ii. Linear regression of predicted frequencies on number of prior H's.

(b) Test for GF at the individual-level: Fraction of individuals who always report 50%, always report weakly less than 50% (at least one frequency strictly), always report weakly less than 50% (at least one frequency strictly).

(c) Calibrate GF: Estimate the Rabin-Vayanos model by non-linear least squares on the first-difference formulation of the model.

I.B.3. Parameter-uncertainty questions (3 Questions)

Ask about the percentage of flips at a randomly-chosen location that are H, given that 1, 2, or 5 other randomly chosen flips came up H. (The form of these questions and their premise are analogous to those for the $N = 10$ case, as discussed in I.A.4., except here we conjecture gambler's fallacy is even less likely, so we suspect responses are even less likely to depart from 50% in either direction.)

Planned analyses:

(a) Test for GF and parameter uncertainty.

i. Test on average whether the judged probability of one random flip being a head, conditional on other random flips being heads, is equal, greater than, or less than 50%.

(b) Calibrate beta-binomial parameter uncertainty.

i. Find the best-fitting parameters of the beta distribution for the 3-bin equalized elicitation histogram (I.B.1.(G)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.B.3).

ii. Find the best-fitting parameters of the beta distribution for the 5-bin equalized elicitation histogram (I.B.1.(H)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.B.3).

I.C. One-million-flip sets

I.C.1. Histogram questions (3 Questions)

(A) 11-bin elicitation (0-50,000 / 50,001-150,000 / 150,001-250,000 / 250,001-350,000 / 350,001-450,000 / 450,001-549,999 / 550,000-649,999 / 650,000-749,999 / 750,000-849,999 / 850,000-949,999 / 950,000-1,000,000 heads) using SPIES.
(True probabilities all 0%, except the middle, which is 100%.)

(B) 3-bin equalized elicitation (0-499,784 / 499,785-500,215 / 500,216-1,000,000 heads) using SPIES.
(True probabilities 33.32%, 33.35%, 33.32%, respectively.)

(C) 5-bin equalized elicitation (0-499,579 / 499,580-499,873 / 499,874-500,126 / 500,127-500,420 / 500,421-1,000,000 heads) using SPIES.
(True probabilities 20.07%, 20.00%, 19.86%, 20.00%, 20.07%, respectively.)

Planned analyses:

(a) Test for bin effects.

i. Compare the mean probability assigned to 550,000-649,999 in (A) with the mean probability assigned to 499,785-500,215 in (B).

(b) Test for NBLLN in N = 1 million.

i. Examine whether (B) generates approximately equal mean elicited probabilities across the three bins.

ii. Examine whether (C) generates approximately equal mean elicited probabilities across the five bins.

I.C.2. Streak questions (10 Questions)

10 sequence questions in a random order that start at a randomly chosen flip *j*th location between 1 and 9,999,990:
"Please estimate the percentage that had Heads come up on flip [*j*]."
"Some of the sets of a million flips had Heads come up on flip [*j*]. For all of these, please estimate the percentage that also came up Heads on the next flip."
"Some of the sets of a million flips had Heads come up on flip [*j* through *j*+1]. For all of these, please estimate the percentage that also came up Heads on the next flip."
And so on, up to *j*+9.

Planned analyses:

(a) Test for GF: How the predicted frequency of H varies with the number of prior H's.

i. t-test of each predicted frequency's difference from 50%.
ii. Linear regression of predicted frequencies on number of prior H's.

(b) Test for GF at the individual-level: Fraction of individuals who always report 50%, always report weakly less than 50% (at least one frequency strictly), always report weakly less than 50% (at least one frequency strictly).

(c) Calibrate GF: Estimate the Rabin-Vayanos model by non-linear least squares on the first-difference formulation of the model.

I.C.3. Parameter-uncertainty questions (3 Questions)

Ask about the percentage of flips at a randomly-chosen location that are H, given that 1, 2, or 5 other randomly chosen flips came up H. (The form of these questions and their premise are analogous to those for the N = 1,000 case, as discussed in I.B.3.)

Planned analyses:

(a) Test for GF and parameter uncertainty.

i. Test on average whether the judged probability of one random flip being a head, conditional on other random flips being heads, is equal, greater than, or less than 50%.

(b) Calibrate beta-binomial parameter uncertainty.

i. Find the best-fitting parameters of the beta distribution for the 3-bin equalized elicitation histogram (I.C.1.(B)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.C.3).

ii. Find the best-fitting parameters of the beta distribution for the 5-bin equalized elicitation histogram (I.C.1.(C)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.C.3).

I.D. Comparisons across different question types

Planned analyses:

(a) Test for Kahneman and Tversky's "sample size neglect" (including for the much larger sample size of 1 million).

(i) Compare the distribution of 0-5%, 5-15%, …, 95-100% heads across the 10-flip, 1000-flip, and 1-million-flip sets.

(ii) Compare the equalized histogram beliefs across the 10-flip, 1000-flip, and 1-million-flip sets.

(b) Test whether participants' sequence beliefs can be reconciled with their histogram beliefs.

(i) Compare the implied probability of 10 heads out of 10 from I.A.3 (the sequence beliefs) to the histogram probability (I.A.2.(A)), which is contaminated by bin effects, and to the normatively correct probability.

(c) Test whether the effect of some flips on another flip is weaker as we go from the $N = 10$ to the $N = 1000$ and $N = 1$ million parameter-uncertainty questions (question types I.A.4. I.B.3, and I.C.3).

(d) Compare the strength of the GF across the 10-flip, 1000-flip, and 1-million-flip sets (question types I.A.3. I.B.2, and I.C.2).

(e) Compare the estimated parameter-uncertainty models from the different question types.

I.E. Overall Test of Bias

As a set of descriptive statistics, we will calculate the earnings foregone by giving wrong answers to certain questions instead of correct answers. The idea here is that, for some of the questions, the right answers are actually simpler and require less effort (in particular, 50% for the conditional probability questions, 0% for the tails of the histogram questions) than the wrong answers (which may rely on complicated theories such as GF). Specifically, we will calculate from the following questions:

Betting questions (I.A.1): Actual earnings / earnings possible from always choosing the larger amount of money.

Streak questions (I.A.3, I.B.2, I.C.2): Actual earnings / earnings possible from always choosing 50%.

Parameter-uncertainty questions (I.A.4, I.B.3, I.C.3): Actual earnings / earnings possible from always choosing 50%.

Histogram questions for $N = 10$:

(a) I.A.2.(A). 11-bin elicitation: Sum of expected earnings from choices for 0, 1, 9, and 10 heads / sum of expected earnings from having put 0% on these bins.

(b) I.A.2.(D). $N = 10$ histogram questions, 11 separate-screen questions: Sum of expected earnings from choices for 0, 1, 9, and 10 heads / sum of expected earnings from having put 0% on these bins.

Histogram questions for $N = 1000$:

(a) I.B.1.(A). 11-bin elicitation: Sum of expected earnings from choices for all bins except the middle three / sum of expected earnings from having put 0% on these bins.
(b) I.B.1.(C). 3-bin elicitation: Sum of expected earnings from choices for 0-450 and 550-1000 heads / sum of expected earnings from having put 0% on these bins.
(c) I.B.1.(D). 11-bin narrowed elicitation: Sum of expected earnings from choices for all bins except the middle three / sum of expected earnings from having put 0% on these bins.
(d) I.B.1.(E). 3-bin distorted-900 elicitation: Sum of expected earnings from choices for 0-909 and 910 heads / sum of expected earnings from having put 0% on these bins.

Histogram questions for $N = 1{,}000{,}000$:

(a) I.C.1.(A). 11-bin elicitation: Sum of expected earnings from choices for all bins / sum of expected earnings from having put 0% for all bins except 450,001-549,999 and 100% on that one.

We have no specific planned analyses for the results of these calculations, but we intend to use the results of the calculations to make the point that the biases we find are not merely the result of participants not putting in the effort to do appropriate calculations.

II. Questions of Secondary Interest

II.A. U.S. ethnic distribution (1 Question)

In order to give subjects some practice with the survey interface and to orient them toward thinking about distributions of outcomes in a larger population, we will give our research participants a practice test item that asks them to estimate the proportion of different ethnic groups in the United States. We provide no monetary rewards for accurate responses on this practice item.

Planned analysis:

Consistent with our other hypotheses, we expect to see "binning" effects that bias responses toward equal probabilities on all bins. Subjects will overestimate population groups that constitute a small percentage of US population (Native Americans) and underestimate the largest groups (non-Hispanic Whites). We plan to report the results for readers to interpret as they will, but not integrate the findings into our formal analysis of the other questions.

II.B. Robustness

(a) Order effects.

i. Do results look different when the question appeared early vs. late in the experiment? Although we try to neutralize any inference about the right answer or possible twists in how we developed the data, in principle some seemingly inconsistent answers over the course of the experiment may be because subjects changed their inference about the process as the questions changed. If we find no significant order effects besides what appears to be pure fatigue, that would make such explanations less likely.

(b) Fatigue.

i. Do results look noisier when the question appeared late in the experiment?

ii. In order to further address our concerns about fatigue, we will ask two questions designed to collect self-reports of effort:

- "Please tell us how much thought and effort you put in to answering the questions on this survey."

- "There were a number of questions on which you had to estimate the frequency of each of a set of possible outcomes (the ones that had to add to 100%). Specifically on these questions, how much effort did you exert?"
    - Responses for both were on a 5-point scale from "I went through as quickly as I could" to "I put a lot of effort into making the best guesses I could"

Because we are concerned that anyone who responds with a 1 on either scale was probably not paying attention, we expect it to make sense to treat these people differently. Our plan is to analyze our data both with and without these people, and report both results.

(c) Calculators.

It is not clear whether use of calculating tools is more or less externally valid. To be able to assess sensitivity of results to use of such tools, we ask the following question at the end of the survey:

"While answering the survey, which of the following did you do? (Check all that apply.)" The answer options were (A) Wrote down calculations on paper, (B) Used a calculator for basic arithmetic, and (C) Used an online tool for calculating binomial probabilities or probability distributions.

We will test robustness of findings to restricting the analysis to subjects who checked none of these.

(d) Incentives.
We ask the following question at the end of the survey to check on the effectiveness of the incentives:

"When answering the questions, did you try to answer as accurately as you could?" The answer options were (A) Yes and (B) No (explain below).

As long as at least 30% of subjects chose A, we will test robustness of findings to restricting the analysis to subjects who chose A.

(e) Self-doubt on "obvious" 50%-answer questions; streak questions: I.A.3, I.B.2, and I.C.2.
Test for subjects who answered the first four questions with 50% and then switched on at least one subsequent question in that section. They were probably wondering what we were after, and not reacting to streaks due to gambler's fallacy. We will test robustness of findings to dropping these subjects. (Our decision to use the first four questions, as opposed to some other number, is arbitrary, but we specify it for concreteness.)

II.C. Individual Differences

The survey will ask:
Gender
Age
Income
Brief 6-item math quiz (with $.25 incentive for each correct answer)
SAT/ACT scores
[relative to earlier experiment, dropped questions about how their statistics/math teacher would have answered; we found no clear differences with what they themselves answered]

Planned analyses:
Conduct some exploratory analyses of associations between GF, NBLLN, and these characteristics. Some specific hypotheses:
(a) Higher math skills are associated with less NBLLN.
(b) Higher math skills are not associated with less GF.
(c) Our intuition is weak on the relationship between math skills and these biases, especially GF: there are a priori reasons to suspect higher skills, attention, etc., are associated with *more* severe GF because of greater (but mistaken) attention to patterns, or weaker for the obvious reason of grasping the right answer better.

In this paper, we will report results about individual differences in an appendix (not in the body of the paper). Dan is interested in these analyses because he thinks they bear on interesting research questions, such as whether or not mathematical

training successfully corrects people's statistical intuitions. (But of course, since the data are correlational, they will be only suggestive.)

Appendix C: Screenshots of Experiment 1

In this experiment, we are going to ask you lots of questions about coin flips. We are interested in what you think about how likely different sequences of heads and tails are.

There will be 54 questions on 47 separate pages. Each one matters. It is important that you read each question carefully and provide the best answer you can.

The survey will give you chances to earn lottery tickets, and you are likely to win several thousand lottery tickets if you perform well. Each lottery ticket (or portion thereof) will increase your chance of winning one of our prizes. Your probability of winning one of the $50 prizes will be equal to the number of lottery tickets you earn divided by 40,000. (We expect that about 1 in 10 participants will win a $50 prize.) Your probability of winning increases in exact proportion to the number of tickets you earn.

When the study is complete, we will conduct the lottery, and if you are one of the winners, we will contact you.

Two notes before you begin:

1) For some of the questions, we will ask you to make judgments using numbers or ranges we provide. In some of these questions, we have chosen the examples and numbers literally randomly. At other times, we have carefully selected the examples to get your judgments in a wide range of scenarios. In fact, you will note that we often ask very similar-sounding questions; these questions may have very similar answers, or very different ones. In all of these cases, the specific numbers in the question are NOT meant to indicate more or less likely answers.

2) You cannot go back to change answers on prior pages. Hitting the back button on your browser will not work properly.

Email address:

We want to make sure we get your email address right, so could you do us a favor and enter it again?

>>

We flipped a coin ten times. Actually, we had a computer simulate the coin flipping, generating exactly the same type of random series real coins do. This was a fair coin, in the sense that coin could come up either heads or tails, and there was an equal chance of each. That generated one ten-flip set. Then we did it again, and again, and again, until...we had 1 million ten-flip sets.

For the next ten questions, we will ask you to bet on which one of two different patterns is more common in the million ten-flip sets. We will show you the outcomes of some of the flips (either H or T) but not others (denoted "x").

For instance, we may ask you:

Which of the following patterns is more frequent?
A) xHHHHHHHHH
B) xTTTTHHHHH

If you answer (A) and the (A) pattern is more common, you will earn 49 lottery tickets.
If you answer (B) and the (B) pattern is more common, you will earn 51 lottery tickets.

The first of these patterns, xHHHHHHHHH, includes two different possibilities: HHHHHHHHHH and THHHHHHHHH. Similarly, the second pattern, xTTTTHHHHH, also includes two different possibilities (TTTTTHHHHH and HTTTTHHHHH).

When you decide which pattern to bet on, you should consider both which pattern you think is more common (in the million sets) and also how many lottery tickets you will earn if you are right.

****Please note: The number of lottery tickets associated with (A) and (B) are chosen randomly between 45 and 55. They do not represent any useful hint toward which pattern is more frequent.****

>>

Which of the following two patterns would you like to bet on?
A) xxHHHHTxTx
B) xxHHHHTxHx

If you answer (A) and the (A) pattern is more frequent, you will earn 47 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 53 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

Which of the following two patterns would you like to bet on?
A) xHxTTTHHxx
B) xTxTTTHHxx

If you answer (A) and the (A) pattern is more frequent, you will earn 55 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 45 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

Which of the following two patterns would you like to bet on?
A) xxHHHHHxTx
B) xxHHHHHxHx

If you answer (A) and the (A) pattern is more frequent, you will earn 45 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 55 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

Which of the following two patterns would you like to bet on?
A) xHxTTTTTxx
B) xTxTTTTTxx

If you answer (A) and the (A) pattern is more frequent, you will earn 49 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 51 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

Which of the following two patterns would you like to bet on?
A) xHxHTHTTxx
B) xTxHTHTTxx

If you answer (A) and the (A) pattern is more frequent, you will earn 49 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 51 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

Which of the following two patterns would you like to bet on?
A) xxTTTTHxHx
B) xxTTTTHxTx

If you answer (A) and the (A) pattern is more frequent, you will earn 47 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 53 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

Which of the following two patterns would you like to bet on?
A) xHxTTHHHxx
B) xTxTTHHHxx

If you answer (A) and the (A) pattern is more frequent, you will earn 47 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 53 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

Which of the following two patterns would you like to bet on?
A) xxHTHTHxHx
B) xxHTHTHxTx

If you answer (A) and the (A) pattern is more frequent, you will earn 55 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 45 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

Which of the following two patterns would you like to bet on?
A) xHxHHHTHxx
B) xTxHHHTHxx

If you answer (A) and the (A) pattern is more frequent, you will earn 47 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 53 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

Which of the following two patterns would you like to bet on?
A) xTxTTTTHxx
B) xHxTTTTHxx

If you answer (A) and the (A) pattern is more frequent, you will earn 47 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 53 lottery tickets.
If both (A) and (B) occur with precisely equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

- A is more frequent
- B is more frequent

>>

For the next set of questions, we will be asking you to estimate the frequency of various possible outcomes of heads and tails in the one million ten-flip sets. Your accuracy will be rewarded as follows.

After you have finished the survey, for each of the following questions we will randomly pick one of the entry boxes and compute the square of the distance between your answer and the truth. If your answer is exactly correct, you will get 100 lottery tickets. The farther your answer is from the truth, the fewer lottery tickets you will win (down to a minimum of zero if you are 100% off) and the lower your chances of winning one of the $50 prizes. We will call this the quadratic scoring rule (since we are squaring the distance between your answer and truth).

The formula for the quadratic scoring rule can be written this way:
Number of tickets you earn = $100 - .01 * (p - t)^2$
where p is the percentage (from 0 to 100) that you report
and t is the actual percentage (from 0 to 100)

Basically, it pays for you to be as accurate as possible.

To give you some practice with the format of the next few questions, please fill in the chart below by guessing what percentage of the U.S. population is of each racial/ethnic group (for example, if you think 5% of the U.S. population is non-Hispanic white, then enter "5"). The closer your answer is to the correct answer (as reported by the U.S. Census Bureau in August 2008), the more you will get paid. Since this one is for practice, it doesn't count for money or lottery tickets.

Non-Hispanic White	0 %
Hispanic	0 %
Non-Hispanic Black	0 %
Asian/Pacific-Islander	0 %
Native American	0 %
Multiracial	0 %
Total	0 %

Note that this practice problem is the only one on which you will receive immediate performance feedback in this experiment.

>>

Please estimate the percentages of ten-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

6-10 heads	0 %
5 heads	0 %
0-4 heads	0 %
Total	0 %

>>

Please estimate the percentages of ten-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

10 tails	0	%
9 tails	0	%
8 tails	0	%
7 tails	0	%
6 tails	0	%
5 tails	0	%
4 tails	0	%
3 tails	0	%
2 tails	0	%
1 tail	0	%
0 tails	0	%
Total	0	%

>>

Please estimate the percentages of ten-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

7-10 tails	0 %
6 tails	0 %
5 tails	0 %
4 tails	0 %
0-3 tails	0 %
Total	0 %

>>

For the next set of questions, we will be asking you to estimate the frequency of various possible outcomes of heads and tails in the one million ten-flip sets. Your accuracy will be rewarded using the quadratic scoring rule.

It pays to be accurate.

>>

What percentage of ten-flip sets include exactly 6 HEADS and 4 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 1 HEAD and 9 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 9 HEADS and 1 TAIL?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include no HEADS and 10 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 8 HEADS and 2 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 3 HEADS and 7 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 10 HEADS and no TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 4 HEADS and 6 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 7 HEADS and 3 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 5 HEADS and 5 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 5 HEADS and 5 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 5 HEADS and 5 TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include exactly 2 HEADS and 8 TAILS?

(Your answer should be a number between 0 and 100.)

>>

For the next set of questions, we will be asking you to estimate the frequency of various possible outcomes of heads and tails in the one million ten-flip sets. Your accuracy will be rewarded using the quadratic scoring rule.

It pays to be accurate.

>>

Among the sets in which the first two flips came up TAILS in what percentage did the third flip also come up TAILS?

(Your answer should be a number between 0 and 100.)

>>

Among the sets in which the first three flips came up TAILS in what percentage did the fourth flip also come up TAILS?

(Your answer should be a number between 0 and 100.)

>>

Among the sets in which the first five flips came up TAILS in what percentage did the sixth flip also come up TAILS?

(Your answer should be a number between 0 and 100.)

>>

What percentage of the ten-flip sets have a TAIL as the result of the first flip?

(Your answer should be a number between 0 and 100.)

>>

Among the sets in which the first eight flips came up TAILS in what percentage did the ninth flip also come up TAILS?

(Your answer should be a number between 0 and 100.)

>>

Among the sets in which the first nine flips came up TAILS in what percentage did the tenth flip also come up TAILS?

(Your answer should be a number between 0 and 100.)

>>

Among the sets in which the first sixth flips came up TAILS in what percentage did the seventh flip also come up TAILS?

(Your answer should be a number between 0 and 100.)

>>

Among the sets in which the first seven flips came up TAILS in what percentage did the eighth flip also come up TAILS?

(Your answer should be a number between 0 and 100.)

>>

Among the sets in which the first flip came up TAILS in what percentage did the second flip also come up TAILS?

(Your answer should be a number between 0 and 100.)

>>

Among the sets in which the first four flips came up TAILS in what percentage did the fifth flip also come up TAILS?

(Your answer should be a number between 0 and 100.)

>>

Good work: you have completed 34 of 54 questions and 39 of 47 pages in this survey.

Okay, we really like flipping coins. In addition to the batch of one million ten-flip sets you have just been answering questions about, we also generated a bigger batch of coin flips. This batch also contains a million sets, but each set has a thousand flips.

The next few survey questions pertain to those thousand-flip sets. For each question, we will be asking you to estimate the frequency of various possible outcomes of heads and tails. We will reward your accuracy using the quadratic scoring rule.

Remember, it pays for you to be as accurate as possible.

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

950 to 1000 heads	0 %
850 to 949 heads	0 %
750 to 849 heads	0 %
650 to 749 heads	0 %
550 to 649 heads	0 %
451 to 549 heads	0 %
351 to 450 heads	0 %
251 to 350 heads	0 %
151 to 250 heads	0 %
51 to 150 heads	0 %
0 to 50 heads	0 %
Total	0 %

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0 to 480 heads.....	0 %
481 to 519 heads.....	0 %
520 to 1000 heads.....	0 %
Total	0 %

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

950 to 1000 heads	0 %
850 to 949 heads	0 %
750 to 849 heads	0 %
650 to 749 heads	0 %
520 to 649 heads	0 %
481 to 519 heads	0 %
351 to 480 heads	0 %
251 to 350 heads	0 %
151 to 250 heads	0 %
51 to 150 heads	0 %
0 to 50 heads	0 %
Total	0 %

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of tails in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

550 to 1000 tails	0 %
451 to 549 tails	0 %
0 to 450 tails	0 %
Total	0 %

>>

For the following questions, please answer the way a mathematics professor would answer them. In fact, we asked a mathematics professor to answer all of these questions. We will reward you for answering the way the math professor answered.

All of the questions pertain to the million ten-flip sets.

Which of the following two patterns would you like to bet on?
A) xxHHHHHxTx
B) xxHHHHHxHx

If you answer (A) and the (A) pattern is more frequent, you will earn 49 lottery tickets.
If you answer (B) and the (B) pattern is more frequent, you will earn 51 lottery tickets.
If both (A) and (B) occur with equal frequency in the million sets we generated, then neither choice earns you lottery tickets.

If you answer the way the math professor answered, you will earn 100 lottery tickets.

- ○ A is more frequent
- ○ B is more frequent

Please estimate the percentage of ten-flip sets with each of the following numbers of heads in them:

The closer your answer is to the math professor's, the more lottery tickets you will earn, using the quadratic scoring rule.

0-4 heads	0 %
5 heads	0 %
6-10 heads	0 %
Total	0 %

Among the sets in which the first two flips came up HEADS in what percentage of these did the third flip also come up HEADS? (The closer your answer is to the math professor's, the more lottery tickets you will earn, using the quadratic scoring rule.)

Your answer should be a number between 0 and 100.

Among the sets in which the first nine flips came up HEADS in what percentage of these did the tenth flip also come up HEADS? (The closer your answer is to the math professor's, the more lottery tickets you will earn, using the quadratic scoring rule.)

Your answer should be a number between 0 and 100.

>>

For the next six questions, you will earn 20 lottery tickets for each correct answer.

If a = 7b, then what does $(5a)/7$ equal?

It takes a small plane an hour to travel 144 miles. How many miles does it travel in 5 minutes?

A dog eats ¾ of a pound of meat every day. How many pounds of meat does the dog eat per week?

If $2/6 = 2/(3+x)$, what is the value of x?

If $8x + 2y = 46$, and $y=7$, what is x?

The volume of a cube is 27 cubic inches. What would be the volume of a cube whose sides were twice as long?

>>

Your responses on the following demographic questions are optional.

What is your gender?

- Male
- Female

How many years old are you?

What is your approximate annual household income?

In case you are the winner of one of our prizes, we need your email address to contact you.

>>

May we email you with information about other research in which you could participate?

- Yes. (If you choose this option you can get off the mailing list at any time.)
- No thank you.

>>

Please tell us how much thought an effort you put in to answering the questions on this survey.

(Note that your response to this question will have no bearing on your probability of being paid or winning the prize.)

I went through quickly				I put a lot of effort into make the best guesses I could
○	○	○	○	○

There were a number of questions on which you had to estimate the freqency of each of a set of possible outcomes (the ones that had to add to 100%). Specifically on these questions, how much effort did you put forth?

(Again, your response to this question will have no bearing on your probability of being paid or winning the prize.)

I did not work too hard to get them exactly right				I put a lot of effort into make the best guesses I could
○	○	○	○	○

>>

Thanks again for participating in this experiment. If you have any questions about the experiment, your performance, or flipping coins for fun and profit, please do not hesitate to contact the primary investigator on this research project, Don Moore: dmoore@cmu.edu

>>

We thank you for your time spent taking this survey.
Your response has been recorded.

Appendix D: Screenshots of Experiment 2

University of California at Berkeley
Consent to Participate in Research

My name is Don Moore. I am a faculty member at the University of California, Berkeley's Haas School of Business. I would like to invite you to take part in my research study which concerns how people think about chance and randomness. You qualify for this study if you are 18 years or older.

If you agree to participate in this research, you will be asked to answer an inordinately large number of questions about coin flips. These questions will ask you to think harder about coin flips than you have probably done ever in your life. There will be lots of questions, and they might get boring. Nevertheless, it will be important for you to pay close attention and do your best to answer accurately. We will reward your accuracy by paying you more money.

If you have any questions about the research, you may contact me, Don Moore, by email: dm@berkeley.edu. If you have any questions regarding your treatment or rights as a participant in this research project, you may contact the University of California at Berkeley's Committee for Protection of Human Subjects at 510-642-7461 or subjects@berkeley.edu.

By selecting 'Accept' below and continuing to the following page, you are stating that you have read this consent form, are eligible (because you are over 18), and agree to take part in this research.

ACCEPT

In this experiment, we are going to ask you lots of questions about coin flips. We are interested in what you think about how likely different sequences of heads and tails are.

******Please read these instructions carefully. We will be testing your understanding of the instructions later in the survey.******

There will be 66 separate questions. Each one matters. It is important that you read each question carefully and provide the best answer you can.

Each question has a prize associated with it, and the more accurately you answer, the more tickets you earn. The more tickets you earn, the more likely you are to win the prize.

Here is how it will work: Each ticket earns you a .05% chance to win. So if you earn 100 tickets on a question, you have a 5% chance of winning a prize for that question. The most tickets you can earn on a question is 100, so your chance of winning a prize on any given question is low. But if you do your best, it is highly likely that you will win at least one prize.

The value of each prize goes up the more prizes you win. If you win 1 prize, it is worth $1. If you win 2 prizes, each pays $2 ($4 total). If you win 3 prizes, each pays $3 ($9 total), and so on. So if you win *N* prizes, they will be worth of total of $*N*x*N*. In the extremely unlikely event that you win prizes on all 66 questions, you will earn $4356 (66 prizes, each worth $66). We're not kidding.

The reason we're paying you this way is that it gives you a strong incentive to do your best on every question: winning a prize on any given question will not only earn you money on that question, but will also make it more valuable to win a prize on every other question.

In addition, you get $5 just for completing the survey, regardless of whether you win any prizes.

When the study is complete, we will score your answers and pay you accordingly.

Three notes before you begin:

1) For some of the questions, we will ask you to make judgments using numbers or ranges we provide. In some of these questions, we have chosen the examples and numbers literally randomly. At other times, we have carefully selected the examples to get your judgments in a wide range of scenarios. In fact, you will note that we often ask very similar-sounding questions; these questions may have very similar answers, or very different ones. In all of these cases, the specific numbers in the question are NOT meant to indicate more or less likely answers.

2) You're welcome to take notes, do calculations on paper, or use a calculator if you like. It may or may not be useful to help you answer more accurately.

3) This survey does not work well with the Firefox web browser. Chrome, Safari, and Internet Explorer work better.

4) The "back" button will not work on your browser to go back to previous pages on the survey.

>>

There will be three groupings of questions on this survey:

Ten-flip sets. For one group of questions, we began by having a computer simulate flipping a coin ten times. That made one set. Then we did it again and again and again and...until we had a million ten-flip sets.

Thousand-flip sets. For another group of questions, we did the same thing except that each set had a thousand flips in it.

Million-flip sets. For the third group of questions, each set had a million flips in it (yes, a million sets of a million coin flips each--we really like flipping coins).

We will randomly determine the order in which you encounter these three question groupings. Each simulated coin flip had equal chances of coming up heads and tails (i.e., it was a fair coin).

Comprehension check: Which of the following will you encounter on this survey (check all that apply)?

- [ ] One million sets of ten coin flips
- [ ] One million sets of one hundred coin flips
- [ ] One million sets of one thousand coin flips
- [ ] One million sets of ten thousand coin flips
- [ ] One million sets of one hundred thousand coin flips
- [ ] One million sets of one million coin flips
- [ ] Ten funny cat videos

>>

Some of the questions will be asking you to estimate the frequency of various possible outcomes of heads and tails. Your accuracy will be rewarded as follows.

After you have finished the survey, we will randomly pick one of the entry boxes and use it to calculate how many tickets you get. If your answer is exactly correct, you will get 100 tickets. The farther your answer is from the truth, the fewer tickets you will get (down to a minimum of zero if you are 100% off) and the lower your chances of winning a prize. In particular, we will start with 100% and subtract from it the square of the distance between your answer and the truth. This is called the **quadratic scoring rule** (since we are squaring the distance between your answer and the truth).

The formula for the quadratic scoring rule can be written this way:

Number of tickets you earn = 100 tickets x $(100\% - (p - t)^2)$

where p is the percentage (from 0 to 100%) that you report

and t is the actual percentage (from 0 to 100%)

The reason we're paying you this way is that it gives you an incentive to give your best guess on every question. Put simply: Even if you are unsure, you earn the most money when you give your best guess.

To give you some practice with the format of the next few questions and the payment rule, please fill in the chart below by guessing what percentage of the U.S. population is of each racial/ethnic group (for example, if you think 5% of the U.S. population is non-Hispanic white, then enter "5"). If this were part of the real test, the closer your answer was to the correct answer (as reported in the 2010 U.S. Census), the more you would get paid. Since this one is for practice, it doesn't count for money or tickets.

Non-Hispanic White	0 %
Hispanic	0 %
Non-Hispanic Black	0 %
Asian/Pacific-Islander	0 %
Native American	0 %
Multiracial	0 %
Total	0 %

>>

The next few survey questions pertain to the group of a million **million-flip** sets.
For each question, we will be asking you to estimate the frequency of various possible outcomes of heads and tails. We will reward your accuracy using the quadratic scoring rule.

While some questions may seem similar to each other, others will seem quite different.

Sometimes we will ask you about things that are very likely; sometimes about things that are very unlikely.

For these, as with all the questions, it pays to give your best guess even if you are unsure.

If you would like to review the instructions, click here to open a new browser tab.

>>

Each set has one million flips in it, numbered 1 through 1,000,000. We chose two flip numbers at random. For those sets that came up **Heads** on both flip numbers, we picked a third flip number at random from the remaining 999,998 flip numbers. In what percentage did the third randomly-chosen flip number also come up **Heads**?

Your answer should be a number between 0 and 100.

>>

Each set has one million flips in it, numbered 1 through 1,000,000. We chose five flip numbers at random. For those sets that came up **Heads** on all five flip numbers, we picked a sixth flip number at random from the remaining 999,995 flip numbers. In what percentage did the sixth randomly-chosen flip number also come up **Heads**?

Your answer should be a number between 0 and 100.

>>

Each set has one million flips in it, numbered 1 through 1,000,000. We chose one flip number at random. For those sets that came up **Heads** on that flip number, we picked a second flip number at random from the remaining 999,999 flip numbers. In what percentage did the second randomly-chosen flip number also come up **Heads**?

Your answer should be a number between 0 and 100.

>>

Please estimate the percentages of million-flip sets (in the 1 million we generated) with each of the following numbers of **tails** in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0 to 499,579 tails	0 %
499,580 to 499,873 tails	0 %
499,874 to 500,126 tails	0 %
500,127 to 500,420 tails	0 %
500,421 to 1,000,000 tails	0 %
Total	0 %

If you would like to review the instructions, click here to open a new browser tab.

Please estimate the percentage of million-flip sets that had Heads come up on flip **597482**.

Your answer should be a number between 0 and 100.

>>

Some of the sets of a million flips had Heads come up on each of flips **597482, 597483, 597484, 597485, 597486, 597487, and 597488**. For all of these sets, please estimate the percentage that also came up Heads on the next flip (flip number **597489**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of a million flips had Heads come up on flips **597482 and 597483**. For all of these sets, please estimate the percentage that also came up Heads on the next flip (flip number **597484**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of a million flips had Heads come up on flip **597482**. For all of these, please estimate the percentage that also came up Heads on the next flip (flip number **597483**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of a million flips had Heads come up on each of flips **597482, 597483, 597484, and 597485**. For all of these sets, please estimate the percentage that also came up Heads on the next flip (flip number **597486**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of a million flips had Heads come up on each of flips **597482, 597483, 597484, 597485, 597486, 597487, 597488, 597489, and 597490**. For all of these sets, please estimate the percentage that also came up Heads on the next flip (flip number **597491**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of a million flips had Heads come up on each of flips **597482, 597483, 597484, 597485, 597486, 597487, 597488, and 597489**. For all of these sets, please estimate the percentage that also came up Heads on the next flip (flip number **597490**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of a million flips had Heads come up on each of flips **597482, 597483, 597484, 597485, 597486, and 597487**. For all of these, please estimate the percentage that also came up Heads on the next flip (flip number **597488)**.

Your answer should be a number between 0 and 100.

>>

Some of the sets of a million flips had Heads come up on each of flips **597482, 597483, 597484, 597485, and 597486**. For all of these sets, please estimate the percentage that also came up Heads on the next flip (flip number **597487**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of a million flips had Heads come up on each of flips **597482, 597483, and 597484**. For all of these sets, please estimate the percentage that also came up Heads on the next flip (flip number **597485**).

Your answer should be a number between 0 and 100.

>>

Please estimate the percentages of million-flip sets (in the 1 million we generated) with each of the following numbers of **heads** in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

500,216 to 1,000,000 heads	0 %
499,785 to 500,215 heads	0 %
0 to 499,784 heads	0 %
Total	0 %

If you would like to review the instructions, click here to open a new browser tab.

>>

Please estimate the percentages of million-flip sets (in the 1 million we generated) with each of the following numbers of **heads** in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0 to 50,000 heads.....	0 %
50,001 to 150,000 heads.....	0 %
150,001 to 250,000 heads.....	0 %
250,001 to 350,000 heads.....	0 %
350,001 to 450,000 heads.....	0 %
450,001 to 549,999 heads.....	0 %
550,000 to 649,999 heads.....	0 %
650,000 to 749,999 heads.....	0 %
750,000 to 849,999 heads.....	0 %
850,000 to 949,999 heads.....	0 %
950,000 to 1,000,000 heads.....	0 %
Total	0 %

If you would like to review the instructions, click here to open a new browser tab.

>>

The next few survey questions pertain to the group of a million **ten-flip** sets.

While some questions may seem similar to each other, others will seem quite different.

Sometimes we will ask you about things that are very likely; sometimes about things that are very unlikely.

For these, as with all the questions, it pays to give your best guess even if you are unsure.

If you would like to review the instructions, click here to open a new browser tab.

>>

Some of the sets of 10 flips had Tails come up on **flip 1, 2, 3, 4, and 5**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (**flip number 6**).

Your answer should be a number between 0 and 100.

>>

Please estimate the percentage of 10-flip sets that had Tails come up on **flip 1**.

Your answer should be a number between 0 and 100.

>>

Some of the sets of 10 flips had Tails come up on **flip 1, 2, and 3**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (**flip number 4**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 10 flips had Tails come up on **flip 1, 2, 3, 4, 5, and 6**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (**flip number 7**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 10 flips had Tails come up on **flip 1, 2, 3, 4, 5, 6, and 7**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (**flip number 8**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 10 flips had Tails come up on **flip 1 and 2**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (**flip number 3**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 10 flips had Tails come up on **flip 1, 2, 3, 4, 5, 6, 7, 8, and 9.** For all of these sets, please estimate the percentage that also came up Tails on the next flip (**flip number 10**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 10 flips had Tails come up on **flip 1, 2, 3, 4, 5, 6, 7, and 8**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (**flip number 9**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 10 flips had Tails come up on **flip 1.** For all of these sets, please estimate the percentage that also came up Tails on the next flip (**flip number 2**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 10 flips had Tails come up on **flip 1, 2, 3, and 4**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (**flip number 5**).

Your answer should be a number between 0 and 100.

>>

Each set has 10 flips in it, numbered 1 through 10. We chose two flip numbers at random. For those sets that came up **Heads** on both flip numbers, we picked a third flip number at random from the remaining 8 flip numbers. In what percentage did the third randomly-chosen flip number also come up **Heads**?

Your answer should be a number between 0 and 100.

>>

Each set has 10 flips in it, numbered 1 through 10. We chose one flip number at random. For those sets that came up **Heads** on that flip number, we picked a second flip number at random from the remaining 9 flip numbers. In what percentage did the second randomly-chosen flip number also come up **Heads**?

Your answer should be a number between 0 and 100.

>>

Each set has 10 flips in it, numbered 1 through 10. We chose five flip numbers at random. For those sets that came up **Heads** on all five flip numbers, we picked a sixth flip number at random from the remaining 5 flip numbers. In what percentage did the sixth randomly-chosen flip number also come up **Heads**?

Your answer should be a number between 0 and 100.

>>

Of the million ten-flip sets, some of them had these first 7 flips (in order): THTTTTT

From among these sets, we will select one at random.

For this particular set, do you bet that the next flip is a:

- Head (in which case you will earn 45 tickets if you are right)
- Tail (in which case you will earn 55 tickets if you are right)

>>

Of the million ten-flip sets, some of them had these first 9 flips (in order): TTHHTHTTT

From among these sets, we will select one at random.

For this particular set, do you bet that the next flip is a:

- Head (in which case you will earn 53 tickets if you are right)
- Tail (in which case you will earn 47 tickets if you are right)

>>

Of the million ten-flip sets, we will select one at random.

For this particular set, do you bet that the first flip is a:

- Head (in which case you will earn 47 tickets if you are right)
- Tail (in which case you will earn 53 tickets if you are right)

>>

Of the million ten-flip sets, some of them had these first 8 flips (in order): HTTTHTHT

From among these sets, we will select one at random.

For this particular set, do you bet that the next flip is a:

- Head (in which case you will earn 55 tickets if you are right)
- Tail (in which case you will earn 45 tickets if you are right)

>>

Of the million ten-flip sets, some of them had these first 3 flips (in order): THH

From among these sets, we will select one at random.

For this particular set, do you bet that the next flip is a:

- Head (in which case you will earn 49 tickets if you are right)
- Tail (in which case you will earn 51 tickets if you are right)

>>

Of the million ten-flip sets, some of them had these first 6 flips (in order): HHTTTT

From among these sets, we will select one at random.

For this particular set, do you bet that the next flip is a:

- Head (in which case you will earn 45 tickets if you are right)
- Tail (in which case you will earn 55 tickets if you are right)

>>

Of the million ten-flip sets, some of them had these first 5 flips (in order): HTTTH

From among these sets, we will select one at random.

For this particular set, do you bet that the next flip is a:

- Head (in which case you will earn 45 tickets if you are right)
- Tail (in which case you will earn 55 tickets if you are right)

>>

Of the million ten-flip sets, some of them had these first 2 flips (in order): TH

From among these sets, we will select one at random.

For this particular set, do you bet that the next flip is a:

- Head (in which case you will earn 51 tickets if you are right)
- Tail (in which case you will earn 49 tickets if you are right)

>>

Of the million ten-flip sets, some of them had this first flip: T

From among these sets, we will select one at random.

For this particular set, do you bet that the next flip is a:

- Head (in which case you will earn 53 tickets if you are right)
- Tail (in which case you will earn 47 tickets if you are right)

>>

Of the million ten-flip sets, some of them had these first 4 flips (in order): TTHT

From among these sets, we will select one at random.

For this particular set, do you bet that the next flip is a:

- Head (in which case you will earn 53 tickets if you are right)
- Tail (in which case you will earn 47 tickets if you are right)

>>

Please estimate the percentages of ten-flip sets (in the 1 million we generated) with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0 heads	0	%
1 head	0	%
2 heads	0	%
3 heads	0	%
4 heads	0	%
5 heads	0	%
6 heads	0	%
7 heads	0	%
8 heads	0	%
9 heads	0	%
10 heads	0	%
Total	0	%

If you would like to review the instructions, click here to open a new browser tab.

>>

Please estimate the percentages of ten-flip sets with each of the following numbers of tails in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0-3 tails	0	%
4 tails	0	%
5 tails	0	%
6 tails	0	%
7-10 tails	0	%
Total	0	%

If you would like to review the instructions, click here to open a new browser tab.

>>

Please estimate the percentages of ten-flip sets with each of the following numbers of tails in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0-4 tails	0	%
5 tails	0	%
6-10 tails	0	%
Total	0	%

If you would like to review the instructions, click here to open a new browser tab.

>>

For the next set of questions, we will be asking you to estimate the frequency of various possible outcomes of heads and tails in the one million ten-flip sets. Your accuracy will be rewarded using the quadratic scoring rule.

For these, as with all the questions, it pays to give your best guess even if you are unsure.

If you would like to review the instructions, click here to open a new browser tab.

>>

What percentage of ten-flip sets include **exactly 10 Heads and no Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **exactly 5 Heads and 5 Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **exactly 8 Heads and 2 Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **exactly 4 Heads and 6 Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **exactly 6 Heads and 4 Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **no Heads and 10 Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **exactly 3 Heads and 7 Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **exactly 1 Head and 9 Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **exactly 2 Heads and 8 Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **exactly 7 Heads and 3 Tails**?

(Your answer should be a number between 0 and 100.)

>>

What percentage of ten-flip sets include **exactly 9 Heads and 1 Tail**?

(Your answer should be a number between 0 and 100.)

>>

The next few survey questions pertain to the group of a million **thousand-flip** sets.
For each question, we will be asking you to estimate the frequency of various possible outcomes of heads and tails. We will reward your accuracy using the quadratic scoring rule.

While some questions may seem similar to each other, others will seem quite different.

Sometimes we will ask you about things that are very likely; sometimes about things that are very unlikely.

For these, as with all the questions, it pays to give your best guess even if you are unsure.

If you would like to review the instructions, click here to open a new browser tab.

>>

Each set has 1000 flips in it, numbered 1 through 1000. We chose two flip numbers at random. For those sets that came up **Tails** on both flip numbers, we picked a third flip number at random from the remaining 998 flip numbers. In what percentage did the third randomly-chosen flip number also come up **Tails**?

Your answer should be a number between 0 and 100.

>>

Each set has 1000 flips in it, numbered 1 through 1000. We chose one flip number at random. For those sets that came up **Tails** on that flip number, we picked a second flip number at random from the remaining 999 flip numbers. In what percentage did the second randomly-chosen flip number also come up **Tails**?

Your answer should be a number between 0 and 100.

>>

Each set has 1000 flips in it, numbered 1 through 1000. We chose five flip numbers at random. For those sets that came up **Tails** on all five flip numbers, we picked a sixth flip number at random from the remaining 995 flip numbers. In what percentage did the sixth randomly-chosen flip number also come up **Tails**?

Your answer should be a number between 0 and 100.

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0 to 493 heads.....	0	%
494 to 506 heads.....	0	%
507 to 1000 heads.....	0	%
Total	0	%

If you would like to review the instructions, click here to open a new browser tab.

>>

Please estimate the percentage of 1000-flip sets that had Tails come up on flip **562**.

Your answer should be a number between 0 and 100.

>>

Some of the sets of 1000 flips had Tails come up on each of flips **562, 563, 564, 565, 566, 567, 568, 569, and 570**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (flip number **571**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 1000 flips had Tails come up on each of flips **562 and 563**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (flip number **564**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 1000 flips had Tails come up on flip **562**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (flip number **563**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 1000 flips had Tails come up on each of flips **562, 563, 564, 565, 566, 567, and 568**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (flip number **569**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 1000 flips had Tails come up on each of flips **562, 563, 564, 565, 566, 567, 568, and 569**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (flip number **570**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 1000 flips had Tails come up on each of flips **562, 563, 564, 565, 566, and 567**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (flip number **568**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 1000 flips had Tails come up on each of flip **562, 563, 564, 565, and 566**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (flip number **567**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 1000 flips had Tails come up on each of flips **562, 563, 564, and 565**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (flip number **566**).

Your answer should be a number between 0 and 100.

>>

Some of the sets of 1000 flips had Tails come up on each of flips **562, 563, and 564**. For all of these sets, please estimate the percentage that also came up Tails on the next flip (flip number **565**).

Your answer should be a number between 0 and 100.

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

550 to 1000 heads.....	0 %
451 to 549 heads.....	0 %
0 to 450 heads.....	0 %
Total	0 %

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of tails in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

950 to 1000 tails	0	%
850 to 949 tails	0	%
750 to 849 tails	0	%
650 to 749 tails	0	%
520 to 649 tails	0	%
481 to 519 tails	0	%
351 to 480 tails	0	%
251 to 350 tails	0	%
151 to 250 tails	0	%
51 to 150 tails	0	%
0 to 50 tails	0	%
Total	0	%

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of tails in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0 to 499 tails.....	0	%
500 tails.....	0	%
501 to 1000 tails.....	0	%
Total	0	%

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

911 to 1000 heads	0	%
910 heads	0	%
0 to 909 heads	0	%
Total	0	%

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of heads in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0 to 50 heads	0	%
51 to 150 heads	0	%
151 to 250 heads	0	%
251 to 350 heads	0	%
351 to 450 heads	0	%
451 to 549 heads	0	%
550 to 649 heads	0	%
650 to 749 heads	0	%
750 to 849 heads	0	%
850 to 949 heads	0	%
950 to 1000 heads	0	%
Total	0	%

If you would like to review the instructions, click here to open a new browser tab.

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of tails in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

513 to 1000 tails	0	%
504 to 512 tails	0	%
497 to 503 tails	0	%
488 to 496 tails	0	%
0 to 487 tails	0	%
Total	0	%

>>

Please estimate the percentage of thousand-flip sets with each of the following numbers of tails in them:

Before you answer, please look at all the categories below. Think about how you will answer all of them before you answer any of them.

0 to 480 tails.....	0 %
481 to 519 tails.....	0 %
520 to 1000 tails.....	0 %
Total	0 %

If you would like to review the instructions, click here to open a new browser tab.

>>

For the next six questions, you will earn 25 cents for each correct answer.

If *a* = 7b, then what does (5*a*)/7 equal?

If 8*x* + 2*y* = 46, and *y* = 7, what is *x*?

A dog eats ¾ of a pound of meat every day. How many pounds of meat does the dog eat per week?

If 2/6 = 2/(3+*x*), what is the value of *x*?

The volume of a cube is 27 cubic inches. What would be the volume (in cubic inches) of a cube whose sides were twice as long?

It takes a small plane an hour to travel 144 miles. How many miles does it travel in 5 minutes?

>>

Your responses on the following demographic questions are optional.

What is your gender?

- Male
- Female

How many years old are you?

What is your approximate annual household income? (If you are a full-time student, you should include yourself in your parents' household.)

If you took it and you remember your SAT score, please enter it here:
(Please give us your score out of 1600 points, excluding the writing section.)

If you took it and you remember your ACT score, please enter it here:

>>

Note that your responses to these questions will have no bearing on your payment.

Please tell us how much thought and effort you put in to answering the questions on this survey.

I went through as quickly as I could — I put a lot of effort into making the best guesses I could

There were a number of questions on which you had to estimate the frequency of each of a set of possible outcomes (the ones that had to add to 100%). Specifically on these questions, how much effort did you exert?

I went through as quickly as I could — I put a lot of effort into making the best guesses I could

While answering the survey, which of the following did you do? (Check all that apply.)

- [ ] Wrote down calculations on paper
- [ ] Used a calculator for basic arithmetic
- [ ] Used an online tool for calculating binomial probabilities or probability distributions

When answering the questions, did you try to answer as accurately as you could?

- ( ) Yes
- ( ) No (explain below)

>>

If you have any comments or feedback regarding this survey, we would be happy to hear your thoughts:

Thanks again for participating in this experiment. If you have any questions about the experiment, your performance, or flipping coins as a hobby, please contact the primary investigator on this research project, Don Moore: dmoore@haas.berkeley.edu

>>

We thank you for your time spent taking this survey.
Your response has been recorded.

Appendix E: Proofs of Propositions

Proposition 1: Assume A1′, and fix any two events and two binnings of those events: $E, E' \in \mathcal{E}, \mathcal{E}'$. Then $\frac{P(E|\mathcal{E})}{P(E'|\mathcal{E})} = \frac{P(E|\mathcal{E}')}{P(E'|\mathcal{E}')}$.

Proof of Proposition 1: $\frac{P(E|\mathcal{E})}{P(E'|\mathcal{E})} = \frac{g(r(E))/\sum_{F\in\mathcal{E}} g(r(F))}{g(r(E'))/\sum_{F\in\mathcal{E}} g(r(F))} = \frac{g(r(E))}{g(r(E'))} = \frac{g(r(E))/\sum_{F\in\mathcal{E}'} g(r(F))}{g(r(E'))/\sum_{F\in\mathcal{E}'} g(r(F))} = \frac{P(E|\mathcal{E}')}{P(E'|\mathcal{E}')}$. ∎

Proposition 2: Assume A1″, and fix any two events $E, E' \in \mathcal{E}$. Then: $\frac{P(E|\mathcal{E})}{P(E'|\mathcal{E})} > 1$ if and only if $\frac{r(E)}{r(E')} > 1$; and $\frac{P(E|\mathcal{E})}{P(E'|\mathcal{E})} = 1$ if and only if $\frac{r(E)}{r(E')} = 1$.

Proof of Proposition 2: As in the proof of Proposition 1, $\frac{P(E|\mathcal{E})}{P(E'|\mathcal{E})} = \frac{g(r(E))}{g(r(E'))}$. Since $g(\cdot)$ is a strictly increasing function, the result follows. ∎

Proposition 3: Assume A1″ and A2′, and fix any two events $E, E' \in \mathcal{E}$. If $\frac{P(E|\mathcal{E})}{P(E'|\mathcal{E})} > 1$, then $\frac{r(E)}{r(E')} > \frac{P(E|\mathcal{E})}{P(E'|\mathcal{E})}$.

Proof of Proposition 3: As in the proof of Proposition 1, $\frac{P(E|\mathcal{E})}{P(E'|\mathcal{E})} = \frac{g(r(E))}{g(r(E'))} > 1$. Note that

$$\frac{g(r(E))}{g(r(E'))} = \frac{g(r(E')) + [g(r(E)) - g(r(E'))]}{g(r(E'))}$$
$$= 1 + \frac{g(r(E)) - g(r(E'))}{g(r(E'))}$$
$$< 1 + \frac{r(E) - r(E')}{r(E')}$$

$$= \frac{r(E)}{r(E')},$$

where the inequality follows because $g(\cdot)$ is strictly increasing and concave. The result follows. ∎

Proposition 4: Suppose the bias of the coin is not known to be 50% but is instead drawn from ν, a (continuous or discrete) nondegenerate distribution on [0,1] that has mean μ_ν. Then $\pi(1,N) > \mu_\nu$, and $\pi(M,N)$ does not depend on N and is strictly increasing in M.

Proof of Proposition 4: Since the bias of the coin, θ, is drawn from the distribution ν, the probability of k heads out of N flips is $f(k) \equiv \int_0^1 \binom{N}{k} \theta^k (1-\theta)^{N-k} d\nu(\theta)$. If we get all heads when we randomly draw M flips from a fixed sample of N flips, then the probability that the next flip is a head, $\pi(M,N)$, is $\frac{P(\text{draw } M+1 \text{ heads})}{P(\text{draw } M \text{ heads})} = \frac{\sum_{k=M+1}^{N} P(\text{draw } M+1 \text{ heads}|\text{fixed sample has } k \text{ heads}) P(\text{fixed sample has } k \text{ heads})}{\sum_{k=M}^{N} P(\text{draw } M \text{ heads}|\text{fixed sample has } k \text{ heads}) P(\text{fixed sample has } k \text{ heads})}$. Hence, $\pi(M,N) = \frac{\sum_{k=M+1}^{N} \left[\frac{k}{N}\cdot\frac{k-1}{N-1}\cdots\frac{k-M+1}{N-M+1}\cdot\frac{k-M}{N-M}\right] f(k)}{\sum_{k=M}^{N} \left[\frac{k}{N}\cdot\frac{k-1}{N-1}\cdots\frac{k-M+1}{N-M+1}\right] f(k)}$.

We now show that $\pi(M,N)$ can be expressed as

$$\pi(M,N) = \frac{\int_0^1 \theta^{M+1} d\nu(\theta)}{\int_0^1 \theta^M d\nu(\theta)}, \tag{A1}$$

from which it follows that $\pi(M,N)$ does not depend on N. To derive equation (A1), notice that the k^{th} summand in the numerator of the above expression for $\pi(M,N)$ can be written as

$$\left[\frac{k}{N}\cdot\frac{k-1}{N-1}\cdots\frac{k-M+1}{N-M+1}\cdot\frac{k-M}{N-M}\right] f(k)$$
$$= \left[\frac{k}{N}\cdot\frac{k-1}{N-1}\cdots\frac{k-M+1}{N-M+1}\cdot\frac{k-M}{N-M}\right] \frac{N!}{k!\,(N-k)!} \int_0^1 \theta^k (1-\theta)^{N-k} d\nu(\theta)$$

$$= \frac{(N-M-1)!}{(k-M-1)!\,(N-k)!}\int_0^1 \theta^k(1-\theta)^{N-k}d\nu(\theta).$$

Therefore the summation in the numerator can be written as

$$\begin{aligned}
&\sum_{k=M+1}^{N} \frac{(N-M-1)!}{(k-M-1)!\,(N-k)!}\int_0^1 \theta^k(1-\theta)^{N-k}d\nu(\theta) \\
&\quad = \int_0^1 \sum_{k=M+1}^{N} \frac{(N-M-1)!}{(k-M-1)!\,(N-k)!}\theta^k(1-\theta)^{N-k}d\nu(\theta) \\
&\quad = \int_0^1 \theta^{M+1} \sum_{k=M+1}^{N} \frac{(N-M-1)!}{(k-M-1)!\,(N-k)!}\theta^{k-M-1}(1-\theta)^{N-k}d\nu(\theta) \\
&\quad = \int_0^1 \theta^{M+1} \sum_{\vartheta=0}^{N-M-1} \frac{(N-M-1)!}{\vartheta!\,(N-M-1-\vartheta)!}\theta^{\vartheta}(1-\theta)^{N-M-1-\vartheta}d\nu(\theta) \\
&\quad = \int_0^1 \theta^{M+1}d\nu(\theta),
\end{aligned}$$

where the 3rd equality follows from a change in variables $\vartheta \equiv k-M-1$. An analogous argument establishes the denominator of equation (A1).

Next, we show that $\pi(1,N) > \mu_\nu$. Define V as the random variable that has the distribution ν. By equation (A1),

$$\pi(1,N) = \frac{\mathrm{E}[V^2]}{\mathrm{E}[V]} = \frac{\mathrm{E}[V]^2 + \mathrm{Var}[V]}{\mathrm{E}[V]} = \mathrm{E}[V] + \frac{\mathrm{Var}[V]}{\mathrm{E}[V]} = \mu_\nu + \frac{\mathrm{Var}[V]}{\mu_\nu}.$$

Since ν is nondegenerate, $\mathrm{Var}[V] > 0$, and therefore $\pi(1,N) > \mu_\nu$.

Finally, we show that $\pi(M,N)$ is increasing in M. To do so, we begin by defining

$$z(\theta) \equiv \frac{\theta^{M+1}\nu(\theta)}{(M+1)\int_0^1 x^M d\nu(x)}.$$

Taking the differential of both sides,

$$dz(\theta) = \frac{\theta^M d\nu(\theta)}{\int_0^1 x^M d\nu(x)}.$$

Because $dz(\theta)$ integrates to 1 over [0,1], we can interpret z as a distribution for some random variable, which we will call Z. Now, by (A1), observe that

$$\begin{aligned}
\frac{\pi(M+1,N)}{\pi(M,N)} &= \frac{\int_0^1 \theta^{M+2}d\nu(\theta)}{\int_0^1 \theta^{M+1}d\nu(\theta)} \cdot \frac{\int_0^1 \theta^{M}d\nu(\theta)}{\int_0^1 \theta^{M+1}d\nu(\theta)} \\
&= \frac{\int_0^1 \theta^{2}dz(\theta)}{\int_0^1 \theta^{1}dz(\theta)} \cdot \frac{\int_0^1 \theta^{0}dz(\theta)}{\int_0^1 \theta^{1}dz(\theta)}
\end{aligned}$$

$$= \frac{\int_0^1 \theta^2 dz(\theta)}{\left[\int_0^1 \theta dz(\theta)\right]^2}$$
$$= \frac{\mathrm{E}[Z^2]}{\mathrm{E}[Z]^2}$$
$$= \frac{\mathrm{E}[Z]^2 + \mathrm{Var}[Z]}{\mathrm{E}[Z]^2} > 1.$$

The inequality follows because ν being nondegenerate implies that z is nondegenerate, and therefore $\mathrm{Var}[Z] > 0$. Hence, $\pi(M+1, N) > \pi(M, N)$. ∎

Appendix F: Robustness and Other Additional Results

(Results from Section II of the Analysis Plan in Appendix B)

Contents

Note that Section II of the analysis plan for Experiment 2 (reproduced in Appendix B) lists a number of robustness and additional planned analyses. Many of these are discussed in Section 7 of the paper. Here we discuss the results of the remaining planned analyses, all of which pertain only to Experiment 2.

A. Self-doubt on "obvious" 50%-answer streak questions

When we asked subjects about the probability of a head after a streak of heads, we were concerned that they might know the correct answer of 50% but give a different answer because they wouldn't want to repeatedly answer 50% to many questions in a row. To address this concern, in our analysis plan, we had pre-specified the following robustness analysis: for each of the ten-flip, thousand-flip, and million-flip streak questions, we would re-run the analysis after excluding participants who answered 50% for the first four flips, and then gave responses other than 50% for at least one of the remaining questions. When we do so, our results remain qualitatively similar (and still consistent with GF). For full results, see Appendix G.II.B.e.

B. Individual Differences

We now turn to reporting the results of pre-specified hypotheses about individual differences, which we considered to be secondary hypotheses.

(a) Higher math skills are associated with less NBLLN.

We test the hypothesis that participants with higher math skills are less likely to exhibit NBLLN. We measure math skills using a six-question math quiz. Since 60% of respondents got all six questions correct, we split the sample into an all-six-correct group ("high math skills") and a less-than-six-correct group ("low math skills"). Our pre-specified analysis focused on the five-bin histogram for the ten-flip set and the eleven-bin histogram for the thousand-flip set. Across these two histograms, we find nearly identical results between those with high math skills as compared to those with low math skills. The one suggestive difference is that those with high math skills assigned higher probability to the middle bin in the eleven-bin histogram ($p = 0.03$). For more details, see Appendix G.II.C.a.

(b) Higher math skills are not associated with less GF.

We hypothesized that math skills were unrelated to GF. We compared the responses of the high vs. low math skill participants to the streak questions in the ten-, thousand-, and million-flip sets. We find essentially no evidence of any difference. Across the thirty individual responses we examined (the ten streak lengths from 0 to 9 across the three sample sizes), only one hypothesis test was significant at the 5% level: in the million-flip set, the high-math-skill respondents assessed the probability of a head after one head at 46% versus 42% for the low-math-skill respondents ($p = 0.002$). For detailed results, see Appendix G.II.C.b.

(c) Miscellaneous demographics

We test for GF differences based on sex and age in Experiment 2. The detailed results are in Appendix G.II.C.c; we summarize them briefly here.

For sex, we compared the responses of women vs. men to the streak questions in the ten-, thousand-, and million-flip sets. We find some evidence suggesting that GF may be stronger for women than for men. For example, in regressions of the probability assigned to heads on the number of prior heads, a male dummy, and an interaction between the two, the coefficient is positive in all three sample sizes and statistically suggestive (at $p = 0.057$) in the ten-flip sets.

For sex, we split age at the median: those younger than 21 years old (54% of the sample) and those 21+ (42%); note that 12 participants (4%) did not report their age. In analogous regressions as for sex, we do not find strong evidence about the relationship between GF and age. The coefficient on the interaction terms between number of prior heads and a dummy for 21+ years old is positive but insignificant across all three sample sizes.

C. Counterbalancing Variations

In the histogram questions for both experiments, we randomized whether participants saw the question framed as "number of heads" or "number of tails." We also randomized whether the participants saw the questions in increasing or decreasing order. In this section, we present results on whether these two randomizations had an effect on responses.

(a) Bias from “heads” vs. “tails”

For both experiments, we run an OLS regression of responses for each histogram question on a dummy variable indicating whether the participant was given the question about number of heads or number of tails and a constant. Across all questions in Experiment 1 and Experiment 2, the coefficient on the dummy variable is insignificant. For detailed results, see Appendix G.I.D.h.

(b) Bias from increasing vs. decreasing order of number of heads/tails

For both experiments, we run an OLS regression of responses for each histogram question on a dummy variable indicating whether the participant was given the number of heads/tail in increasing or decreasing order and a constant. For the following questions, the coefficient on the dummy variable was significant at the 5% level: Experiment 1 ten-flip 3-bin ($p = 0.004$), Experiment 1 thousand-flip 3-bin (middle bin = 451-549) ($p = 0.011$), Experiment 2 million-flip 11-bin ($p = 0.046$), and Experiment 2 million-flip 3-bin ($p = 0.033$). Across the 11 other histogram questions in total across the two experiments, we do not find significant coefficients on the dummy variable. For detailed results, see Appendix G.I.D.i.

Appendix G: Analysis Log

(Organized in accordance with the Analysis Plan in Appendix B)

Table of Contents

I.A. 10-flip sets

I.A.1. Betting questions (10 Questions)

In question q = 1,…,10 (in a random order), flips 1, 2,…, q-1 are shown (and determined randomly), and the participant must make a bet about whether the q^{th} flip is H or T. The participant is told that we will pick one sequence at random from the million sets (matching the sequence shown), and they will be paid if the sequence we draw matches what they picked. If the participant picks H, then the payoff is x ϵ {45, 47, 49, 51, 53, 55} lottery tickets (the value of x is randomly chosen), and if the participant picks T, then the payoff is the symmetric (around 50) number of lottery tickets.

(a) Testing gambler's fallacy (GF): How the frequency of predicting H varies with the prior number of heads.

i. Frequency of predicting H (Experiment 2).

# of prior Heads	***q***									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	58.12*** (2.82) [n=308]	59.06** (4.04) [n=149]	63.44*** (5.02) [n=93]	72.09*** (6.92) [n=43]	56.76 (8.26) [n=37]	57.14 (20.20) [n=7]	88.89*** (11.11) [n=9]	60.00 (24.49) [n=5]	- [n=0]	- [n=0]
1		49.06 (4.00) [n=159]	46.21 (4.15) [n=145]	49.50 (5.00) [n=101]	52.38 (6.34) [n=63]	69.77*** (7.09) [n=43]	62.50 (8.70) [n=32]	61.54 (14.04) [n=13]	44.44 (17.57) [n=9]	40.00 (16.33) [n=10]
2			37.14** (5.82) [n=70]	42.24** (4.61) [n=116]	52.68 (4.74) [n=112]	46.08 (4.96) [n=102]	55.56 (6.31) [n=63]	51.95 (5.73) [n=77]	51.52 (8.83) [n=33]	56.00 (10.13) [n=25]
3				33.33* (6.88) [n=48]	46.25 (5.61) [n=80]	48.51 (5.00) [n=101]	46.79 (4.80) [n=109]	60.71 (5.36) [n=84]	63.49 (6.11) [n=63]	63.79** (6.37) [n=58]
4					43.75 (12.81) [n=16]	51.22 (7.90) [n=41]	45.59 (6.08) [n=68]	42.65 (6.04) [n=68]	58.75 (5.54) [n=80]	50.72 (6.06) [n=69]
5						42.86 (13.73) [n=14]	54.17 (10.39) [n=24]	54.17 (7.27) [n=48]	45.21 (5.87) [n=73]	45.21 (5.87) [n=73]
6							0 (0) [n=3]	60.00 (16.33) [n=10]	51.16 (7.71) [n=43]	45.28 (6.90) [n=53]
7								33.33 (33.33) [n=3]	42.86 (20.20) [n=7]	27.78* (10.86) [n=18]
8									- [n=0]	50.00 (50.00) [n=2]
9										- [n=0]

Notes: Table shows mean frequency (%) of predicting H for each *q* and each number of prior Heads. Standard errors are in parentheses, and sample size in brackets. Standard

errors and significance levels are from a one-sample *t*-test of difference from 50%. *** Significant at 1% significance level; ** 5% significance level; * 10% significance level.

Ex post follow-up analyses: Why is the frequency of betting on H 58% for 0 prior H's? This might be because of imperfect randomization of payoffs. For example, for 0 prior Heads, 54.5% (355 out of 651) received higher payoffs for betting on Heads. More specifically:

Payoffs for betting on H	Number of occurrences	Frequency of betting on H
45	90	24.4%
47	103	35.0%
49	103	33.0%
51	103	80.6%
53	139	86.3%
55	113	86.7%

When there were 0 prior Heads in the first question the participant saw (q = 1), 52.6% (162 out of 308) received higher payoffs for betting on Heads.

Payoffs for betting on H	Number of occurrences	Frequency of betting on H
45	40	15.0%
47	53	37.7%
49	53	30.2%
51	52	76.9%
53	46	89.1%
55	64	87.5%

For 0 prior Heads and q = 2, 53.0% (79 out of 149) received higher payoffs for betting on Heads.

Payoffs for betting on H	Number of occurrences	Frequency of betting on H
45	26	30.8%
47	27	29.6%
49	17	23.5%
51	19	68.4%
53	37	89.2%
55	23	95.7%

For 0 prior Heads and q = 3, 54.84% (51 out of 93) received higher payoffs for betting on Heads.

Payoffs for betting on H	Number of occurrences	Frequency of betting on H
45	11	36.4%
47	12	33.3%
49	19	42.1%
51	19	94.7%
53	21	85.7%
55	11	63.6%

For 0 prior Heads and q = 4, 65.1% (28 out of 43) received higher payoffs for betting on Heads.

Payoffs for betting on H	Number of occurrences	Frequency of betting on H
45	9	44.4%
47	1	100%
49	5	20%
51	4	100%
53	21	85.7%
55	3	100%

For 0 prior Heads and q = 5, 45.9% (17 out of 37) received higher payoffs for betting on Heads.

Payoffs for betting on H	Number of occurrences	Frequency of betting on H
45	2	0%
47	10	30%
49	8	50%
51	4	100%
53	7	71.4%
55	6	83.3%

For 0 prior Heads and q = 6, 65.12% (28 out of 43) received higher payoffs for betting on Heads.

Payoffs for betting on H	Number of occurrences	Frequency of betting on H
45	1	0%
47	0	-
49	0	-
51	3	66.7%
53	0	-
55	3	66.7%

For 0 prior Heads and q = 7, 65.12% (28 out of 43) received higher payoffs for betting on Heads.

Payoffs for betting on H	Number of occurrences	Frequency of betting on H
45	1	0%
47	0	-
49	1	100%
51	2	100%
53	2	100%
55	3	100%

For 0 prior Heads and q = 8, 65.12% (28 out of 43) received higher payoffs for betting on Heads.

Payoffs for betting on H	Number of occurrences	Frequency of betting on H
45	0	-
47	0	-

49	0	-
51	0	-
53	5	60%
55	0	-

(b) How the frequency of predicting H varies with the payoff of H: Including payoffs as regressors in the model from (a) above.

i. Linear regression (linear probability model) of prediction on number of prior H's (Experiment 2).

	(a)	**(b)**	**(c)**
# Heads - # Tails	-0.0229*** (0.0051)	-0.0289*** (0.0049)	-0.0287*** (0.0049)
Payoff of H (relative to 50)		0.0688*** (0.0035)	0.0688*** (0.0035)
Dummy for (#H - #T)=-1, i.e., "middle-bin effect"			0.0073 (0.0219)
Constant	0.5168*** (0.0104)	0.5197*** (0.0095)	0.5185*** (0.0101)
Obs	3080	3080	3080

Notes: Payoff of H is the payoff given for betting on Heads relative to 50 lottery tickets, and takes on values {-5, -3, -1, 1, 3, 5}. Clustered standard errors in parentheses. *** Significant at 1% significance level; ** 5% significance level; * 10% significance level.

More flexible dummy specification:

	(a)	**(b)**
# Heads - # Tails (base = 0)		
-7	-0.0678 (0.1313)	0.0091 (0.1353)
-6	0.1322 (0.1123)	0.1136 (0.1073)
-5	0.0433 (0.0754)	0.0439 (0.0599)
-4	0.0341 (0.0533)	0.0530 (0.0458)
-3	0.0900** (0.0393)	0.1276*** (0.0335)
-2	0.0577* (0.0338)	0.0648** (0.0309)
-1	0.0022 (0.0291)	0.0183 (0.0246)
1	-0.0741***	-0.0561**

	(0.0287)	(0.0235)
2	-0.0981*** (0.0337)	-0.1087*** (0.0308)
3	-0.0766* (0.0425)	-0.0986*** (0.0370)
4	-0.0285 (0.0569)	-0.0761 (0.0514)
5	-0.1297* (0.0732)	-0.1915** (0.0774)
6	-0.2345 (0.1467)	-0.4546*** (0.1495)
7	-0.1345 (0.2224)	-0.1401 (0.2544)
Payoff of H (base = 49)		
45		-0.0540** (0.0270)
47		-0.0182 (0.0280)
51		0.4073*** (0.0378)
53		0.4852*** (0.0340)
55		0.5271*** (0.0346)
Constant	0.5345*** (0.0190)	0.3102*** (0.0346)
Obs	3080	3080

Regression using Closest-is-H dummies (Experiment 2)

	(1)	(2)
Closest flip is H	-0.1190** (0.0190)	-0.1124** (0.0168)
2nd closest flip is H	-0.0728** (0.0206)	-0.0719** (0.0183)
3rd closest flip is H	-0.0080 (0.0194)	-0.0041 (0.0176)
4th closest flip is H	0.0109 (0.0205)	0.0091 (0.0183)
5th closest flip is H	0.0217 (0.0228)	0.0055 (0.0203)
6th closest flip is H	-0.0243 (0.0245)	-0.0226 (0.0221)
7th closest flip is H	0.0538 (0.0290)	0.0372 (0.0255)

8th closest flip is H	0.0186 (0.0325)	0.0095 (0.0278)
9th closest flip is H	-0.1020* (0.0412)	-0.0616 (0.0376)
Prize difference = -6		0.0273 (0.0271)
Prize difference = -2		0.0536* (0.0271)
Prize difference = 2		0.4451** (0.0370)
Prize difference = 6		0.5269** (0.0340)
Prize difference = 10		0.5643** (0.0334)
Constant	0.5962** (0.0170)	0.3319** (0.0273)
Obs	3080	3080

Note: Linear probability regressions. Standard errors are clustered by participant and are in parentheses. The omitted category for prize difference is -10 cents. * $p < 0.05$, ** $p < 0.01$

ii. Linear regression (linear probability model) of prediction on number of prior H's (Experiment 1).

	(1)	(2)	(3)
# Heads - # Tails	-0.0010	-0.0099	-0.0142*
	(0.0076)	(0.0076)	(0.0076)
Payoff of H		0.0250***	0.0248***
(relative to 50)		(0.0061)	(0.0060)
Dummy for (#H -			-0.0660*
#T)=-1, i.e. "middle-			(0.0360)
bin effect"			
Constant	0.5096***	-0.7420**	-0.7107**
	(0.0160)	(0.3016)	(0.3025)
# Obs	987	987	987

Note: "Payoff of H" is the payoff given for betting on heads relative to 50 lotteries, and takes on values {-5, -3, -1, 1, 3, 5}. Clustered standard errors in parentheses.
* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$.

Regression using Closest-is-H dummies (Experiment 1)

	(1) Full Sample	(2) Prize Diff.=±2	(3) Prize Diff.=±6	(4) Prize Diff.=±10
Closest Flip is H	-0.0095	-0.0119	-0.0047	-0.0171
	(0.0326)	(0.0527)	(0.0552)	(0.0555)
2nd Closest Flip is H	-0.0755*	-0.1293*	-0.0553	-0.0407
	(0.0319)	(0.0546)	(0.0586)	(0.0639)
3rd Closest Flip is H	-0.0196	-0.0838	0.0000	0.0322
	(0.0360)	(0.0556)	(0.0544)	(0.0631)
4th Closest Flip is H	-0.0460	-0.0925	-0.1181*	0.0720
	(0.0329)	(0.0561)	(0.0548)	(0.0607)
5th Closest Flip is H	0.0435	-0.0690	0.0583	0.1391*
	(0.0293)	(0.0541)	(0.0568)	(0.0581)
Target-later		0.0480	0.0870	0.0337
		(0.0520)	(0.0538)	(0.0598)
Head Option on Top		-0.0144	0.0528	0.0586
		(0.0562)	(0.0470)	(0.0501)
Constant	0.5624**	0.6830**	0.4880**	0.3776**
	(0.0383)	(0.0699)	(0.0846)	(0.0806)
Number of choices	987	347	342	298

(c) Consistency of GF across questions: Comparing people's predictions to the Rabin-Vayanos model as fitted on data from I.A.3 (streak questions): Both with and without the payoff information included as an input to the model's predictions (Experiment 2).

Recall that R-V model is as follows:

$$q_t = \omega - \alpha \sum_{k=0}^{\infty} \delta^{k+1} y_{t-1-k}$$

Where q_t represents the agent's belief regarding the t^{th} flip; y_{t-1-k} represents the outcome of the $(t-1-k)^{th}$ flip; ω parameterizes the bias of the coin; α parameterizes the magnitude of the GF effect on the next flip; and δ parameterizes the rate of decay of GF. The outcome variable y is equal to +1 if Head and -1 if Tail. The agent's perceived probability that the t^{th} flip will be a head is $p_t = \frac{q_t+1}{2}$.

An R-V agent bets on a head if:

$$\text{payoff(Heads)} * p_t > \text{payoff(Tails)} * (1 - p_t)$$

With α = 0.1601187, δ = 0.6211256.

Participants' responses coincided with what an R-V agent would have done 65.3% of the time (2011 out of 3080), and the error rate was 34.7%. The standard deviation of a

binary variable that equals 1 if participants' bets coincide with what an R-V agent would do is 0.476.

The table below reports error rates by payoff for Heads, and as expected, the error rates are lower when payoffs are more extreme.

Payoff for Heads	**Error rate**
45	33.9%
47	37.5%
49	39.0%
51	39.9%
53	31.9%
55	25.9%
Overall	**34.7%**

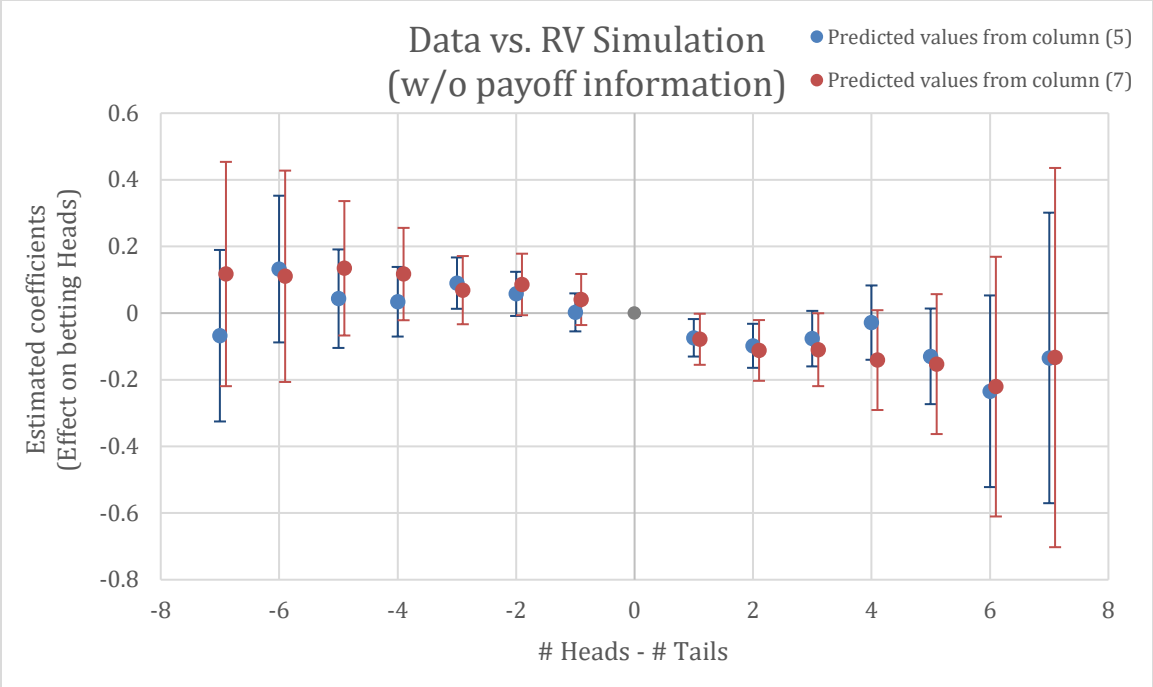
Next, we simulate 1000 datasets by taking what the R-V agent would have done when faced with the participants' options, and we flip the choice according to the respective error rates by payoffs (e.g., when the payoff for Heads is 47, we flip the choice 37.5% of the time). For each simulated dataset, we re-run the regression. Mean coefficients are reported below, and standard errors are in parentheses. The standard errors are the square-root of the sum of two terms: mean of the squared standard errors across the 1000 simulations, and the variance of the estimated coefficient across the 1000 simulations.

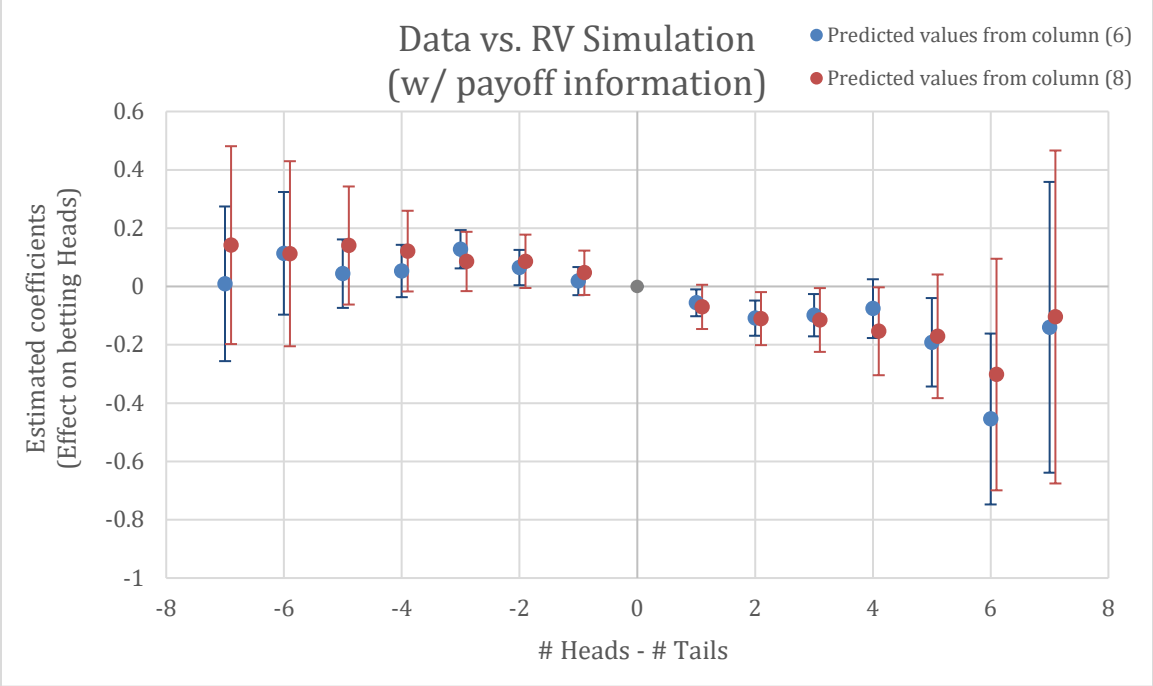
	(1) **Full Sample**	**(2)** **Full Sample**	**(3)** **RV Simulation**	**(4)** **RV Simulation**
# Heads - # Tails	-0.0229***	-0.0289***	-0.0344***	-0.0365***
	(0.0051)	(0.0049)	(0.0057)	(0.0057)
Payoff of H		0.0688***		0.0237***
(relative to 50)		(0.0035)		(0.0035)
Constant	0.5168***	0.5197***	0.5085***	0.5095***
	(0.0104)	(0.0095)	(0.0123)	(0.0123)
Obs	3080	3080	3080	3080

	(5) **Full Sample**	**(6)** **Full Sample**	**(7)** **RV Simulation**	**(8)** **RV Simulation**
# Heads - # Tails				
(base = 0)				
-7	-0.0678	0.0091	0.1173	0.1417
	(0.1313)	(0.1353)	(0.1717)	(0.1731)
-6	0.1322	0.1136	0.1105	0.1122
	(0.1123)	(0.1073)	(0.1618)	(0.1619)
-5	0.0433	0.0439	0.1345	0.1404
	(0.0754)	(0.0599)	(0.1029)	(0.1033)
-4	0.0341	0.0530	0.1172	0.1212*
	(0.0533)	(0.0458)	(0.0707)	(0.0706)
-3	0.0900**	0.1276***	0.0689	0.0857*

	(0.0393)	(0.0335)	(0.0522)	(0.0519)
-2	0.0577*	0.0648**	0.0860*	0.0862*
	(0.0338)	(0.0309)	(0.0471)	(0.0467)
-1	0.0022	0.0183	0.0408	0.0469
	(0.0291)	(0.0246)	(0.0391)	(0.0388)
1	-0.0741***	-0.0561**	-0.0785**	-0.0701*
	(0.0287)	(0.0235)	(0.0391)	(0.0387)
2	-0.0981***	-0.1087***	-0.1120**	-0.1103**
	(0.0337)	(0.0308)	(0.0465)	(0.0464)
3	-0.0766*	-0.0986***	-0.1098**	-0.1150**
	(0.0425)	(0.0370)	(0.0558)	(0.0558)
4	-0.0285	-0.0761	-0.1410*	-0.1537**
	(0.0569)	(0.0514)	(0.0764)	(0.0769)
5	-0.1297*	-0.1915**	-0.1530	-0.1710
	(0.0732)	(0.0774)	(0.1071)	(0.1082)
6	-0.2345	-0.4546***	-0.2206	-0.3021
	(0.1467)	(0.1495)	(0.1988)	(0.2026)
7	-0.1345	-0.1401	-0.1334	-0.1047
	(0.2224)	(0.2544)	(0.2903)	(0.2913)
Payoff of H (base = 49)				
45		-0.0540**		-0.0694
		(0.0270)		(0.0423)
47		-0.0182		-0.0254
		(0.0280)		(0.0428)
51		0.4073***		0.0522
		(0.0378)		(0.0434)
53		0.4852***		0.1052**
		(0.0340)		(0.0423)
55		0.5271***		0.1736***
		(0.0346)		(0.0421)
Constant	0.5345***	0.3102***	0.5224***	0.4814***
	(0.0190)	(0.0346)	(0.0250)	(0.0371)
Obs	3080	3080	3080	3080

Below are graphs comparing the regression results (error bars indicate 95% confidence intervals).





(d) Consistency of GF across questions (Experiment 1)

In this experiment, the estimated Rabin-Vayanos parameters were:

$$\alpha = 0.031, \delta = 0.947.$$

The table below reports error rates by payoff for Heads:

Payoff for Heads	**Error rate**

45	42.8%
47	38.9%
49	46.7%
51	43.3%
53	40.6%
55	42.5%
Overall	**42.5%**

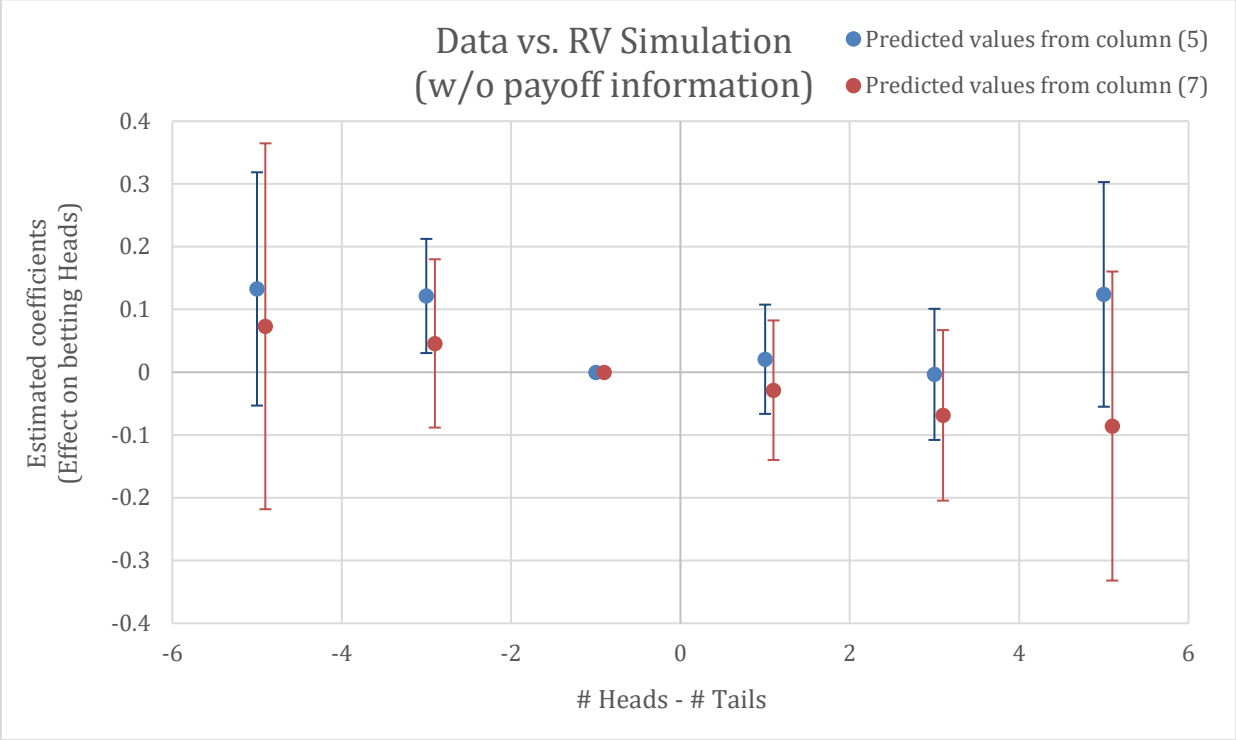
Next, we simulate 1000 datasets by taking what the R-V agent would have done when faced with the participants' options, with the choice flipping according to the respective error rates by payoffs (e.g. when the payoff for Heads is 47, the choice flips 38.9% of the time). Mean coefficients are reported below, and standard errors are in parentheses. The standard errors are square-root of the sum of two terms: mean of the squared standard errors across the 1000 simulations; and the variance of the estimated coefficient across the 1000 simulations.

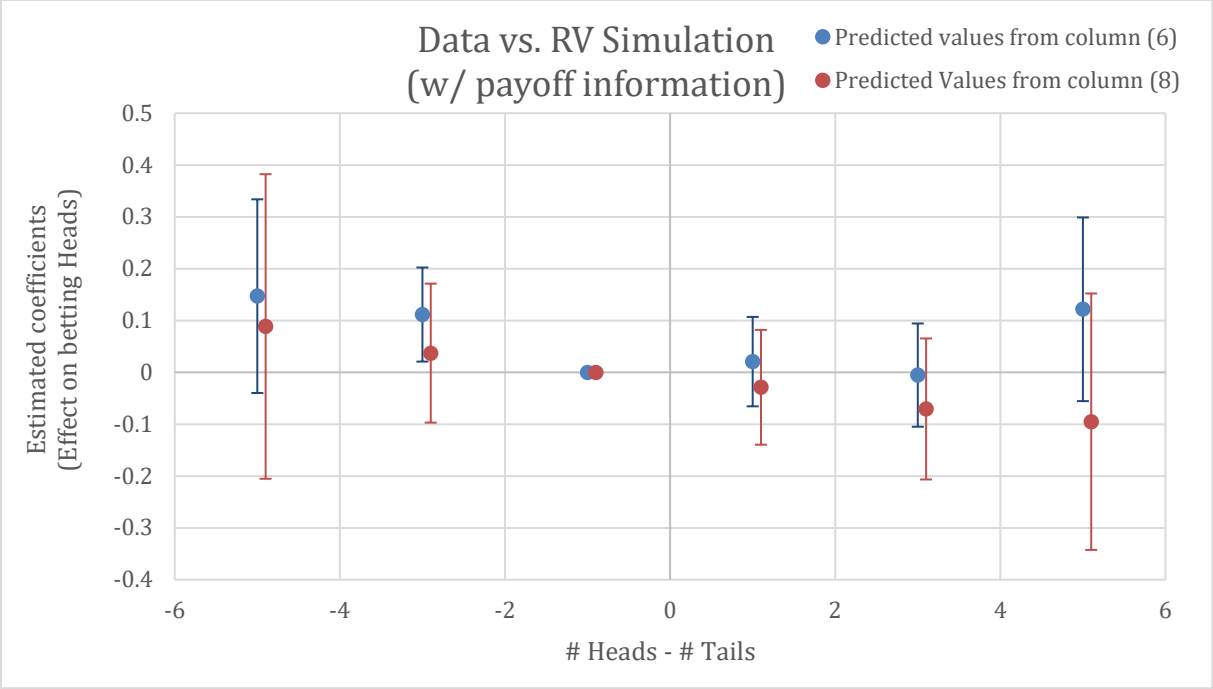
	(1)	**(2)**	**(3)**	**(4)**
	Full sample	**Full sample**	**RV Simulation**	**RV Simulation**
# Heads - # Tails	-0.0100	-0.0099	-0.0175	-0.0174
	(0.0076)	(0.0076)	(0.0102)	(0.0102)
Payoff of H		0.0250		0.0151
(relative to 50)		(0.0061)		(0.0067)
Constant	0.5096	0.5077	0.4990	0.4978
	(0.0160)	(0.0157)	(0.0221)	(0.0220)
Obs	987	987	987	987

	(5)	**(6)**	**(7)**	**(8)**
	Full sample	**Full sample**	**RV Simulation**	**RV Simulation**
# Heads - # Tails (base = -1)				
-5	0.1327	0.1470	0.0732	0.0885
	(0.0948)	(0.0953)	(0.1487)	(0.1499)
-3	0.1215	0.1115	0.0459	0.0371
	(0.0464)	(0.0463)	(0.0684)	(0.0684)
1	0.0207	0.0207	-0.0286	-0.0288
	(0.0444)	(0.0440)	(0.0567)	(0.0565)
3	-0.0034	-0.0053	-0.0687	-0.0706
	(0.0533)	(0.0508)	(0.0693)	(0.0694)
5	0.1240	0.1217	-0.0857	-0.0952
	(0.0913)	(0.0904)	(0.1256)	(0.1262)
Payoff of H (base = 49)				
45		0.0042		-0.0674
		(0.0542)		(0.0806)
47		0.0001		-0.0707
		(0.0516)		(0.0762)

51		0.1897		0.0078
		(0.0665)		(0.0756)
53		0.1733		0.0528
		(0.0604)		(0.0754)
55		0.1973		0.0670
		(0.0609)		(0.0785)
Constant	0.4760	0.3812	0.5118	0.5142
	(0.0282)	(0.0390)	(0.0398)	(0.0635)
Obs	987	987	987	987

Below are graphs comparing the regression results (error bars indicate 95% confidence intervals).





I.A.2. Histogram questions (14 Questions)

Experiment 1 Results

Bins (A)	True	Mean	Bins (B)	True	Mean	Bins (C)	True	Mean
0	0.1%	6.1%	0-3	17.2%	18.3%	0-4	37.7%	33.9%
1	1.0%	6.4%	4	20.5%	21.5%	5	24.6%	36.2%
2	4.4%	8.0%	5	24.6%	28.1%	6-10	37.7%	29.9%
3	11.7%	9.0%	6	20.5%	18.3%			
4	20.5%	12.3%	7-10	17.2%	13.8%			
5	24.6%	20.0%						
6	20.5%	12.7%						
7	11.7%	8.9%						
8	4.4%	7.3%						
9	1.0%	6.5%						
10	0.1%	2.7%						

Experiment 2 Results

Bins (A)	True	Mean	Bins (B)	True	Mean	Bins (C)	True	Mean
0	0.1%	2.2%	0-3	17.2%	15.9%	0-4	37.7%	34.0%
1	1.0%	3.8%	4	20.5%	18.3%	5	24.6%	32.9%
2	4.4%	5.5%	5	24.6%	32.1%	6-10	37.7%	33.2%
3	11.7%	9.3%	6	20.5%	18.1%			
4	20.5%	15.1%	7-10	17.2%	15.6%			
5	24.6%	28.3%						

6	20.5%	14.9%
7	11.7%	9.2%
8	4.4%	5.5%
9	1.0%	3.9%
10	0.1%	2.4%

(a) Are median responses significantly different from the truth?

i. Experiment 1

Figure 6

Bins	N = 10, True	N = 10, Median	*p* value
0-3	17.2%	15.0%	0.78
4	20.5%	20.0%	0.55
5	24.6%	25.0%	0.10
6	20.5%	18.0%	0.0001
7-10	17.2%	10.0%	0.0002

Figure 17

Bins	N = 1000, True	N = 1000, Median	*p* value
0-480	10.8%	33.0%	0.0000
481-519	78.3%	35.0%	0.0000
520-1000	10.8%	25.5%	0.0000

Figure 14

Bins	N = 1000, True	N = 1000, Median	*p* value
0-450	0.1%	33.0%	0.0000
451-549	99.8%	40.0%	0.0000
550-1000	0.1%	26.0%	0.0000

Figure 8

Bins	N = 10, True	N = 10, Median	*p* value
0	0.10%	5%	0.0000
1	0.98%	30%	0.0000
2	4.39%	30%	0.0000
3	11.72%	30%	0.0000
4	20.51%	29%	0.0001
5	24.61%	36%	0.0000
6	20.51%	36%	0.0000
7	11.72%	28%	0.0000
8	4.39%	20%	0.0000
9	0.98%	11%	0.0000
10	0.10%	1%	0.0000

ii. Experiment 2

Figure 10

Bins	True prob	Median estimate	*p* value
0-3	17.2%	13.8%	0.0000
4	20.5%	20.0%	0.0000
5	24.6%	30.0%	0.0000
6	20.5%	20.0%	0.0000
7-10	17.2%	13.0%	0.0000

Figure 12

Bins	True prob	Median estimate (separate)	*p* value
0	0.1%	1.0%	0.0000
1	1.0%	3.5%	0.0000
2	4.4%	9.0%	0.0000
3	11.7%	12.0%	0.0000
4	20.5%	23.0%	0.0000
5	24.6%	35.0%	0.0000
6	20.5%	20.5%	0.0000
7	11.7%	12.5%	0.0000
8	4.4%	8.4%	0.0000
9	1.0%	3.0%	0.0000
10	0.1%	1.0%	0.0000

Figure 16

Bins	True prob	Median estimate	*p* value
0-450	0.1%	30.0%	0.0000
451-549	99.8%	40.0%	0.0000
550-1000	0.1%	30.0%	0.0000

Figure 19

Bins	True prob	Median estimate	*p* value
0-480	11.0%	32.8%	0.0000
481-519	78.0%	34.0%	0.0000
520-1000	11.0%	33.0%	0.0000

Figure 24

Bins	True prob	Median estimate	*p* value
0-499	48.7%	46.0%	0.0000
500	2.5%	10.0%	0.0000
501-1000	48.7%	45.5%	0.0000

Figure 25

Bins	True prob	Median estimate	*p* value
0-493	34.1%	38.0%	0.0000
494-506	31.9%	25.0%	0.0000
507-1000	34.1%	35.0%	0.0000

Figure 26

Bins	True prob	Median estimate	*p* value
0-909	100.0%	90.9%	0.0000
910	0.0%	1.0%	0.0000
911-1000	0.0%	6.3%	0.0000

Figure 27

Bins	True prob	Median estimate	*p* value
0-487	21.5%	22.8%	0.0000
488-496	19.8%	11.0%	0.0000
497-503	17.5%	18.0%	0.11
504-512	19.8%	12.5%	0.0000
513-1000	21.5%	24.8%	0.0000

Figure 28

Bins	True prob	Median estimate	*p* value
0-499784	33.3%	33.4%	0.79
499785-500215	33.4%	33.0%	0.23
500216-1000000	33.3%	33.8%	0.47

Figure 29

Bins	True prob	Median estimate	*p* value
0-499579	20.1%	20.0%	0.0000
499580-499873	20.0%	12.0%	0.0000
499874-500126	19.9%	20.0%	0.0000
500127-500420	20.0%	12.0%	0.0000
500421-1000000	20.1%	20.0%	0.0000

(b) Test for bin effects.

i. Compare the mean probability assigned to 0-3 in (B) with the sum of the mean probabilities assigned to 0, 1, 2, 3 in (A); similarly, compare 7-10 with the sum of 7, 8, 9, 10 (Experiment 2).

Mean probability assigned to 0-3 in (B) = 15.9%
Sum of mean probabilities assigned to 0, 1, 2, 3 in (A) = 20.8%
(Two-tailed t-test p-value < 0.0001)

Mean probability assigned to 7-10 in (B) = 15.6%
Sum of mean probabilities assigned to 7, 8, 9, 10 in (A) = 20.9%

(p-value < 0.0001)

Follow-up analyses: Among people who saw that histogram first:
Mean probability assigned to 0-3 in (B) = 15.8% [Sample = 105 out of 308]
Sum of mean probabilities assigned to 0, 1, 2, 3 in (A) = 19.1% [Sample = 95 out of 308]
(p-value = 0.0423)

Mean probability assigned to 7-10 in (B) = 15.0% [Sample = 105 out of 308]
Sum of mean probabilities assigned to 7, 8, 9, 10 in (A) = 21.1% [Sample = 95 out of 308]
(p-value = 0.0017)

ii. Compare the mean probability assigned to 0-4 in (C) with the sum of the mean probabilities assigned to 0, 1, 2, 3, 4 in (A); similarly, compare 6-10 with the sum of 6, 7, 8, 9, 10 (Experiment 2).
Mean probability assigned to 0-4 in (C) = 34.0%
Sum of mean probabilities assigned to 0, 1, 2, 3, 4 in (A) = 35.9%
(p-value = 0.0086)

Mean probability assigned to 6-10 in (C) = 33.2%
Sum of mean probabilities assigned to 6, 7, 8, 9, 10 in (A) = 35.9%
(p-value = 0.0002)

Follow-up analyses: Among people who saw that histogram first:
Mean probability assigned to 0-4 in (C) = 31.4% [Sample = 108 out of 308]
Sum of mean probabilities assigned to 0, 1, 2, 3, 4 in (A) = 35.2% [Sample = 95 out of 308]
(p-value = 0.0123)

Mean probability assigned to 6-10 in (C) = 32.2% [Sample = 108 out of 308]
Sum of mean probabilities assigned to 6, 7, 8, 9, 10 in (A) = 36.8% [Sample = 95 out of 308]
(p-value = 0.0053)

iii. Compare the mean probability assigned to 0-4 in (C) with the sum of the mean probabilities assigned to 0-3, 4 in (B); similarly, compare 6-10 with the sum of 6, 7-10 (Experiment 2).
Mean probability assigned to 0-4 in (C) = 34.0%
Sum of mean probabilities assigned to 0-3, 4 in (B) = 34.2%
(p-value = 0.6774)

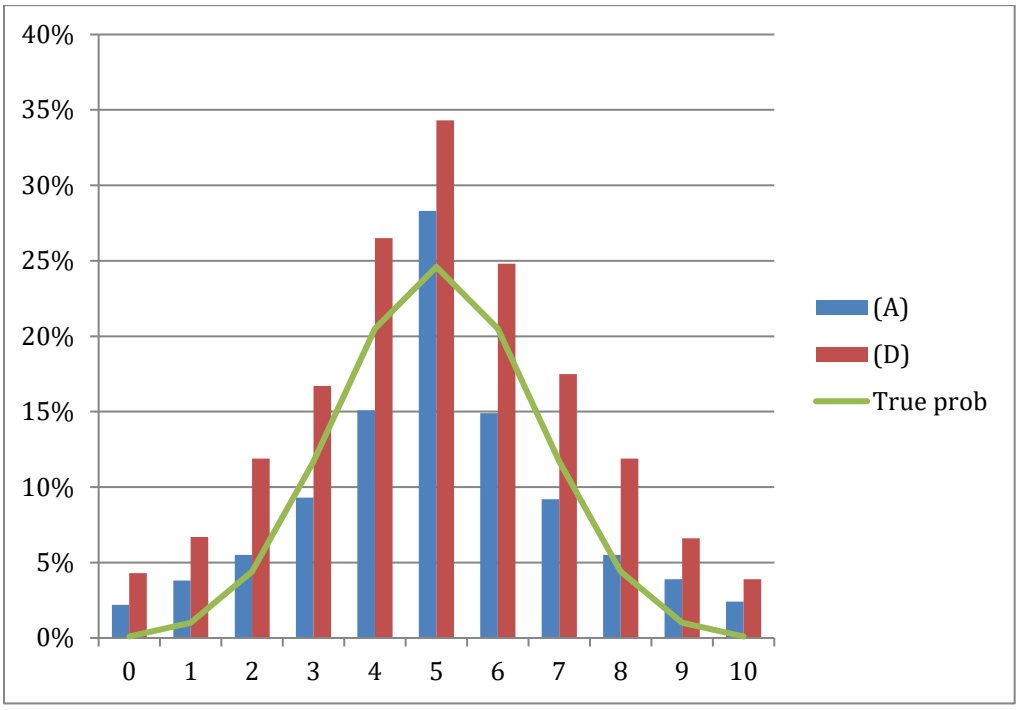
Mean probability assigned to 6-10 in (C) = 33.2%
Sum of mean probabilities assigned to 6, 7-10 in (B) = 33.7%
(p-value = 0.3913)

Follow-up analysis: Among people who saw that histogram first:
Mean probability assigned to 0-4 in (C) = 31.4% [Sample = 108 out of 308]
Sum of mean probabilities assigned to 0-3, 4 in (B) = 34.4% [Sample = 105 out of 308]
(p-value = 0.0523)

Mean probability assigned to 6-10 in (C) = 32.2% [Sample = 108 out of 308]
Sum of mean probabilities assigned to 6, 7-10 in (B) = 33.3% [Sample = 105 out of 308]
(p-value = 0.4504)

iv. Compare the mean probabilities from (D) to those from (A) (Experiment 2).

Bins	**True probability**	**(A)**	**(D)**	**p-value (A) vs. (D)**
0	0.1%	2.2%	4.3%	0.0008
1	1.0%	3.8%	6.7%	< 0.0001
2	4.4%	5.5%	11.9%	< 0.0001
3	11.7%	9.3%	16.7%	< 0.0001
4	20.5%	15.1%	26.5%	< 0.0001
5	24.6%	28.3%	34.3%	< 0.0001
6	20.5%	14.9%	24.8%	< 0.0001
7	11.7%	9.2%	17.5%	< 0.0001
8	4.4%	5.5%	11.9%	< 0.0001
9	1.0%	3.9%	6.6%	< 0.0001
10	0.1%	2.4%	3.9%	0.0043
Sum	100%	100%	165.2%	



v. Predicting 5 heads out of 10-flips by 2- to 11-bin elicitations

Experiment 1

Number of Bins	Mean beliefs of 5 heads
2-bin	39.45
3-bin	36.20
5-bin	28.12
11-bin	19.99

Pairwise *p*-values for equality of mean beliefs of 5 heads out of 10-flips

	2-bin	3-bin	5-bin
2-bin			
3-bin	0.3431		
5-bin	0.0012	0.0000	
11-bin	0.0000	0.0000	0.0000

Follow-up Analysis: How many people gave weakly increasing probabilities from 2- to 11-bin elicitations?

42 (40.38%) out of 104 participants gave weakly decreasing probabilities.
10 (9.62%) out of 104 participants gave weakly increasing probabilities.

Experiment 2

Number of Bins	Mean beliefs of 5 heads
2-bin	34.33
3-bin	32.87
5-bin	32.09
11-bin	28.27

Pair-wise *p*-values for equality of mean beliefs of 5 heads out of 10-flips

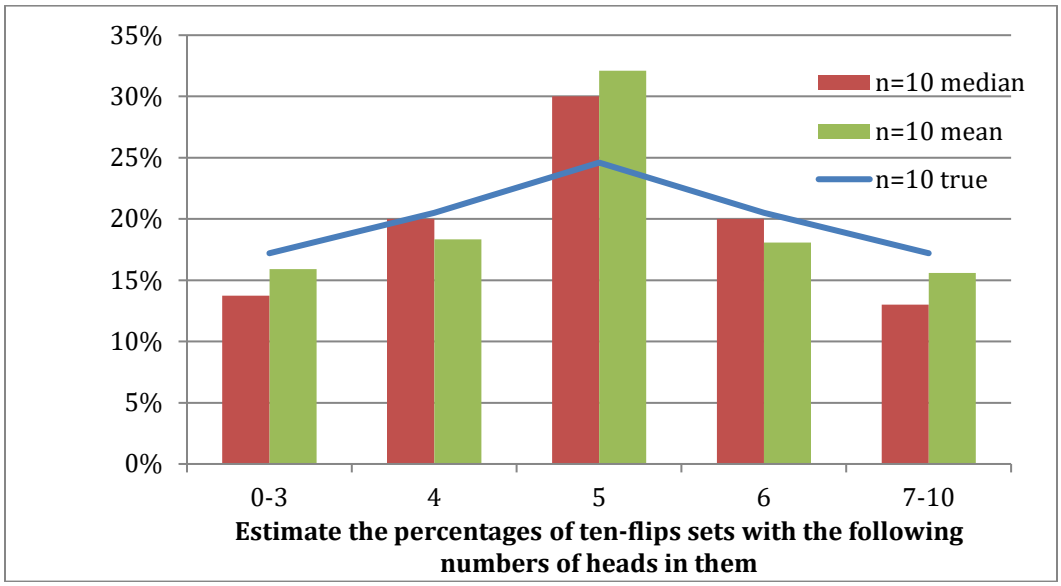
	2-bin	3-bin	5-bin
2-bin			
3-bin	0.2209		
5-bin	0.0605	0.2944	
11-bin	0.0000	0.0000	0.0000

Follow-up Analysis: How many people gave weakly increasing probabilities from 2- to 11-bin elicitations?

102 (33.12%) out of 308 participants gave weakly decreasing probabilities.
60 (19.48%) out of 308 participants gave weakly increasing probabilities.

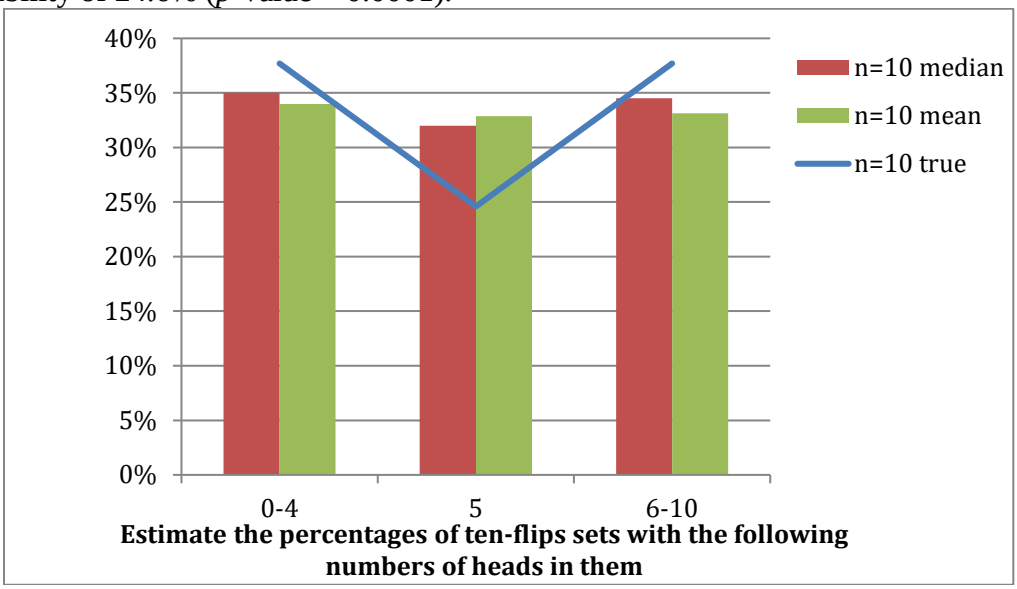
(c) Test for NBLLN in N = 10.

i. Examine whether (B) generates approximately equal mean elicited probabilities across the five bins (which is close to normatively correct) (Experiment 2).



Follow-up analyses: How many subjects gave approx equal weights to all the bins?
6.2% (19 out of 308) gave weight 20% to all the bins.
7.5% (23 out of 308) gave weights between 17-23% to all the bins.

ii. Examine whether (C) generates overweighting of 5 heads (whereas NBLLN + bin effects would predict underweighting of 5 heads) (Experiment 2).
Mean probability assigned to 5 was 32.9%, which is significantly larger than true probability of 24.6% (p-value < 0.0001).



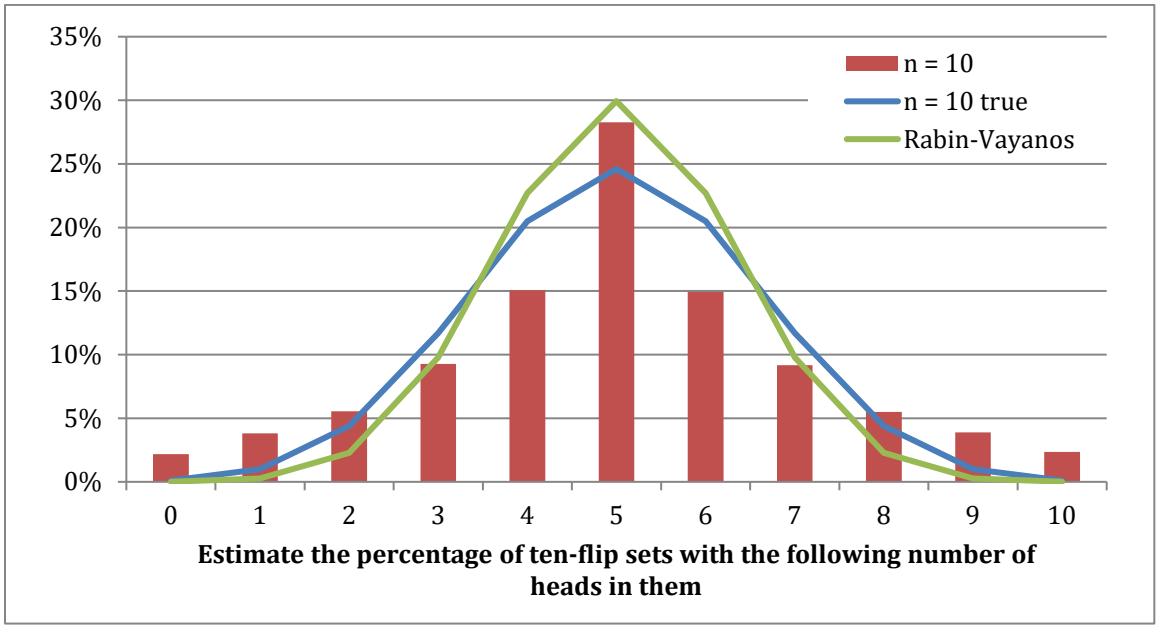
Follow-up analyses: How many subjects gave approx equal weights to all the bins?

4.9% (15 subjects out of 308) gave weights between 33-34% to all the bins.
12.0% (37 subjects out of 308) gave weights between 30-36% to all the bins.

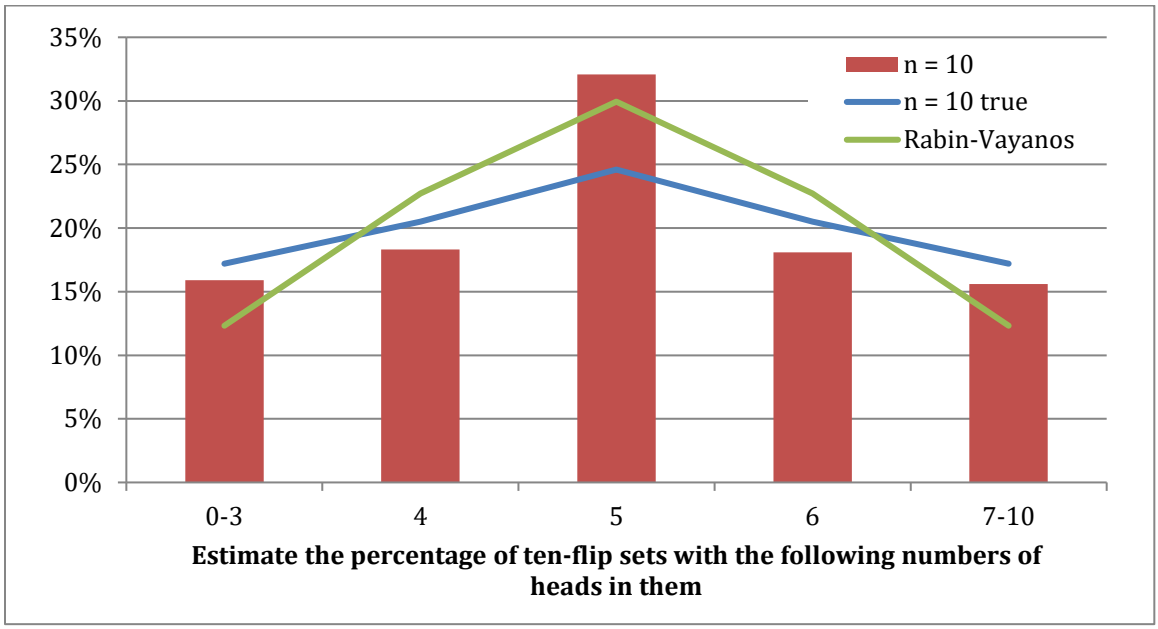
(d) Test for the presence of exact representativeness.

Examine whether the Rabin-Vayanos model estimated in I.A.1.d falls short of predicting respondents' predictions about 5 heads out of 10 flips (Experiment 2).

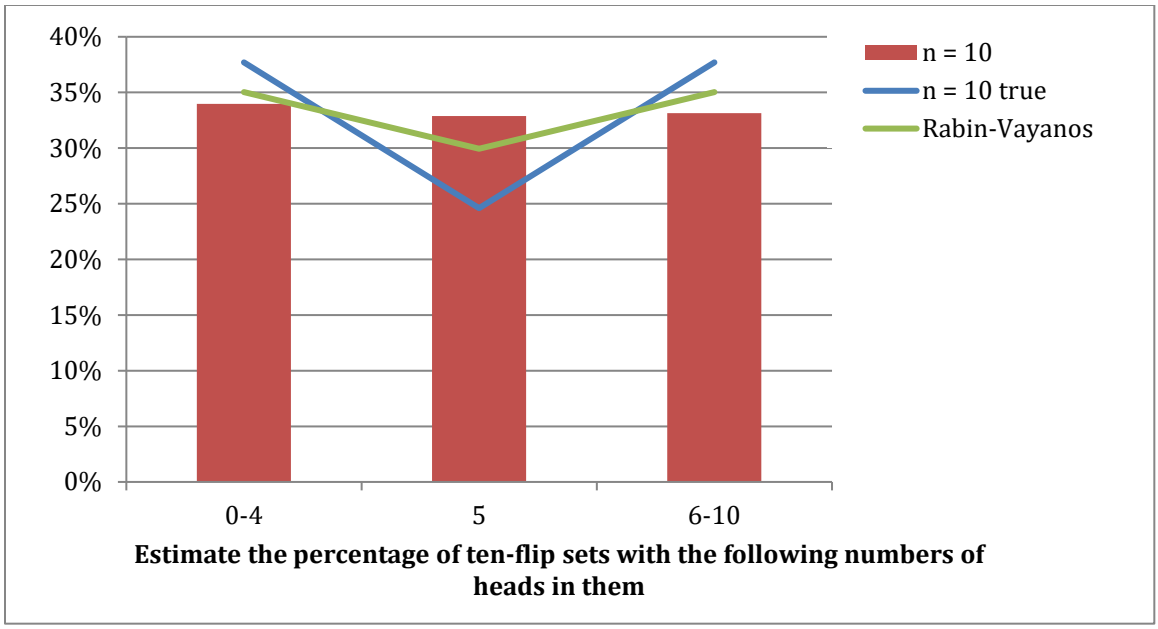
Comparing Rabin-Vayanos predictions to mean histogram responses:
11-bin:



5-bin:

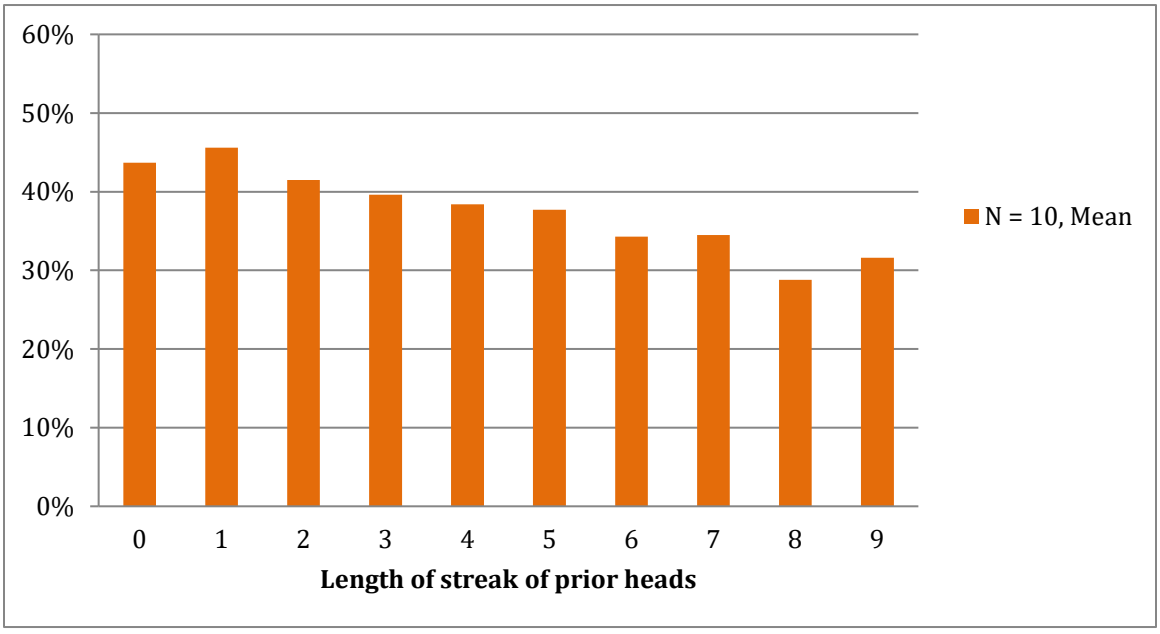


3-bin:

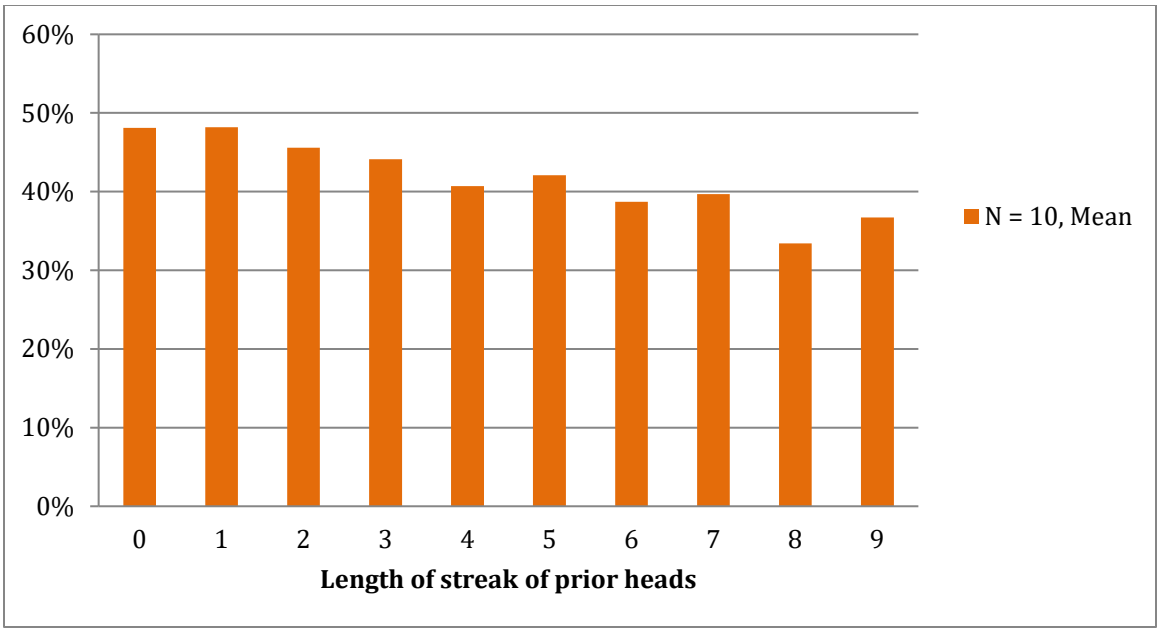


I.A.3. Streak questions (10 Questions)

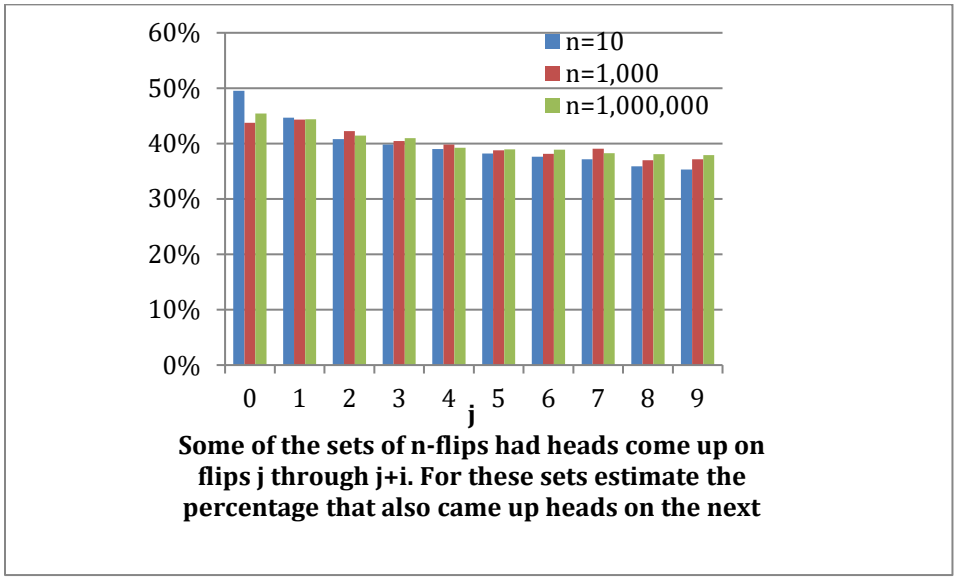
Experiment 1 Results



Follow-up analyses: Why is the elicited probability of heads less than 50% when there are zero prior heads? If we eliminate responses = 10% (when participants may have misunderstood the question), does this result disappear?



Experiment 2 Results



Follow-up analyses: Did some subjects switch away from 50%, perhaps feeling weird about entering 50% all the time?

The table below shows the number of subjects (out of 308) who entered "50%" for the first *x* streak questions:

x	**10-flips**	**1,000-flips**	**1,000,000-flips**
10	137	133	140
9	141	137	144
8	143	138	146
7	145	139	149

6	147	143	151
5	150	146	158
4	152	149	165
3	160	157	168
2	174	171	178
1	199	194	196

Misunderstanding the question about probability of a head on the first flip

As noted, mean responses to the streak-0 question ("unconditional probability") are less than 50%:

10-flips: 49.53%
1000-flips: 43.76%
Million-flips: 45.43%

If participants misunderstood the question, then we might expect to see less departure from 50% if the unconditional probability was elicited after some of the conditional probabilities were elicited (because having answered the conditional probability questions already might have made the unconditional probability question clearer due to the contrast).

Below, we report the mean reported unconditional probabilities for 10-flips depending on when it was asked:

Order	**Mean**	**Standard deviation**	**Observations**	**p-value (vs. First)**
First	47.75	1.49	32	-
Second	51.03	0.77	31	0.057
Third	47.51	1.48	39	0.911
Fourth	49.34	1.48	29	0.452
Fifth	50.66	0.70	29	0.093
Sixth	50.87	0.59	30	0.063
Seventh	48.59	1.70	27	0.709
Eighth	49.92	1.37	38	0.287
Ninth	48.70	0.98	27	0.610
Last	51.59	2.50	26	0.175

First (mean = 47.75) vs. all others (mean = 49.73): $p = 0.172$

Below, we report the mean reported unconditional probabilities for 1000-flips depending when (among the 10 streak conditional and unconditional probability questions) it was asked:

Order	**Mean**	**Standard deviation**	**Observations**	**p-value (vs. First)**
First	37.01	3.82	28	-
Second	43.77	2.98	35	0.0299
Third	44.25	4.58	29	0.1253
Fourth	42.97	3.32	31	0.0830

Fifth	41.03	3.21	29	0.2213
Sixth	49.94	0.18	32	<0.0001
Seventh	42.02	2.77	35	0.0796
Eighth	44.42	2.54	31	0.0067
Ninth	48.76	1.00	33	<0.0001
Last	42.00	3.37	25	0.1517

First (mean = 37.01) vs. all others (mean = 44.44): $p = 0.024$

0-streak responses for million-flips depending on when it was asked among the ten streak questions:

Order	**Mean**	**Standard deviation**	**Observations**	**p-value (vs. First)**
First	44.21	2.31	39	-
Second	45.19	3.58	31	0.7872
Third	47.86	1.97	30	0.0738
Fourth	44.82	3.13	27	0.8468
Fifth	49.07	1.31	29	0.0009
Sixth	40.05	3.59	29	0.2564
Seventh	49.23	0.48	30	<0.0001
Eighth	44.24	3.32	33	0.9923
Ninth	47.69	1.77	29	0.0591
Last	42.52	3.18	31	0.5985

First (mean = 44.21) vs. all others (mean = 45.61): $p = 0.583$

(a) Test for GF: How the predicted frequency of H varies with the number of prior H's.

i. t-test of each predicted frequency's difference from 50%.

Experiment 1

Number of prior H's	**Mean predicted freq of H (%)**	***t*-statistic (two-tailed *p*-value)**
0	43.7%	-2.8737 (0.0050)
1	45.6%	-1.9973 (0.0485)
2	41.5%	-3.9908 (0.0001)
3	39.6%	-4.7275 (<0.0001)
4	38.4%	-5.0910 (<0.0001)
5	37.7%	-5.5185 (<0.0001)
6	34.3%	-7.2766

		(<0.0001)
7	34.5%	-6.2625 (<0.0001)
8	28.8%	-9.0837 (<0.0001)
9	31.6%	-7.1778 (<0.0001)

Experiment 2

Number of prior H's	**Mean predicted freq of H (%)**	***t*-statistic (two-tailed *p*-value)**
0	49.53	-1.07 (0.2860)
1	44.65*	-6.95 (<0.0001)
2	40.83*	-9.15 (<0.0001)
3	39.82*	-9.79 (<0.0001)
4	39.01*	-10.16 (<0.0001)
5	38.18*	-10.47 (<0.0001)
6	37.60*	-10.32 (<0.0001)
7	37.15*	-10.67 (<0.0001)
8	35.90*	-11.15 (<0.0001)
9	35.32*	-11.04 (<0.0001)

Note: 35.32 is significantly different from 44.65 with p-value < 0.0001.
44.65 is significantly different from 49.53 with p-value < 0.0001.

ii. Linear regression of predicted frequencies on number of prior H's.

Experiment 1

	(a.i)
Number of prior H's	-1.69*** (0.250)
Constant	45.16 (1.336)
Obs	1010

Experiment 2

	(a.ii)
Number of prior H's	-1.30*** (0.119)
Constant	45.66 (0.634)
Obs	3080

(b) Test for GF at the individual level: Fraction of individuals who always report 50%, always report weakly less than 50% (at least one frequency strictly), always report weakly more than 50% (at least one frequency strictly).

Experiment 1
Fraction of individuals who always report 50%: 19/104 = 18.3%
Fraction of individuals who always report weakly less than 50% (at least one frequency strictly): 34/104 = 32.7%
Fraction of individuals who always report weakly more than 50% (at least one frequency strictly): 3/104 = 2.9%.

Experiment 2
Fraction of individuals who always report 50%: 137/308 = 44.5%
Fraction of individuals who always report weakly less than 50% (at least one frequency strictly): 113/308 = 36.7%
Fraction of individuals who always report weakly more than 50% (at least one frequency strictly): 0/308 = 0%.

(c) Calibrate GF: Estimate the Rabin-Vayanos model by non-linear least squares on the first-difference formulation of the model (Experiment 2).

	Estimates using individual-level data	**Estimates using aggregate data**
alpha	0.16*** (0.035) [0.090, 0.230]	0.16*** (0.031)
delta	0.62*** (0.056) [0.512, 0.730]	0.62*** (0.063)
Obs	2772	9

Note: Standard errors in parentheses (clustered for individual-level regression), and 95% confidence intervals in brackets. * Significant at 10% ** 5% *** 1%.

(d) Within-Subject Correlation between GF and NBLLN

Experiment 1

GF predicts that participants' beliefs about the frequency of a head on the m^{th} flip, given that the first m-1 flips were all heads, will be decreasing in m. Consistent with GF, the mean belief following 9 heads was 32%, significantly below ($p < 0.01$) the mean belief of 44% following one head.[1] The mean beliefs are monotonically decreasing as m increases from 2 to 10. A linear regression of participants' mean beliefs on m yields a coefficient of -1.7 percentage points (SE = 0.2). Even after excluding participants who reported 10% for some value of m, participants' mean beliefs appear to be consistent with GF for $m \geq 2$, and the slope of the regression of mean belief on m is -1.5 percentage points (SE = 0.3), similar to before.

We then predict the slope for each individual. A measure of a participant's degree of GF is the participant-specific slope from a regression of the participants' beliefs about the frequency of a head, given that the first m flips were a head, on m. 58 participants (57.43%) out of 101 participants had a negative slope.

Experiment 2

Pooling the data across the three sample sizes, a linear regression of participants' belief on m gives a coefficient of -1.3 percentage points (SE = 0.1). The coefficient is smaller in magnitude than the corresponding coefficient from Experiment 1, suggesting that the Berkeley students in Experiment 2 exhibit a weaker GF on average than the shoppers in Experiment 1.

We then predict the slope for each individual across all three sample sizes. A measure of a participant's degree of GF is the participant-specific slope from a regression of the participants' beliefs about the frequency of a head, given that the first m flips were a head, on m. 130 (42.21%) out of 308 participants had a negative slope.

Follow-up Analyses: Measuring NBLLN as sum of 0-5% Heads and 95-100% Heads in 1000-flips 11-bin elicitations

[1] Here and hereafter, unless stated otherwise, all p-values are from two-tailed t-tests when comparing mean beliefs and from Wilcoxon signed rank sum tests when comparing median beliefs. These are one-sample tests when comparing participants' beliefs to a specific value (such as 50%) and paired tests when comparing participants' beliefs across two situations (such as after 1 head vs. after 9 heads).

A measure of a participant's degree of NBLLN for both experiments is the sum of the reported probabilities of "0-5% heads" and "95-100% heads" in the 11-bin histogram for the sample size of 1,000. The correlation between this measure of NBLLN and our regression-based measure of GF is 0.26 (p = 0.02) in Experiment 1 and 0.09 (p = 0.12) in Experiment 2.

I.A.4. Parameter-uncertainty questions (3 Questions) (Experiment 2)

(a) Test for parameter uncertainty.

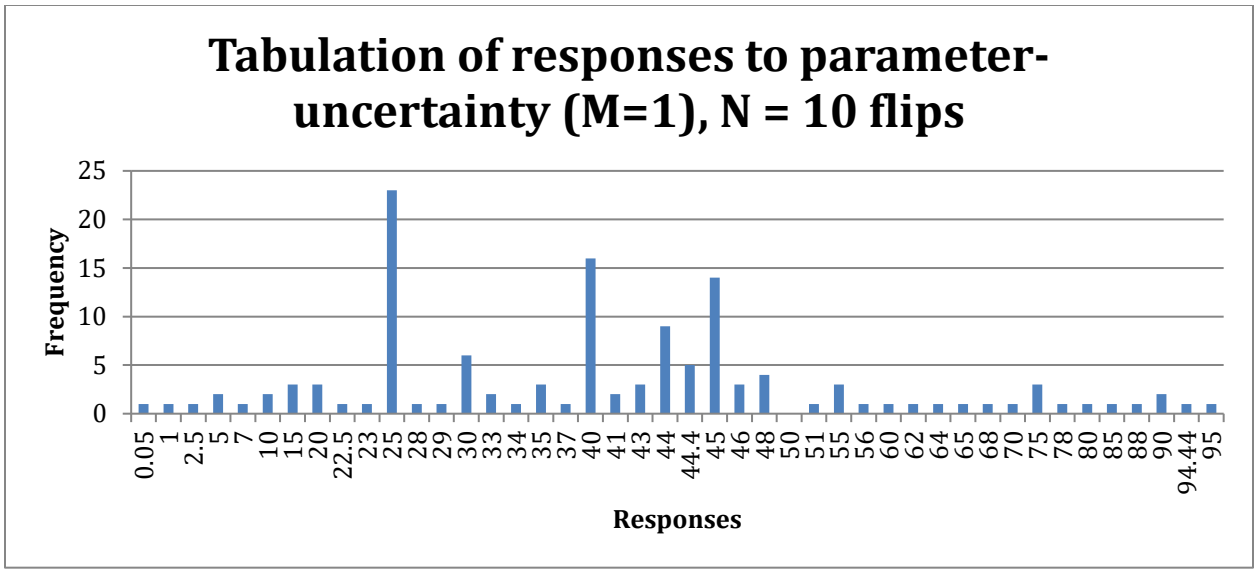
i. Test on average whether the judged probability of one random flip being a head, conditional on other random flips being heads, is equal, greater than, or less than 50%.

M	**Mean predicted freq of H (%)**	**Difference from 50% *t*-statistic (two-tailed *p*-value)**	**Difference from 45.65% *t*-statistic (two-tailed *p*-value)**
1	45.65	-5.68 (<0.0001)	-
2	40.66	-10.33 (<0.0001)	-5.52 (<0.0001)
5	36.71	-11.36 (<0.0001)	-7.64 (<0.0001)

Note that we test the difference from 45.65% because that is the mean belief reported by participants for the likelihood of the first flip being a Head.

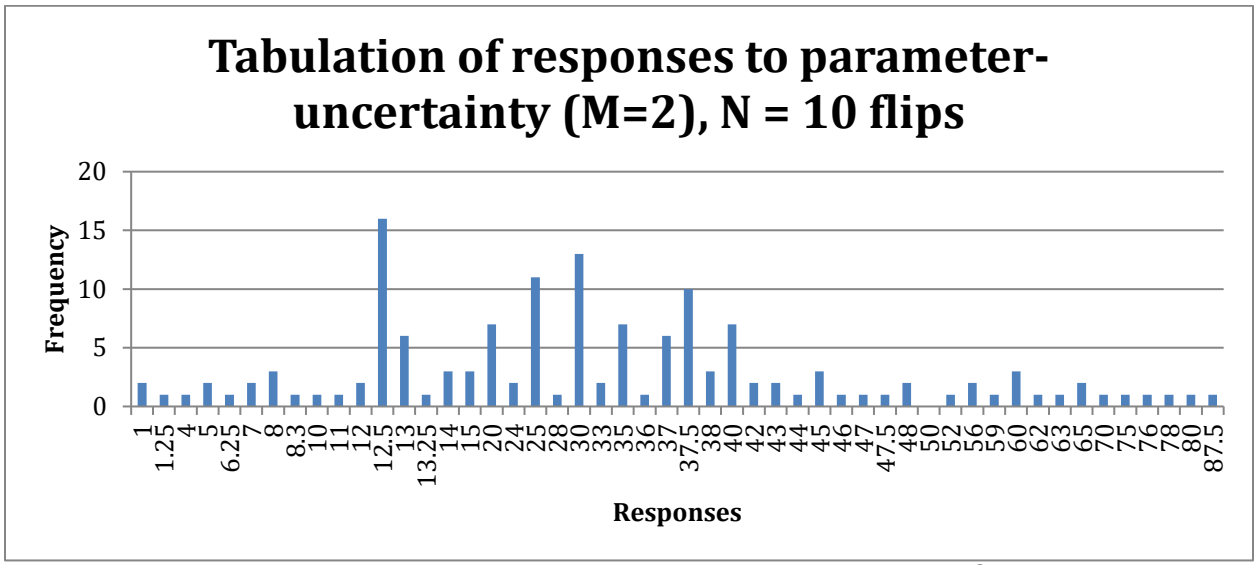
Follow-up analyses: Are there confused subjects?

Tabulation of responses to M=1 question (176 responded "50%"; omitted):

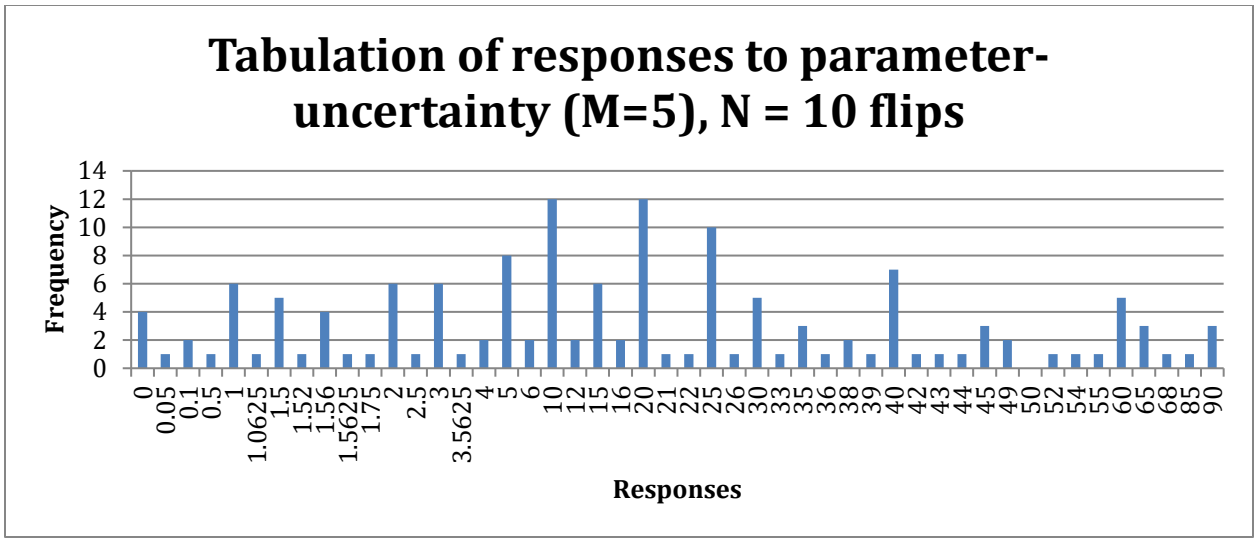


23 subjects responded "25%". One possibility is that they may have mistakenly thought the question asked: "What is the probability that first ***and*** second randomly-chosen flip number both came up Heads?" In other words, they may have thought: $0.5^2 = 0.25$.

When we tabulate responses to M=2 questions (162 responded "50%"), we see results consistent with this confusion:



16 subjects said "12.5%", which is consistent with this confusion, where $0.5^3 = 0.125$. 10 of these subjects entered "25%" for the M=1 question.
Tabulations for M=5 question (164 responded "50%):



Follow-up analyses: Excluding subjects who may be confused

Here, we exclude those whose responses to M=5 were less than or equal to 10%.

M	**Mean predicted freq of H (%)**	**Difference from 50% *t*-statistic (two-tailed *p*-value)**	**Difference from 47.74% *t*-statistic (two-tailed *p*-value)**
1	47.74%	-2.95 (0.0034)	-
2	45.14%	-5.85 (<0.0001)	-3.13 (0.0020)
5	45.53%	-5.48 (<0.0001)	-2.71 (0.0073)

Note that we test the difference from 47.74% because that is the mean belief reported by participants for the likelihood of the first flip being a Head.

(b) Calibrate beta-binomial parameter uncertainty.

i. Find the best-fitting parameters of the beta distribution for the 5-bin elicitation histogram (I.A.2.(B)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.A.4).

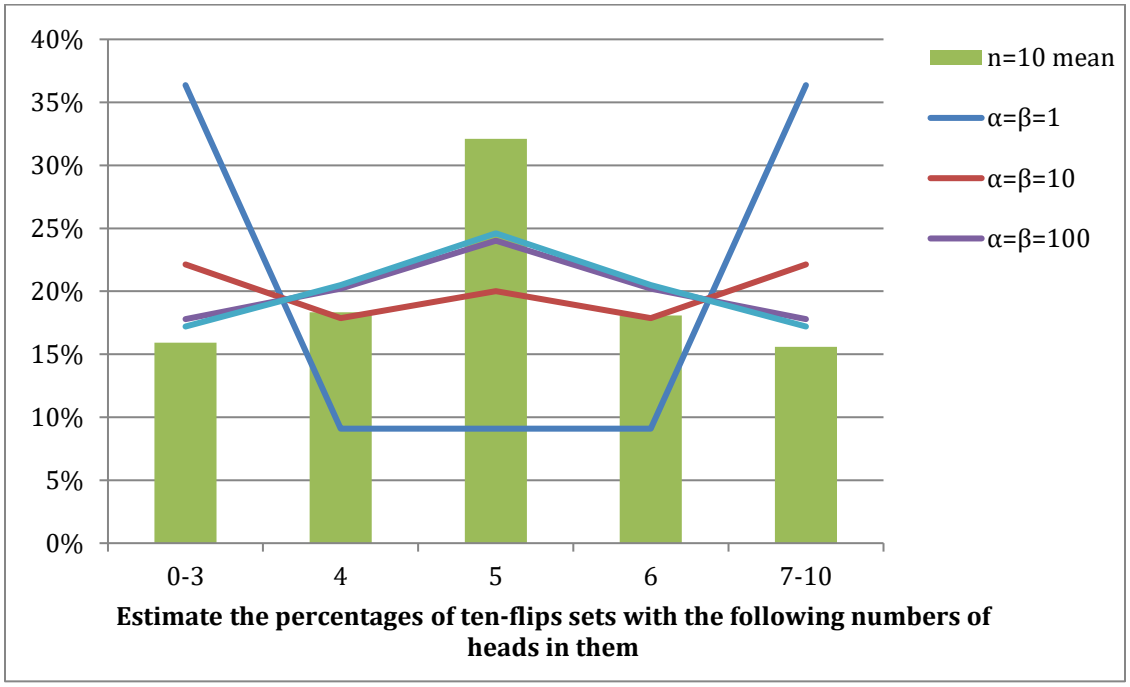
The p.d.f. of the beta-binomial distribution is

$$Beta(x,\alpha,\beta) = \binom{n}{x}\frac{B(x+\alpha)B(n-x+\beta)}{B(\alpha+\beta)}$$

Where B is the beta function and n = 10. We set α = β for symmetry.

The best-fitting α and β appear to be at the limit approaching infinity. Table below shows some example results:
Observed and Theoretical Probability Densities

	0-3	4	5	6	7-10	Sum of squared residuals
Observed	0.16	0.18	0.32	0.18	0.16	-
$\alpha = 1$	0.36	0.09	0.09	0.09	0.36	0.155
$\alpha = 5$	0.25	0.16	0.17	0.16	0.25	0.042
$\alpha = 10$	0.22	0.18	0.20	0.18	0.22	0.023
$\alpha = 100$	0.18	0.20	0.24	0.20	0.18	0.008
$\alpha = 500$	0.17	0.20	0.24	0.20	0.17	0.007
$\alpha = \infty$	0.17	0.21	0.25	0.21	0.17	0.007



I.B. 1,000-flip sets

I.B.1. Histogram questions (8 Questions)

Experiment 2

Bins (A)	True	Mean	Bins (B)	True	Mean	Bins (C)	True	Mean
0-50	0.0%	2.9%	0-480	11.0%	31.4%	0-450	0.1%	29.8%
51-150	0.0%	4.0%	481-519	78.0%	36.3%	451-549	99.8%	41.1%
151-250	0.0%	5.2%	520-1000	11.0%	32.3%	550-1000	0.1%	29.1%
251-350	0.0%	7.8%						
351-450	0.1%	13.5%						
451-549	99.8%	34.0%						

550-649	0.1%	13.3%
650-749	0.0%	7.8%
750-849	0.0%	5.2%
850-949	0.0%	3.9%
950-1000	0.0%	2.5%

Bins (D)	**True**	**Mean**
0-50	0.0%	2.8%
51-150	0.0%	4.2%
151-250	0.0%	5.4%
251-350	0.0%	8.1%
351-480	11.0%	14.5%
481-519	78.0%	30.3%
520-649	11.0%	15.0%
650-749	0.0%	8.0%
750-849	0.0%	5.3%
850-949	0.0%	3.9%
950-1000	0.0%	2.5%

Bins (E)	**True**	**Mean**
0-909	100.0%	79.1%
910	0.0%†	7.4%
911-1000	0.0%	13.5%

Bins (F)	**True**	**Mean**
0-499	48.7%	40.6%
500	2.5%	19.3%
501-1000	48.7%	40.1%

Bins (G)	**True**	**Mean**
0-493	34.1%	35.7%
494-506	31.9%	29.1%
507-1000	34.1%	35.2%

Bins (H)	**True**	**Mean**
0-487	21.5%	26.2%
488-496	19.8%	12.5%
497-503	17.5%	21.2%
504-512	19.8%	13.5%
513-1000	21.5%	26.5%

† The probability of 910 heads is 1.0101×10^{-171}; the probability of 911-1000 heads is 1.1057×10^{-172}.

(a) Test for bin effects (Experiment 2).

i. Compare the mean probability assigned to 0-450 in (C) with the sum of the mean probabilities assigned to 0-50, …, 351-450

Mean probability assigned to 0-450 in (C) = 29.77%
Sum of mean probabilities assigned to 0-50, ..., 351-450 in (A) = 33.27%
(p-value < 0.0001)

Mean probability assigned to 550-1000 in (C) = 29.11%
Sum of mean probabilities assigned to 550-649, ..., 950-1000 in (A) = 32.73%
(p-value < 0.0001)

ii. Compare the probability assigned to 910 heads in (E) to the probability assigned to 850-949 heads in (A).

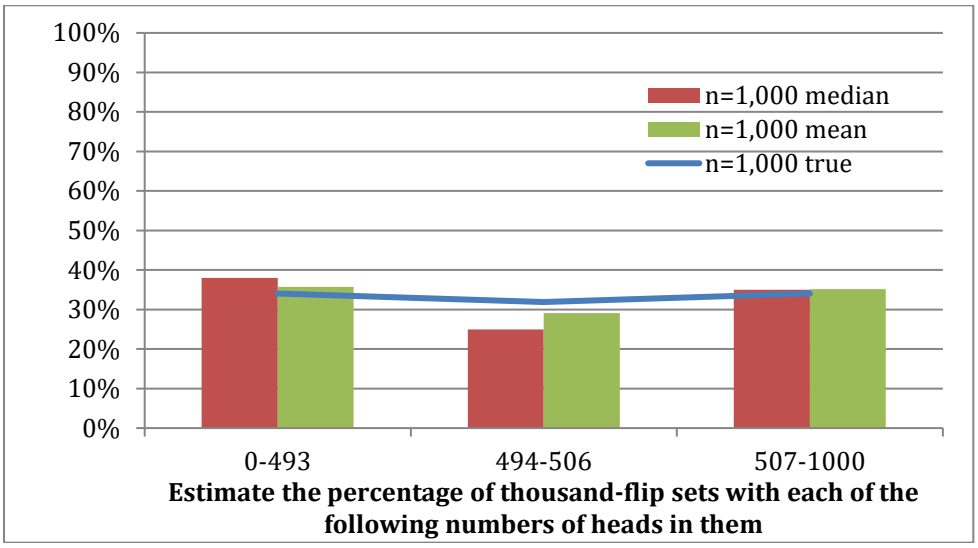
Mean probability assigned to 910 in (E) = 7.36%

Mean probability assigned to 850-949 in (A) = 3.89%
(p-value < 0.0001)

iii. Compare the probability assigned to 500 heads in (F) to the probability assigned to 451-549 heads in (A).
Mean probability assigned to 500 in (F) = 19.29%
Mean probability assigned to 451-549 in (A) = 34.19%
(p-value < 0.0001)

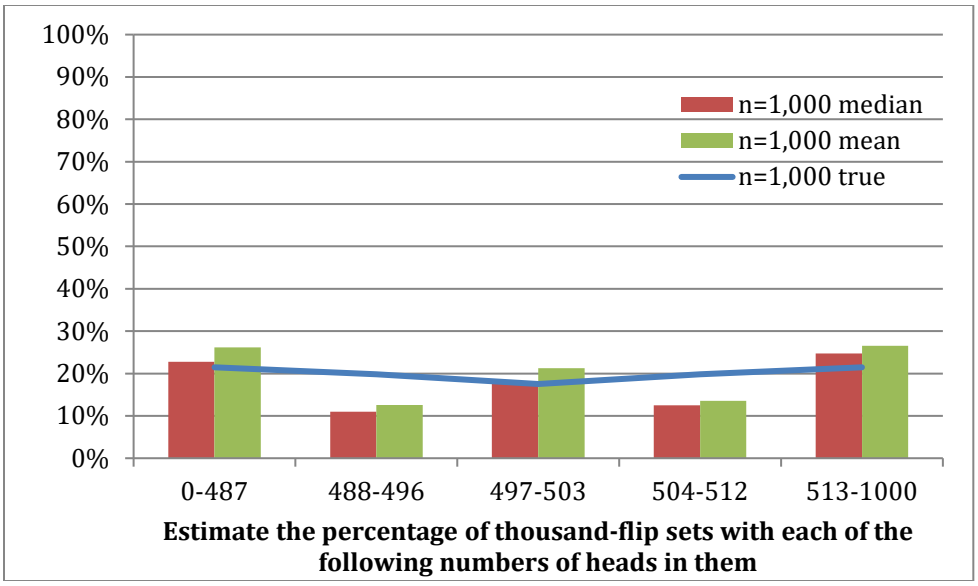
(b) Test for NBLLN in N = 1000 (Experiment 2).

i. Examine whether (G) generates approximately equal mean elicited probabilities across the three bins.



Follow-up analyses: How many subjects gave approx equal weights to all the bins?
5.5% (17 out of 308) gave weights between 33-34% to all the bins.
10.4% (32 out of 308) subjects gave weights between 30-36% to all the bins.

ii. Examine whether (H) generates approximately equal mean elicited probabilities across the five bins.



Follow-up analyses: How many subjects gave approx equal weights to all the bins?
8.1% (25 out of 308) gave weight 20% to all the bins.
8.8% (27 out of 308) gave weights between 17-27% to all the bins.

Follow-up analyses: How many subjects overweight both extreme and middle bins?
4.5% overweight the first, middle, and last bins.
5.2% overweight the middle bin and overweight the extreme bins (i.e., for 5.2% of participants, the middle bin is greater than the true probability, and the mean of the first and last bin is greater than the true probability of the first bin).

Follow-up analyses: How many subjects overweight each bin?
51.3% overweight the first bin.
51.9% overweight the last bin.
51.9% overweight the first and last bin (i.e., for 51.9% of participants, the mean of the first and last bin is greater than the true probability of the first bin)
50.3% overweight the middle bin.

Follow-up analyses: How many subjects overweight both extreme bins?
48.4% (149 out of 308) overweight both extreme bins.

(c) Test for the presence of "exact representativeness" (Experiment 2).

i. Test whether the average probability assigned to 481-519 in (B) exceeds the average probability assigned to 451-549 in (C).
Mean probability assigned to 481-519 in (B) = 36.27%
Mean probability assigned to 451-549 in (C) = 41.12%
(p-value < 0.0001)

I.B.2. Streak questions (10 Questions)

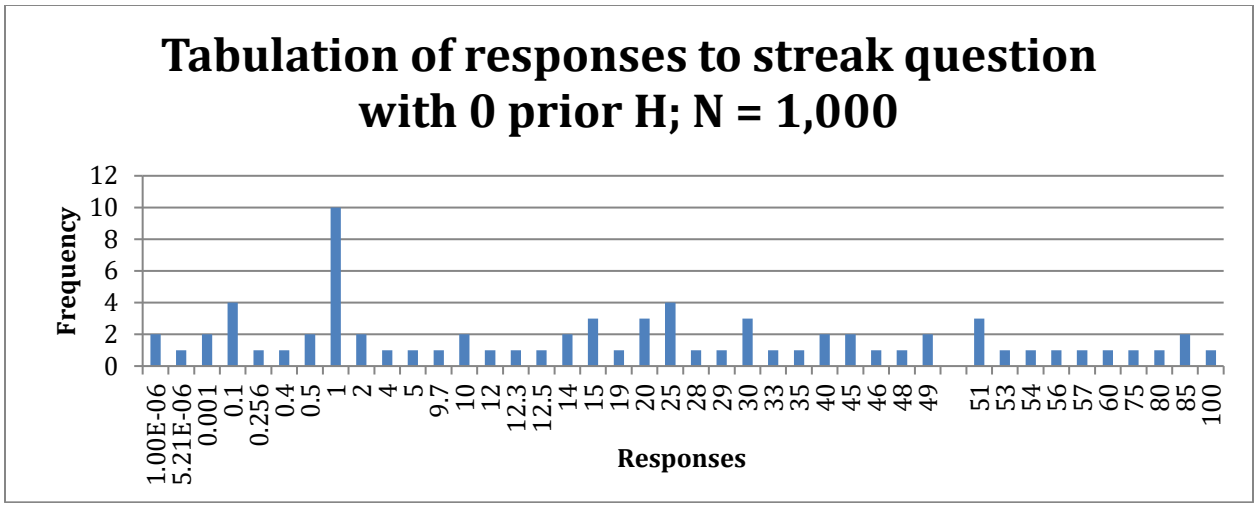
(a) Test for GF: How the predicted frequency of H varies with the number of prior H's (Experiment 2).

i. t-test of each predicted frequency's difference from 50%.

Number of prior H's	Mean predicted freq of H (%)	Difference from 50% *t*-statistic (two-tailed p-value)	Difference from 43.76% *t*-statistic (two-tailed *p*-value)
0	43.76	-6.61 (<0.0001)	
1	44.34	-6.96 (<0.0001)	0.71 (0.4773)
2	42.23	-7.96 (<0.0001)	-1.5628 (0.1191)
3	40.45	-8.90 (<0.0001)	-3.0826 (0.0022)
4	39.85	-9.39 (<0.0001)	-3.6191 (0.0003)
5	38.75	-10.38 (<0.0001)	-4.62 (<0.0001)
6	38.17	-10.16 (<0.0001)	-4.80 (<0.0001)
7	39.06	-9.52 (<0.0001)	-4.09 (0.0001)
8	36.99	-10.80 (<0.0001)	-5.62 (<0.0001)
9	37.18	-10.49 (<0.0001)	-5.38 (<0.0001)

Note that we test the difference from 43.76% because that is the mean belief reported by participants for the likelihood of the first flip being a Head.
Note: 37.18 is significantly different from 44.34 with p-value < 0.0001.

Follow-up analyses: Why is the mean unconditional probability less than 50%?
Tabulation of responses to this question (234 responded "50%"; omitted from figure):



30 people entered values ≤ 10%. To see whether they may have been confused, the table below shows their responses to the questions eliciting the likelihood of a flip landing on heads without being given the results of any prior flips (what we call "streak-0 questions") for N=10 and N=1,000,000:

N = 10		N = 1,000,000	
Responses	Frequency	Responses	Frequency
5	1	1.00E-09	1
10	3	1.00E-07	1
15	1	1.00E-06	3
25	1	0.00001	1
40	1	0.0001	2
45	1	0.021	1
50	19	0.14536	1
55	1	0.5	1
60	1	1	2
80	1	2	2
		5	1
		9.8	1
		10	1
		14	1
		25	2
		30	1
		40	1
		50	7

These "confused" subjects seemed to have gotten 1k-flip and 1m-flip questions wrong but correctly answered the 10-flip streak-0 question (19 out of 30). This perhaps suggests that some subjects got confused with the big numbers "1k" and "1m" in the question.

Follow-up analyses: Breakdown of Streak-0 responses by those who saw this question first vs. those who saw this question later.
Those who saw this question first: Mean = 37.01% [Sample = 28 out of 308]
Others: Mean = 44.44% [Sample = 280 out of 308]
(p-value = 0.0626)

ii. Linear regression of predicted frequencies on number of prior H's.

	(a.ii)
Number of prior H's	-0.81*** (0.119)
Constant	43.75 (0.633)
Obs	3080

(b) Test for GF at the individual-level: Fraction of individuals who always report 50%, always report weakly less than 50% (at least one frequency strictly), always report weakly less than 50% (at least one frequency strictly) (Experiment 2).

Fraction of individuals who always report 50% = 133/308 = 43.18%
Fraction of individuals who always report weakly less than 50% (at least one frequency strictly) = 119/308 = 38.64%
Fraction of individuals who always report weakly more than 50% (at least one frequency strictly) = 0/308 = 0%.

(c) Calibrate GF: Estimate the Rabin-Vayanos model by non-linear least squares on the first-difference formulation of the model.

	Estimates using individual-level data	**Estimates using aggregate data**
alpha	0.020* (0.012) [-0.004, 0.044]	0.02 (0.021)
delta	0.94*** (0.098) [0.743, 1.129]	0.94*** (0.200)
Obs	2772	9

Note: Standard errors in parentheses (clustered for individual-level regression), and 95% confidence intervals in brackets. * Significant at 10% ** 5% *** 1%.

I.B.3. Parameter-uncertainty questions (3 Questions) (Experiment 2)

(a) Test for GF and parameter uncertainty.

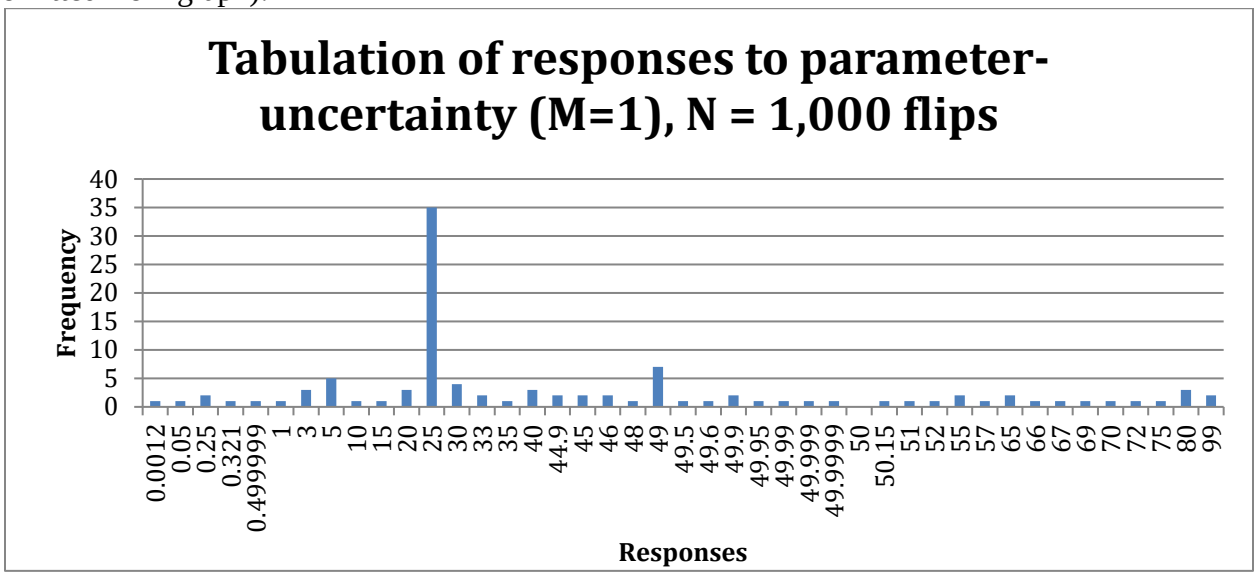
i. Test on average whether the judged probability of one random flip being a head, conditional on other random flips being heads, is equal, greater than, or less than 50%.

M	Mean predicted freq of H (%)	Difference from 50% *t*-statistic (two-tailed *p*-value)	Difference from 44.83% *t*-statistic (two-tailed *p*-value)
1	44.83*	-6.17 (<0.0001)	-
2	41.98*	-8.25 (<0.0001)	-2.93 (0.0037)
5	39.39*	-9.89 (<0.0001)	-5.07 (<0.0001)

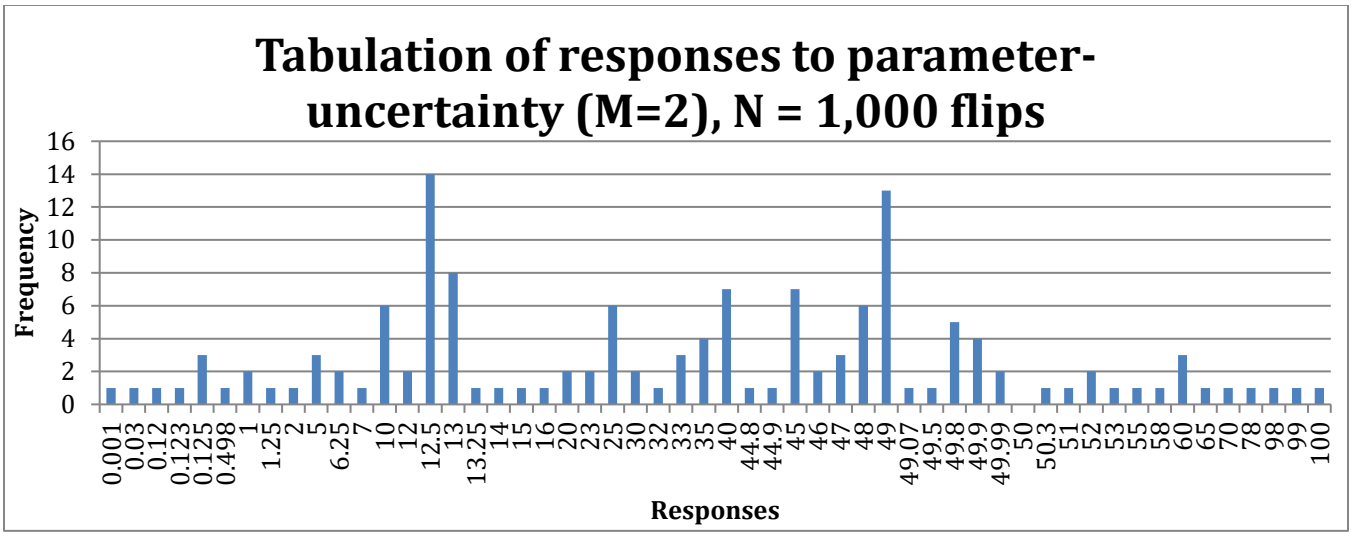
Note that we test the difference from 44.83% because that is the mean belief reported by participants for the likelihood of the first flip being a Head.

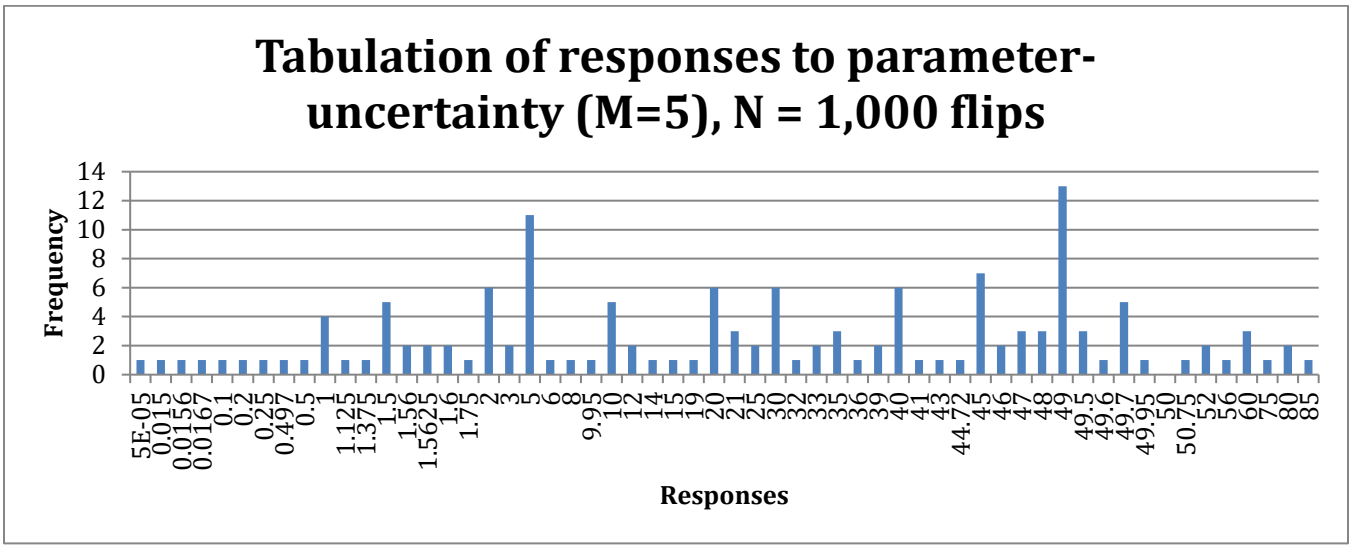
Follow-up analyses: Are there confused subjects?

Figure below shows tabulation of responses to M=1 question (202 responded "50%"; omitted from graph):



35 subjects responded "25%". Figures below show responses to M=2 question (167 responded "50%") and M=5 question (165 responded "50%"):





Follow-up analyses: Excluding subjects who may be confused

Here, we exclude those whose responses to M=5 were less than or equal to 10%.

M	Mean predicted freq of H (%)	Difference from 50% *t*-statistic (two-tailed *p*-value)	Difference from 48.41% *t*-statistic (two-tailed *p*-value)
1	48.41%	-2.26 (0.025)	-
2	46.74%	-4.56 (<0.0001)	-2.34 (0.020)
5	47.07%	-4.91 (<0.0001)	-2.24 (0.026)

Note that we test the difference from 48.41% because that is the mean belief reported by participants for the likelihood of the first flip being a Head.

(b) Calibrate beta-binomial parameter uncertainty.

i. Find the best-fitting parameters of the beta distribution for the 3-bin equalized elicitation histogram (I.B.1.(G)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.B.3).

(Omitted because we know that the answer will be infinite parameter values, as in Section I.A.4.b above)

ii. Find the best-fitting parameters of the beta distribution for the 5-bin equalized elicitation histogram (I.B.1.(H)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.B.3).

(Omitted because we know that the answer will be infinite parameter values, as in Section I.A.4.b above)

I.C. One-million-flip sets (Experiment 2 only)

I.C.1. Histogram questions (3 Questions)

Full results:

Bins	True	Mean	Bins	True	Mean	Bins	True	Mean
0-50000	0.0%	3.7%	0-499784	33.3%	32.8%	0-499579	20.1%	24.4%
50001-150000	0.0%	4.2%	499785-500215	33.4%	34.3%	499580-499873	20.0%	12.6%
150001-250000	0.0%	5.2%	500216-1000000	33.3%	32.9%	499874-500126	19.9%	25.7%
250001-350000	0.0%	7.6%				500127-500420	20.0%	12.8%
350001-450000	0.0%	13.3%				500421-1000000	20.1%	24.6%
450001-549999	100.0%	34.7%						
550000-649999	0.0%	12.4%						
650000-749999	0.0%	7.3%						
750000-849999	0.0%	5.2%						
850000-949999	0.0%	3.9%						
950000-1000000	0.0%	2.6%						

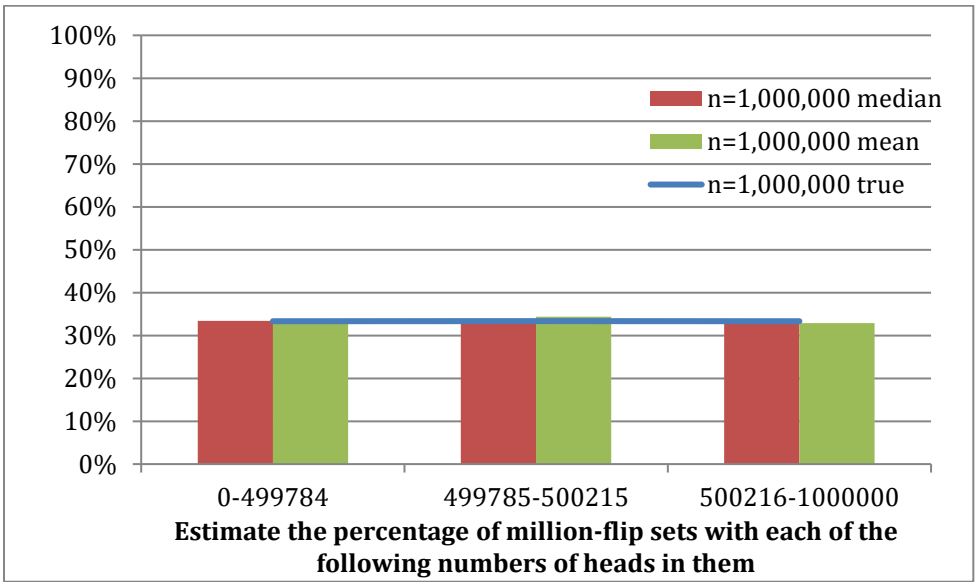
(a) Test for bin effects.

i. Compare the mean probability assigned to 550,000-649,999 in (A) with the mean probability assigned to 499,785-500,215 in (B).
Mean probability assigned to 550,000-649,999 in (A) = 12.35%
Mean probability assigned to 499,785-500,215 in (B) = 34.34%

(p-value < 0.0001)

(b) Test for NBLLN in N = 1 million.

i. Examine whether (B) generates approximately equal mean elicited probabilities across the three bins.



Follow-up analyses: How many subjects gave approx equal weights to all the bins?
7.8% (24 out of 308) gave weights between 33-34% to all the bins.
14.3% (44 out of 308) gave weights between 30-36% to all the bins.

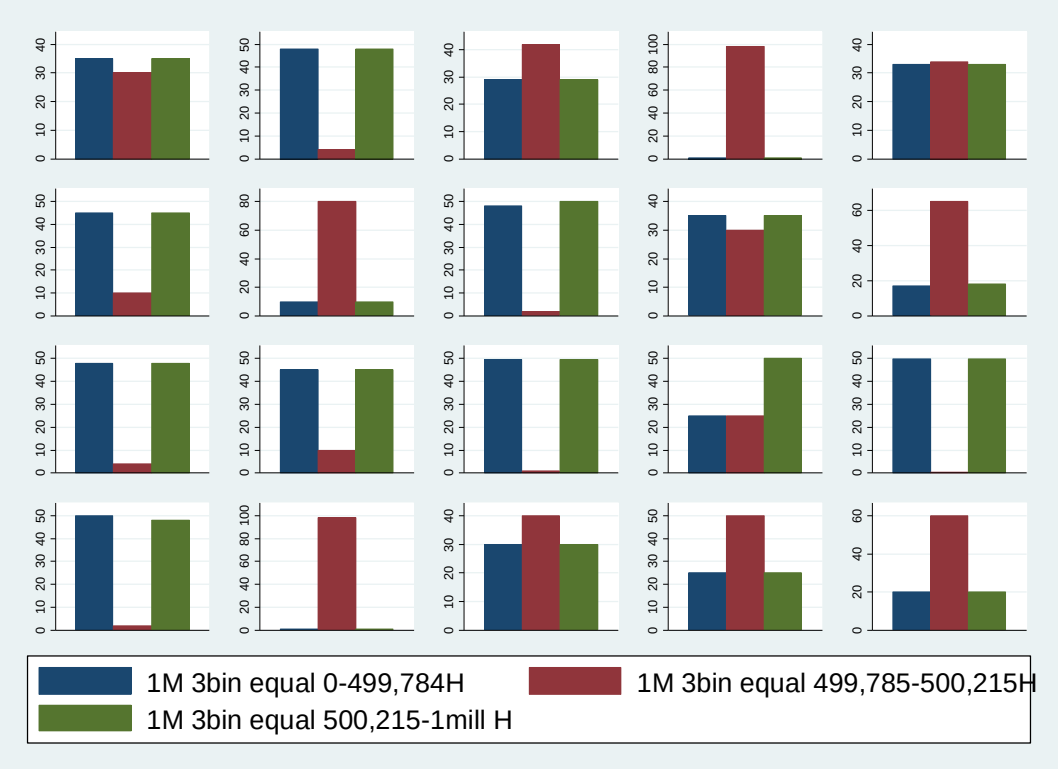
44.2% (136 out of 308) gave higher weight to the middle-bin than the outer bins.
43.8% (135 out of 308) gave lower weight to the middle-bin than the outer bins.

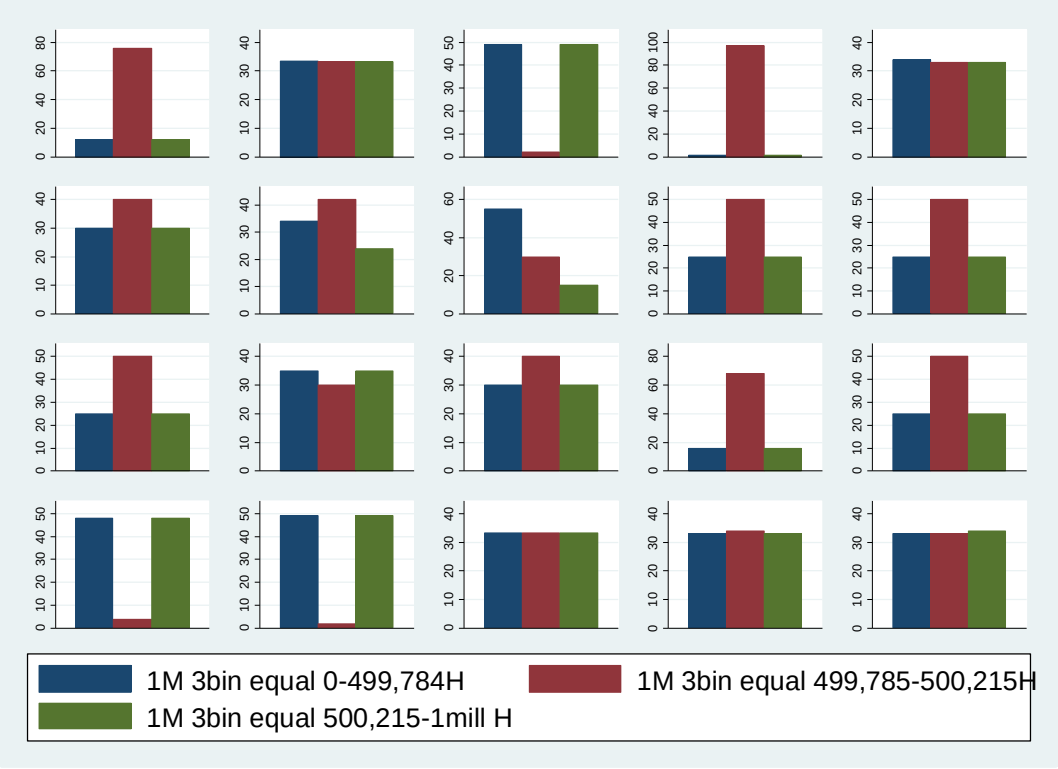
	Higher weight to middle bin	Lower weight to middle bin	Others
Average Math Quiz Score	5.42	5.47	4.69

Do people who put more weights on tails do the same thing for other questions?
We find evidence to suggest participants that gave lower weight to the middle-bin than the outer bins in this question are also giving lower weight to the middle-bin in the other 1 million flip-set questions.

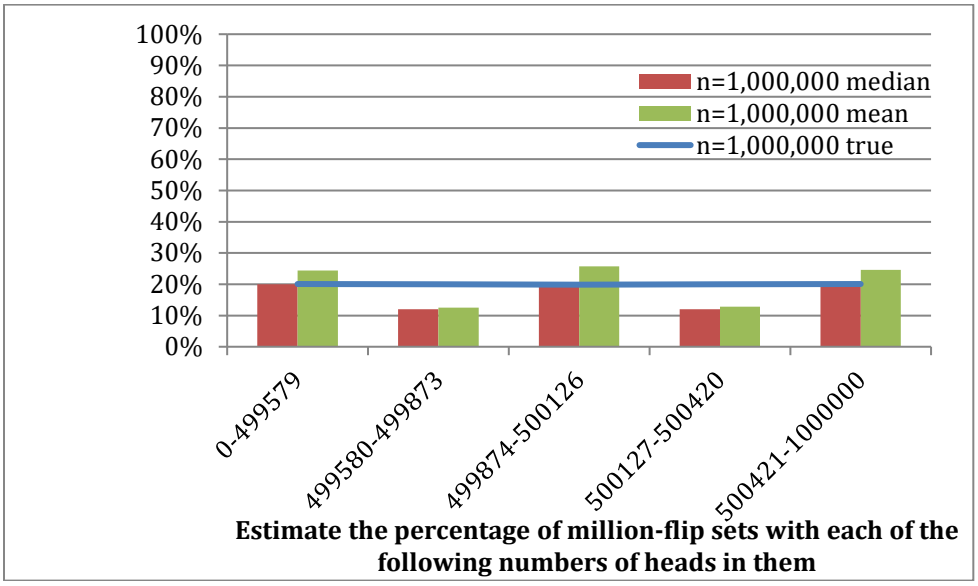
	Lower weight to middle bin in 3-bin 1m flip set	Others	*p*-value
Average middle-bin response: 5-bin 1m flips	10.25 (1.01)	37.61 (2.00)	0.00
Average middle-bin response: 11-bin 1m flips	26.81 (1.62)	41.20 (2.38)	0.00

To give a sense of the individual-level data, here are the histograms of the first 40 participants:





ii. Examine whether (C) generates approximately equal mean elicited probabilities across the five bins.



Follow-up analyses: How many subjects gave approx equal weights to all the bins?
8.8% (27 out of 308) gave weight 20% to all the bins.
10.4% (32 out of 308) gave weights between 17-23% to all the bins.

Follow-up analyses: How many subjects overweight the extreme bins?
40% (124 out of 308) gave weight greater than 20.1% for both extreme bins.

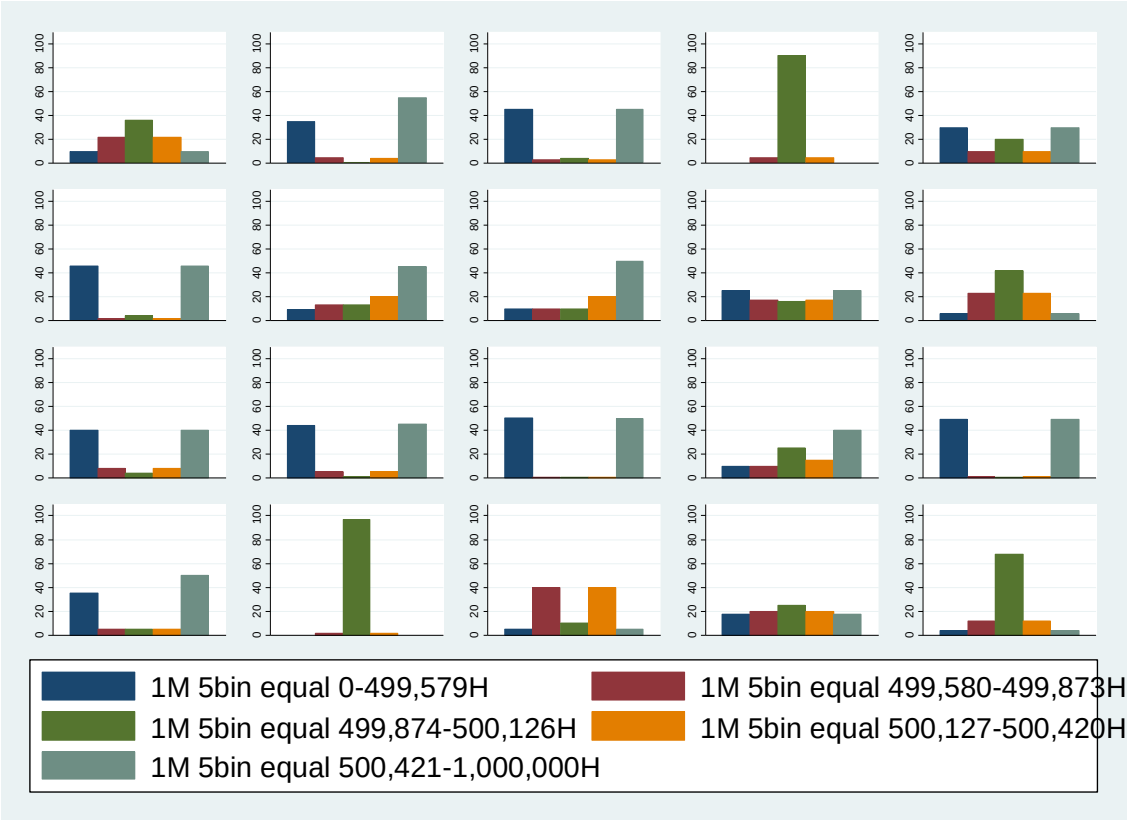
Follow-up analyses: How many subjects overweight the middle bin?
56.5% (174 out of 308) gave weight greater than 19.9% to the middle bin.

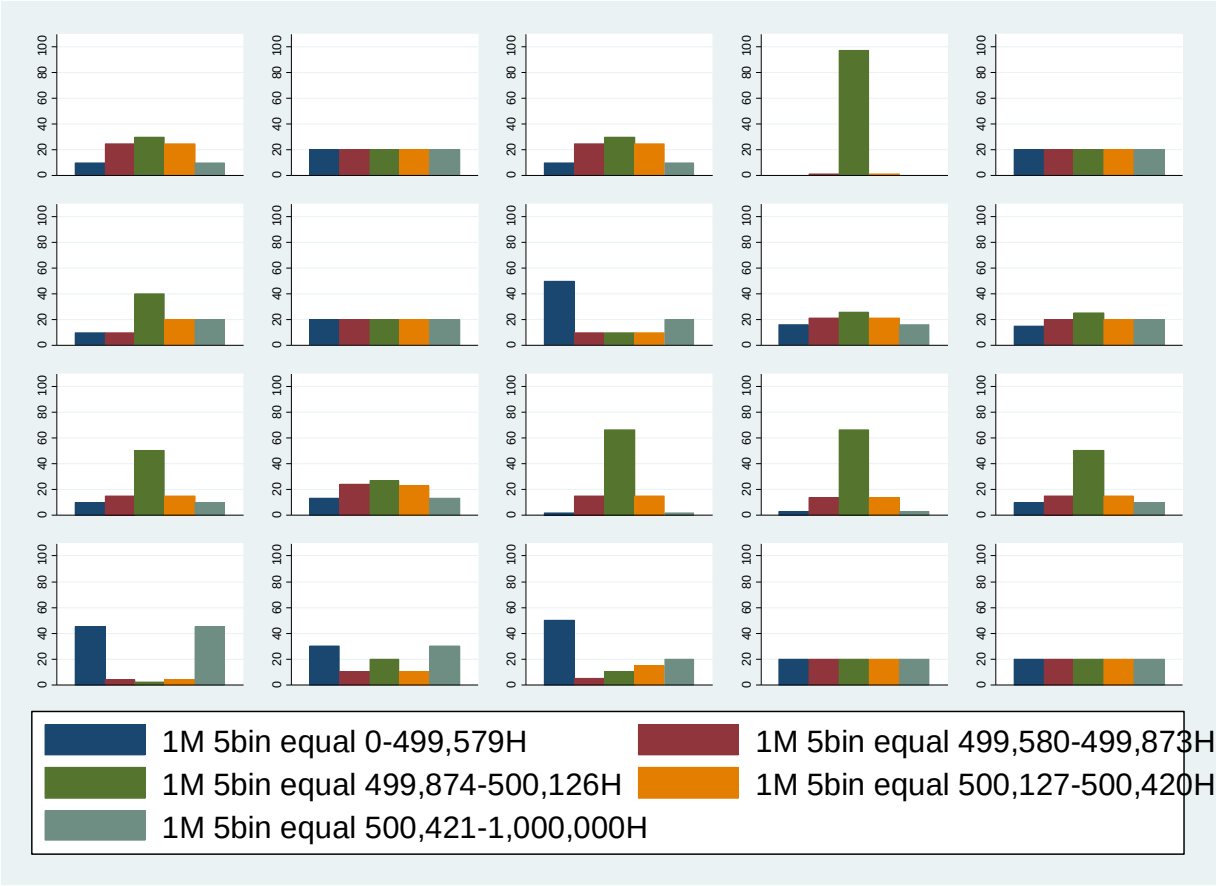
Follow-up analyses: How many subjects overweight the middle and extreme bins?
4.5% (14 out of 308) overweight both the middle and extreme bins.

Follow-up analyses: How many subjects had higher inner-tails than outer-tails?
(i) For the left tail:
39.6% (122 out of 308) gave greater weight to the inner-tail than the outer-tail.
13.6% (42 out of 308) gave equal weight to both.
46.8% (144 out of 308) gave greater weight to the outer-tail than the inner-tail.
(ii) Right tail:
38.6% (119 out of 308) gave greater weight to the inner-tail than the outer-tail.
11.4% (35 out of 308) gave equal weight to both.
50% (154 out of 308) gave greater weight to the higher outer-tail than the inner-tail.
(iii) Both tails:
34.7% (107 out of 308) gave greater weight to the inner-tails than the corresponding outer-tails.
10.7% (33 out of 308) gave equal weight to the inner-tails and corresponding outer-tails.
44.4% (137 out of 308) gave greater weight to the outer-tails than the corresponding inner-tails.

	Higher inner-tails for both left and right	Same inner- and outer-tails for both left and right	Higher outer-tails for both left and right
Average Math Quiz Score	5.4	5.2	5.5

Below are histograms of first 40 participants:





I.C.2. Streak questions (10 Questions)

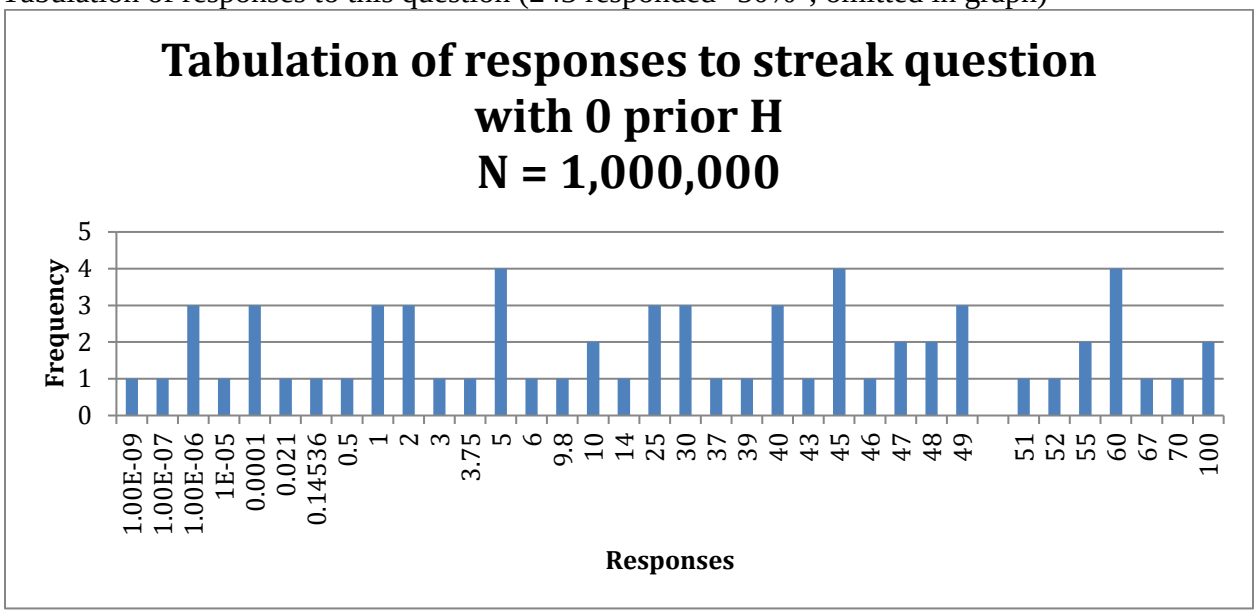
(a) Test for GF: How the predicted frequency of H varies with the number of prior H's.

i. t-test of each predicted frequency's difference from 50%.

Number of prior H's	Mean predicted freq of H (%)	Difference from 50% *t*-statistic (two-tailed p-value)	Difference from 45.43% *t*-statistic (two-tailed *p*-value)
0	45.43	-5.38 (<0.0001)	
1	44.38	-6.97 (<0.0001)	-1.30 (0.1953)
2	41.44	-8.80 (<0.0001)	-4.10 (0.0001)
3	40.98	8.58 (<0.0001)	-4.23 (<0.0001)
4	39.26	-9.48 (<0.0001)	-5.45 (<0.0001)
5	38.93	-9.65 (<0.0001)	-5.66 (<0.0001)
6	38.87	-9.55 (<0.0001)	-5.63 (<0.0001)
7	38.27	-10.14 (<0.0001)	-6.19 (<0.0001)
8	38.08	-9.92 (<0.0001)	-6.12 (<0.0001)
9	37.91	-9.51 (<0.0001)	-5.92 (<0.0001)

Note that we test the difference from 45.43% because that is the mean belief reported by participants for the likelihood of the first flip being a Head.
37.91 is significantly different from 44.38 with p-value < 0.0001.

Why is mean for 0 prior H less than 50%?
Tabulation of responses to this question (243 responded "50%"; omitted in graph)



28 people entered values ≤ 10%. The table below shows tabulations of their responses to the streak-0 questions for N=10 and N=1,000.

N=10		N=1,000	
Responses	**Frequency**	**Responses**	**Frequency**
5	1	1.00E-06	2
10	2	5.21E-06	1
15	1	0.001	2
33	1	0.1	3
40	1	0.256	1
48	1	0.4	1
50	20	0.5	1
60	1	1	6
		9.7	1
		12.5	1
		14	1
		15	3
		20	1
		50	3
		85	1

These "confused" subjects seemed to have gotten 1k-flip and 1m-flip questions wrong but correctly answered the 10-flip streak-0 question (20 out of 28). This perhaps suggests that some subjects got confused with the big numbers "1k" and "1m" in the question.

Follow-up analyses: Breakdown of Streak-0 responses by those who saw this question first vs. those who saw this question later.
Those who saw this question first: Mean = 44.21% [Sample = 39]
Others: Mean = 45.61% [Sample = 269]

ii. Linear regression of predicted frequencies on number of prior H's.

	(a.ii)
Number of prior H's	-0.81*** (0.119)
Constant	44.02 (0.64)
Obs	3080

(b) Test for GF at the individual-level: Fraction of individuals who always report 50%, always report weakly less than 50% (at least one frequency strictly), always report weakly more than 50% (at least one frequency strictly).

Fraction of individuals who always report 50% = 140/308 = 45.45%
Fraction of individuals who always report weakly less than 50% (at least one frequency strictly) = 123/308 = 39.94%
Fraction of individuals who always report weakly more than 50% (at least one frequency strictly) = 1/308 = 0.32%

(c) Calibrate GF: Estimate the Rabin-Vayanos model by non-linear least squares on the first-difference formulation of the model.

	Estimates using individual-level data	**Estimates using aggregate data**
alpha	0.05** (0.019) [0.010, 0.084]	0.05* (0.02)
delta	0.788*** (0.077) [0.636, 0.940]	0.79*** (0.118)
Obs	2772	

Note: Standard errors in parentheses (clustered for individual-level data), and 95% confidence intervals in brackets. * Significant at 10% ** 5% *** 1%.

I.C.3. Parameter-uncertainty questions (3 Questions)

(a) Test for GF and parameter uncertainty.

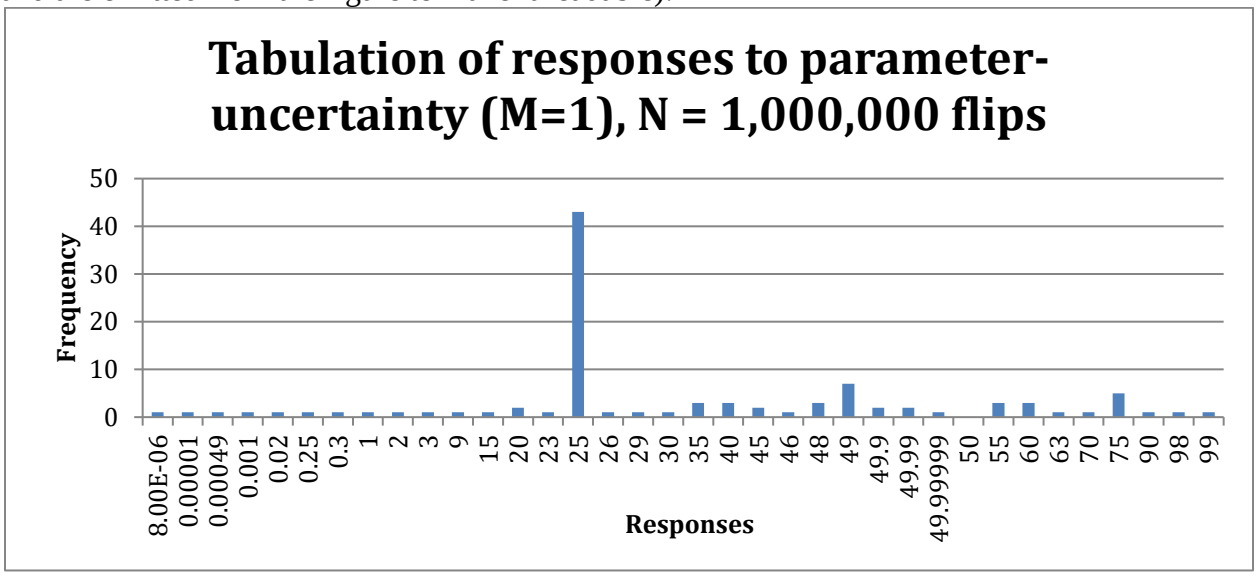
i. Test on average whether the judged probability of one random flip being a head, conditional on other random flips being heads, is equal, greater than, or less than 50%.

M	Mean predicted freq of H (%)	Difference from 50% *t*-statistic (two-tailed *p*-value)	Difference from 44.94% *t*-statistic (two-tailed *p*-value)
1	44.94	-6.27 (<0.0001)	-
2	41.70	-8.56 (<0.0001)	-3.34 (0.0009)
5	40.55	-8.96 (<0.0001)	-4.16 (<0.0001)

Note that we test the difference from 44.94% because that is the mean belief reported by participants for the likelihood of the first flip being a Head.

Follow-up analyses: Are there confused subjects?

The figure below shows tabulation of responses to M=1 question (202 responded "50%" and are omitted from the figure to make it readable):



Follow-up analyses: Excluding subjects who may be confused

Here, I exclude those whose responses to M=5 were less than or equal to 10%.

M	Mean predicted freq of H (%)	Difference from 50% *t*-statistic (two-tailed *p*-value)	Difference from 46.55% *t*-statistic (two-tailed *p*-value)
1	46.55	-5.12 (<0.0001)	-

2	43.02	-7.71 (<0.0001)	-3.90 (0.0001)
5	41.60	-8.21 (<0.0001)	-4.84 (<0.0001)

Note that we test the difference from 46.55% because that is the mean belief reported by participants for the likelihood of the first flip being a Head.

(b) Calibrate beta-binomial parameter uncertainty.

i. Find the best-fitting parameters of the beta distribution for the 3-bin equalized elicitation histogram (I.C.1.(B)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.C.3).
(Omitted)

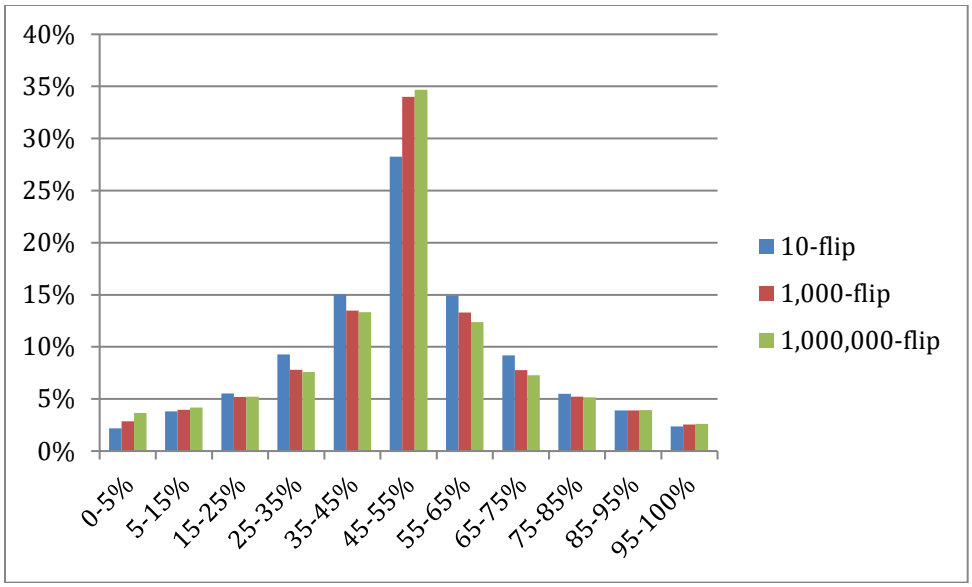
ii. Find the best-fitting parameters of the beta distribution for the 5-bin equalized elicitation histogram (I.C.1.(C)) according to the least-squares criterion. Examine how well the model fits the data from the parameter uncertainty questions (I.C.3).
(Omitted)

I.D. Comparisons across different question types

(a) Test for Kahneman and Tversky's "sample size neglect" (including for the much larger sample size of 1 million) (Experiment 2).

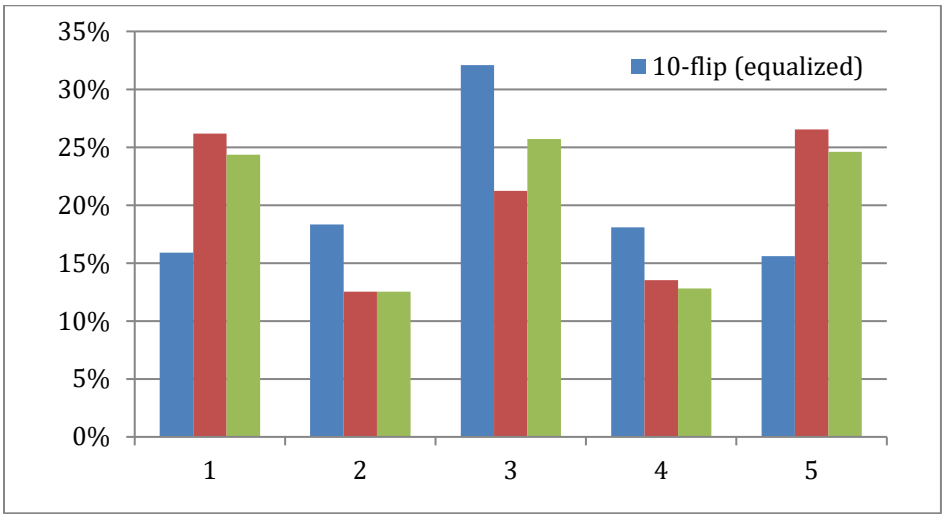
(i) Compare the distribution of 0-5%, 5-15%, …, 95-100% heads across the 10-flip, 1000-flip, and 1-million-flip sets.

Bins	10 flips	1,000 flips	1,000,000 flips
0-5%	2.2%	2.9%	3.7%
5-15%	3.8%	4.0%	4.2%
15-25%	5.5%	5.2%	5.2%
25-35%	9.3%	7.8%	7.6%
35-45%	15.1%	13.5%	13.3%
45-55%	28.3%	34.0%	34.7%
55-65%	14.9%	13.3%	12.4%
65-75%	9.2%	7.8%	7.3%
75-85%	5.5%	5.2%	5.2%
85-95%	3.9%	3.9%	3.9%
95-100%	2.4%	2.5%	2.6%



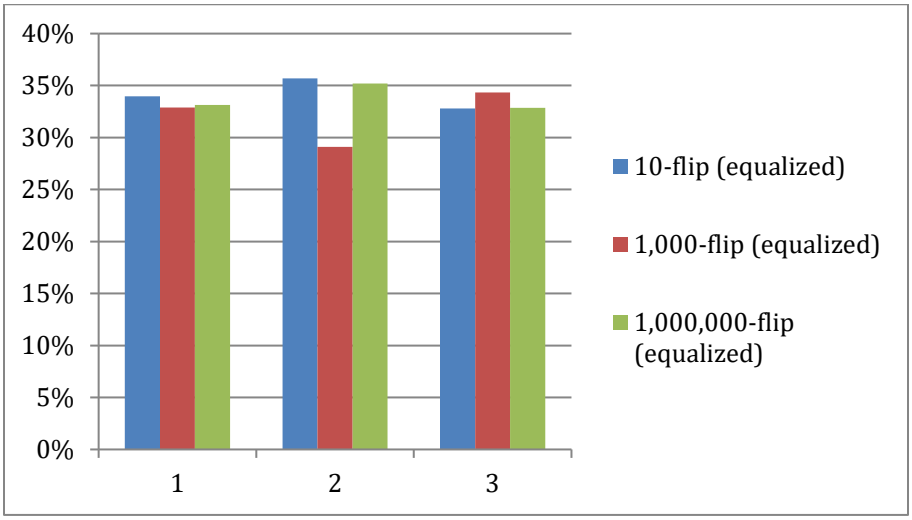
(ii) Compare the equalized histogram beliefs across the 10-flip, 1000-flip, and 1-million-flip sets.

Bin	10 flips	1,000 flips	1,000,000 flips
1	15.9%	26.2%	24.4%
2	18.3%	12.5%	12.6%
3	32.1%	21.2%	25.7%
4	18.1%	13.5%	12.8%
5	15.6%	26.5%	24.6%



Bin	10 flips	1,000 flips	1,000,000 flips
1	34.0%	35.7%	32.8%
2	32.9%	29.1%	34.3%

3	33.2%	35.2%	32.9%



(b) Do individuals whose histogram distributions are more spread out for samples of size 10 also tend to be spread out for sample sizes of 1,000?

Experiment 1:
Yes, we find a significant (*p* - value = 0.0000) positive correlation of 0.668 between the distributions for sample sizes 10 and 1,000.

Experiment 2:
Yes, we find a significant (*p* - value = 0.0000) positive correlation of 0.425 between the distributions for sample sizes 10 and 1,000.

(c) Test whether participants' sequence beliefs can be reconciled with their histogram beliefs (Experiment 2).

(i) Compare the implied probability of 10 heads out of 10 from I.A.3 (the sequence beliefs) to the histogram probability (I.A.2.(A)), which is contaminated by bin effects, and to the normatively correct probability.
Implied probability of 10 heads out of 10 from I.A.3 = 0.00948%

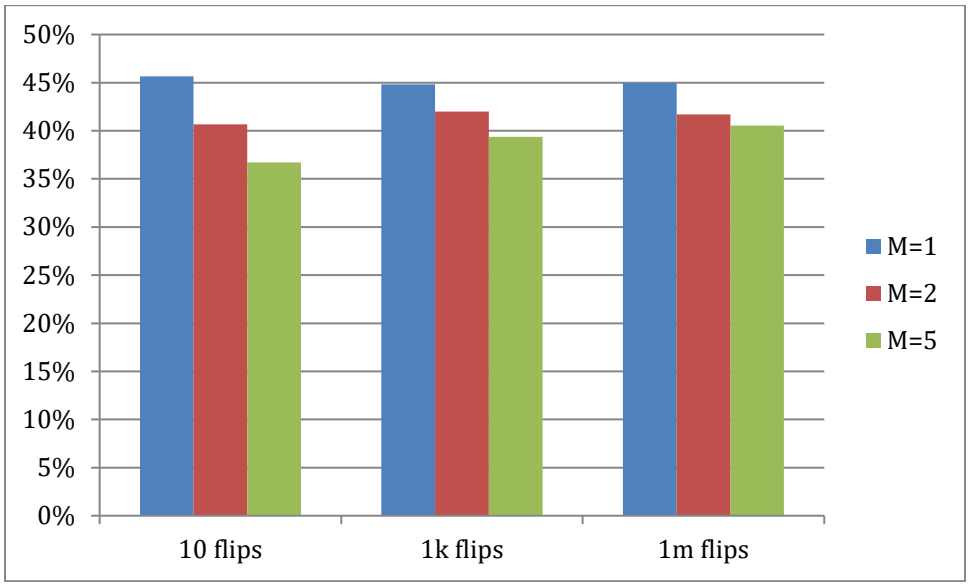
Histogram probability from I.A.2.A = 2.37%
True probability = 0.09766%
(p-value < 0.0001)

(d) Test whether the effect of some flips on another flip is weaker as we go from the N = 10 to the N = 1000 and N = 1 million parameter-uncertainty questions (question types I.A.4, I.B.3, and I.C.3) (Experiment 2).

	(1)

M=2	-5.00*** (0.832)
M=5	-8.94*** (1.207)
N=1000	-0.83 (0.921)
N=1m	-0.71 (0.909)
(M=2)×(N=1000)	2.15** (1.069)
(M=2)×(N=1m)	1.76* (1.017)
(M=5)×(N=1000)	3.50** (1.436)
(M=5)×(N=1m)	4.54*** (1.357)
Constant	45.65*** (0.767)
Obs	2772

Notes: Clustered standard errors in parentheses. *** Significant at 1% significance level; ** 5%; * 10%.



(e) Compare the strength of the GF across the 10-flip, 1000-flip, and 1-million-flip sets (question types I.A.3. I.B.2, and I.C.2) (Experiment 2).

	10-flip	**1,000-flip**	**1,000,000-flip**
Number of prior H's	-1.30*** (0.119)	-0.81*** (0.119)	-0.81*** (0.119)
Constant	46.97	44.56	44.83

	(0.737)	(0.736)	(0.741)
Obs	3080	3080	3080

Note: Standard errors in parentheses. *** Significant at 1% significance level; ** 5%; * 10%.

(f) Compare the estimated parameter-uncertainty models from the different question types.

(Beta-binomial calibrations were omitted because the parameters are identical)

M	**10-flip**	**1,000-flips**	**1,000,000-flips**
1	45.65% (0.765)	44.83% (0.839)	44.94% (0.807)
2	40.66% (0.905)	41.98% (0.972)	41.70% (0.969)
5	36.71% (1.170)	39.39% (1.073)	40.55% (1.056)
1 vs. 2 *p*-value	<0.0001	0.0037	0.0009
2 vs. 5 *p*-value	0.0008	0.0162	0.2738

Note: Standard errors in parentheses.

(g) Bin order effects across all sample sizes

i. Do participants' beliefs about the first and last bins differ from each other (pooling across histograms)?

We test whether participants weight the first and last bins differently across histograms where the first and last bins are symmetric.

Our procedure is as follows: To enable us to compare first vs. last bin responses across the various questions in our survey, we first normalize responses by question. That is, for each first bin response, we subtract the mean response for the given question across all respondents and across all bins. We then divide this value by the standard deviation of responses within-question. We do the same procedure for the last bin responses in each question. Once we have these normalized first and last bin responses, we subtract the last bin response from the first bin response for each question for each person. We test whether this difference is significantly different from zero by running a linear regression of the difference on a constant, clustering standard errors by participant.

In Experiment 1, we see evidence of such bin-order effects. That is, there is significant evidence ($p = 0.0000$) that participants put more weight on the first bin than the symmetric last bin.

We do not see these bin order effects in Experiment 2. That is, the difference between weight on the first bin is not statistically different from the weight on the last bin ($p = 0.156$).

ii. Do participants' beliefs about the first and last bins differ from the true probabilities?

We report results for each histogram question individually, and for pooled analysis across all histogram questions, for Experiment 1 and Experiment 2.

(a) To test whether the first and last bin responses were significantly different from the truth, we ran paired t-tests for each question, separately for first bins vs truth and last bin vs truth. We report the results for each question below.

Experiment 1 results:

In Experiment 1, we find that the reported belief about the *first* bin is significantly at $p < 0.005$ larger than the truth for the following questions: 10-flip 11-bin, 1000-flip 11-bin (where the middle bin was 451-549), 1000-flip 3-bin (where the middle bin was 451-549), 1000-flip 3-bin (where the middle bin was 481-519), and 1000-flip 11-bin (where the middle bin was 481-519). The reported belief is not significantly different ($p = 0.445$) than the true probability for the first bin of the 10-flip 5-bin histogram. Interestingly, the reported belief is significantly *lower* ($p = 0.005$) for the first bin than the truth for the 10-flip 3-bin histogram.

In Experiment 1, we find that the reported belief about the *last* bin is significantly at $p < 0.005$ larger than the truth for the following questions: 10-flip 11-bin, 1000-flip 11-bin (where the middle bin was 451-549), 1000-flip 3-bin (where the middle bin was 451-549), 1000-flip 3-bin (where the middle bin was 481-519), and 1000-flip 11-bin (where the middle bin was 481-519). The reported belief is significantly *lower* for the last bin than the truth for the 10-flip 3-bin histogram ($p = 0.0000$) and for the 10-flip 5-bin histogram ($p = 0.001$).

Experiment 2 results:

In Experiment 2, we find that the reported belief about the first bin is significantly at $p < 0.005$ larger than the truth for the following questions: 10-flip 11-bin, 1000-flip 11-bin (where the middle bin was 451-549), 1000-flip 3-bin (where the middle bin was 451-549), 1000-flip 3-bin (where the middle bin was 481-519), and 1000-flip 11-bin (where the middle bin was 481-519). The predicted probabilities are less significantly different than the true probabilities for the first bin of the 10-flip 5-bin histogram ($p = 0.045$) and the 1000-flip 3-bin equalized histogram ($p = 0.042$). Interestingly, the predicted probability is significantly lower for the first bin than the truth for the 10-flip 3-bin histogram ($p = 0.0000$).

In Experiment 2, the belief about the following last bins are significantly at $p < 0.005$ larger than the truth: 10-flip 11-bin, 1000-flip 11-bin, 1000-flip 3-bin narrowed, 1000-flip 3-bin, 1000-flip 11-bin narrowed, 1000-flip 5-bin equalized, 1 million-flip 11-bin, 1 million-flip 5-bin equalized. The truth and predicted probabilities are less significantly different for the last bin of the 10-flip 5-bin histogram ($p = 0.011$) and not significantly different for the 1-million-flip 3-bin equalized histogram ($p = 0.574$) and the 1000-flip 3-bin equalized histograms ($p = 0.166$). The truth is significantly higher than the predicted probability for the last bin of the 10-flip 3-bin histogram ($p = 0.0000$) and 1000-flip 3-bin histogram (where the middle bin = 500) ($p = 0.0000$).

(b) Our approach for pooling across all histograms is as follows: First, we subtract the true probability from the first bin response for each histogram for each person. In order to make these differences comparable across histograms, we divide by the standard deviation of these differences within histogram. This mitigates the concern that questions with a larger true probability would skew the comparison. We then run a linear regression of these values on a constant, and cluster standard errors by participant. We run this regression separately for first and last bins. We report the results for Experiment 1 and Experiment 2 below.

Experiment 1 Results:

Pooling across histograms, we find significant evidence ($p = 0.0000$) that the belief about the first bin is larger than the truth. Similarly we find significant evidence ($p = 0.0000$) that the belief about the last bin is larger than the truth.

Experiment 2 Results:

Pooling across histograms, we find significant evidence ($p = 0.0000$) that the belief about the first bin is larger than the truth. Similarly we find significant evidence ($p = 0.0000$) that belief about the last bin is larger than truth.

(h) Heads vs tails bias

In this analysis, we test whether participants' beliefs suffered from biases about the probabilities of heads and tails being different. To test this, we run an OLS regression of participants' responses to the middle bin for each histogram question on a dummy variable indicating whether the participants saw the question framed as number of heads (= 1) or number of tails (= 0).

Experiment 1 Results

	10-flip 11-bin	**10-flip 5-bin**	**10-flip 3-bin**	**1000-flip 11-bin**	**1000-flip 3-bin (481-519)**	**1000-flip 3-bin (451-549)**
Dummy for Heads	-2.04 (3.412)	1.42 (3.344)	-2.94 (3.417)	1.49 (3.515)	1.31 (3.41)	0.01 (3.625)
Constant	21.10*** (2.518)	27.29*** (2.556)	37.71*** (2.452)	17.83*** (2.423)	34.87*** (2.504)	39.58*** (2.512)
Observations	101	101	101	101	102	102

Note: Standard errors in parentheses. *** Significant at 1% significance level; ** 5%; * 10%.

Experiment 2 Results

	10-flip 11-bin	10-flip 5-bin	10-flip 3-bin	1000-flip 11-bin	1000-flip 3-bin (481-519)	1000-flip 3-bin (451-549)
Dummy for Heads	-1.18 (1.855)	0.93 (1.944)	3.24 (2.075)	-2.35 (2.923)	1.56 (2.965)	2.36 (3.149)
Constant	28.90*** (1.357)	31.66*** (1.32)	31.18*** (1.501)	35.31*** (2.026)	35.47*** (2.117)	39.88*** (2.277)
Observations	308	308	308	308	308	308

Note: Standard errors in parentheses. *** Significant at 1% significance level; ** 5%; * 10%.

	1m-flip 11-bin	1m-flip 5-bin	1m-flip 3-bin
Dummy for Heads	-2.21 (3.172)	-0.05 (2.876)	3.23 (3.056)
Constant	35.83*** (2.080)	25.64*** (1.966)	32.76*** (2.140)
Observations	307	308	308

Note: Standard errors in parentheses. *** Significant at 1% significance level; ** 5%; * 10%.

(i) Order of increasing vs decreasing number of heads bias

In this analysis, we test whether participants' beliefs suffered from biases related to whether the number of heads in the histogram questions were increasing or decreasing. To test this, we run an OLS regression of participants' responses to the middle bin for each histogram question on a dummy variable indicating whether the participants saw the question in increasing order of heads (= 1) or in decreasing order of heads (= 0).

Experiment 1 Results

	10-flip 11-bin	10-flip 5-bin	10-flip 3-bin	1000-flip 11-bin	1000-flip 3-bin (481-519)	1000-flip 3-bin (451-549)
Dummy for Increasing	-1.20 (3.403)	-1.74 (3.308)	-9.73*** (3.300)	-4.61 (3.484)	2.25 (3.397)	-9.10** (3.517)
Constant	20.61*** (2.442)	29.07*** (2.441)	41.50*** (2.435)	20.91*** (2.500)	34.41*** (2.449)	43.78*** (2.388)

Observations	101	101	101	101	102	102

Note: Standard errors in parentheses. *** Significant at 1% significance level; ** 5%; * 10%.

Experiment 2 Results

	10-flip 11-bin	**10-flip 5-bin**	**10-flip 3-bin**	**1000-flip 11-bin**	**1000-flip 3-bin (481-519)**	**1000-flip 3-bin (451-549)**
Dummy for Increasing	-1.498 (1.85)	-0.818 (1.939)	-3.196 (2.074)	4.184 (2.927)	4.239 (2.963)	-1.512 (3.161)
Constant	29.04*** (1.325)	32.51*** (1.384)	34.52*** (1.486)	32.29*** (1.973)	34.30*** (2.019)	41.81*** (2.131)
Observations	308	308	308	308	308	308

Note: Standard errors in parentheses. *** Significant at 1% significance level; ** 5%; * 10%.

	1m-flip 11-bin	**1m-flip 5-bin**	**1m-flip 3-bin**
Dummy for Increasing	6.273** (3.137)	1.198 (2.876)	6.491** (3.039)
Constant	31.44*** (2.321)	25.06*** (1.96)	31.10*** (2.149)
Observations	307	308	308

Note: Standard errors in parentheses. *** Significant at 1% significance level; ** 5%; * 10%.

I.E. Overall Tests of Bias

(a) Betting questions (I.A.1): Actual earnings / earnings possible from always choosing the larger amount of money.

Note: As we discussed in the text, we planned to measure the expected utility cost from deviations from correct beliefs, and tried to design the experiment (subject to the constraints of limited budget and asking lots of questions) to maximize those costs. We

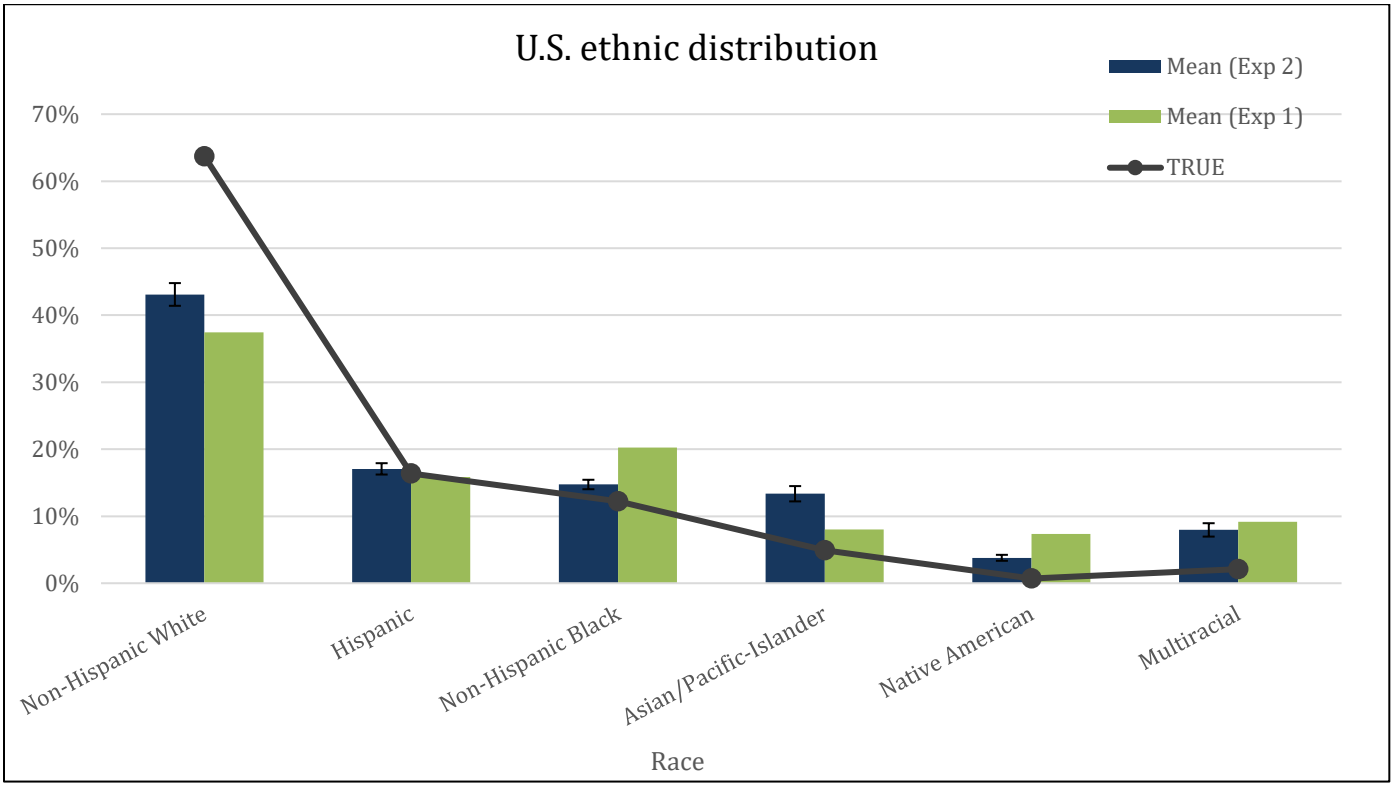
calculate the expected amount of payoff foregone from participants' mean responses in each question relative to the correct responses.

The results are shown in Appendix Tables H.6.A and H.6.B.

Section II. Questions of Secondary Interest

II.A. U.S. ethnic distribution

Consistent with our other hypotheses, we expect to see "binning" effects that bias responses toward equal probabilities on all bins. Subjects will overestimate population groups that constitute a small percentage of US population (Native Americans) and underestimate the largest groups (non-Hispanic Whites). We plan to report the results for readers to interpret as they will, but not integrate the findings into our formal analysis of the other questions.



Note: True U.S. ethnic distribution obtained from the 2010 US Census Bureau (link: http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf)

II.B. Robustness

(a) Order effects (Experiment 2).

i. Do results look different when the question appeared early vs. late in the experiment? Although we try to neutralize any inference about the right answer or possible twists in how we developed the data, in principle some seemingly inconsistent answers over the course of the experiment may be because subjects changed their inference about the process as the questions changed. If we find no significant order effects besides what appears to be pure fatigue, that would make such explanations less likely.

Order effects across the experiment

We compare answers to streak questions by whether the subject saw this question early, middle, or late in the experiment. We do not observe any systematic order effects.

10-flip:

Number of prior H's	**Early***	**Middle****	**Late*****	***p*-value ANOVA equal means**	***p*-value ANOVA equal variance**
0	49.96 (5.37)	50.08 (7.63)	48.57 (9.77)	0.30	<0.001
1	45.43 (11.95)	44.57 (14.24)	43.90 (14.45)	0.71	0.10
2	43.49 (18.02)	39.62 (18.69)	39.03 (15.91)	0.13	0.25
3	41.95 (17.70)	38.97 (19.86)	38.27 (17.29)	0.29	0.34
4	40.99 (18.08)	37.73 (20.11)	38.01 (18.94)	0.38	0.57
5	40.69 (19.00)	36.05 (20.74)	37.36 (19.76)	0.22	0.68
6	40.91 (20.34)	35.03 (22.55)	36.29 (20.29)	0.10	0.49
7	40.26 (20.22)	33.20 (21.71)	37.28 (21.24)	0.06	0.76
8	38.32 (20.80)	33.14 (22.68)	35.74 (23.08)	0.25	0.52
9	37.18 (22.21)	33.94 (23.76)	34.53 (24.25)	0.56	0.64
Obs	**112**	**92**	**104**		

Note: Standard deviations in parentheses. *P*-value ANOVA equal variance is calculated using Bartlett's test for equal variances (chi-square test).

* "Early" defined as 10-flip questions appearing before 1,000-flip and 1,000,000-flip questions

** "Middle" defined as 10-flip questions appearing between 1,000-flip and 1,000,000-flip questions

*** "Late" defined as 10-flip questions appearing after 1,000-flip and 1,000,000-flip questions.

1,000-flip:

Number of prior H's	**Early**	**Middle**	**Late**	***p*-value ANOVA equal**	***p*-value ANOVA equal**

				means	variance
0	43.66 (16.89)	43.93 (17.59)	43.68 (15.32)	0.99	0.35
1	41.24 (16.09)	45.50 (13.72)	45.78 (12.86)	0.05	0.07
2	39.18 (18.35)	43.47 (16.63)	43.56 (16.38)	0.13	0.48
3	36.84 (21.41)	41.56 (18.30)	42.37 (16.68)	0.09	0.04
4	38.97 (19.11)	39.64 (19.37)	40.80 (18.58)	0.79	0.91
5	35.27 (20.58)	39.99 (19.75)	40.45 (16.51)	0.11	0.07
6	36.58 (21.26)	38.55 (20.20)	39.12 (20.05)	0.66	0.82
7	36.82 (21.54)	40.29 (19.48)	39.69 (19.69)	0.44	0.55
8	33.84 (22.47)	38.59 (20.80)	38.02 (20.21)	0.23	0.56
9	35.10 (22.68)	38.58 (21.03)	37.54 (20.85)	0.51	0.66
Obs	**91**	**109**	**108**		

1,000,000-flip:

Number of prior H's	**Early**	**Middle**	**Late**	***p*-value ANOVA equal means**	***p*-value ANOVA equal variance**
0	47.70 (13.94)	43.28 (16.27)	45.35 (14.07)	0.10	0.20
1	44.54 (14.98)	43.62 (14.42)	45.07 (12.92)	0.76	0.33
2	39.88 (18.31)	42.16 (16.28)	42.35 (16.56)	0.51	0.43
3	41.24 (19.42)	40.83 (18.52)	40.87 (17.43)	0.98	0.56
4	39.17 (20.42)	39.76 (19.37)	38.79 (20.07)	0.94	0.86
5	39.65 (21.03)	38.45 (19.94)	38.67 (19.57)	0.90	0.75
6	41.78 (21.42)	38.56 (18.68)	36.05 (21.04)	0.14	0.33
7	39.61 (20.70)	37.91 (19.84)	37.19 (20.53)	0.68	0.90
8	39.61 (21.57)	38.25 (19.77)	36.21 (22.04)	0.52	0.51

9	38.46 (24.04)	39.35 (21.26)	35.71 (21.50)	0.49	0.38
Obs	**105**	**107**	**96**		

We similarly look at the 11-bin histogram questions, and find no systematic evidence of fatigue:

10-flip 11-bin Histogram

Bins	**Early**	**Middle**	**Late**	***p*-value ANOVA equal means**	***p*-value ANOVA equal variance**
1	2.15 (3.33)	2.22 (2.97)	2.16 (3.37)	0.99	0.41
2	3.91 (5.44)	3.79 (3.45)	3.72 (3.55)	0.94	<0.001
3	5.55 (3.18)	5.61 (3.34)	5.41 (3.38)	0.91	0.79
4	9.20 (3.48)	9.31 (4.02)	9.28 (5.35)	0.98	<0.001
5	14.67 (5.16)	14.05 (5.62)	16.48 (7.06)	0.01	0.003
6	29.57 (16.68)	27.58 (16.25)	27.49 (15.74)	0.57	0.84
7	14.85 (5.20)	14.65 (5.61)	15.43 (6.12)	0.59	0.24
8	9.23 (3.19)	9.72 (4.42)	8.59 (3.43)	0.10	0.002
9	5.53 (3.22)	5.63 (3.32)	5.29 (3.20)	0.74	0.93
10	3.35 (2.92)	4.37 (5.95)	4.04 (5.80)	0.33	<0.001
11	3.35 (2.92)	3.08 (6.11)	2.15 (3.95)	0.17	<0.001
Obs	**112**	**92**	**104**		

1k-flip 11-bin Histogram

Bins	**Early**	**Middle**	**Late**	***p*-value ANOVA equal means**	***p*-value ANOVA equal variance**
1	2.29 (3.29)	2.56 (3.30)	3.61 (8.13)	0.19	<0.001
2	3.45	3.96	4.34	0.25	0.53

	(4.00)	(3.59)	(3.65)		
3	4.41 (3.59)	5.26 (3.37)	5.75 (3.22)	0.02	0.55
4	7.12 (5.64)	7.94 (3.78)	8.13 (3.62)	0.23	<0.001
5	12.49 (8.49)	14.50 (9.95)	13.21 (5.71)	0.21	<0.001
6	40.84 (29.90)	32.44 (23.74)	30.35 (22.51)	0.01	0.01
7	12.41 (8.13)	13.53 (5.83)	13.72 (6.98)	0.37	0.004
8	6.78 (6.10)	8.00 (3.76)	8.33 (3.35)	0.04	<0.001
9	4.10 (3.60)	5.38 (3.49)	6.00 (3.43)	0.0007	0.89
10	3.50 (4.03)	3.88 (3.49)	4.22 (3.67)	0.39	0.35
11	2.60 (3.88)	2.56 (3.33)	2.45 (2.88)	0.95	0.01
Obs	**91**	**109**	**108**		

1m-flip 11-bin Histogram

Bins	**Early**	**Middle**	**Late**	***p*-value ANOVA equal means**	***p*-value ANOVA equal variance**
1	4.21 (9.80)	3.26 (10.00)	3.48 (10.84)	0.78	0.56
2	4.79 (4.86)	3.80 (4.75)	3.86 (4.01)	0.22	0.13
3	5.67 (4.11)	4.99 (3.75)	4.94 (3.46)	0.30	0.23
4	7.84 (4.44)	7.40 (4.35)	7.54 (4.27)	0.75	0.93
5	12.26 (7.83)	14.36 (10.48)	13.27 (7.01)	0.21	<0.001
6	32.83 (27.71)	36.40 (27.76)	35.48 (26.97)	0.62	0.95
7	11.64 (6.72)	12.34 (7.71)	13.15 (6.45)	0.31	0.16
8	7.46 (4.38)	6.95 (4.00)	7.43 (4.17)	0.61	0.66
9	5.54 (3.80)	4.88 (3.66)	5.02 (3.56)	0.40	0.81
10	4.47 (4.07)	3.63 (4.13)	3.66 (3.42)	0.22	0.13

11	3.35 (6.85)	2.23 (4.71)	2.17 (2.83)	0.18	<0.001
Obs	**105**	**107**	**96**		

(b) Self-reports of effort

Experiment 2

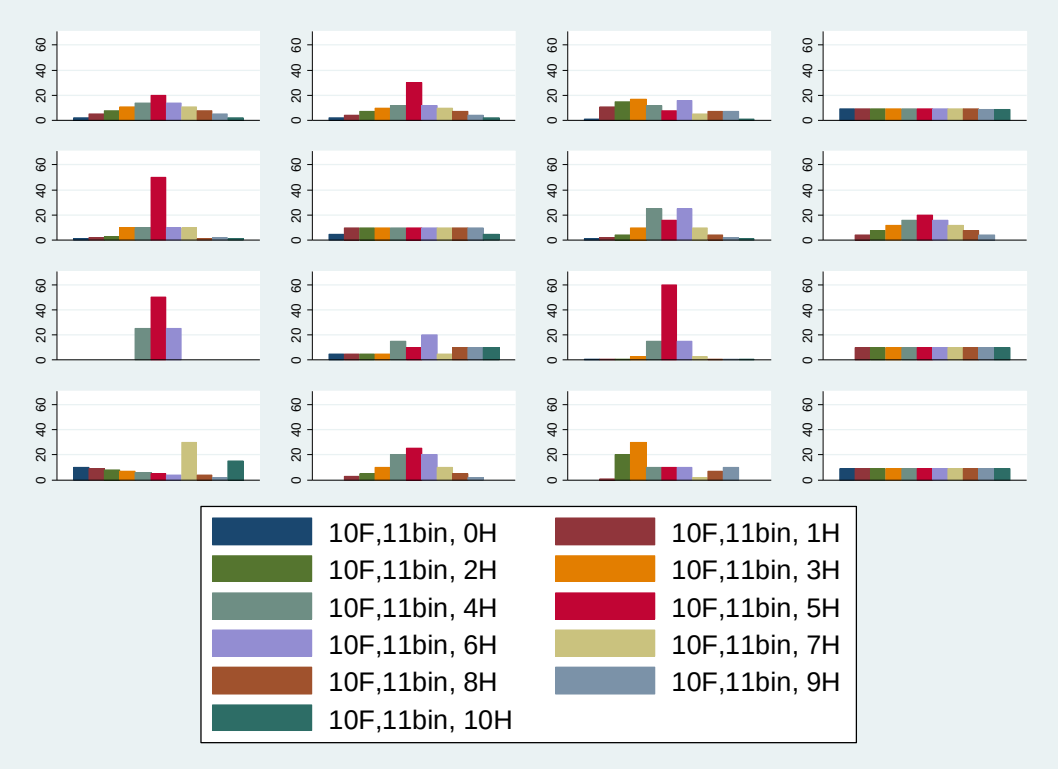
Below are tabulations of self-reported effort questions [1 = "I went through as quickly as I could" to 5 = "I put a lot of effort into making the best guesses I could"]. Very few people report that they went through each question as fast as they could.

	"Please tell us how much thought and effort you put in to answering the questions on this survey"	"There were a number of questions on which you had to estimate the frequency of each of a set of possible outcomes. How much effort did you exert?"
1	10	16
2	26	32
3	86	72
4	108	102
5	71	79
Missing	7	7

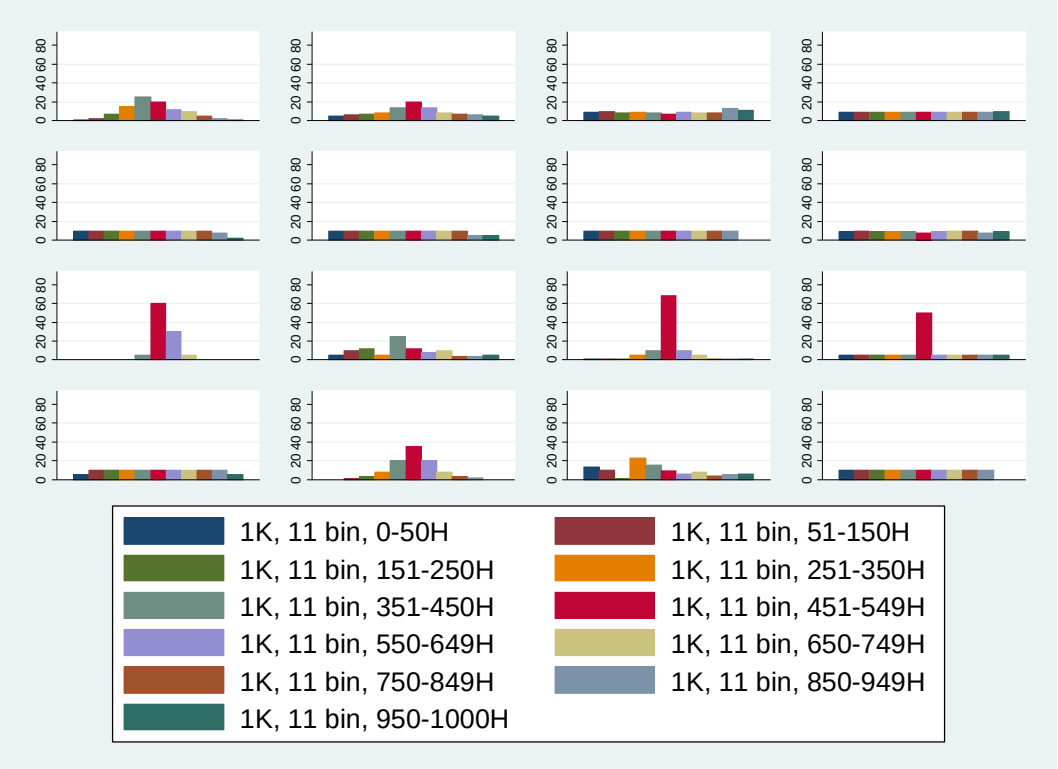
What we notice is that minimum-effort subjects' responses for histogram questions tend to be more "evened out".

Effort for frequency questions (i.e. "there were a number of questions on which…")	Approximately equal answers to 10-flip 11-bin questions (all bins given 8-10%)	Approximately equal answers to 1k-flip 11-bin questions (all bins given 8-10%)	Approximately equal answers to 1m-flip 11-bin questions (all bins given 8-10%)
1-3	8.3%	8.3%	8.3%
4-5	4.4%	3.3%	2.2%
p-value	0.162	0.058	0.053

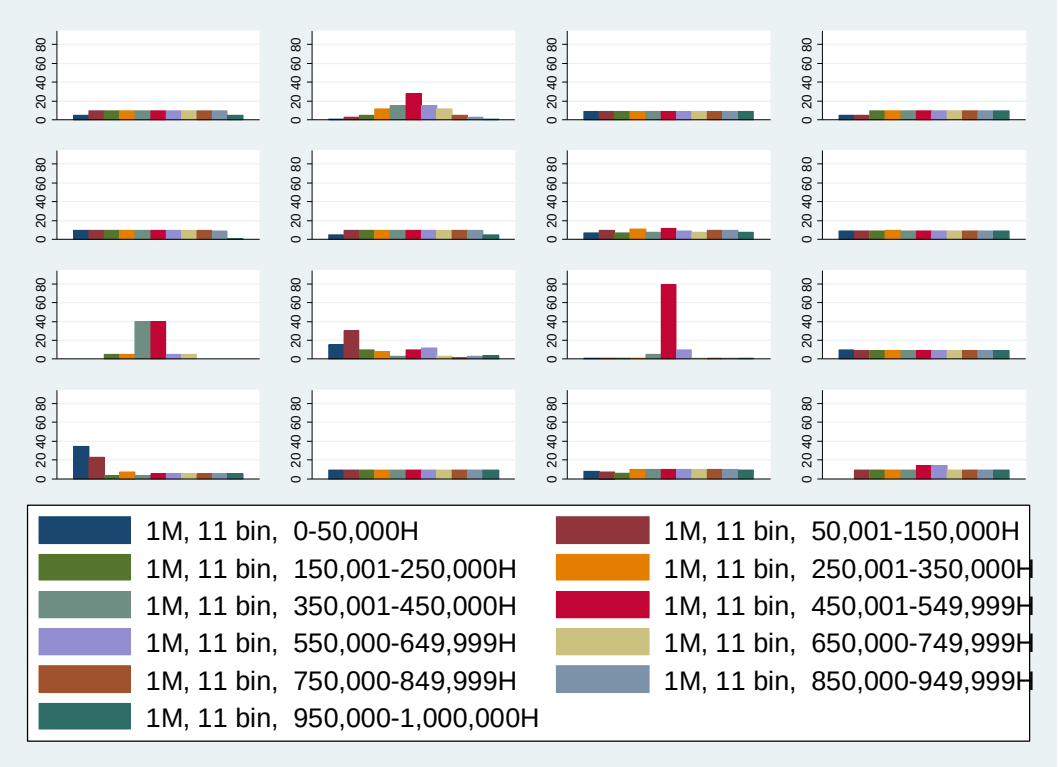
Below are 10-flip 11-bin histograms of 16 minimum-effort subjects:



Below are 1k-flip 11-bin histograms of 16 minimum-effort subjects:

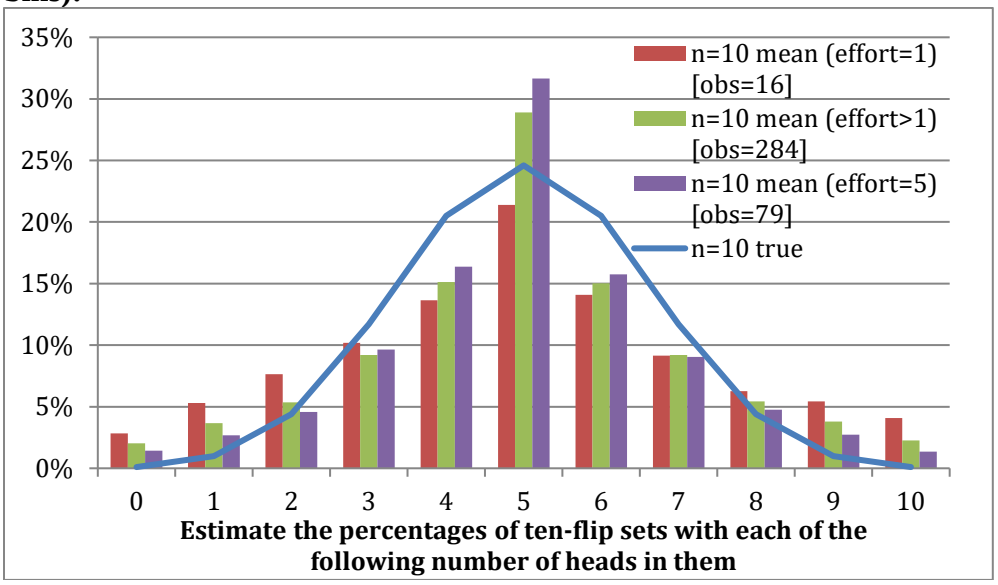


Below are million-flip 11-bin histograms of 16 minimum-effort subjects:

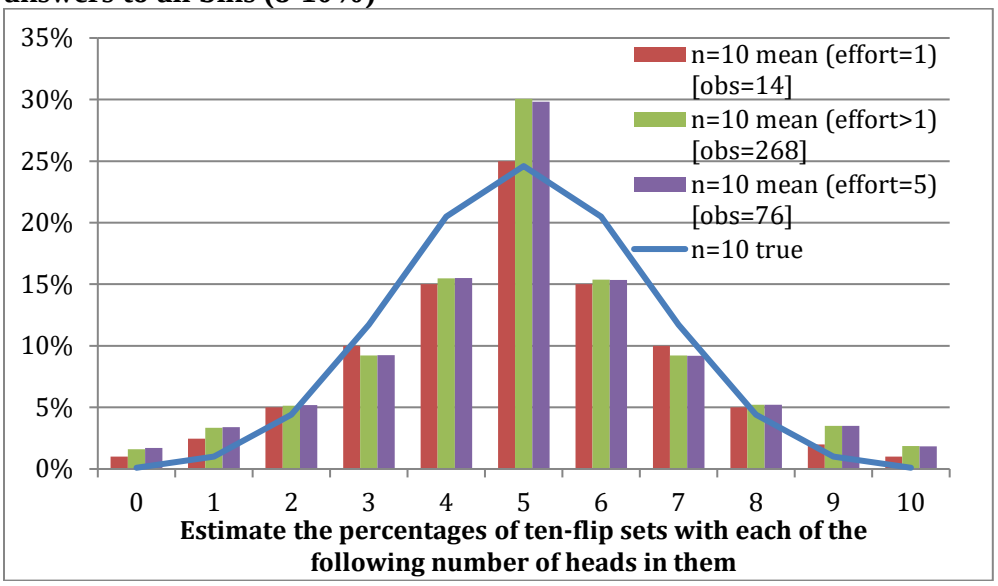


Next, we compare the responses to various histogram questions between low- and high-effort subjects.

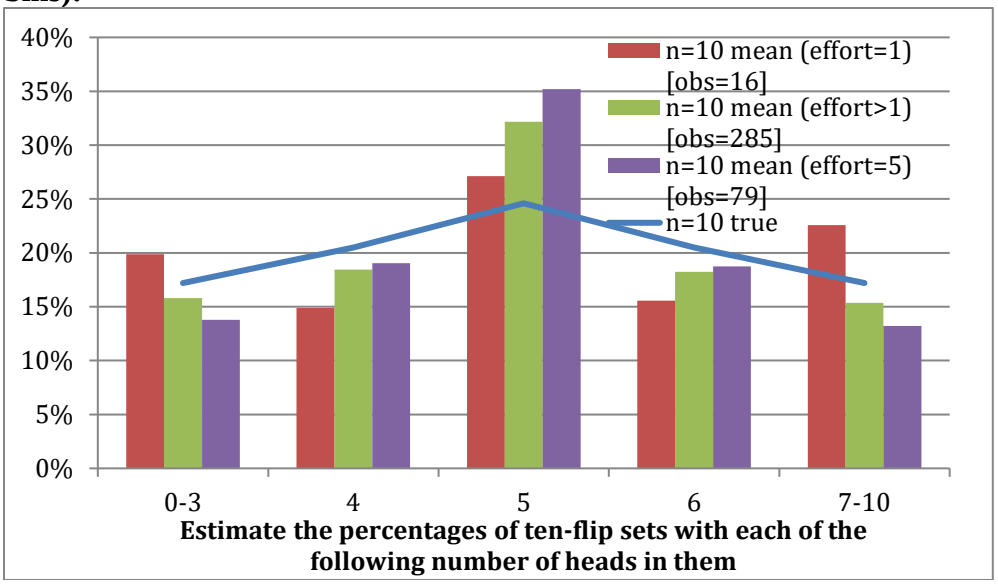
10-flip 11-bin Histogram (including subjects who put approx. equal answers to all bins):



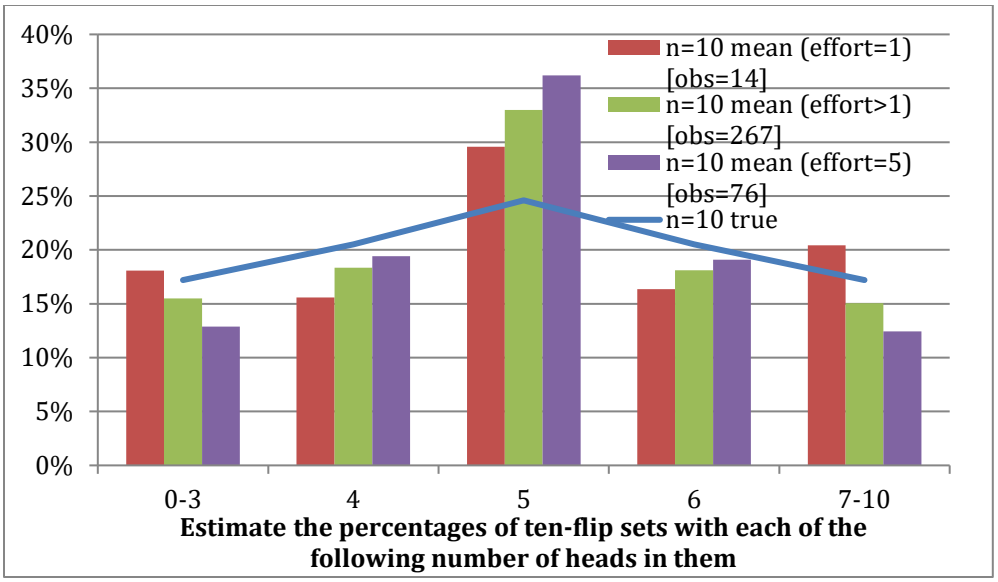
10-flip 11-bin Histogram after dropping subjects who put approximately equal answers to all bins (8-10%)



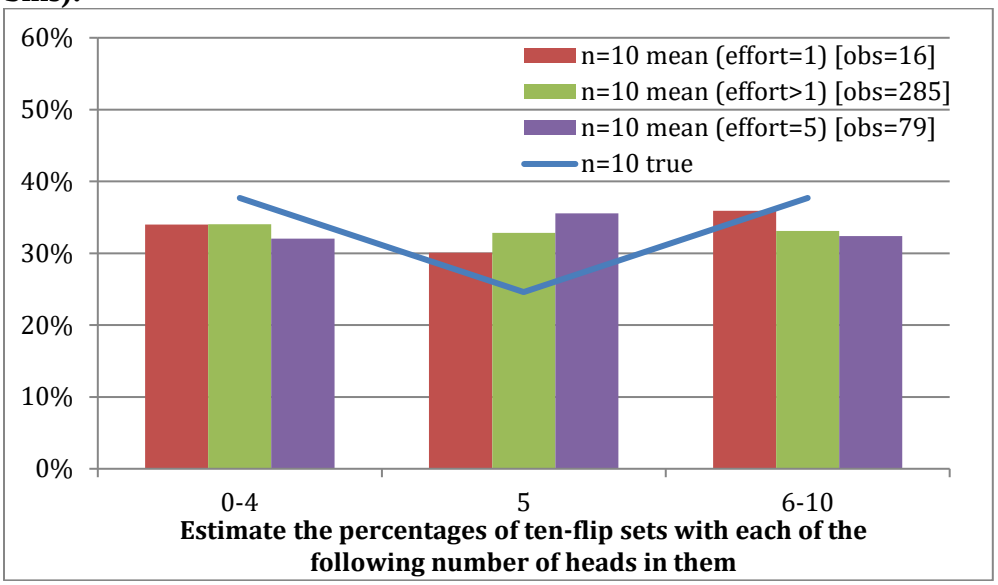
10-flip 5-bin Histogram (including subjects who put approx. equal answers to all bins):



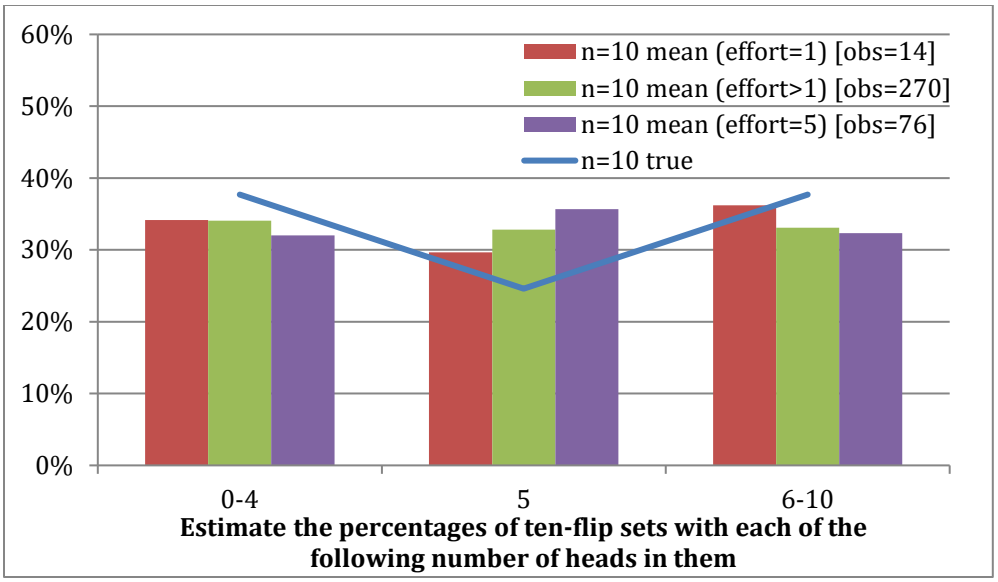
10-flip 5-bin Histogram after dropping subjects who put approximately equal answers to all bins (19-21%)



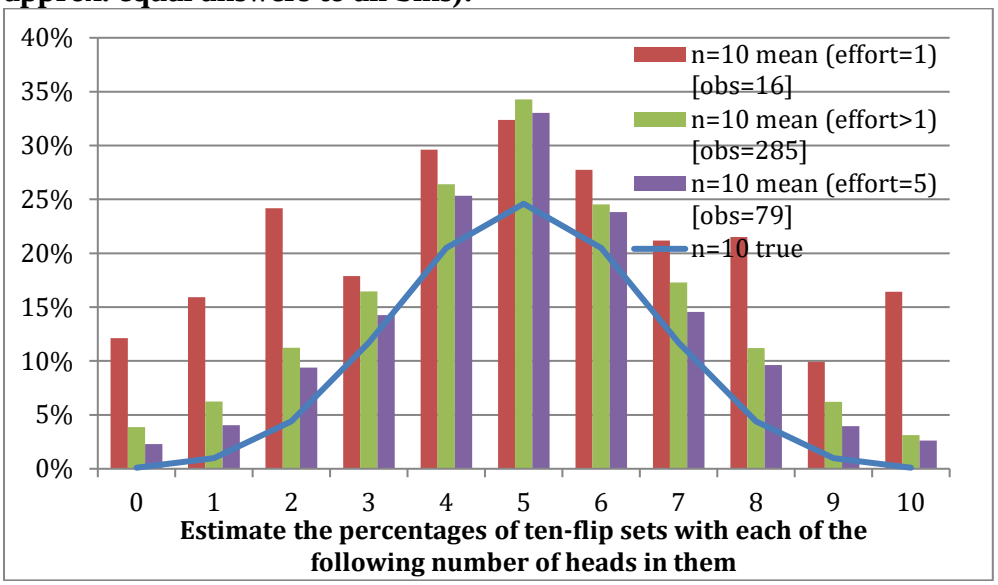
10-flip 3-bin Histogram (including subjects who put approx. equal answers to all bins):



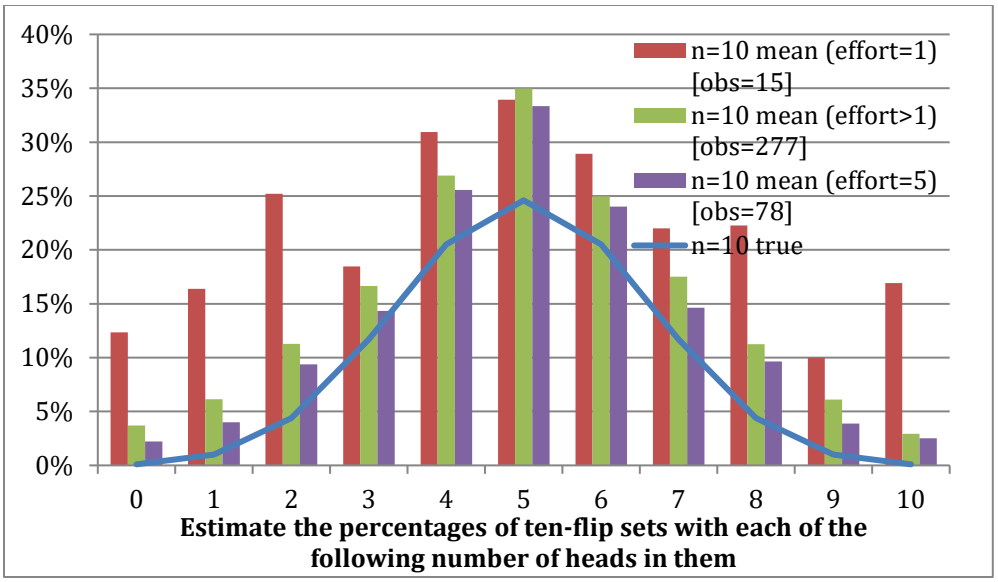
10-flip 3-bin Histogram after dropping subjects who put approximately equal answers to all bins (32-34%)



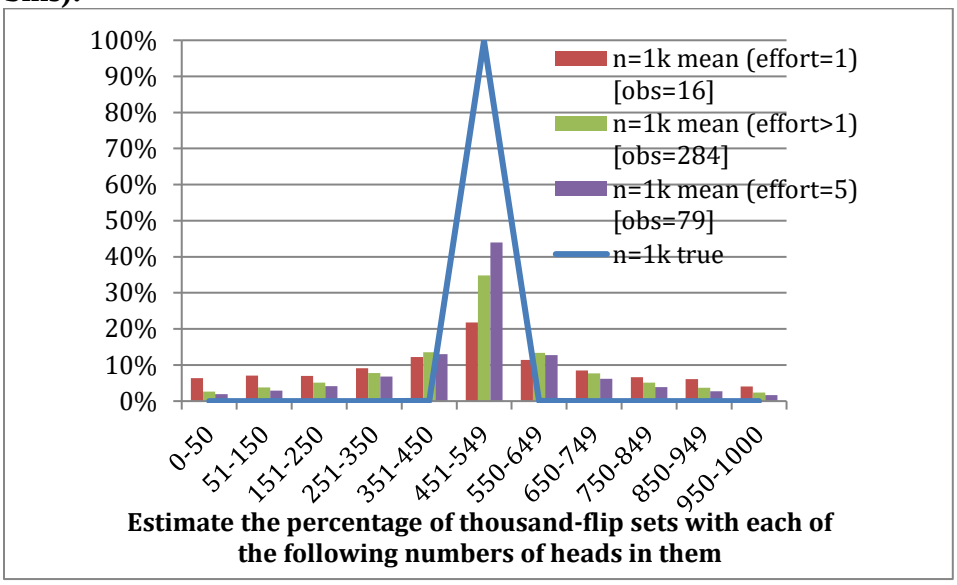
10-flip 11-bin separate-bin-elicitation histogram (including subjects who put approx. equal answers to all bins):



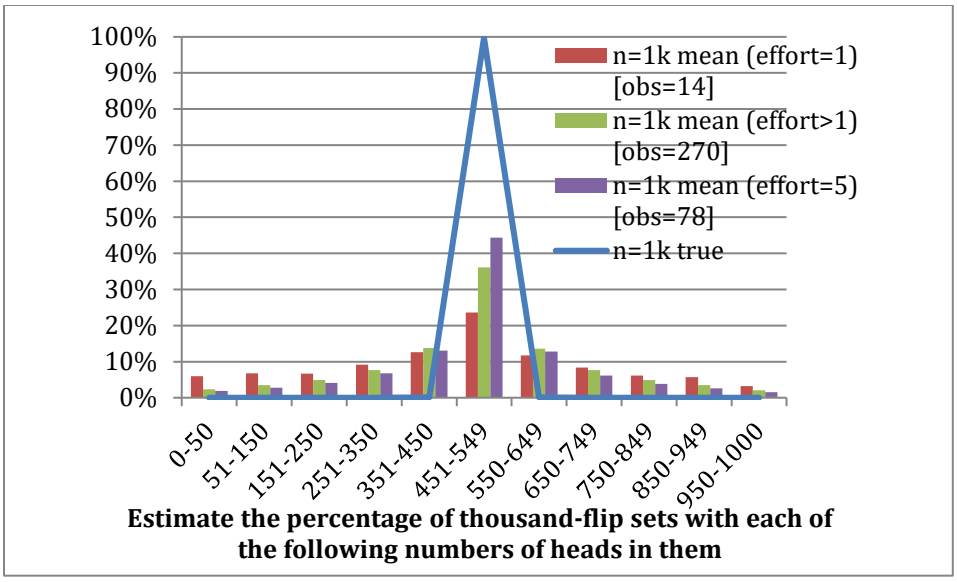
10-flip 11-bin separate-bin-elicitation histogram after dropping subjects who put approximately equal answers to all bins (8-10%):



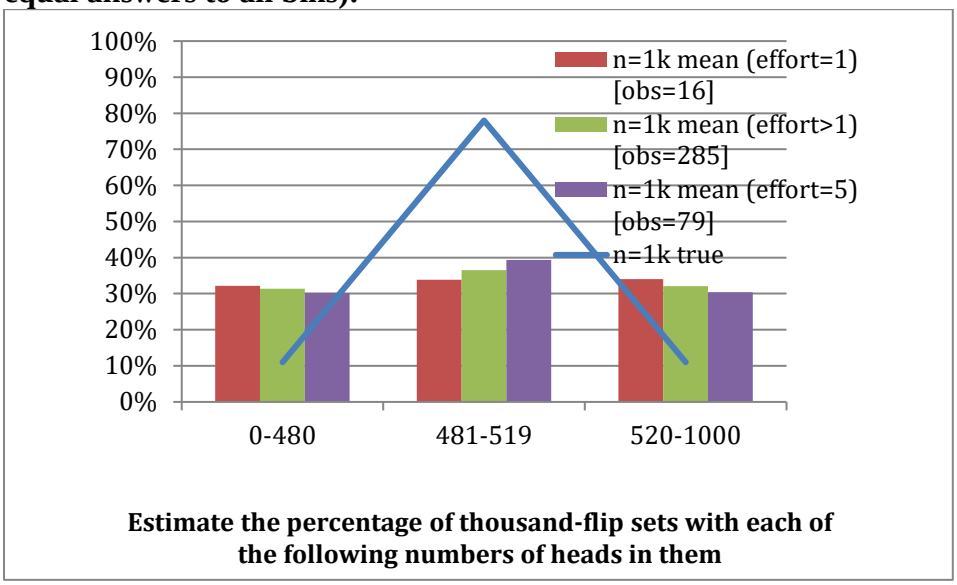
1k-flip 11-bin histogram (including subjects who put approx. equal answers to all bins):



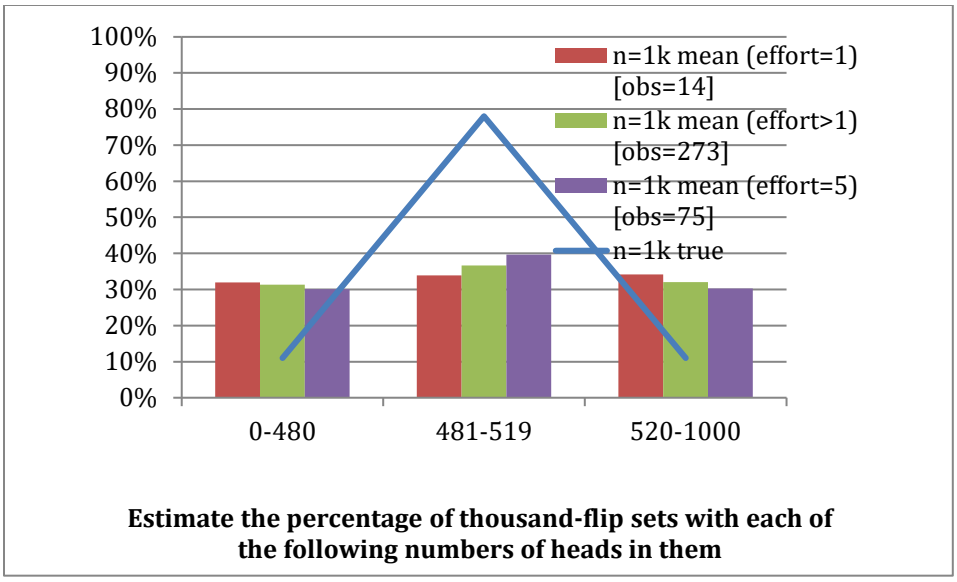
1k-flip 11-bin histogram after dropping subjects who put approximately equal answers to all bins (8-10%):



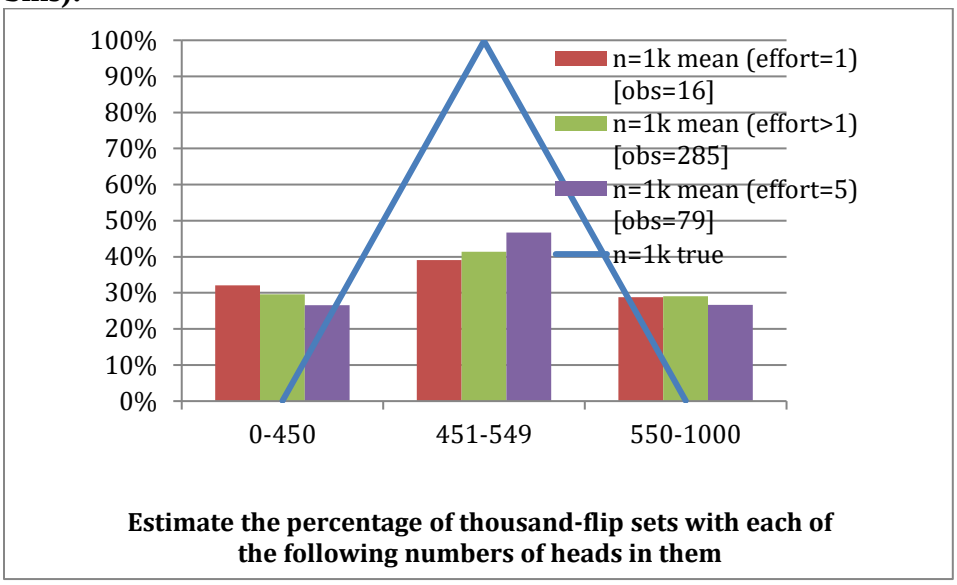
1k-flip 3-bin (middle bin = 481-519) Histogram (including subjects who put approx. equal answers to all bins):



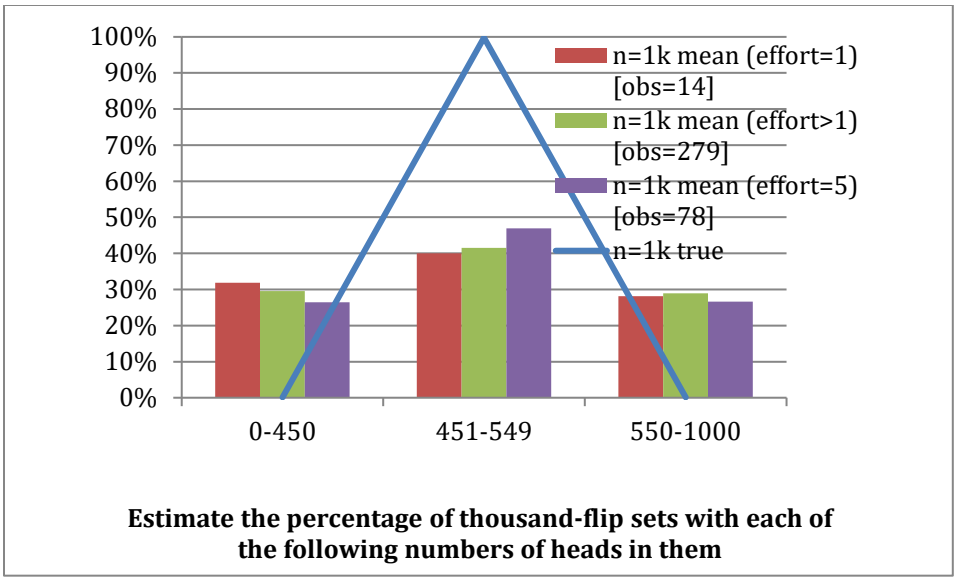
1k-flip 3-bin Narrow Histogram after dropping subjects who put approximately equal answers to all bins (32-34%):



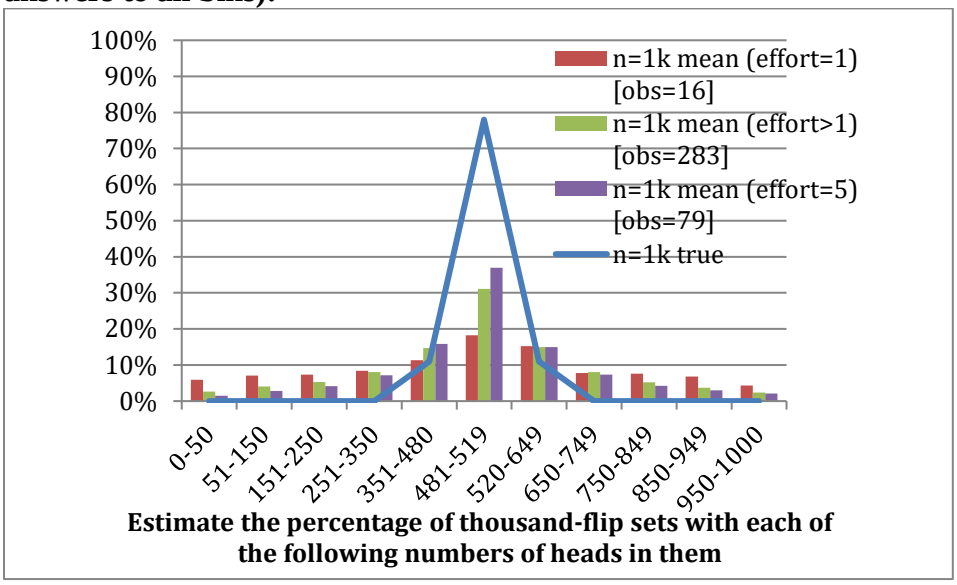
1k-flip 3-bin Histogram (including subjects who put approx. equal answers to all bins):



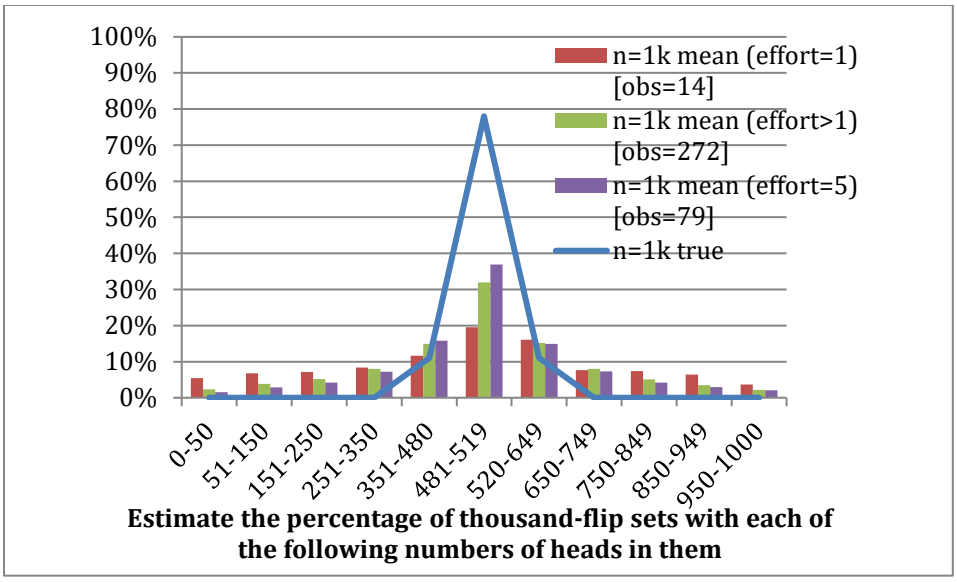
1k-flip 3-bin Histogram after dropping subjects who put approximately equal answers to all bins (32-34%):



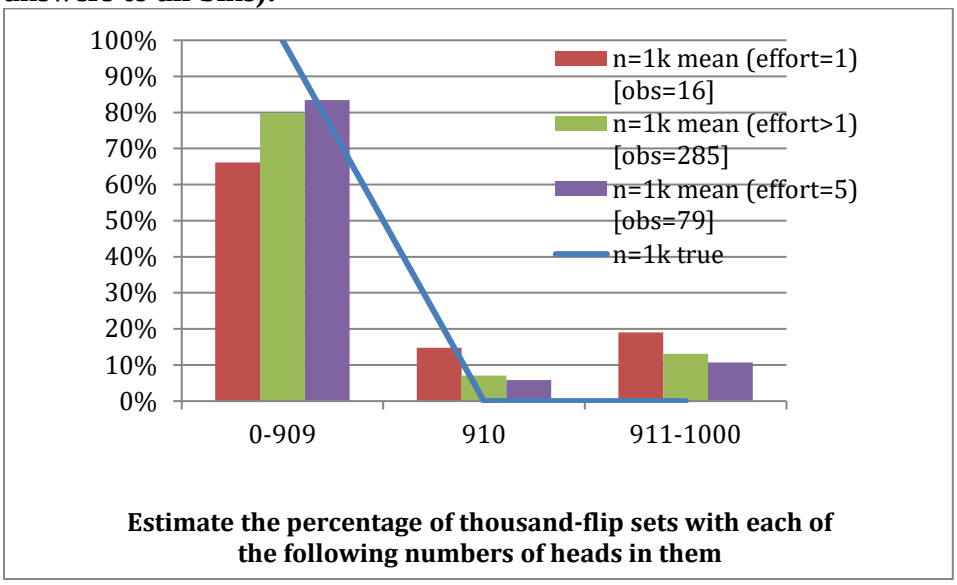
1k-flip 11-bin Narrow Histogram (including subjects who put approx. equal answers to all bins):



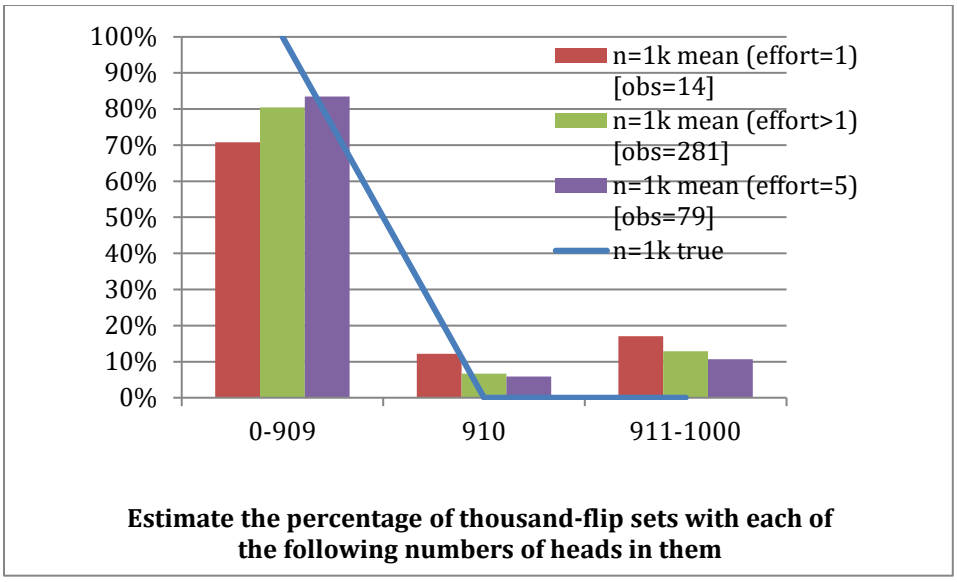
1k-flip 11-bin Narrow Histogram after dropping subjects who put approximately equal answers to all bins (8-10%):



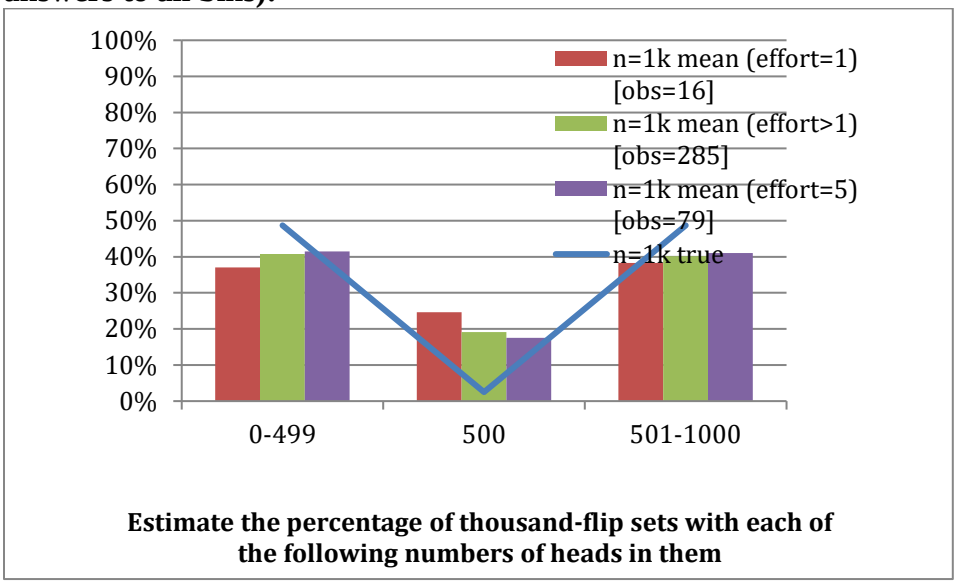
1k-flip 3-bin Distorted-900 Histogram (including subjects who put approx. equal answers to all bins):



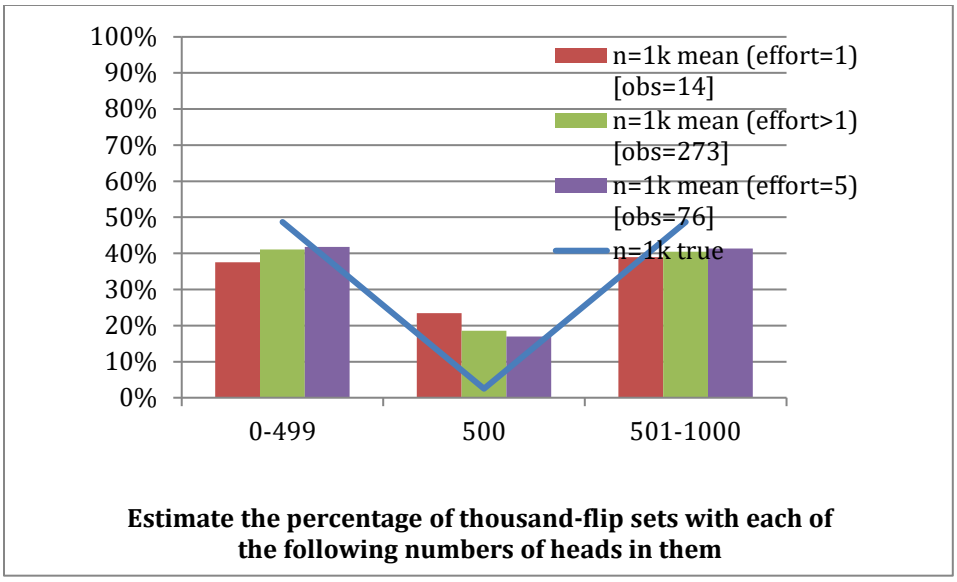
1k-flip 3-bin Distorted-900 Histogram after dropping subjects who put approximately equal answers to all bins (8-10%):



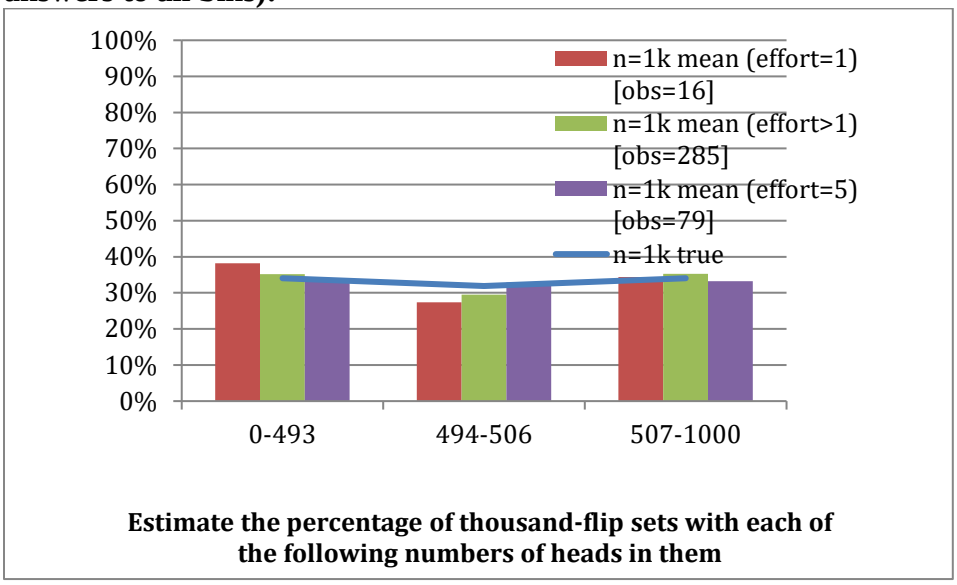
1k-flip 3-bin Distorted-500 Histogram (including subjects who put approx. equal answers to all bins):



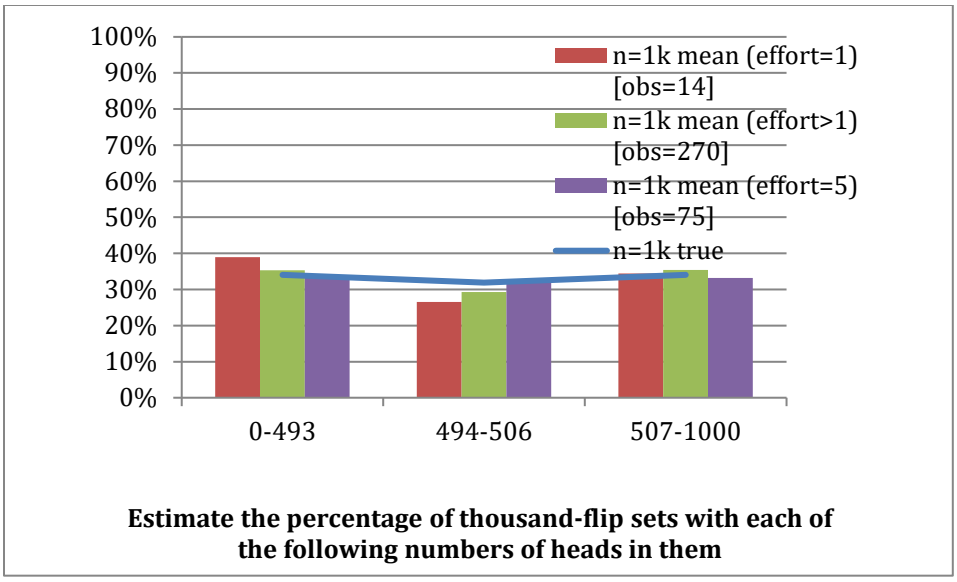
1k-flip 3-bin Distorted-500 Histogram after dropping subjects who put approximately equal answers to all bins (8-10%):



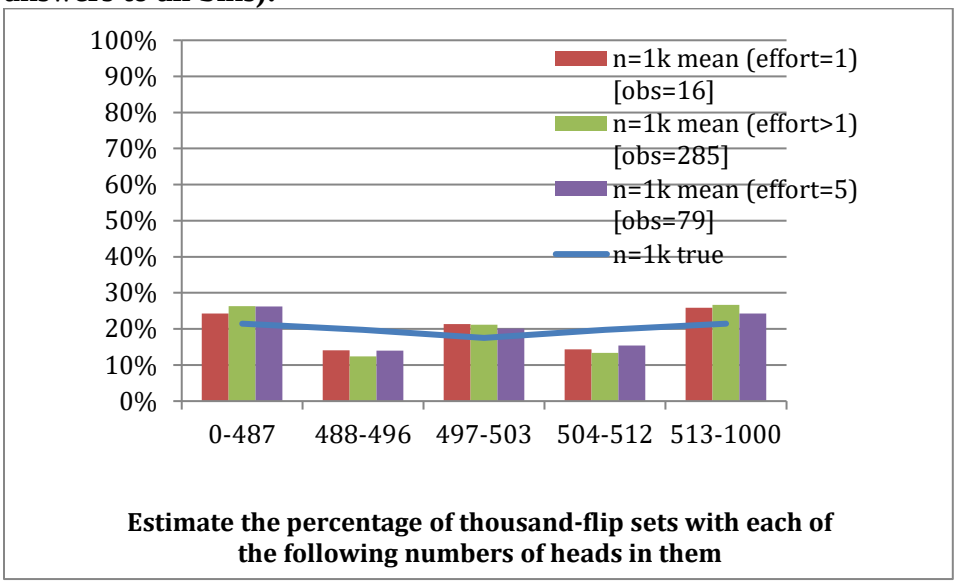
1k-flip 3-bin Equalized Histogram (including subjects who put approx. equal answers to all bins):



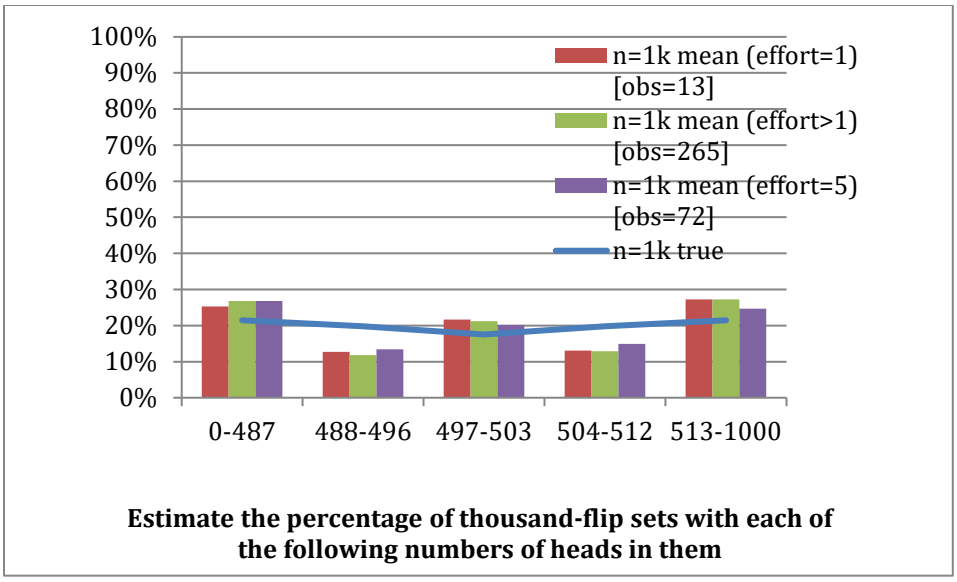
1k-flip 3-bin Equalized Histogram after dropping subjects who put approximately equal answers to all bins (32-34%):



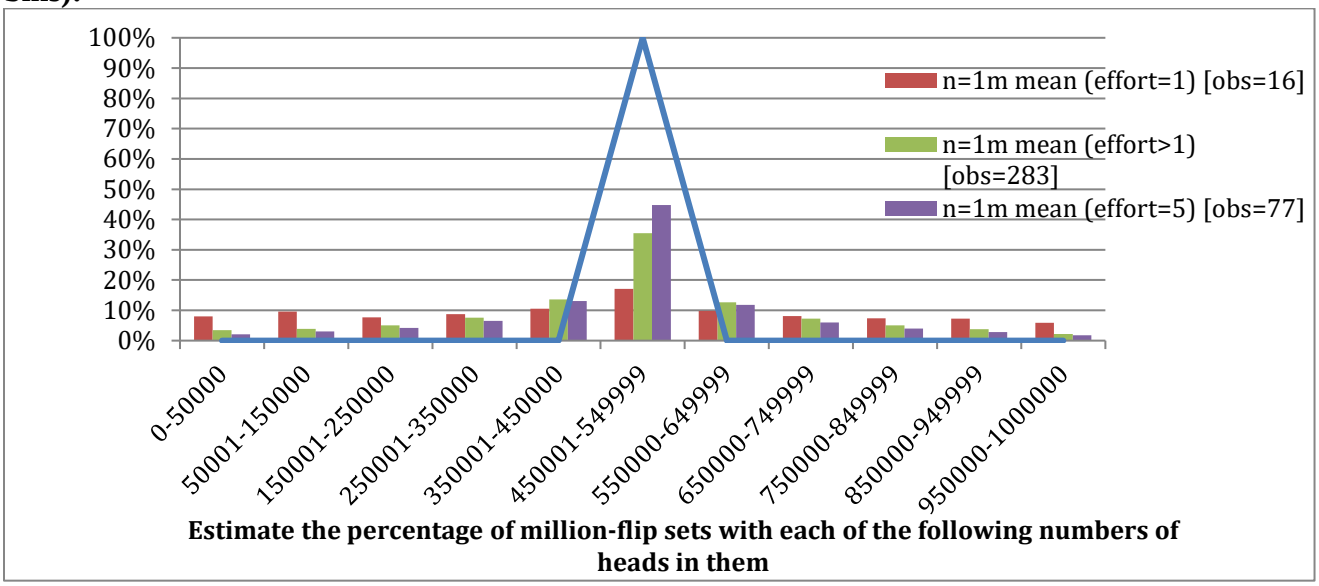
1k-flip 5-bin Equalized Histogram (including subjects who put approx. equal answers to all bins):



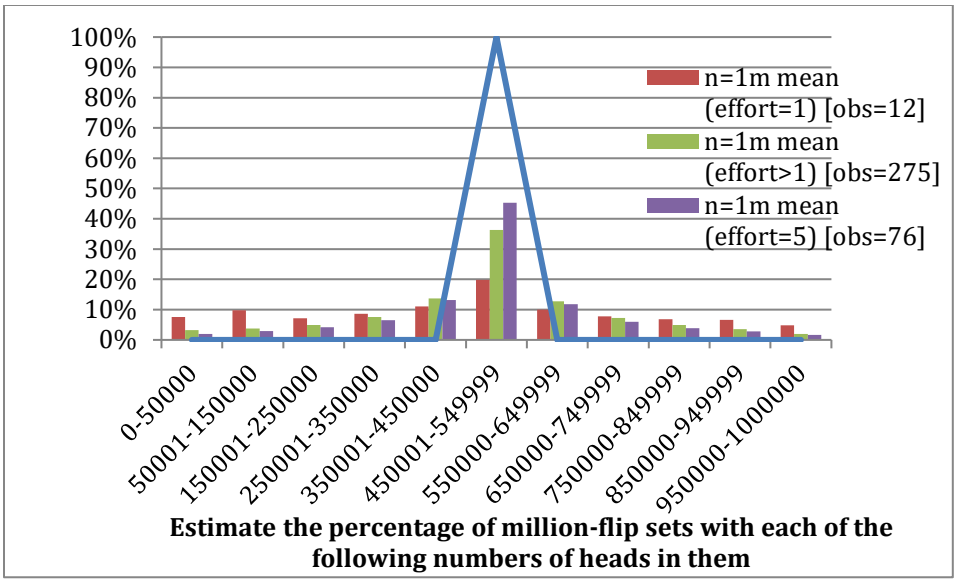
1k-flip 5-bin Histogram after dropping subjects who put approximately equal answers to all bins (19-21%):



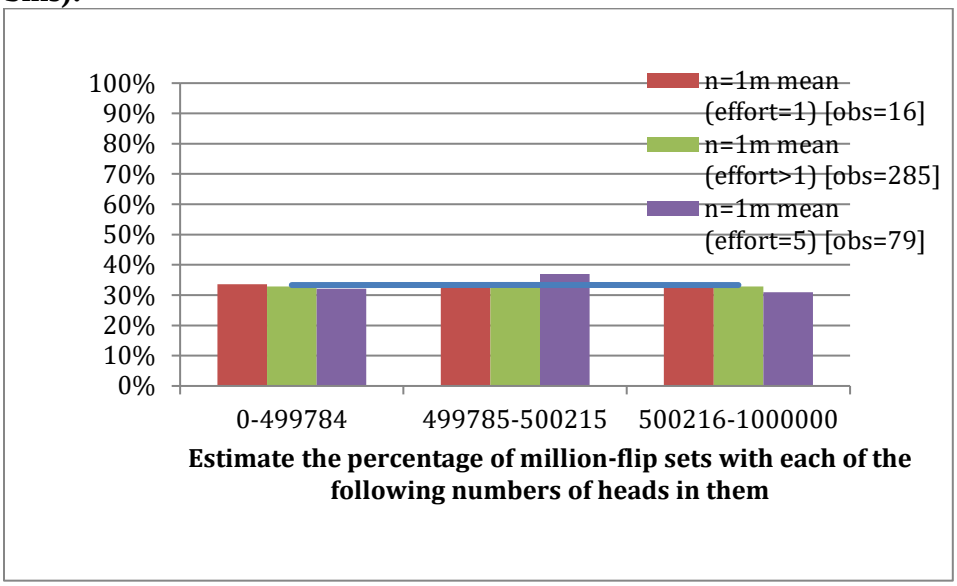
1m-flip 11-bin Histogram (including subjects who put approx. equal answers to all bins):



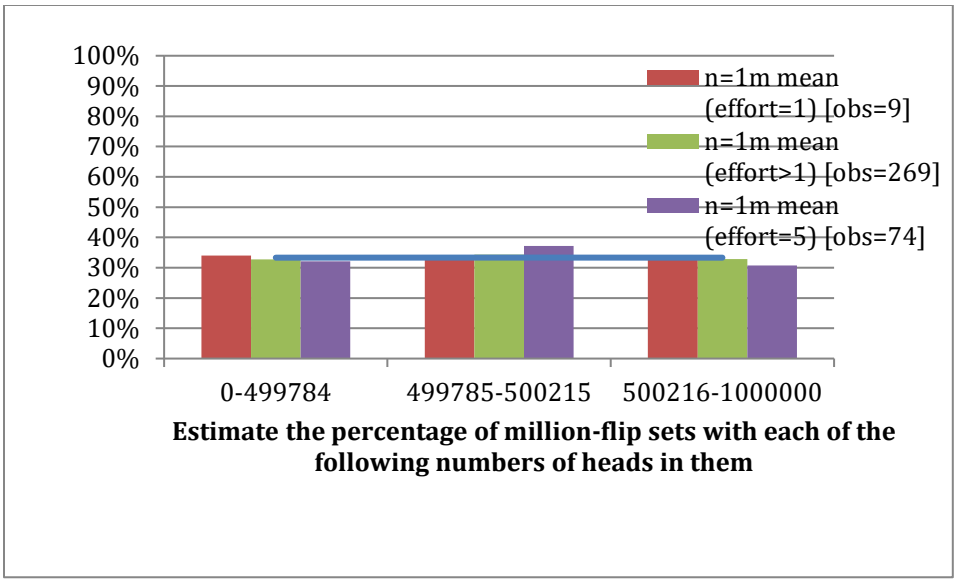
1m-flip 11-bin Histogram after dropping subjects who put approximately equal answers to all bins (8-10%):



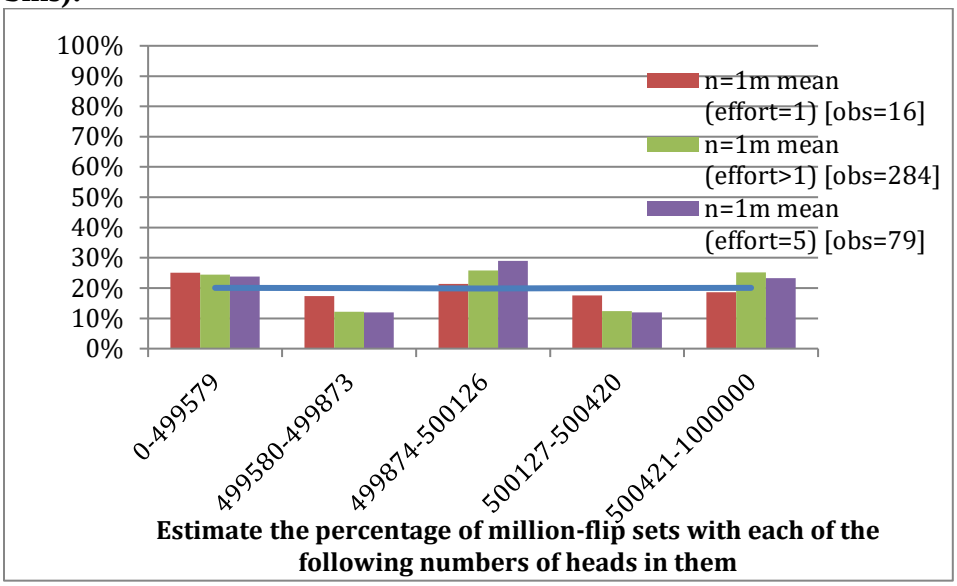
1m-flip 3-bin Histogram (including subjects who put approx. equal answers to all bins):



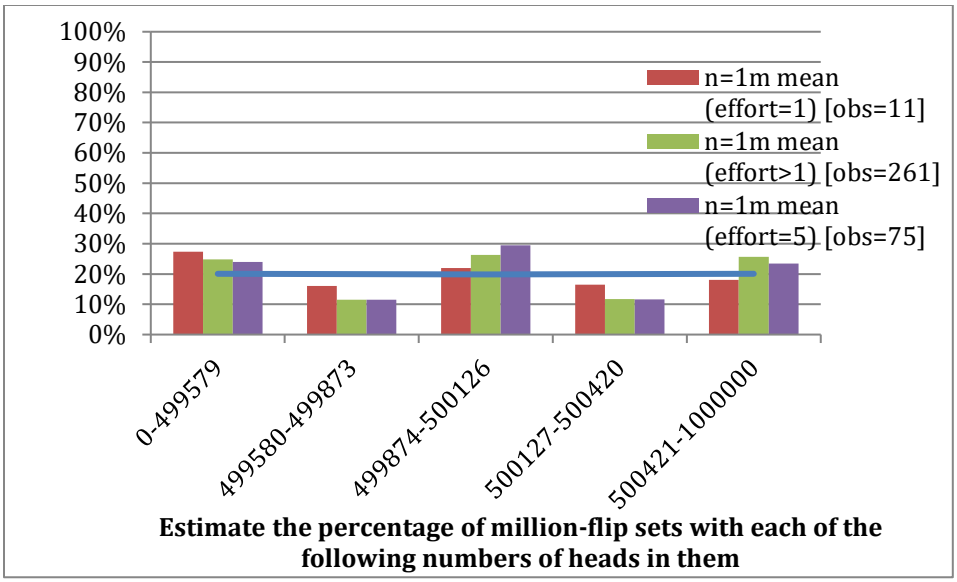
1m-flip 3-bin Histogram after dropping subjects who put approximately equal answers to all bins (32-34%):



1m-flip 5-bin Histogram (including subjects who put approx. equal answers to all bins):



1m-flip 5-bin Histogram after dropping subjects who put approximately equal answers to all bins (19-21%):



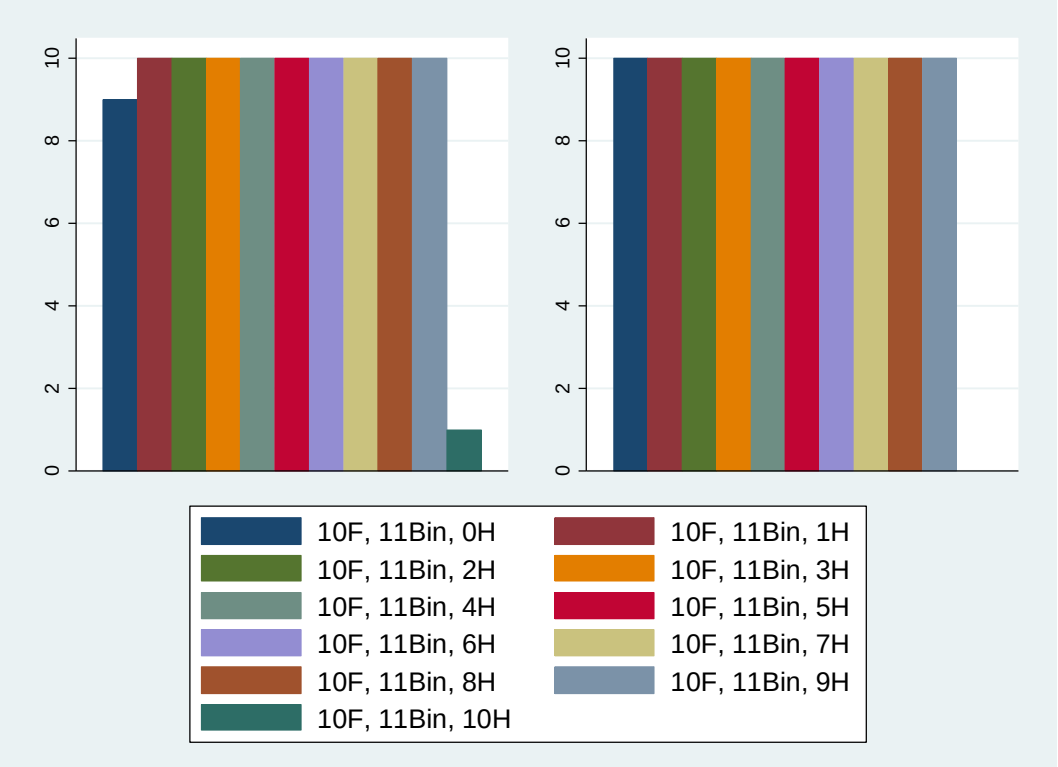
Experiment 1

Below is the repeated previous analyses using Experiment 1 data.

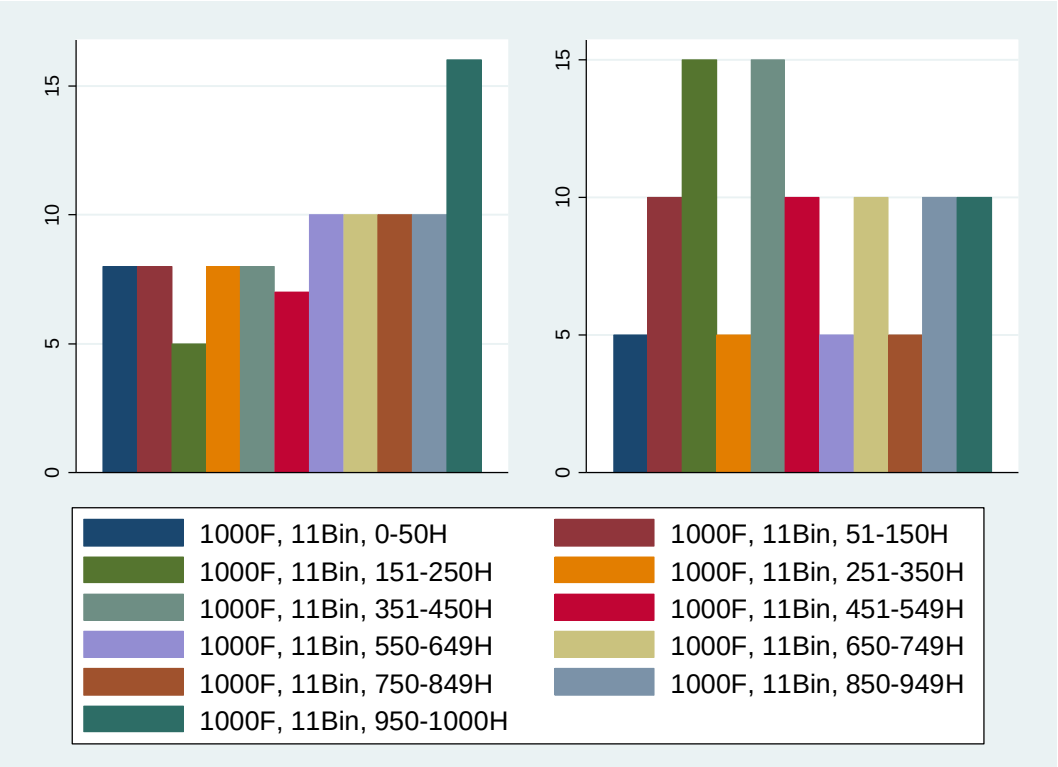
	"Please tell us how much thought and effort you put in to answering the questions on this survey"	"There were a number of questions on which you had to estimate the frequency of each of a set of possible outcomes. How much effort did you exert?"
1	3	2
2	5	4
3	20	30
4	26	21
5	22	20
Missing	28	27

Effort for frequency questions (i.e. "there were a number of questions on which…")	Approximately equal answers to 10-flip 11-bin questions (all bins given 8-10%)	Approximately equal answers to 1k-flip 11-bin questions (all bins given 8-10%)
1-3		0%
4-5	1.47%	7.35%

Below are 10-flip 11-bin histograms of 2 minimum-effort subjects:

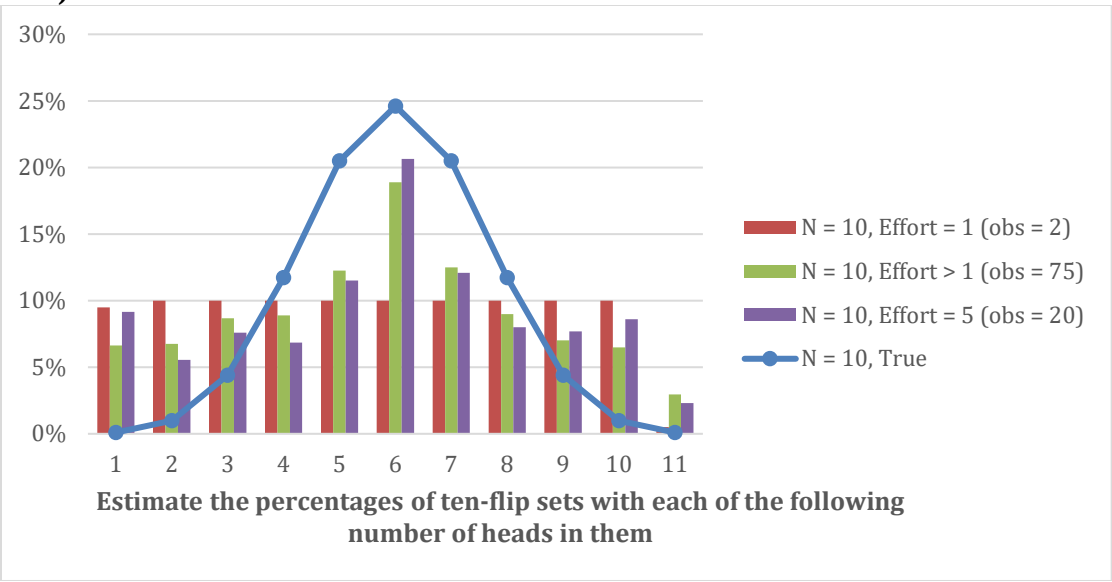


Below are 1k-flip 11-bin histograms of 2 minimum-effort subjects:

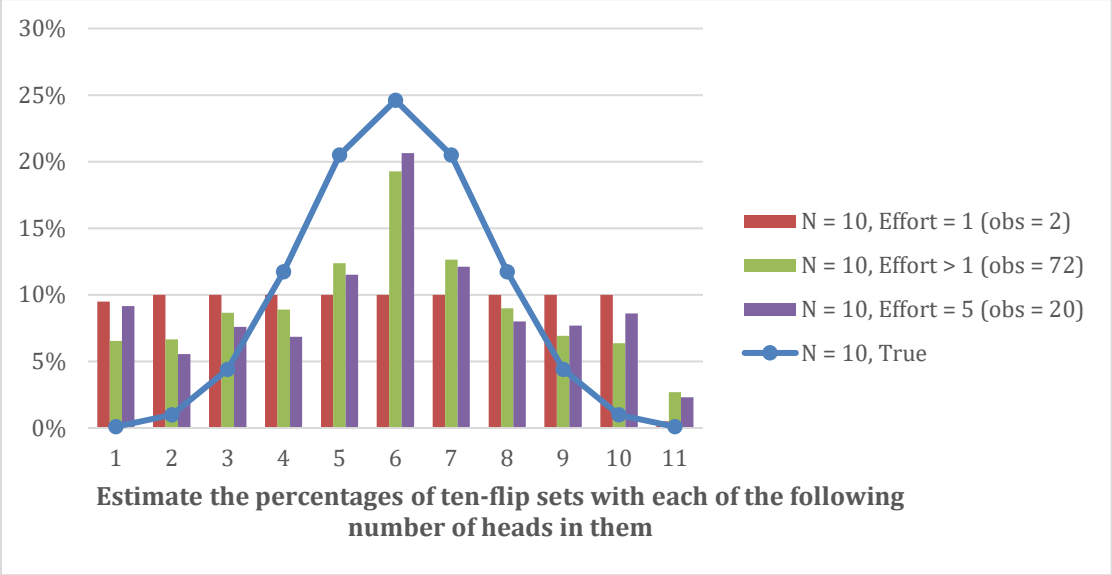


Next, we compare the responses to various histogram questions between low- and high-effort subjects.

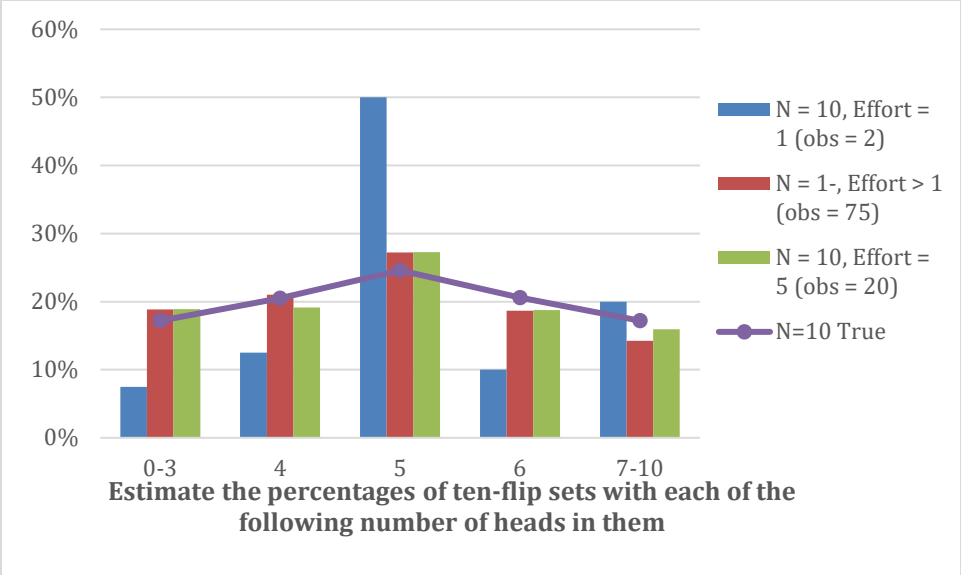
10-flip 11-bin Histogram (including subjects who put approx. equal answers to all bins):



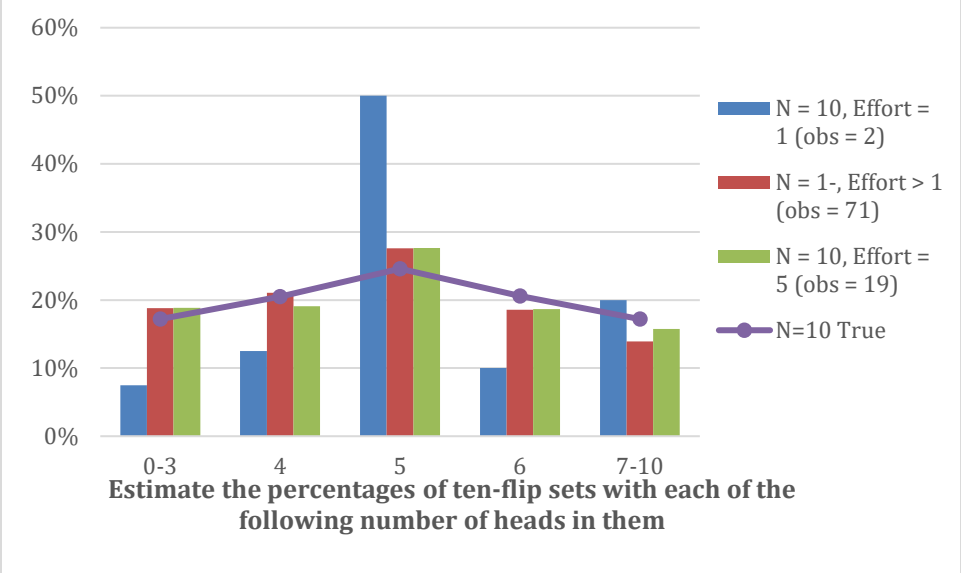
10-flip 11-bin Histogram after dropping subjects who put approximately equal answers to all bins (8-10%)



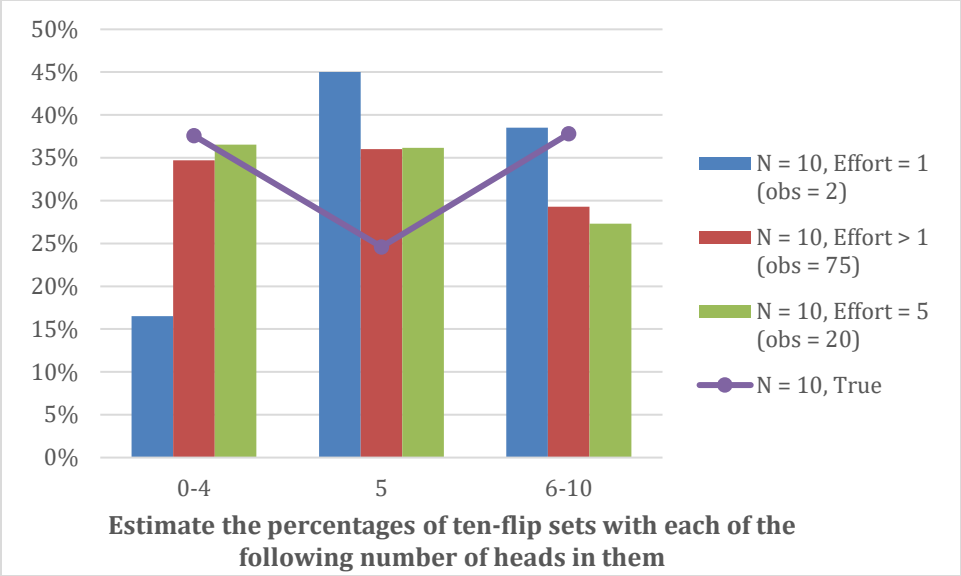
10-flip 5-bin Histogram (including subjects who put approx. equal answers to all bins):



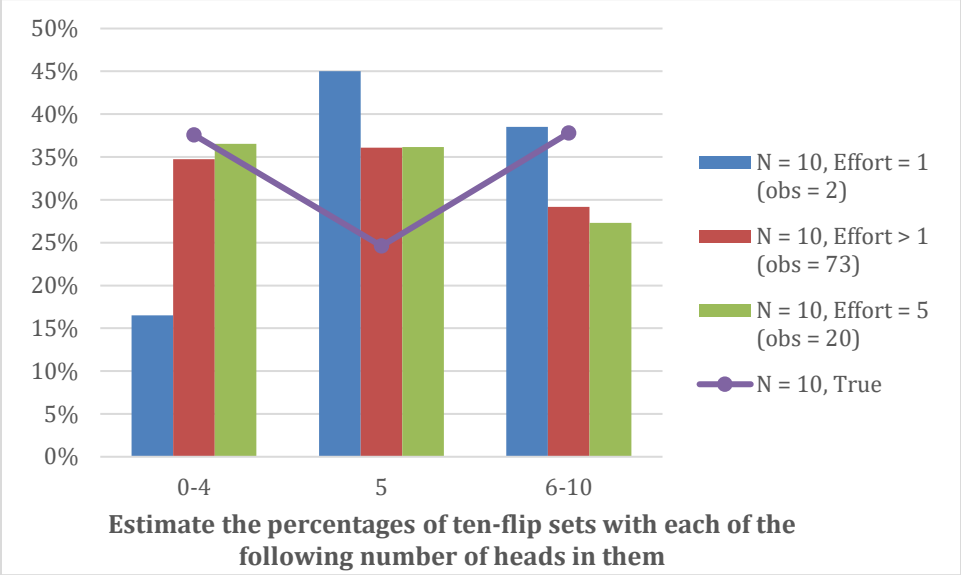
10-flip 5-bin Histogram after dropping subjects who put approximately equal answers to all bins (19-21%)



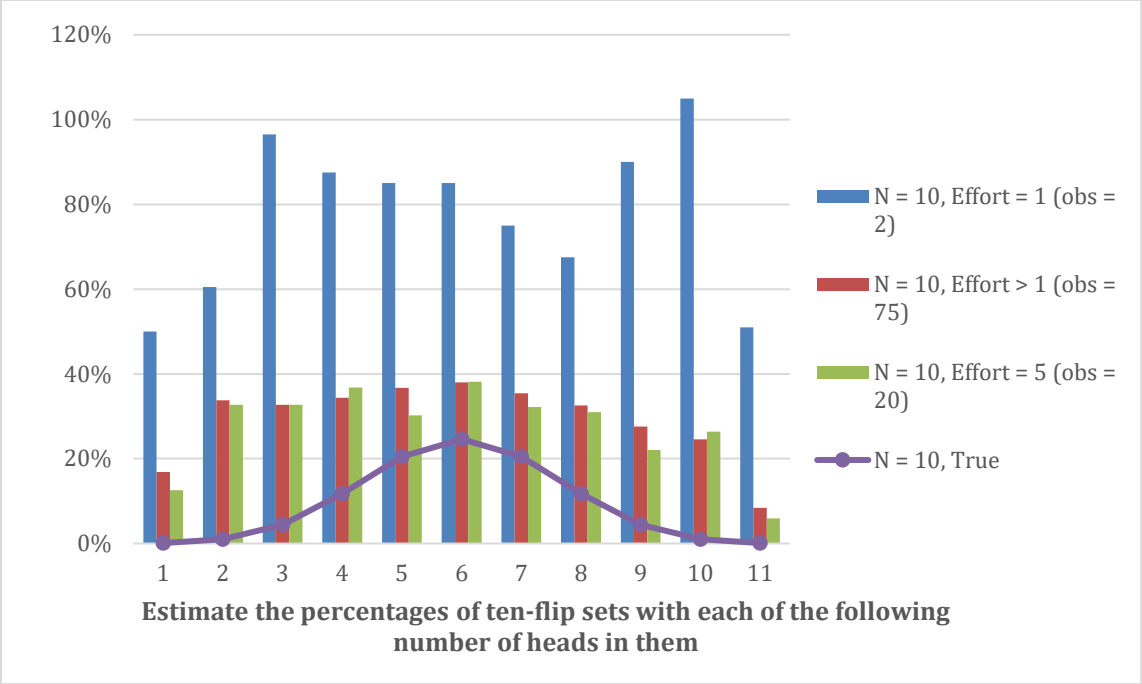
10-flip 3-bin Histogram (including subjects who put approx. equal answers to all bins):



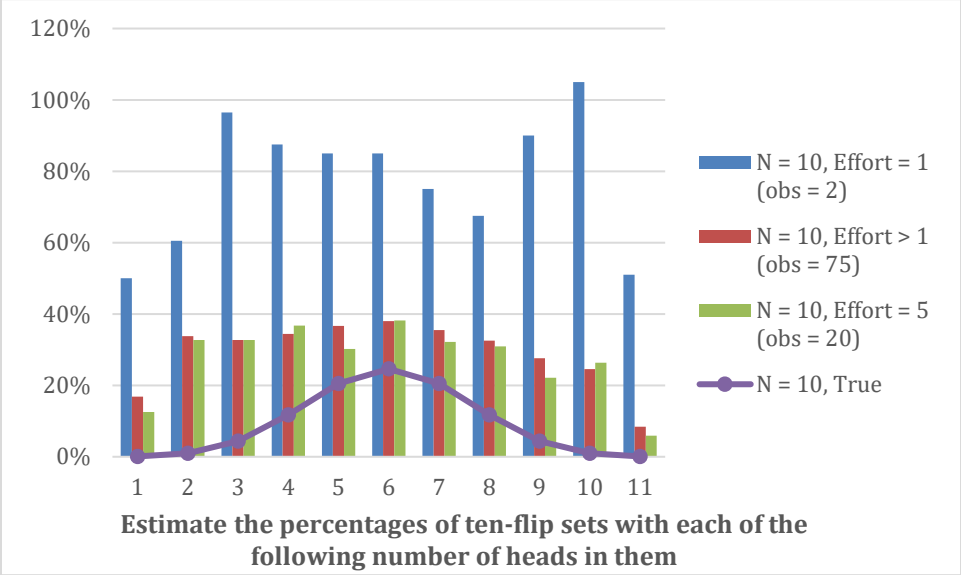
10-flip 3-bin Histogram after dropping subjects who put approximately equal answers to all bins (32-34%)



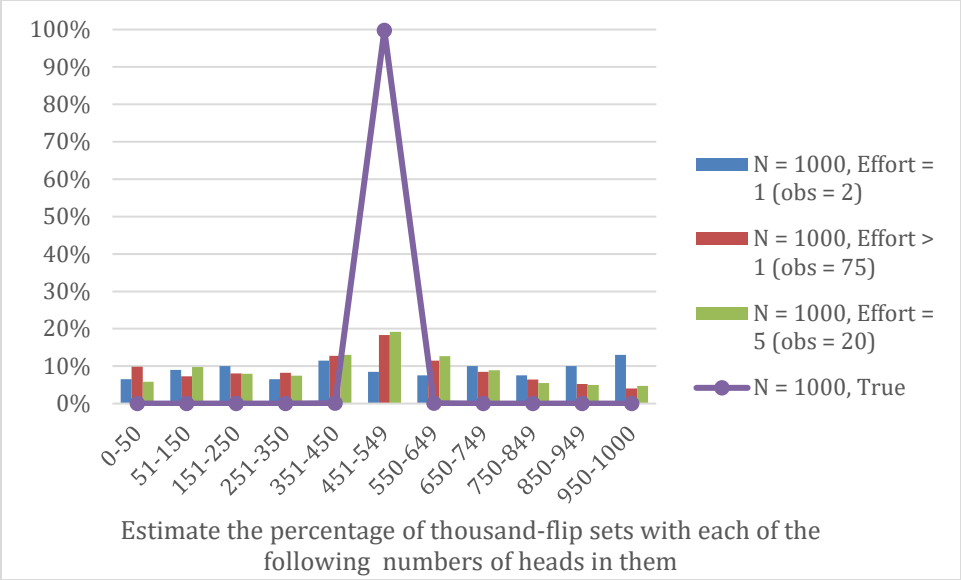
10-flip 11-bin Separate Histogram (including subjects who put approx. equal answers to all bins):



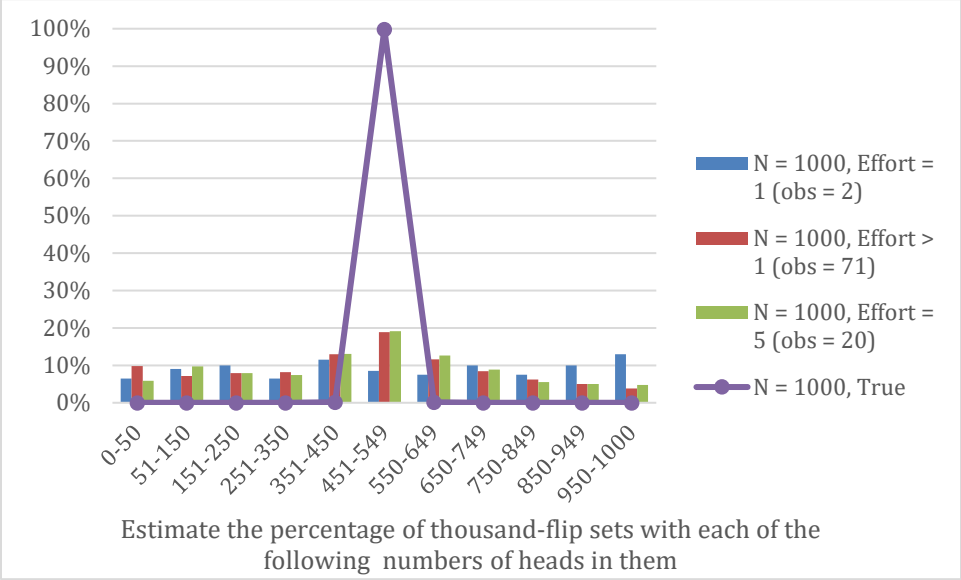
10-flip 11-bin Separate Histogram after dropping subjects who put approximately equal answers to all bins (8-10%):



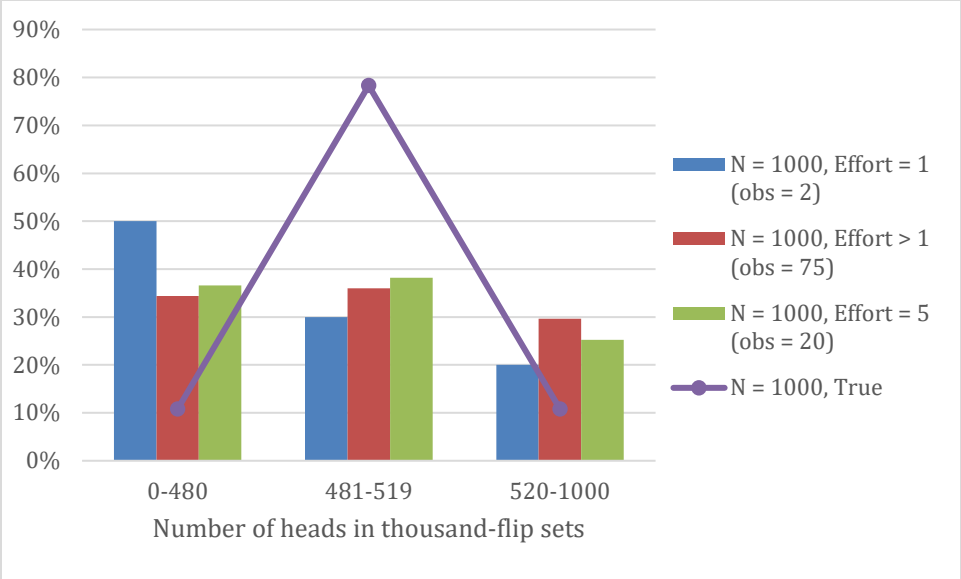
1k-flip 11-bin Histogram (including subjects who put approx. equal answers to all bins):



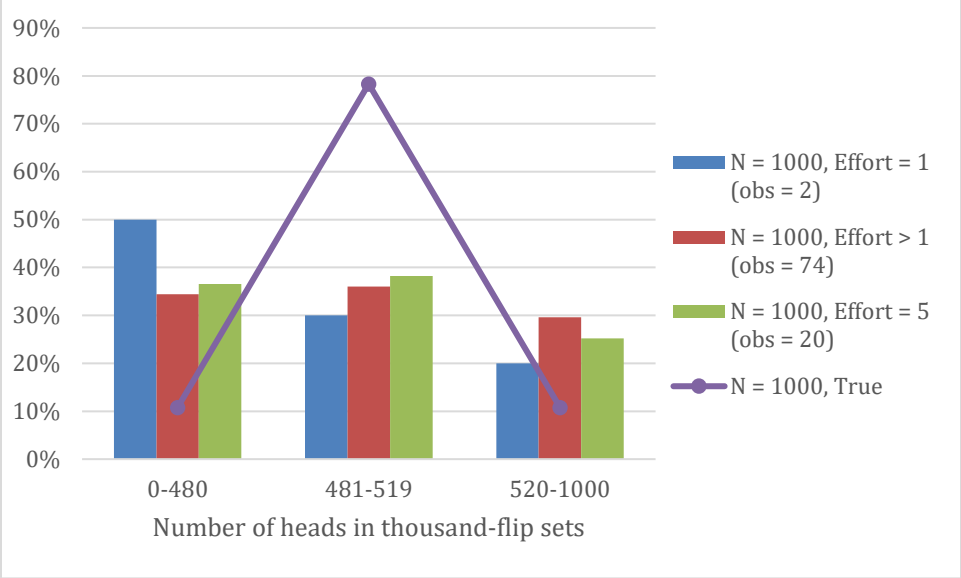
1k-flip 11-bin Histogram after dropping subjects who put approximately equal answers to all bins (8-10%):



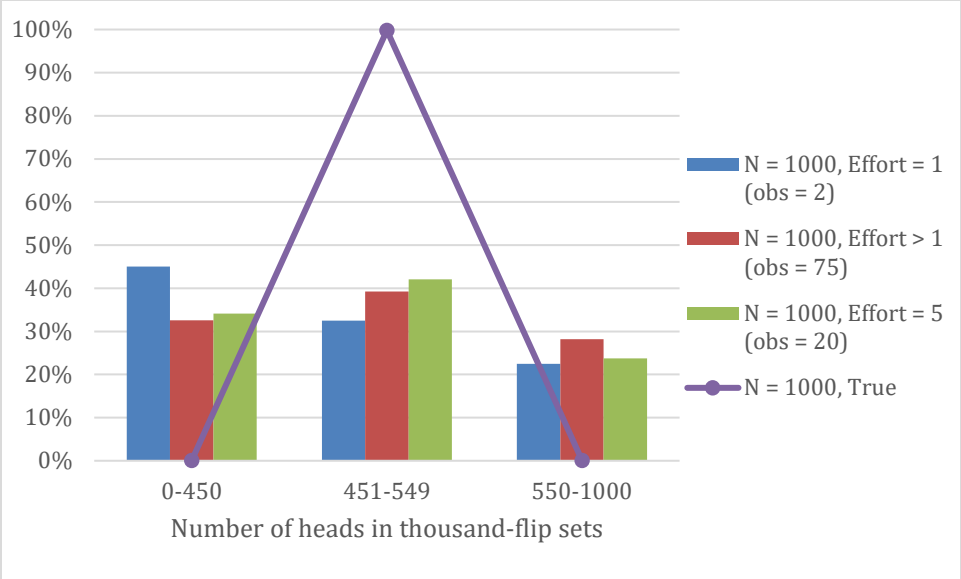
1k-flip 3-bin Narrow Histogram (including subjects who put approx. equal answers to all bins):



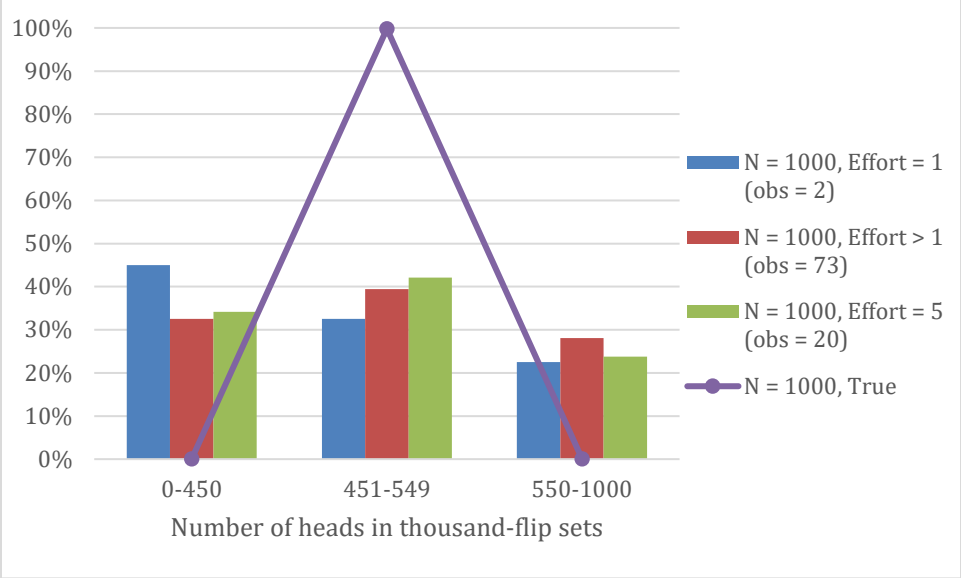
1k-flip 3-bin Narrow Histogram after dropping subjects who put approximately equal answers to all bins (32-34%):



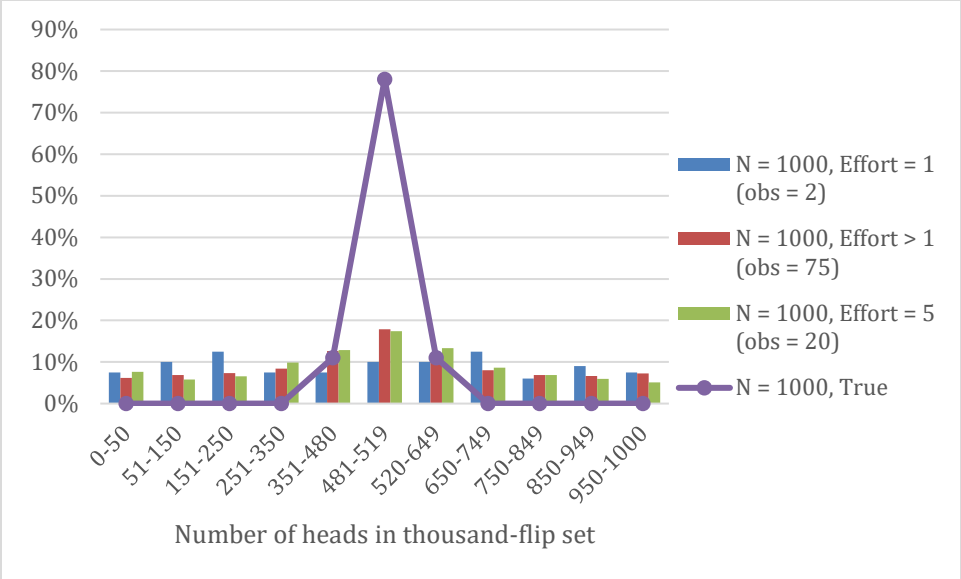
1k-flip 3-bin Histogram (including subjects who put approx. equal answers to all bins):



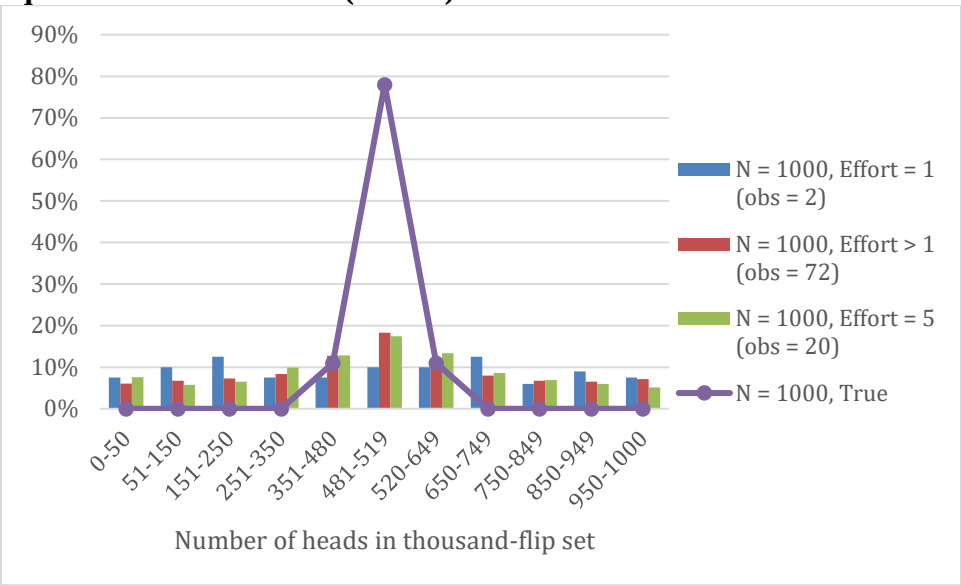
1k-flip 3-bin Histogram after dropping subjects who put approximately equal answers to all bins (32-34%):



1k-flip 11-bin Narrow Histogram (including subjects who put approx. equal answers to all bins):



1k-flip 11-bin Narrow Histogram after dropping subjects who put approximately equal answers to all bins (8-10%):



(ii) Follow-up analysis: Can these minimum-effort participants explain why we have <50% responses for the streak-0 questions?

Experiment 2:
Minimum-effort subjects' answers generally tended to be lower than others, but even with these 16 subjects excluded, we still observe <50% responses for streak-0 questions for the 1k-flip and 1m-flip.

Streak questions 10-flip:

Number of prior H's	Full Sample	Effort>1 [Sample=292]	Effort=1 [Sample=16]
0	49.53	49.86	43.44
1	44.65	44.59	45.88
2	40.83	40.98	37.96
3	39.82	40.23	32.27
4	39.01	39.05	38.26
5	38.18	38.28	36.23
6	37.60	37.12	46.31
7	37.15	37.38	32.98
8	35.90	36.08	32.72
9	35.32	35.25	36.48

Streak questions 1k-flip:

Number of prior H's	Full Sample	Effort>1 [Sample=292]	Effort=1 [Sample=16]
0	43.76	44.36	32.82
1	44.34	44.30	45.00
2	42.23	42.35	40.19
3	40.45	40.29	43.36
4	39.85	39.94	38.09
5	38.75	38.63	40.94
6	38.17	37.96	41.97
7	39.06	38.98	40.47
8	36.99	36.84	39.63
9	37.18	36.98	40.90

Streak questions 1m-flip:

Number of prior H's	Full Sample	Effort>1 [Sample=292]	Effort=1 [Sample=16]
0	45.43	46.00	35.07
1	44.38	44.90	34.91
2	41.44	41.94	32.35
3	40.98	41.54	30.70
4	39.26	39.75	30.28
5	38.93	39.81	22.83
6	38.87	39.34	30.34
7	38.27	38.94	25.98
8	38.08	38.71	26.52
9	37.91	38.19	32.79

Experiment 1:

Both of the minimum effort participants answered 50% for the streak-0 questions, which means that some participants who believed they put in effort believed there was less than

50% chance of getting heads. Note that the sample size of participants whose self-reported effort was 1 is small (only two participants).

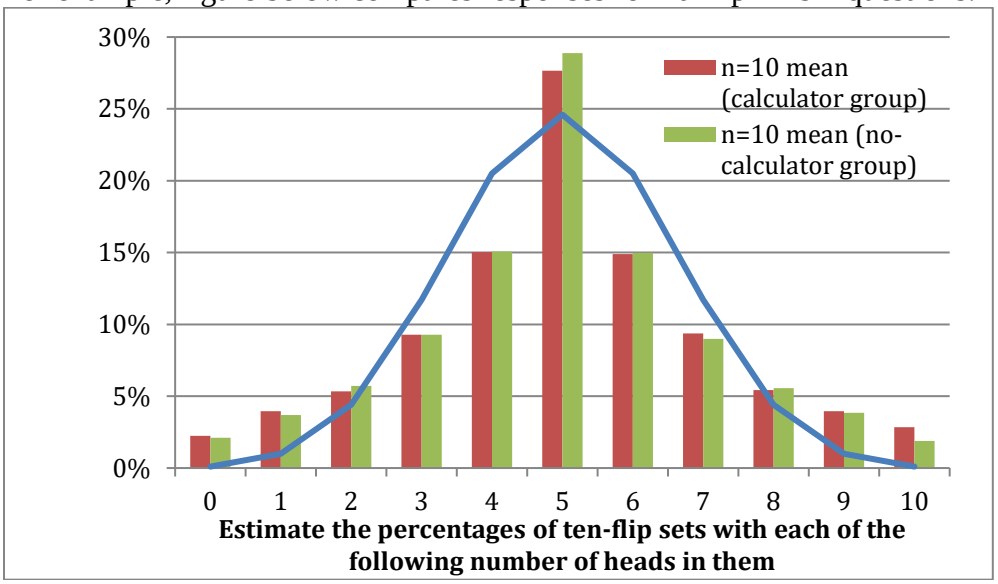
Streak questions 10-flip:

Number of prior H's	Full Sample	Effort>1 [Sample=75]	Effort=1 [Sample=2]
0	43.70	42.84	50
1	45.64	44.52	50
2	41.54	40.86	31.5
3	39.56	39.23	37.5
4	38.36	37.97	35
5	37.68	36.76	35
6	34.28	33.67	15
7	34.53	34.72	20
8	28.77	28.19	10
9	31.64	31.17	10

(c) Calculators (Experiment 2)

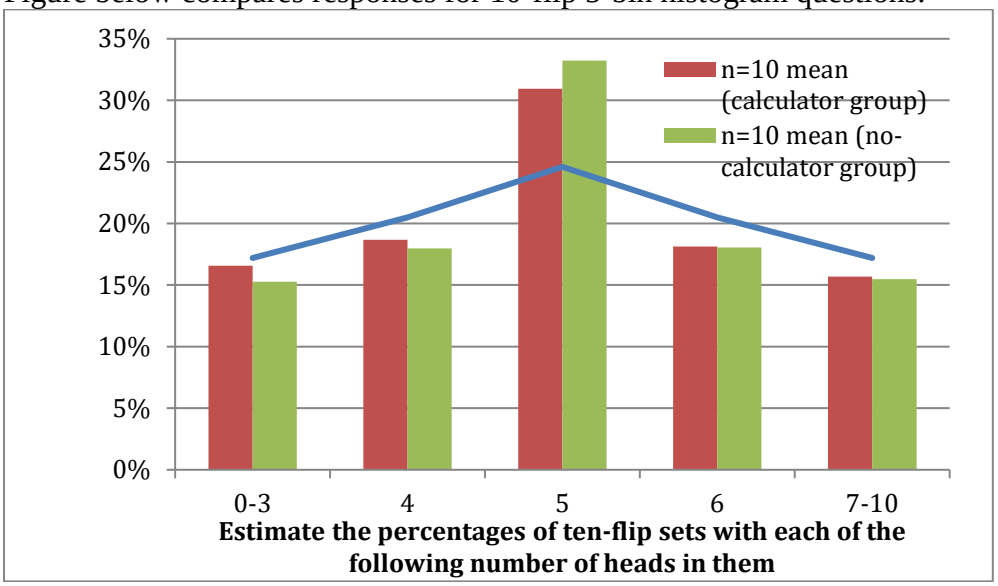
31.5% said they wrote down calculations on paper, 30.8% said they used calculators, and 1.9% used online tools. 50.3% (155 out of 308) subjects checked none of the above. These "no-calculator" subjects' responses were not significantly different.

For example, figure below compares responses for 10-flip 11-bin questions:



Middle-bin calculator (mean = 27.7%) vs. no-calculator (mean = 28.9): $p = 0.516$

Figure below compares responses for 10-flip 5-bin histogram questions:



Middle-bin calculator (mean = 30.9%) vs. no-calculator (mean = 33.2%): $p = 0.235$

(d) Incentives (Experiment 2)

To the question "When answering the questions, did you try to answer as accurately as you could?", 92.4% (279 out of 302) responses "Yes" and 7.6% (23 out of 302) responded "No" (6 entries missing).

(e) Self-doubt on "obvious" 50%-answer questions (streak questions I.A.3, I.B.2, and I.C.2) (Experiment 2)

Test for subjects who answered the first four questions with 50% and then switched on at least one subsequent question in that section. They may have been wondering what we were after, and not reacting to streaks due to gambler's fallacy.

In general, our GF results remain intact even with these subjects excluded from the analysis.

For the 10-flip streak questions, 152 subjects answered the first four questions with 50%, and 9.9% (15 out of 152) switched on at least one subsequent question.

Number of prior H's	Full Sample	Without the 15 subjects
0	49.53	49.50
1	44.65	44.59
2	40.83	40.61
3	39.82	39.42
4	39.01	38.48

5	38.18	37.71
6	37.60	37.19
7	37.15	36.74
8	35.90	35.89
9	35.32	35.46

For the 1k-flip streak questions, 149 subjects answered the first four questions with 50%, and 10.7% (16 out of 152) switched on at least one subsequent question.

Number of prior H's	**Full Sample**	**Without the 16 subjects**
0	43.76	43.84
1	44.34	44.12
2	42.23	41.95
3	40.45	40.10
4	39.85	39.43
5	38.75	38.21
6	38.17	38.10
7	39.06	38.63
8	36.99	36.67
9	37.18	36.96

For the 1m-flip streak questions, 165 subjects answered the first four questions with 50%, and 15.2% (25 out of 165) switched on at least one subsequent question.

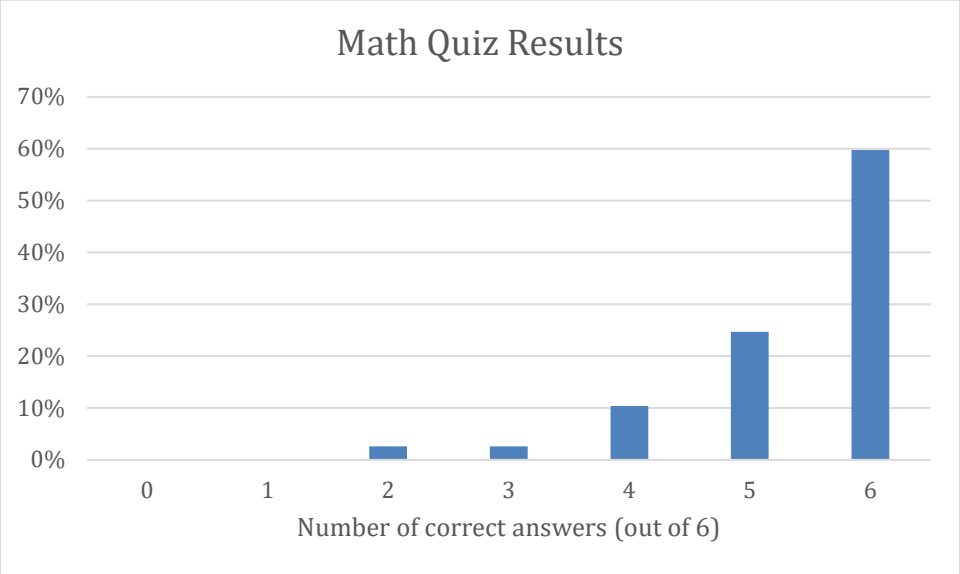
Number of prior H's	**Full Sample**	**Without the 25 subjects**
0	45.43	45.71
1	44.38	43.75
2	41.44	40.78
3	40.98	40.32
4	39.26	38.57
5	38.93	38.00
6	38.87	38.25
7	38.27	37.59
8	38.08	38.10
9	37.91	36.71

II.C. Individual Differences

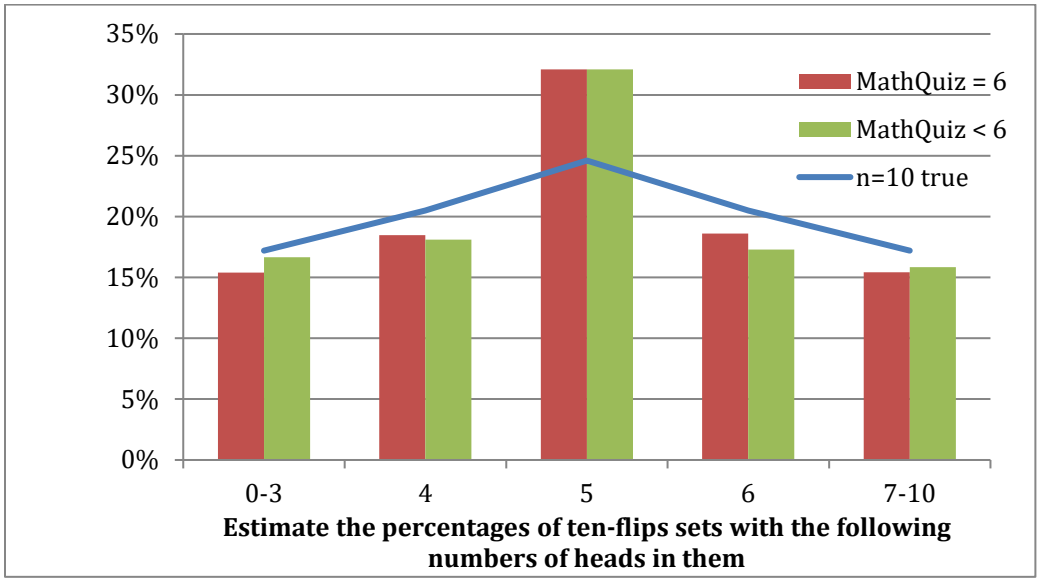
Conduct some exploratory analyses of associations between GF, NBLLN, and these characteristics. Some specific hypotheses:

(a) Higher math skills are associated with less NBLLN (Experiment 2).

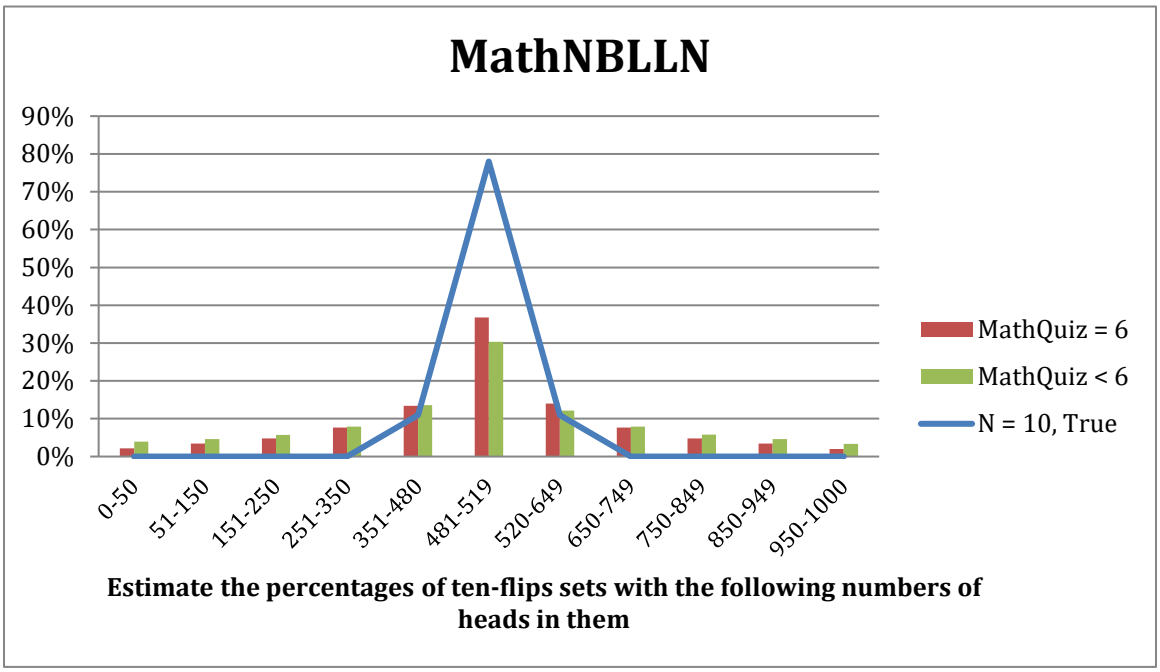
We find no evidence that higher math skills are associated with less NBLLN. First, below are results of the math quiz.



Below is a test for NBLLN in N = 10: Whether (B) generates approximately equal mean elicited probabilities across the five bins (which is close to normatively correct). We observe no significant differences between MathQuiz=6 and MathQuiz<6 subjects (*p*-value = 0.9991 for the middle bin).



We repeat this analysis for the 11-bin histogram for the 1000-flip sets. In this case, we observe significant (at the 5% level) differences for those between MathQuiz=6 and MathQuiz<6 subjects (*p*-value of the middle bin = 0.034). Those of the MathQuiz=6 group predicted significantly higher probabilities for the middle bin.



(b) Higher math skills are not associated with less GF.

We find no significant evidence that math skills are associated with less GF. Below shows how the predicted frequency of H varies with the number of prior H's in the streak questions, along with regression estimates (10-flip):

Number of prior H's	Full Sample	MathQuiz = 6 [184 out of 308]	MathQuiz < 6 [124 out of 308]	p-value (6 vs. <6)
0	49.53	49.76	49.19	0.5204
1	44.65	44.88	44.33	0.6900
2	40.83	40.60	41.16	0.7209
3	39.82	39.73	39.95	0.8964
4	39.01	39.01	39.01	0.9994
5	38.18	37.39	39.34	0.2691
6	37.60	37.46	37.80	0.8586
7	37.15	36.55	38.04	0.4439
8	35.90	35.42	36.61	0.5545
9	35.32	33.55	37.94	0.0444

	Predicted Freq of H
Number of prior H's	-1.0723*** (0.1872)
MathQuiz = 6	0.8354 (1.293)
(Number of prior H's) * (MathQuiz = 6)	-0.3865 (0.2422)

Constant	45.1628*** (0.9995)
Obs	3080

Note: Dependent variable is predicted frequency of H. Standard errors in parentheses. *** Significant at 1% significance level; ** 5% significance level; * 10% significance level.

For 1,000-flips:

Number of prior H's	**Full Sample**	**MathQuiz = 6 [184 out of 308]**	**MathQuiz < 6 [124 out of 308]**	**p-value (6 vs. <6)**
0	43.76	44.39	42.83	0.3332
1	44.34	45.13	42.17	0.1666
2	42.23	43.23	40.76	0.1495
3	40.45	41.01	39.62	0.4616
4	39.85	41.17	37.89	0.0805
5	38.75	39.34	37.89	0.4285
6	38.17	38.96	37.00	0.3343
7	39.06	39.77	38.00	0.3699
8	36.99	37.36	36.44	0.6463
9	37.18	37.86	36.18	0.4303

	Predicted Freq of H
Number of prior H's	-0.7795*** (0.1868)
MathQuiz = 6	2.1110 (1.290)
(Number of prior H's) * (MathQuiz = 6)	-0.0595 (0.2417)
Constant	42.4851*** (0.9972)
Obs	3080

For 1,000,000-flips:

Number of prior H's	**Full Sample**	**MathQuiz = 6 [184 out of 308]**	**MathQuiz < 6 [124 out of 308]**	**p-value (6 vs. <6)**
0	45.43	45.63	45.13	0.7469
1	44.38	46.32	41.51	0.0023
2	41.44	42.74	39.51	0.0593
3	40.98	41.33	40.46	0.6448
4	39.26	39.79	38.46	0.4893
5	38.93	40.58	36.47	0.0374
6	38.87	40.71	36.14	0.0219
7	38.27	39.59	36.30	0.0969
8	38.08	38.29	37.76	0.7965

9	37.91	38.57	36.93	0.4777

	Predicted Freq of H
Number of prior H's	-0.7936*** (0.1880)
MathQuiz = 6	2.6434** (1.2985)
(Number of prior H's) * (MathQuiz = 6)	-0.0347 (0.2432)
Constant	42.4403*** (1.0036)
Obs	3080

Our intuition is weak on the relationship between math skills and these biases, especially GF: there are a priori reasons to suspect higher skills, attention, etc., are associated with more severe GF because of greater (but mistaken) attention to patterns, or weaker for the obvious reason of grasping the right answer better.

(c) Miscellaneous demographics (Experiment 2)

(i) Gender

There were 107 males and 195 females (6 entries missing) in Experiment 2. Gambler's fallacy appears to be stronger among females than males. For example, below are responses to 10-flip streak questions:

Number of prior H's	**Full Sample**	**Male [Sample=107]**	**Female [Sample=195]**	**p-value (M vs. F)**
0	49.53	49.90	49.31	0.5343
1	44.65	45.45	44.44	0.5317
2	40.83	44.20	39.27	0.0191
3	39.82	43.95	37.91	0.0053
4	39.01	43.26	36.82	0.0044
5	38.18	41.82	36.56	0.0257
6	37.60	41.83	35.64	0.0138
7	37.15	41.99	34.61	0.0035
8	35.90	40.42	33.76	0.0119
9	35.32	37.81	34.26	0.2045

	Predicted Freq of H
Number of prior H's	-1.46*** (0.147)
Male	2.69** (1.322)
(Number of prior H's) * (Male)	0.47* (0.248)
Constant	44.82***

	(0.787)
Obs	3020

1,000-flip

Number of prior H's	Full Sample	Male [Sample=107]	Female [Sample=195]	p-value (M vs. F)
0	43.76	45.69	42.51	0.1131
1	44.34	45.56	44.14	0.4009
2	42.23	44.17	41.90	0.2626
3	40.45	43.26	39.52	0.0940
4	39.85	42.72	38.93	0.0910
5	38.75	41.36	37.97	0.1323
6	38.17	41.03	37.25	0.1190
7	39.06	43.22	37.45	0.0155
8	36.99	40.63	35.85	0.0565
9	37.18	40.60	35.93	0.0674

	Predicted Freq of H
Number of prior H's	-0.89*** (0.146)
Male	2.21* (1.314)
(Number of prior H's) * (Male)	0.33 (0.246)
Constant	43.16*** (0.782)
Obs	3020

1,000,000-flip

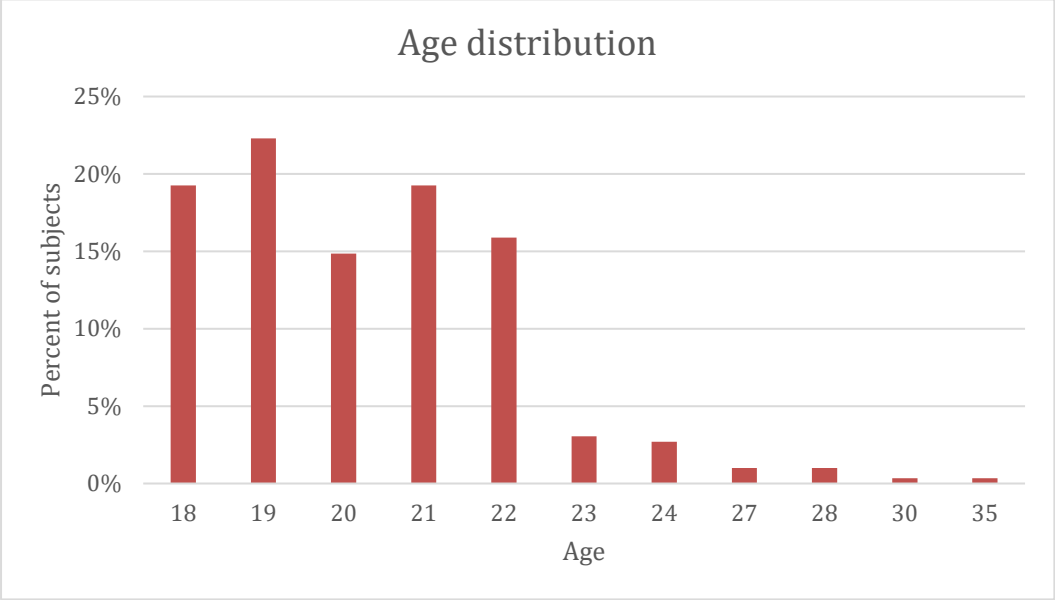
Number of prior H's	Full Sample	Male [Sample=107]	Female [Sample=195]	p-value (M vs. F)
0	45.43	46.79	44.55	0.2147
1	44.38	45.72	43.99	0.3076
2	41.44	44.08	40.50	0.0784
3	40.98	42.44	41.03	0.5152
4	39.26	41.50	38.91	0.2706
5	38.93	41.48	38.43	0.2009
6	38.87	40.74	38.50	0.3563
7	38.27	42.07	37.09	0.0372
8	38.08	41.07	37.34	0.1355
9	37.91	40.75	37.26	0.1871

	Predicted Freq of H
Number of prior H's	-0.83*** (0.148)

Male	1.95 (1.323)
(Number of prior H's) * (Male)	0.21 (0.248)
Constant	43.50*** (0.788)
Obs	3020

(ii) Age

Below is the age distribution of 296 subjects (12 entries missing) for Experiment 2:



We split subjects into "old" (≥ 21 years old) and "young" (≤ 20 years old). We find no significant difference in GF between these two groups. For example, 10-flip streak questions:

Number of prior H's	**Full Sample**	**Old [Sample=129]**	**Young [Sample=167]**	**p-value (O vs. Y)**
0	49.53	49.19	49.82	0.4657
1	44.65	45.36	44.85	0.7392
2	40.83	42.90	39.89	0.1425
3	39.82	41.12	39.31	0.3952
4	39.01	40.51	38.23	0.3052
5	38.18	40.21	37.11	0.1791
6	37.60	39.11	36.98	0.3848
7	37.15	38.92	35.75	0.1997
8	35.90	37.51	35.02	0.3341
9	35.32	35.78	35.02	0.7782

	Predicted Freq of H

Number of prior H's	-1.40*** (0.161)
Old	0.37 (1.273)
(Number of prior H's) * (Old)	0.21 (0.238)
Constant	45.49*** (0.861)
Obs	3080

1,000-flip

Number of prior H's	**Full Sample**	**Old [Sample=129]**	**Young [Sample=167]**	**p-value (O vs. Y)**
0	43.76	42.59	45.02	0.2071
1	44.34	43.25	46.24	0.0629
2	42.23	42.40	43.78	0.4756
3	40.45	39.91	42.12	0.3061
4	39.85	40.27	40.57	0.8893
5	38.75	38.22	40.41	0.3162
6	38.17	37.72	39.82	0.3703
7	39.06	38.58	40.39	0.4397
8	36.99	36.47	39.17	0.2640
9	37.18	37.68	37.99	0.9006

	Predicted Freq of H
Number of prior H's	-0.83*** (0.161)
Old	-3.35*** (1.267)
(Number of prior H's) * (Old)	0.03 (0.237)
Constant	45.28*** (0.857)
Obs	3080

1,000,000-flip

Number of prior H's	**Full Sample**	**Old [Sample=129]**	**Young [Sample=167]**	**p-value (O vs. Y)**
0	45.43	45.22	45.39	0.9274
1	44.38	44.40	45.02	0.7057
2	41.44	43.07	41.25	0.3538
3	40.98	41.71	41.77	0.9766
4	39.26	40.59	39.73	0.7071
5	38.93	40.12	39.15	0.6766
6	38.87	39.72	39.30	0.8578

7	38.27	39.64	38.40	0.5939
8	38.08	40.21	37.08	0.1930
9	37.91	40.22	37.34	0.2629

	Predicted Freq of H
Number of prior H's	-0.91*** (0.162)
Old	-1.13 (1.281)
(Number of prior H's) * (Old)	0.21 (0.240)
Constant	44.54*** (0.866)
Obs	3080

Appendix H: Appendix Tables

Table H.1. Demographics of participants in each experiment (% by category).

Age	Exp 1	Exp 2
18-20	13.5	54.2
21-25	29.8	39.3
26-30	9.6	2.3
31-35	8.7	0.3
36-40	4.8	0.0
41-45	6.7	0.0
46-50	2.9	0.0
51-55	8.7	0.0
56-60	8.7	0.0
61+	2.9	0.0
Missing	3.8	3.9

Sex	Exp 1	Exp 2
Male	51.0	34.7
Female	41.4	63.3
Missing	7.7	2.0

Household Income	Exp 1	Exp 2
Below $50k	67.3	23.1
$50k-100k	11.5	19.2
$100k-150k	9.62	17.5
$150k-200k	1.0	10.4
$200k or more	3.9	14.9
Missing	6.7	14.9

Table H.2. Estimates of the parameters of the Rabin-Vayanos using individual and aggregate data.

	Estimates using individual-level data		Estimates using aggregate data	
	(1) Experiment 1	(2) Experiment 2	(3) Experiment 1	(4) Experiment 2
α	0.0314	0.1601***	0.0314	0.1601***
	(0.02165)	(0.0354)	(0.04258)	(0.0312)
δ	0.9468***	0.6211***	0.9468**	0.6211***
	(0.15188)	(0.0555)	(0.31063)	(0.0633)
# Obs	621	2772	9	9

Note: Estimates of the Rabin-Vayanos model by non-linear least squares on the first difference formulation of the model, as described in the text. Participants who gave exactly 10% for all streak questions were excluded from Experiment 1 results. Standard errors in parentheses (clustered for individual-level regression), and 95% confidence intervals in brackets. * $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$.

Table H.3.A. Regression of betting-on-heads on realizations of other flips and payoff in Experiment 1.

	(1)	(2)	(3)
# Heads - # Tails	-0.0010	-0.0099	-0.0142*
	(0.0076)	(0.0076)	(0.0076)
Payoff of H (relative to 50)		0.0250***	0.0248***
		(0.0061)	(0.0060)
Dummy for (#H - #T)=-1, i.e. "middle-bin effect"			-0.0660*
			(0.0360)
Constant	0.5096***	-0.7420**	-0.7107**
	(0.0160)	(0.3016)	(0.3025)
# Obs	987	987	987

Note: "Payoff of H" is the payoff given for betting on heads relative to 50 lotteries, and takes on values {-5, -3, -1, 1, 3, 5}. Clustered standard errors in parentheses.
* $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$.

Table H.3.B. Regression of betting-on-heads on realizations of other flips and payoff in Experiment 2.

	(1)	(2)	(3)
# Heads - # Tails	-0.0229***	-0.0289***	-0.0287***
	(0.0051)	(0.0049)	(0.0049)
Payoff of H (relative to 50)		0.0688***	0.0688***
		(0.0035)	(0.0035)
Dummy for (#H - #T)=-1, i.e. "middle-bin effect"			0.0073
			(0.0219)
Constant	0.5168***	0.5197***	0.5185***
	(0.0104)	(0.0095)	(0.0101)
# Obs	3080	3080	3080

Note: "Payoff of H" is the payoff given for betting on Heads relative to 50 lotteries, and takes on values {-5, -3, -1, 1, 3, 5}. Clustered standard errors in parentheses. * $p < 10\%$, ** $p < 5\%$, *** $p < 1\%$.

Table H.4.A. Comparison of 10-flip streak question responses by when subjects saw the question in Experiment 2.

(1) Number of prior H's	(2) Early	(3) Middle	(4) Late	(5) *p*-value ANOVA equal means	(6) *p*-value ANOVA equal variance
0	49.96	50.08	48.57	0.30	<0.001
	(5.37)	(7.63)	(9.77)		
1	45.43	44.57	43.90	0.71	0.10
	(11.95)	(14.24)	(14.45)		
2	43.49	39.62	39.03	0.13	0.25
	(18.02)	(18.69)	(15.91)		
3	41.95	38.97	38.27	0.29	0.34
	(17.70)	(19.86)	(17.29)		
4	40.99	37.73	38.01	0.38	0.57
	(18.08)	(20.11)	(18.94)		
5	40.69	36.05	37.36	0.22	0.68
	(19.00)	(20.74)	(19.76)		
6	40.91	35.03	36.29	0.10	0.49
	(20.34)	(22.55)	(20.29)		
7	40.26	33.20	37.28	0.06	0.76
	(20.22)	(21.71)	(21.24)		
8	38.32	33.14	35.74	0.25	0.52
	(20.80)	(22.68)	(23.08)		
9	37.18	33.94	34.53	0.56	0.64
	(22.21)	(23.76)	(24.25)		
# Obs	112	92	104		

Note: Standard deviations in parentheses. "Early" is defined as 10-flip questions appearing before 1,000-flip and 1,000,000-flip questions. "Middle" is defined as 10-flip questions appearing between 1,000-flip and 1,000,000-flip questions. "Late" is defined as 10-flip questions appearing after 1,000-flip and 1,000,000-flip questions.

Table H.4.B. Comparison of 1,000-flip streak question responses by when subjects saw the question in Experiment 2.

(1) Number of prior H's	(2) Early	(3) Middle	(4) Late	(5) *p*-value ANOVA equal means	(6) *p*-value ANOVA equal variance
0	43.66	43.93	43.68	0.99	0.35
	(16.89)	(17.59)	(15.32)		
1	41.24	45.50	45.78	0.05	0.07
	(16.09)	(13.72)	(12.86)		
2	39.18	43.47	43.56	0.13	0.48
	(18.35)	(16.63)	(16.38)		
3	36.84	41.56	42.37	0.09	0.04
	(21.41)	(18.30)	(16.68)		
4	38.97	39.64	40.80	0.79	0.91
	(19.11)	(19.37)	(18.58)		
5	35.27	39.99	40.45	0.11	0.07
	(20.58)	(19.75)	(16.51)		
6	36.58	38.55	39.12	0.66	0.82
	(21.26)	(20.20)	(20.05)		
7	36.82	40.29	39.69	0.44	0.55
	(21.54)	(19.48)	(19.69)		
8	33.84	38.59	38.02	0.23	0.56
	(22.47)	(20.80)	(20.21)		
9	35.10	38.58	37.54	0.51	0.66
	(22.68)	(21.03)	(20.85)		
# Obs	91	109	108		

Note: Standard deviations in parentheses. "Early" is defined as 10-flip questions appearing before 1,000-flip and 1,000,000-flip questions. "Middle" is defined as 10-flip questions appearing between 1,000-flip and 1,000,000-flip questions. "Late" is defined as 10-flip questions appearing after 1,000-flip and 1,000,000-flip questions.

Table H.4.C. Comparison of 1,000,000-flip streak question responses by when subjects saw the question in Experiment 2.

(1) Number of prior H's	(2) Early	(3) Middle	(4) Late	(5) *p*-value ANOVA equal means	(6) *p*-value ANOVA equal variance
0	47.70	43.28	45.35	0.10	0.20
	(13.94)	(16.27)	(14.07)		
1	44.54	43.62	45.07	0.76	0.33
	(14.98)	(14.42)	(12.92)		
2	39.88	42.16	42.35	0.51	0.43
	(18.31)	(16.28)	(16.56)		
3	41.24	40.83	40.87	0.98	0.56
	(19.42)	(18.52)	(17.43)		
4	39.17	39.76	38.79	0.94	0.86
	(20.42)	(19.37)	(20.07)		
5	39.65	38.45	38.67	0.90	0.75
	(21.03)	(19.94)	(19.57)		
6	41.78	38.56	36.05	0.14	0.33
	(21.42)	(18.68)	(21.04)		
7	39.61	37.91	37.19	0.68	0.90
	(20.70)	(19.84)	(20.53)		
8	39.61	38.25	36.21	0.52	0.51
	(21.57)	(19.77)	(22.04)		
9	38.46	39.35	35.71	0.49	0.38
	(24.04)	(21.26)	(21.50)		
# Obs	105	107	96		

Note: Standard deviations in parentheses. "Early" is defined as 10-flip questions appearing before 1,000-flip and 1,000,000-flip questions. "Middle" is defined as 10-flip questions appearing between 1,000-flip and 1,000,000-flip questions. "Late" is defined as 10-flip questions appearing after 1,000-flip and 1,000,000-flip questions.

Table H.4.D. Comparison of 10-flip 11-bin histogram responses by when subjects saw the question in Experiment 2.

(1) Bins	(2) Early	(3) Middle	(4) Late	(5) *p*-value ANOVA equal means	(6) *p*-value ANOVA equal variance
1	2.15	2.22	2.16	0.99	0.41
	(3.33)	(2.97)	(3.37)		
2	3.91	3.79	3.72	0.94	<0.001
	(5.44)	(3.45)	(3.55)		
3	5.55	5.61	5.41	0.91	0.79
	(3.18)	(3.34)	(3.38)		
4	9.20	9.31	9.28	0.98	<0.001
	(3.48)	(4.02)	(5.35)		
5	14.67	14.05	16.48	0.01	0.003
	(5.16)	(5.62)	(7.06)		
6	29.57	27.58	27.49	0.57	0.84
	(16.68)	(16.25)	(15.74)		
7	14.85	14.65	15.43	0.59	0.24
	(5.20)	(5.61)	(6.12)		
8	9.23	9.72	8.59	0.10	0.002
	(3.19)	(4.42)	(3.43)		
9	5.53	5.63	5.29	0.74	0.93
	(3.22)	(3.32)	(3.20)		
10	3.35	4.37	4.04	0.33	<0.001
	(2.92)	(5.95)	(5.80)		
11	3.35	3.08	2.15	0.17	<0.001
	(2.92)	(6.11)	(3.95)		
# Obs	112	92	104		

Note: Standard deviations in parentheses. "Early" is defined as 10-flip questions appearing before 1,000-flip and 1,000,000-flip questions. "Middle" is defined as 10-flip questions appearing between 1,000-flip and 1,000,000-flip questions. "Late" is defined as 10-flip questions appearing after 1,000-flip and 1,000,000-flip questions.

Table H.4.E. Comparison of 1,000-flip 11-bin histogram responses by when subjects saw the question in Experiment 2.

(1) Bins	(2) Early	(3) Middle	(4) Late	(5) *p*-value ANOVA equal means	(6) *p*-value ANOVA equal variance
1	2.29	2.56	3.61	0.19	<0.001
	(3.29)	(3.30)	(8.13)		
2	3.45	3.96	4.34	0.25	0.53
	(4.00)	(3.59)	(3.65)		
3	4.41	5.26	5.75	0.02	0.55
	(3.59)	(3.37)	(3.22)		
4	7.12	7.94	8.13	0.23	<0.001
	(5.64)	(3.78)	(3.62)		
5	12.49	14.50	13.21	0.21	<0.001
	(8.49)	(9.95)	(5.71)		
6	40.84	32.44	30.35	0.01	0.01
	(29.90)	(23.74)	(22.51)		
7	12.41	13.53	13.72	0.37	0.004
	(8.13)	(5.83)	(6.98)		
8	6.78	8.00	8.33	0.04	<0.001
	(6.10)	(3.76)	(3.35)		
9	4.10	5.38	6.00	0.0007	0.89
	(3.60)	(3.49)	(3.43)		
10	3.50	3.88	4.22	0.39	0.35
	(4.03)	(3.49)	(3.67)		
11	2.60	2.56	2.45	0.95	0.01
	(3.88)	(3.33)	(2.88)		
# Obs	91	109	108		

Note: Standard deviations in parentheses. "Early" is defined as 10-flip questions appearing before 1,000-flip and 1,000,000-flip questions. "Middle" is defined as 10-flip questions appearing between 1,000-flip and 1,000,000-flip questions. "Late" is defined as 10-flip questions appearing after 1,000-flip and 1,000,000-flip questions.

Table H.4.F. Comparison of 1,000,000-flip 11-bin histogram responses by when subjects saw the question in Experiment 2.

(1) Bins	(2) Early	(3) Middle	(4) Late	(5) *p*-value ANOVA equal means	(6) *p*-value ANOVA equal variance
1	4.21	3.26	3.48	0.78	0.56
	(9.80)	(10.00)	(10.84)		
2	4.79	3.80	3.86	0.22	0.13
	(4.86)	(4.75)	(4.01)		
3	5.67	4.99	4.94	0.30	0.23
	(4.11)	(3.75)	(3.46)		
4	7.84	7.40	7.54	0.75	0.93
	(4.44)	(4.35)	(4.27)		
5	12.26	14.36	13.27	0.21	<0.001
	(7.83)	(10.48)	(7.01)		
6	32.83	36.40	35.48	0.62	0.95
	(27.71)	(27.76)	(26.97)		
7	11.64	12.34	13.15	0.31	0.16
	(6.72)	(7.71)	(6.45)		
8	7.46	6.95	7.43	0.61	0.66
	(4.38)	(4.00)	(4.17)		
9	5.54	4.88	5.02	0.40	0.81
	(3.80)	(3.66)	(3.56)		
10	4.47	3.63	3.66	0.22	0.13
	(4.07)	(4.13)	(3.42)		
11	3.35	2.23	2.17	0.18	<0.001
	(6.85)	(4.71)	(2.83)		
# Obs	105	107	96		

Note: Standard deviations in parentheses. "Early" is defined as 10-flip questions appearing before 1,000-flip and 1,000,000-flip questions. "Middle" is defined as 10-flip questions appearing between 1,000-flip and 1,000,000-flip questions. "Late" is defined as 10-flip questions appearing after 1,000-flip and 1,000,000-flip questions.

Table H.5.A. 10-flip 11-bin responses by participants who used calculators/tools compared to those who used no tools.

(1) Bins	(2) True prob	(3) Mean (calculator/tools)	(4) Mean (no calculators/tools)	(5) *p*-value
0	0.1%	2.3%	2.1%	0.69
1	1.0%	4.0%	3.7%	0.61
2	4.4%	5.3%	5.7%	0.27
3	11.7%	9.3%	9.3%	0.95
4	20.5%	15.0%	15.1%	0.95
5	24.6%	27.7%	28.9%	0.52
6	20.5%	14.9%	15.0%	0.99
7	11.7%	9.4%	9.0%	0.43
8	4.4%	5.4%	5.6%	0.65
9	1.0%	4.0%	3.8%	0.84
10	0.1%	2.9%	1.9%	0.05

Table H.5.B. 10-flip 5-bin responses by participants who used calculators/tools compared to those who used no tools.

(1) Bins	(2) True prob	(3) Mean (calculator/tools)	(4) Mean (no calculators/tools)	(5) *p*-value
0-3	17.2%	16.6%	15.3%	0.31
4	20.5%	18.7%	18.0%	0.36
5	24.6%	30.9%	33.2%	0.23
6	20.5%	18.1%	18.1%	0.92
7-10	17.2%	15.7%	15.5%	0.87

Table H.6.A. Cost of biases in expected prizes in Experiment 1.

(1) Question Type	(2) Expected earnings from actual choices	(3) Expected earnings from choosing optimally	(4) Expected earnings foregone
Betting	\$0.3159	\$0.3313	\$0.0154
10-flip streak	\$1.2275	\$1.2500	\$0.0225
Total	\$1.5434	\$1.5813	\$0.0379

Note: "Expected earnings from actual choices" is calculated as the expected number of tickets earned (given the participants' choices and the actual probabilities) divided by 40,000 (which gives the probability of winning \$50) multiplied by \$50. "Choosing optimally" means: always choosing the option that pays off more in the betting questions and always choosing the correct answer of 50% in the streak questions. "Expected earnings foregone" is column (3) minus column (2).

Table H.6.B. Cost of biases in expected prizes in Experiment 2.

(1) Question Type	(2) Expected earnings from actual choices	(3) Expected earnings from choosing optimally	(4) Expected earnings foregone
Betting	\$0.7669	\$0.8294	\$0.0625
10-flip streak	\$2.8077	\$2.9600	\$0.1523
1,000-flip streak	\$2.8116	\$2.9593	\$0.1477
1,000,000-flip streak	\$2.8117	\$2.9593	\$0.1476
Total	\$9.1979	\$9.7080	\$0.5101

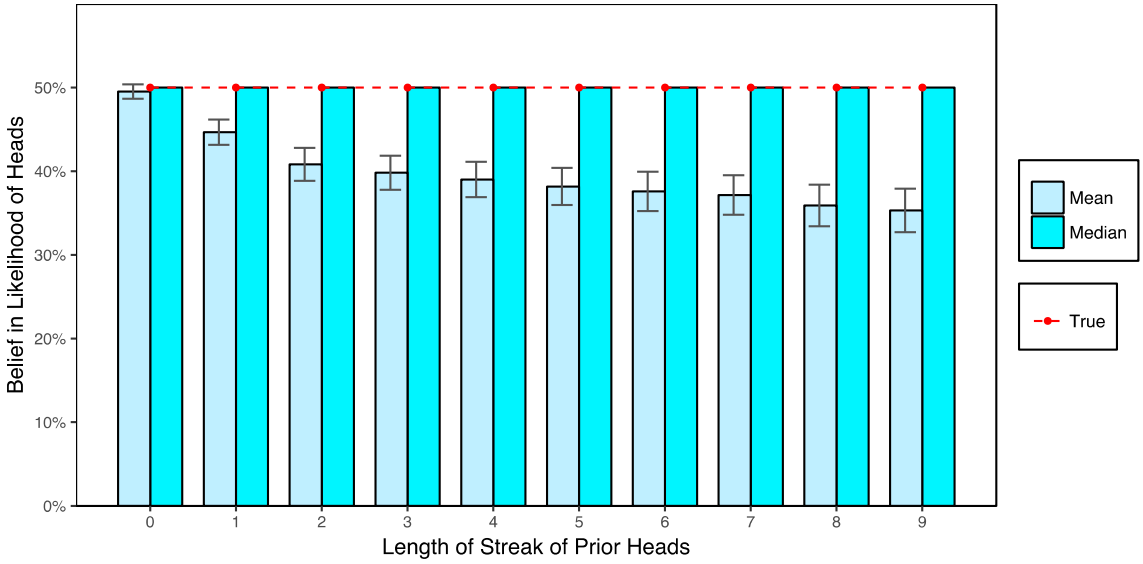
Note: "Expected earnings from actual choices" is calculated as $(x + y)^2 - y^2$, where x is the number of prizes won from a given question type and y is the number of prizes won from all other questions in the experiment. (We use this formula because a participant's total payout is $(x + y)^2$, whereas if she had not earned any prizes from this question type, her payout would be y^2.) We set x equal to the mean number of prizes won from the given question type: 0.1227 for the betting questions and 0.4762 for each of the 10-flip, 1,000-flip, and 1,000,000-flip streak questions. We set y equal to the mean number of prizes won from the rest of the experiment, excluding the given question type: 3.06 for the betting questions and 2.71 for each of the 10-flip, 1,000-flip, and 1,000,000-flip streak questions. "Expected earnings from choosing optimally" is calculated the same way (setting y equal to the same numbers as above), except that we set x equal to the expected number of prizes from choosing optimally: 0.1325 for the betting questions and 0.5 for each of the 10-flip, 1,000-flip, and 1,000,000-flip streak questions. "Expected earnings foregone" is column (3) minus column (2).

Table H.7. Self-reported effort (% by category).

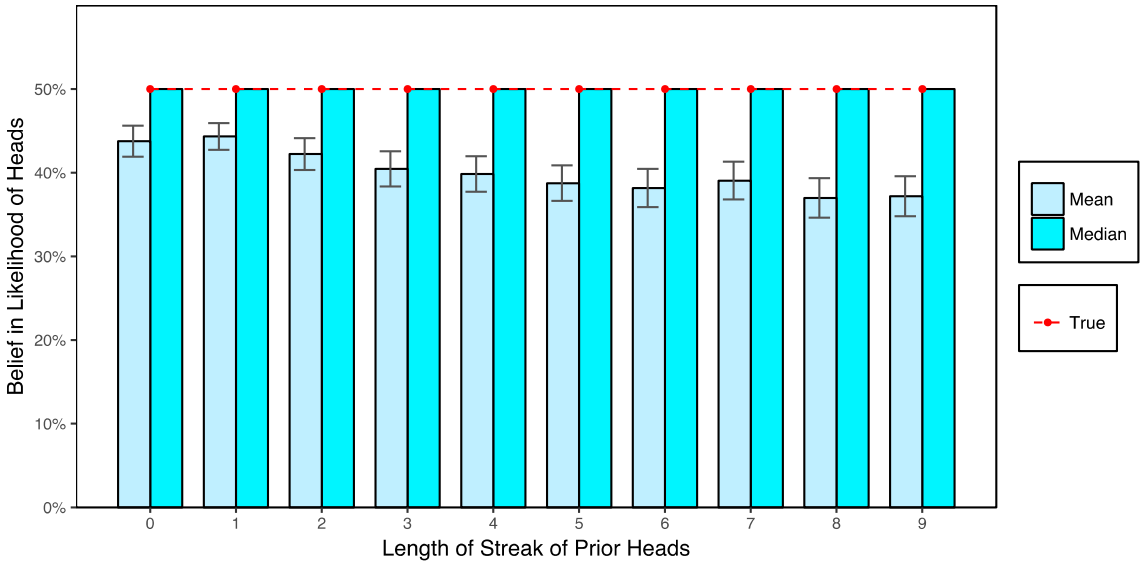
	"Please tell us how much thought and effort you put in to answering the questions on this survey"		"There were a number of questions on which you had to estimate the frequency of each of a set of possible outcomes. How much effort did you exert?"	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2
1	3	10	2	16
2	5	26	4	32
3	20	86	30	72
4	26	108	21	102
5	22	71	20	79
Missing	28	7	27	7

Appendix I: Appendix Figures

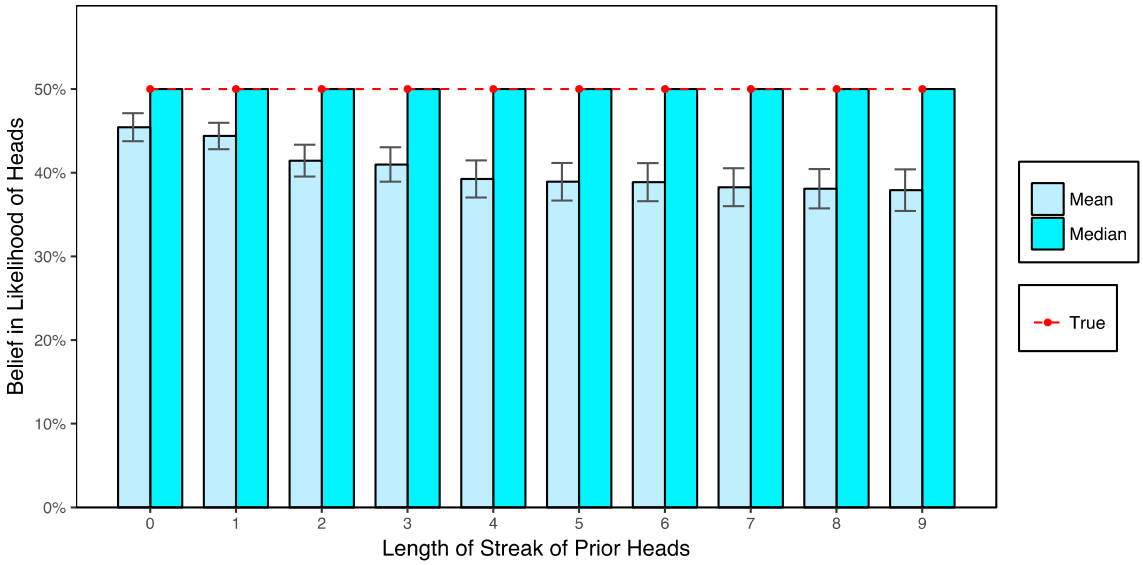
Appendix Figure I.1. Belief in likelihood of heads after a streak of heads in a sample size of *N* = 10 (Experiment 2).



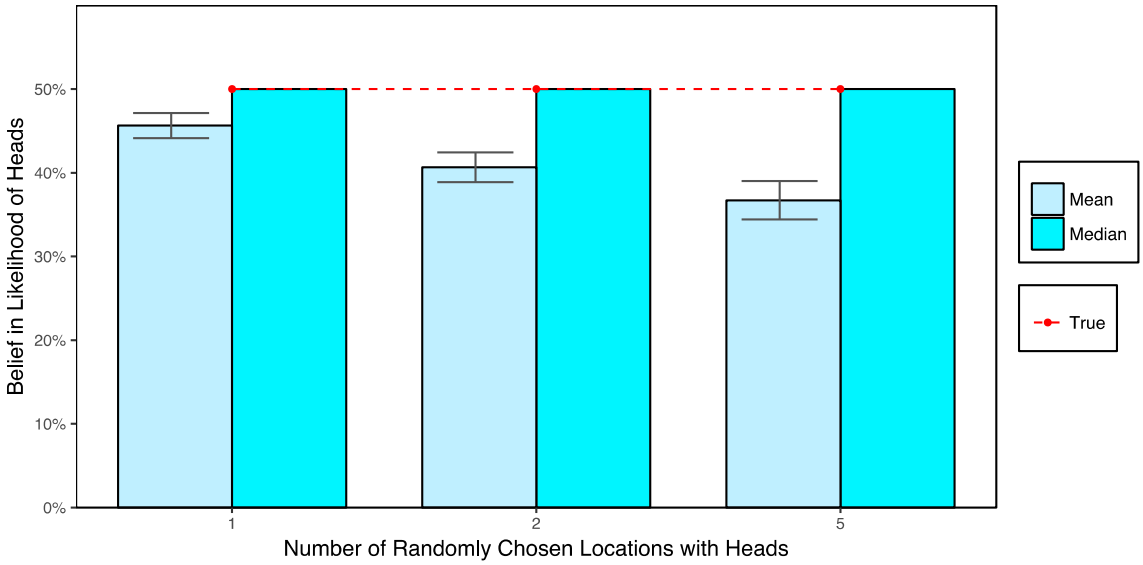
Appendix Figure I.2. Belief in likelihood of heads after a streak of heads in a sample size of *N* = 1000 (Experiment 2).



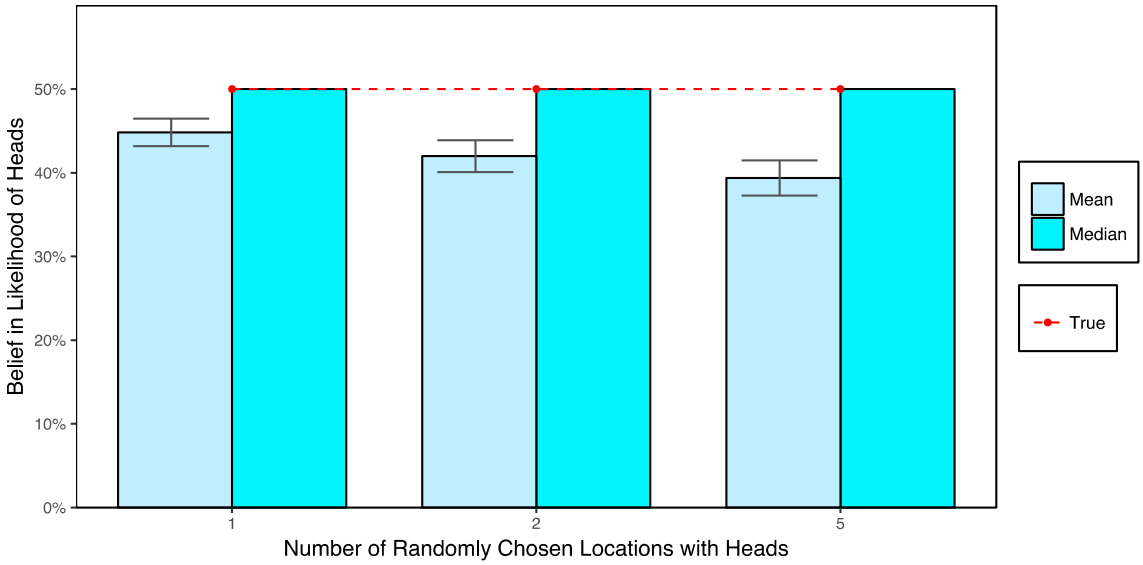
Appendix Figure I.3. Belief in likelihood of heads after a streak of heads in a sample size of *N* = 1 million (Experiment 2).



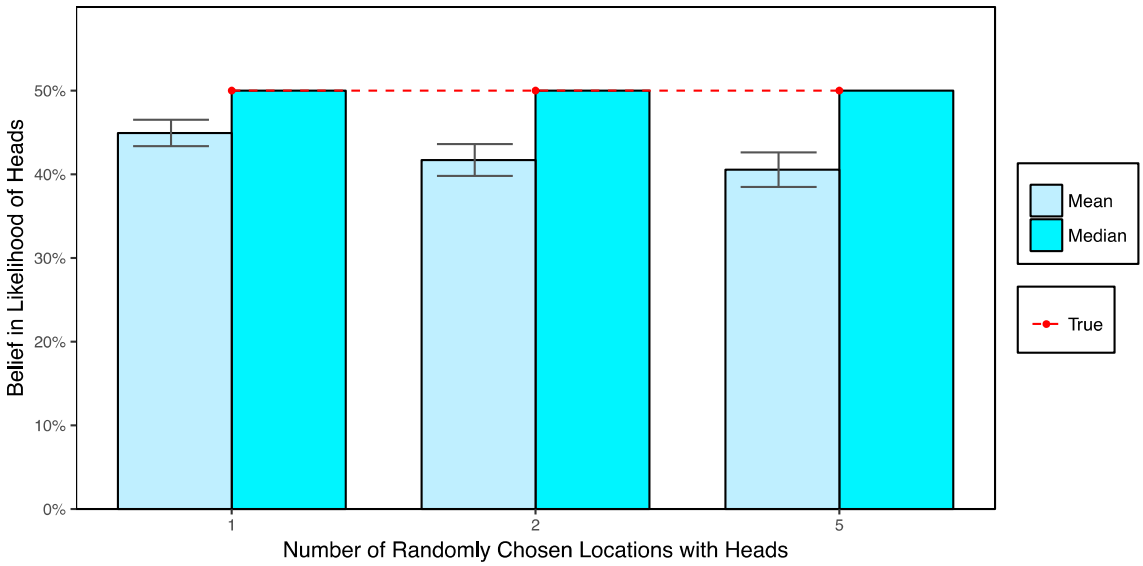
Appendix Figure I.4. Belief in likelihood of heads after randomly chosen flip locations are known to be heads in a sample size of *N* = 10 (Experiment 2).



Appendix Figure I.5. Belief in likelihood of heads after randomly chosen flip locations are known to be heads in a sample size of *N* = 1000 (Experiment 2).



Appendix Figure I.6. Belief in likelihood of heads after randomly chosen flip locations are known to be heads in a sample size of *N* = 1 million (Experiment 2).



Appendix J: Numbers Underlying the Figures

Figure 1. Belief in likelihood of heads after a streak of heads in sample size of *N* = 10 (Experiment 1).

(1) Prior Heads	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0	50.0%	43.7%	(39.4%, 48.1%)	50.0%	43.7%
1	50.0%	45.6%	(41.3%, 50.0%)	50.0%	42.2%
2	50.0%	41.5%	(37.3%, 47.7%)	50.0%	40.8%
3	50.0%	39.6%	(35.2%, 43.9%)	50.0%	39.5%
4	50.0%	38.4%	(33.8%, 42.9%)	50.0%	38.2%
5	50.0%	37.7%	(33.3%, 42.1%)	42.0%	37.0%
6	50.0%	34.3%	(30.0%, 38.6%)	40.0%	35.9%
7	50.0%	34.5%	(29.6%, 39.4%)	40.0%	34.8%
8	50.0%	28.8%	(24.1%, 33.4%)	25.0%	33.8%
9	50.0%	31.6%	(26.6%, 36.7%)	30.0%	32.9%

Figure 2. Belief in likelihood of heads after a streak of heads in sample size of *N* = 10, after excluding participants who reported 10% for any streak length (Experiment 1).

(1) Prior Heads	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0	50.0%	48.1%	(43.1%, 53.0%)	50.0%	48.1%
1	50.0%	48.2%	(43.0%, 53.5%)	50.0%	46.6%
2	50.0%	45.6%	(40.7%, 50.4%)	50.0%	45.2%
3	50.0%	44.1%	(38.9%, 49.4%)	50.0%	43.9%
4	50.0%	40.7%	(35.4%, 46.0%)	50.0%	42.6%
5	50.0%	42.1%	(36.6%, 47.6%)	50.0%	41.4%
6	50.0%	38.7%	(33.6%, 43.8%)	50.0%	40.3%
7	50.0%	39.7%	(33.6%, 45.9%)	50.0%	39.2%
8	50.0%	33.4%	(27.7%, 39.2%)	45.0%	38.2%
9	50.0%	36.7%	(30.5%, 42.8%)	50.0%	37.3%

Figure 3. Belief in likelihood of heads after a streak of heads in sample sizes of *N* = 10, 1000, and 1 million (Experiment 2).

(1) Prior Heads	(2) True prob	(3) *N* = 10, mean	(4) *N* = 10, 95% CI	(4) *N* = 1000, mean	(5) *N* = 1000, 95% CI	(6) *N* = 1 million, mean	(7) *N* = 1 million, 95% CI
0	50.0%	49.5%	(48.7%, 50.4%)	43.8%	(41.9%, 45.6%)	45.4%	(43.8%, 47.1%)
1	50.0%	44.7%	(43.1%, 46.2%)	44.3%	(42.7%, 45.9%)	44.4%	(42.8%, 46.0%)
2	50.0%	40.8%	(38.9%, 42.8%)	42.2%	(40.3%, 44.2%)	41.4%	(39.5%, 43.4%)
3	50.0%	39.8%	(37.8%, 41.9%)	40.5%	(38.3%, 42.6%)	41.0%	(38.9%, 43.1%)
4	50.0%	39.0%	(36.9%, 41.1%)	39.8%	(37.7%, 42.0%)	39.3%	(37.0%, 41.5%)
5	50.0%	38.2%	(36.0%, 40.4%)	38.8%	(36.6%, 40.9%)	38.9%	(36.7%, 41.2%)
6	50.0%	37.6%	(35.2%, 40.0%)	38.2%	(35.9%, 40.5%)	38.9%	(36.6%, 41.2%)
7	50.0%	37.1%	(34.8%, 39.5%)	39.1%	(36.8%, 41.3%)	38.3%	(36.0%, 40.5%)
8	50.0%	35.9%	(33.4%, 38.4%)	37.0%	(34.6%, 39.4%)	38.1%	(35.7%, 40.4%)
9	50.0%	35.3%	(32.7%, 37.9%)	37.2%	(34.8%, 39.6%)	37.9%	(35.4%, 40.4%)

Figure 4. Belief in likelihood of heads after randomly chosen flip locations are known to be heads in sample sizes of *N* = 10, 1000, and 1 million (Experiment 2).

(1) Locat -ions	(2) True prob	(3) *N* = 10, mean	(4) *N* = 10, 95% CI	(4) *N* = 1000, mean	(5) *N* = 1000, 95% CI	(6) *N* = 1 million, mean	(7) *N* = 1 million, 95% CI
1	50.0%	45.7%	(44.1%, 47.2%)	44.8%	(43.2%, 46.5%)	44.9%	(43.4%, 46.5%)
2	50.0%	40.7%	(38.9%, 42.4%)	42.0%	(40.1%, 43.9%)	41.7%	(39.8%, 43.6%)
5	50.0%	36.7%	(34.4%, 39.0%)	39.4%	(37.3%, 41.5%)	40.5%	(38.5%, 42.6%)

Figure 5. Eleven-bin histogram beliefs in sample size of *N* = 10 (Experiment 1).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0	0.1%	6.1%	(4.1%, 8.2%)	2.0%	0.0%
1	1.0%	6.4%	(5.2%, 7.5%)	5.0%	0.5%
2	4.4%	8.0%	(6.7%, 9.3%)	8.0%	3.0%
3	11.7%	9.0%	(8.1%, 9.9%)	10.0%	10.6%
4	20.5%	12.3%	(10.9%, 13.7%)	10.0%	21.9%
5	24.6%	20.0%	(16.6%, 23.4%)	14.0%	27.8%
6	20.5%	12.7%	(11.2%, 14.2%)	10.0%	21.9%
7	11.7%	8.9%	(7.8%, 10.0%)	10.0%	10.6%
8	4.4%	7.3%	(6.0%, 8.7%)	6.0%	3.0%
9	1.0%	6.5%	(4.9%, 8.2%)	5.0%	0.5%
10	0.1%	2.7%	(2.1%, 3.4%)	1.0%	0.0%

Figure 6. Five-bin histogram beliefs in sample size of *N* = 10 (Experiment 1).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0-3	17.2%	18.3%	(15.4%, 21.3%)	15.0%	14.1%
4	20.5%	21.5%	(19.2%, 23.8%)	20.0%	21.9%
5	24.6%	28.1%	(24.9%, 31.4%)	25.0%	27.8%
6	20.5%	18.3%	(16.0%, 20.6%)	18.0%	21.9%
7-10	17.2%	13.8%	(11.9%, 15.7%)	10.0%	14.1%

Figure 7. Three-bin histogram beliefs in sample size of *N* = 10 (Experiment 1).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0-4	37.6%	33.9%	(31.3%, 36.5%)	30.0%	36.1%
5	24.6%	36.2%	(32.8%, 39.6%)	39.0%	27.8%
6-10	37.8%	29.9%	(27.2%, 32.6%)	29.0%	36.1%

Figure 8. Separate-bin-elicitation histogram beliefs in sample size of *N* = 10 (Experiment 1).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0	0.1%	18.0%	(12.5%, 23.6%)	5.0%	0.0%
1	1.0%	36.1%	(29.3%, 42.9%)	30.0%	0.5%
2	4.4%	35.9%	(29.4%, 42.5%)	30.0%	3.0%
3	11.7%	36.7%	(30.2%, 43.3%)	30.0%	10.6%
4	20.5%	38.2%	(30.7%, 45.6%)	29.0%	21.9%
5	24.6%	39.5%	(32.8%, 46.1%)	36.0%	27.8%
6	20.5%	37.7%	(31.1%, 44.3%)	36.0%	21.9%
7	11.7%	34.2%	(27.7%, 40.7%)	28.0%	10.6%
8	4.4%	29.7%	(23.3%, 36.2%)	20.0%	3.0%
9	1.0%	27.9%	(20.7%, 35.1%)	11.0%	0.5%
10	0.1%	11.1%	(6.4%, 15.9%)	1.0%	0.0%

Figure 9. Eleven-bin histogram beliefs in sample size of *N* = 10 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0	0.1%	2.2%	(1.8%, 2.5%)	1.0%	0.0%
1	1.0%	3.8%	(3.3%, 4.3%)	2.5%	0.3%
2	4.4%	5.5%	(5.2%, 5.9%)	5.0%	2.3%
3	11.7%	9.3%	(8.8%, 9.8%)	10.0%	9.8%
4	20.5%	15.1%	(14.4%, 15.7%)	15.0%	22.7%
5	24.6%	28.3%	(26.5%, 30.1%)	25.0%	29.9%
6	20.5%	14.9%	(14.3%, 15.6%)	15.0%	22.7%
7	11.7%	9.2%	(8.8%, 9.6%)	10.0%	9.8%
8	4.4%	5.5%	(5.1%, 5.9%)	5.0%	2.3%
9	1.0%	3.9%	(3.3%, 4.5%)	2.0%	0.3%
10	0.1%	2.4%	(1.9%, 2.9%)	1.0%	0.0%

Figure 10. Five-bin histogram beliefs in sample size of *N* = 10 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0-3	17.2%	15.9%	(14.7%, 17.2%)	13.8%	12.3%
4	20.5%	18.3%	(17.6%, 19.1%)	20.0%	22.7%
5	24.6%	32.1%	(30.2%, 34.0%)	30.0%	29.9%
6	20.5%	18.1%	(17.3%, 18.8%)	20.0%	22.7%
7-10	17.2%	15.6%	(14.4%, 16.8%)	13.0%	12.3%

Figure 11. Three-bin histogram beliefs in sample size of *N* = 10 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0-4	37.7%	34.0%	(32.8%, 35.1%)	35.0%	35.0%
5	24.6%	32.9%	(30.8%, 34.9%)	32.0%	29.9%
6-10	37.7%	33.1%	(32.0%, 34.3%)	34.5%	35.0%

Figure 12. Separate-bin-elicitation histogram beliefs in sample size of *N* = 10 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0	0.1%	4.3%	(3.1%, 5.6%)	1.0%	0.0%
1	1.0%	6.7%	(5.5%, 8.0%)	3.5%	0.3%
2	4.4%	11.9%	(10.5%, 13.3%)	9.0%	2.3%
3	11.7%	16.7%	(15.1%, 18.3%)	12.0%	9.8%
4	20.5%	26.5%	(24.6%, 28.4%)	23.0%	22.7%
5	24.6%	34.3%	(32.0%, 36.6%)	35.0%	29.9%
6	20.5%	24.8%	(22.9%, 26.8%)	20.5%	22.7%
7	11.7%	17.5%	(15.9%, 19.1%)	12.5%	9.8%
8	4.4%	11.9%	(10.5%, 13.4%)	8.4%	2.3%
9	1.0%	6.6%	(5.5%, 7.7%)	3.0%	0.3%
10	0.1%	3.9%	(2.8%, 4.9%)	1.0%	0.0%

Figure 13. Eleven-bin histogram beliefs in sample size of *N* = 1000 (Experiment 1).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-50	0.0%	9.2%	(6.3%, 12.0%)	5.0%
51-150	0.0%	7.2%	(5.6%, 8.8%)	5.0%
151-250	0.0%	8.5%	(6.8%, 10.3%)	9.0%
251-350	0.0%	8.4%	(7.5%, 9.3%)	10.0%
351-450	0.1%	12.9%	(11.0%, 14.8%)	10.0%
451-549	99.8%	18.5%	(15.1%, 22.0%)	11.0%
550-649	0.1%	11.4%	(9.8%, 12.9%)	10.0%
650-749	0.0%	8.2%	(7.2%, 9.2%)	9.0%
750-849	0.0%	6.3%	(5.4%, 7.2%)	5.0%
850-949	0.0%	5.2%	(4.2%, 6.1%)	5.0%
950-1000	0.0%	4.2%	(3.1%, 5.4%)	2.5%

Figure 14. Three-bin histogram beliefs in sample size of *N* = 1000 (Experiment 1).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0-450	0.1%	32.2%	(29.1%, 35.4%)	33.0%	0.01%
451-549	99.8%	39.6%	(36.0%, 43.2%)	40.0%	99.99%
550-1000	0.1%	28.2%	(24.8%, 31.5%)	26.0%	0.01%

Figure 15. Eleven-bin histogram beliefs in sample size of *N* = 1000 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-50	0.0%	2.9%	(2.2%, 3.5%)	1.0%
51-150	0.0%	4.0%	(3.5%, 4.4%)	3.0%
151-250	0.0%	5.2%	(4.8%, 5.6%)	5.0%
251-350	0.0%	7.8%	(7.3%, 8.3%)	9.0%
351-450	0.1%	13.5%	(12.6%, 14.4%)	13.0%
451-549	99.8%	34.0%	(31.1%, 36.9%)	28.0%
550-649	0.1%	13.3%	(12.5%, 14.1%)	13.0%
650-749	0.0%	7.8%	(7.3%, 8.3%)	9.0%
750-849	0.0%	5.2%	(4.8%, 5.6%)	5.0%
850-949	0.0%	3.9%	(3.5%, 4.3%)	2.3%
950-1000	0.0%	2.5%	(2.2%, 2.9%)	1.0%

Figure 16. Three-bin histogram beliefs in sample size of *N* = 1000 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-450	0.1%	29.8%	(28.1%, 31.5%)	30.0%
451-549	99.8%	41.1%	(38.0%, 44.2%)	40.0%
550-1000	0.1%	29.1%	(27.5%, 30.8%)	30.0%

Figure 17. Three-bin histogram beliefs in sample size of *N* = 1000, with the middle bin 481-519 (Experiment 1).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median	(5) Best-Fit RV Model
0-480	10.8%	35.1%	(31.7%, 38.5%)	33.0%	6.7%
481-519	78.3%	35.6%	(32.2%, 38.9%)	35.0%	86.5%
520-1000	10.8%	29.3%	(25.7%, 32.9%)	25.5%	6.7%

Figure 18. Eleven-bin histogram beliefs in sample size of *N* = 1000, with the middle bin 481-519 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-50	0.0%	2.8%	(2.3%, 3.3%)	1.0%
51-150	0.0%	4.2%	(3.7%, 4.7%)	3.0%
151-250	0.0%	5.4%	(5.0%, 5.8%)	5.0%
251-350	0.0%	8.1%	(7.6%, 8.6%)	9.1%
351-480	11.0%	14.5%	(13.7%, 15.3%)	15.0%
481-519	78.0%	30.3%	(27.8%, 32.9%)	25.0%
520-649	11.0%	15.0%	(14.0%, 15.9%)	15.0%
650-749	0.0%	8.0%	(7.6%, 8.5%)	9.0%
750-849	0.0%	5.3%	(4.9%, 5.7%)	5.0%
850-949	0.0%	3.9%	(3.5%, 4.3%)	3.0%
950-1000	0.0%	2.5%	(2.0%, 3.0%)	1.0%

Figure 19. Three-bin histogram beliefs in sample size of *N* = 1000, with the middle bin 481-519 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-480	11.0%	31.4%	(29.9%, 33.0%)	32.8%
481-519	78.0%	36.3%	(33.4%, 39.2%)	34.0%
520-1000	11.0%	32.3%	(30.7%, 33.9%)	33.0%

Figure 20. Sample-size neglect for median beliefs in sample sizes of *N* = 10 and 1000 (Experiment 1).

(1) Bins	(2) *N* = 10, True prob	(3) *N* = 10, Median	(4) *N* = 1000, True prob	(5) *N* = 1000, Median
0-5%	0.1%	2.0%	0.0%	5.0%
5-15%	1.0%	5.0%	0.0%	5.0%
15-25%	4.4%	8.0%	0.0%	9.0%
25-35%	11.7%	10.0%	0.0%	10.0%
35-45%	20.5%	10.0%	0.1%	10.0%
45-55%	24.6%	14.0%	99.8%	11.0%
55-65%	20.5%	10.0%	0.1%	10.0%
65-75%	11.7%	10.0%	0.0%	9.0%
75-85%	4.4%	6.0%	0.0%	5.0%
85-95%	1.0%	5.0%	0.0%	5.0%
95-100%	0.1%	1.0%	0.0%	2.5%

Figure 21. Sample-size neglect for median beliefs in sample sizes of *N* = 10, 1000, and 1 million (Experiment 2).

(1) Bins	(2) *N* = 10, True prob	(3) *N* = 10, Median	(4) *N* = 1000, True prob	(5) *N* = 1000, Median	(6) *N* = 1 million, True prob	(7) *N* = 1 million, Median
0-5%	0.1%	1.0%	0.0%	1.0%	0.0%	1.0%
5-15%	1.0%	2.5%	0.0%	3.0%	0.0%	2.3%
15-25%	4.4%	5.0%	0.0%	5.0%	0.0%	5.0%
25-35%	11.7%	10.0%	0.0%	9.0%	0.0%	9.0%
35-45%	20.5%	15.0%	0.1%	13.0%	0.0%	12.0%
45-55%	24.6%	25.0%	99.8%	28.0%	100.0%	26.0%
55-65%	20.5%	15.0%	0.1%	13.0%	0.0%	10.0%
65-75%	11.7%	10.0%	0.0%	9.0%	0.0%	9.0%
75-85%	4.4%	5.0%	0.0%	5.0%	0.0%	5.0%
85-95%	1.0%	2.0%	0.0%	2.3%	0.0%	2.5%
95-100%	0.1%	1.0%	0.0%	1.0%	0.0%	1.0%

Figure 22. Sample-size neglect for mean beliefs in sample sizes of *N* = 10 and 1000 (Experiment 1).

(1) Bins	(2) *N* = 10, True prob	(3) *N* = 10, Mean	(4) *N* = 10, 95% CI	(5) *N* = 1000, True prob	(6) *N* = 1000, Mean	(7) *N* = 1000, 95% CI
0-5%	0.1%	6.1%	(4.1%, 8.2%)	0.0%	9.2%	(6.3%, 12.0%)
5-15%	1.0%	6.4%	(5.2%, 7.5%)	0.0%	7.2%	(5.6%, 8.8%)
15-25%	4.4%	8.0%	(6.7%, 9.3%)	0.0%	8.5%	(6.8%, 10.3%)
25-35%	11.7%	9.0%	(8.1%, 9.9%)	0.0%	8.4%	(7.5%, 9.3%)
35-45%	20.5%	12.3%	(10.9%, 13.7%)	0.1%	12.9%	(11.0%, 14.8%)
45-55%	24.6%	20.0%	(16.6%, 23.4%)	99.8%	18.5%	(15.1%, 22.0%)
55-65%	20.5%	12.7%	(11.2%, 14.2%)	0.1%	11.4%	(9.8%, 12.9%)
65-75%	11.7%	8.9%	(7.8%, 10.0%)	0.0%	8.2%	(7.2%, 9.2%)
75-85%	4.4%	7.3%	(6.0%, 8.7%)	0.0%	6.3%	(5.4%, 7.2%)
85-95%	1.0%	6.5%	(4.9%, 8.2%)	0.0%	5.2%	(4.2%, 6.1%)
95-100%	0.1%	2.7%	(2.1%, 3.4%)	0.0%	4.2%	(3.1%, 5.4%)

Figure 23. Sample-size neglect for mean beliefs in sample sizes of *N* = 10, 1000, and 1 million (Experiment 2).

(1) Bins	(2) *N* = 10, True prob	(3) *N* = 10, Mean	(4) N = 10, 95% CI	(5) *N* = 1000, True prob	(6) *N* = 1000, Mean	(7) *N* = 1000, 95% CI
0-5%	0.1%	2.2%	(1.8%, 2.5%)	0.0%	2.9%	(2.2%, 3.5%)
5-15%	1.0%	3.8%	(3.3%, 4.3%)	0.0%	4.0%	(3.5%, 4.4%)
15-25%	4.4%	5.5%	(5.2%, 5.9%)	0.0%	5.2%	(4.8%, 5.6%)
25-35%	11.7%	9.3%	(8.8%, 9.8%)	0.0%	7.8%	(7.3%, 8.3%)
35-45%	20.5%	15.1%	(14.4%, 15.7%)	0.1%	13.5%	(12.6%, 14.4%)
45-55%	24.6%	28.3%	(26.5%, 30.1%)	99.8%	34.0%	(31.1%, 36.9%)
55-65%	20.5%	14.9%	(14.3%, 15.6%)	0.1%	13.3%	(12.5%, 14.1%)
65-75%	11.7%	9.2%	(8.8%, 9.6%)	0.0%	7.8%	(7.3%, 8.3%)
75-85%	4.4%	5.5%	(5.1%, 5.9%)	0.0%	5.2%	(4.8%, 5.6%)
85-95%	1.0%	3.9%	(3.3%, 4.5%)	0.0%	3.9%	(3.5%, 4.3%)
95-100%	0.1%	2.4%	(1.9%, 2.9%)	0.0%	2.5%	(2.2%, 2.9%)

(1) Bins	(8) *N* = 1 million, True prob	(9) *N* = 1 million, Mean	(10) *N* = 1 million, 95% CI
0-5%	0.0%	3.7%	(2.5%, 4.8%)
5-15%	0.0%	4.2%	(3.7%, 4.7%)
15-25%	0.0%	5.2%	(4.8%, 5.6%)
25-35%	0.0%	7.6%	(7.1%, 8.1%)
35-45%	0.0%	13.3%	(12.4%, 14.3%)
45-55%	100.0%	34.7%	(31.6%, 37.7%)
55-65%	0.0%	12.4%	(11.6%, 13.2%)
65-75%	0.0%	7.3%	(6.8%, 7.7%)
75-85%	0.0%	5.2%	(4.7%, 5.6%)
85-95%	0.0%	3.9%	(3.5%, 4.4%)
95-100%	0.0%	2.6%	(2.0%, 3.2%)

Figure 24. Three-bin histogram beliefs in sample size of *N* = 1000, with the middle bin 500 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-499	48.7%	40.6%	(39.1%, 42.0%)	46.0%
500	2.5%	19.3%	(16.8%, 21.8%)	10.0%
501-1000	48.7%	40.1%	(38.8%, 41.5%)	45.5%

Figure 25. Three-bin histogram beliefs in sample size of *N* = 1000, with the middle bin 494-506 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-493	34.1%	35.7%	(34.1%, 37.3%)	38.0%
494-506	31.9%	29.1%	(26.4%, 31.9%)	25.0%
507-1000	34.1%	35.2%	(33.6%, 36.8%)	35.0%

Figure 26. Three-bin histogram beliefs in sample size of *N* = 1000, with the middle bin 910 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-909	100.0%	79.1%	(76.2%, 82.1%)	90.9%
910	0.0%	7.4%	(5.7%, 9.0%)	1.0%
911-1000	0.0%	13.5%	(11.6%, 15.5%)	6.3%

Figure 27. Five-bin histogram beliefs in sample size of *N* = 1000 (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-487	21.5%	26.2%	(24.3%, 28.0%)	22.8%
488-496	19.8%	12.5%	(11.5%, 13.6%)	11.0%
497-503	17.5%	21.2%	(18.9%, 23.6%)	18.0%
504-512	19.8%	13.5%	(12.2%, 14.8%)	12.5%
513-1000	21.5%	26.5%	(24.7%, 28.4%)	24.8%

Figure 28. Three-bin histogram beliefs in sample size of *N* = 1 million (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-499,784	33.3%	32.8%	(31.1%, 34.5%)	33.4%
499,785-500,215	33.4%	34.3%	(31.3%, 37.4%)	33.0%
500,216-1,000,000	33.3%	32.9%	(31.2%, 34.5%)	33.8%

Figure 29. Five-bin histogram beliefs in sample size of *N* = 1 million (Experiment 2).

(1) Bins	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0-499,579	20.1%	24.3%	(22.4%, 26.3%)	20.0%
499,580-499,873	20.0%	12.5%	(11.4%, 13.7%)	12.0%
499,874-500,126	19.9%	25.7%	(22.9%, 28.5%)	20.0%
500,127-500,420	20.0%	12.8%	(11.6%, 14.1%)	12.0%
500,421-1,000,000	20.1%	24.6%	(22.6%, 26.6%)	20.0%

Figure 30. Beliefs about the U.S. ethnic distribution (Experiment 1).

(1) Race	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
Non-Hispanic White	63.7%	37.4%	(33.4%, 41.4%)	35.0%
Hispanic	16.4%	15.9%	(14.1%, 17.7%)	15.0%
Non-Hispanic Black	12.2%	20.3%	(17.9%, 22.7%)	10.0%
Asian/Pacific-Islander	4.9%	8.0%	(6.8%, 9.2%)	6.0%
Native American	0.7%	7.3%	(5.6%, 9.1%)	5.0%
Multiracial	2.1%	9.2%	(7.1%, 11.2%)	5.0%

Figure 31. Beliefs about the U.S. ethnic distribution (Experiment 2).

(1) Race	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
Non-Hispanic White	63.7%	43.1%	(41.4%, 44.8%)	42.0%
Hispanic	16.4%	17.1%	(16.2%, 17.9%)	15.0%
Non-Hispanic Black	12.2%	14.7%	(14.0%, 15.4%)	15.0%
Asian/Pacific-Islander	4.9%	13.4%	(12.2%, 14.5%)	10.0%
Native American	0.7%	3.8%	(3.3%, 4.2%)	3.0%
Multiracial	2.1%	8.0%	(7.0%, 9.0%)	5.0%

Appendix Figure I.1. Belief in likelihood of heads after a streak of heads in a sample size of *N* = 10 (Experiment 2).

(1) Prior Heads	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0	50.0%	49.5%	(48.7%, 50.4%)	50.0%
1	50.0%	44.7%	(43.1%, 46.2%)	50.0%
2	50.0%	40.8%	(38.9%, 42.8%)	50.0%
3	50.0%	39.8%	(37.8%, 41.9%)	50.0%
4	50.0%	39.0%	(36.9%, 41.1%)	50.0%
5	50.0%	38.2%	(36.0%, 40.4%)	50.0%
6	50.0%	37.6%	(35.2%, 40.0%)	50.0%
7	50.0%	37.1%	(34.8%, 39.5%)	50.0%
8	50.0%	35.9%	(33.4%, 38.4%)	50.0%
9	50.0%	35.3%	(32.7%, 37.9%)	50.0%

Appendix Figure I.2. Belief in likelihood of heads after a streak of heads in a sample size of *N* = 1000 (Experiment 2).

(1) Prior Heads	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0	50.0%	43.8%	(41.9%, 45.6%)	50.0%
1	50.0%	44.3%	(42.7%, 45.9%)	50.0%
2	50.0%	42.2%	(40.3%, 44.2%)	50.0%
3	50.0%	40.5%	(38.3%, 42.6%)	50.0%
4	50.0%	39.8%	(37.7%, 42.0%)	50.0%
5	50.0%	38.8%	(36.6%, 40.9%)	50.0%
6	50.0%	38.2%	(35.9%, 40.5%)	50.0%
7	50.0%	39.1%	(36.8%, 41.3%)	50.0%
8	50.0%	37.0%	(34.6%, 39.4%)	50.0%
9	50.0%	37.2%	(34.8%, 39.6%)	50.0%

Appendix Figure I.3. Belief in likelihood of heads after a streak of heads in a sample size of *N* = 1 million (Experiment 2).

(1) Prior Heads	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
0	50.0%	45.4%	(43.8%, 47.1%)	50.0%
1	50.0%	44.4%	(42.8%, 46.0%)	50.0%
2	50.0%	41.4%	(39.5%, 43.4%)	50.0%
3	50.0%	41.0%	(38.9%, 43.1%)	50.0%
4	50.0%	39.3%	(37.0%, 41.5%)	50.0%
5	50.0%	38.9%	(36.7%, 41.2%)	50.0%
6	50.0%	38.9%	(36.6%, 41.2%)	50.0%
7	50.0%	38.3%	(36.0%, 40.5%)	50.0%
8	50.0%	38.1%	(35.7%, 40.4%)	50.0%
9	50.0%	37.9%	(35.4%, 40.4%)	50.0%

Appendix Figure I.4. Belief in likelihood of heads after randomly chosen flip locations are known to be heads in a sample size of *N* = 10 (Experiment 2).

(1) Locations	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
1	50.0%	45.7%	(44.1%, 47.2%)	50.0%
2	50.0%	40.7%	(38.9%, 42.4%)	50.0%
5	50.0%	36.7%	(34.4%, 39.0%)	50.0%

Appendix Figure I.5. Belief in likelihood of heads after randomly chosen flip locations are known to be heads in a sample size of *N* = 1000 (Experiment 2).

(1) Locations	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
1	50.0%	44.8%	(43.2%, 46.5%)	50.0%
2	50.0%	42.0%	(40.1%, 43.9%)	50.0%
5	50.0%	39.4%	(37.3%, 41.5%)	50.0%

Appendix Figure I.6. Belief in likelihood of heads after randomly chosen flip locations are known to be heads in a sample size of *N* = 1 million (Experiment 2).

(1) Locations	(2) True prob	(3) Mean	(4) 95% CI	(4) Median
1	50.0%	44.9%	(43.5%, 46.5%)	50.0%
2	50.0%	41.7%	(39.8%, 43.6%)	50.0%
5	50.0%	40.5%	(38.5%, 42.6%)	50.0%