Computational Cognitive Science - Problem Set 1

Name: Chance Jiajie Li
Email: jiajie@mit.edu

1.Coin Flipping

Part A

I. Attach your modified cover story and all four data sets. What effect did you anticipate from the two cover stories you used?
   A. Cover story:
      Dear Participant,
      Welcome to our intriguing experiment! Today, you are going to explore the mysteries of the ordinary coins that we use daily. These coins are minted by the national mint and are a common form of currency in our everyday transactions. They are designed to have a standard weight and balance, with an equal probability of landing on either side.
   B. Datasets:
      1. First Cover Story + Myself - 3, 5, 7, 2, 6, 7, 2, 6, 7
      2. First Cover Story + Friend - 3, 4, 5, 4, 6, 7, 3, 6, 7
      3. Second Cover Story + Myself - 4, 5, 5, 2, 5, 6, 1, 5, 7
      4. Second Cover Story + Friend - 3, 3, 5, 2, 3, 6, 1, 6, 6
   C. Anticipated Effect:
      1. First Condition: This story is likely to induce a sense of curiosity and uncertainty about the coins due to the mention of strange looking. Participants might exhibit a bias towards deeming the coins as unfair, and their judgments might vary more widely across different coin sequences.
      2. Second Condition: This story sets a prior expectation that the coins are regular, standard, and commonly used, suggesting that they are likely to be fair. Participants might be more likely to rate the coins as fair, and there might be a tendency to rationalise any perceived imbalance as a result of chance.

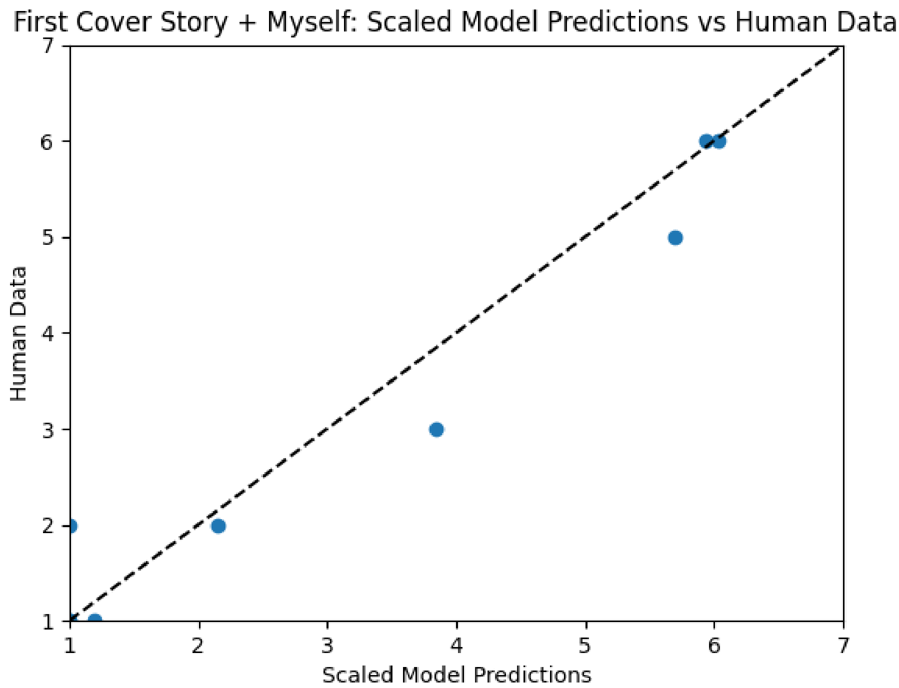
II. Do you see any systematic difference in the ratings between the two conditions of the experiment?
   1. The ratings in "First Cover Story + Myself" and "First Cover Story + Friend" tend to be higher, with more occurrences of 6 and 7, compared to the ratings in "Second Cover Story + Myself" and "Second Cover Story + Friend".
   2. The "Second Cover Story + Myself" and "Second Cover Story + Friend" datasets have more occurrences of lower ratings, such as 1 and 2, compared to the first cover story datasets.
   3. Based on a preliminary observation, it seems like there might be a systematic difference between the two conditions, with the first cover story condition receiving generally higher ratings compared to the second cover story condition. However, to conclusively determine the presence of a systematic difference, a more rigorous statistical analysis, such as a t-test or ANOVA, would be necessary to compare the

means of the ratings between the different conditions and assess the statistical significance of any observed differences.

III. Describe how the differences in the data across the conditions or lack thereof compared to your expectations.
   A. The differences observed were generally in line with my expectations, albeit the variations were somewhat smaller than I had anticipated.

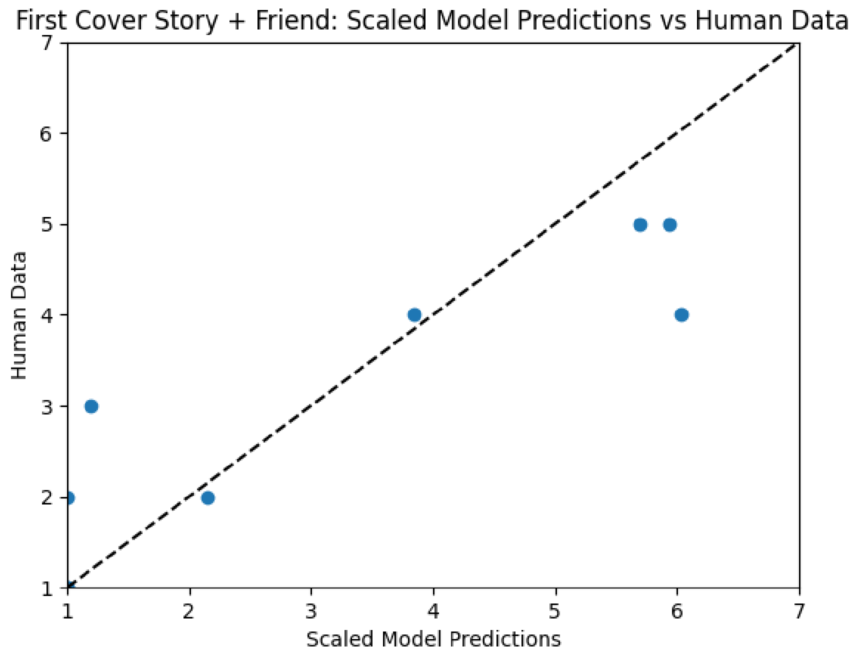
Part B

I. Plot the transformed model predictions against the human data and report a correlation (use np.corrcoef in Python or corrcoef in MATLAB) for each cover story.
   A. First Cover Story + Myself - 3, 5, 7, 2, 6, 7, 2, 6, 7



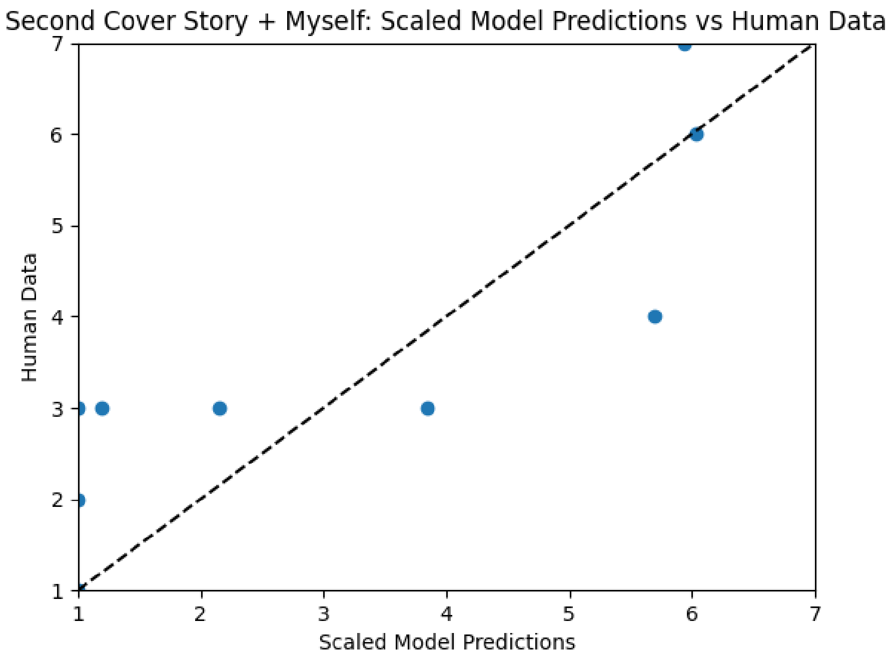
Correlation Coefficient: 0.9747975374698268

   B. First Cover Story + Friend - 3, 4, 5, 4, 6, 7, 3, 6, 7



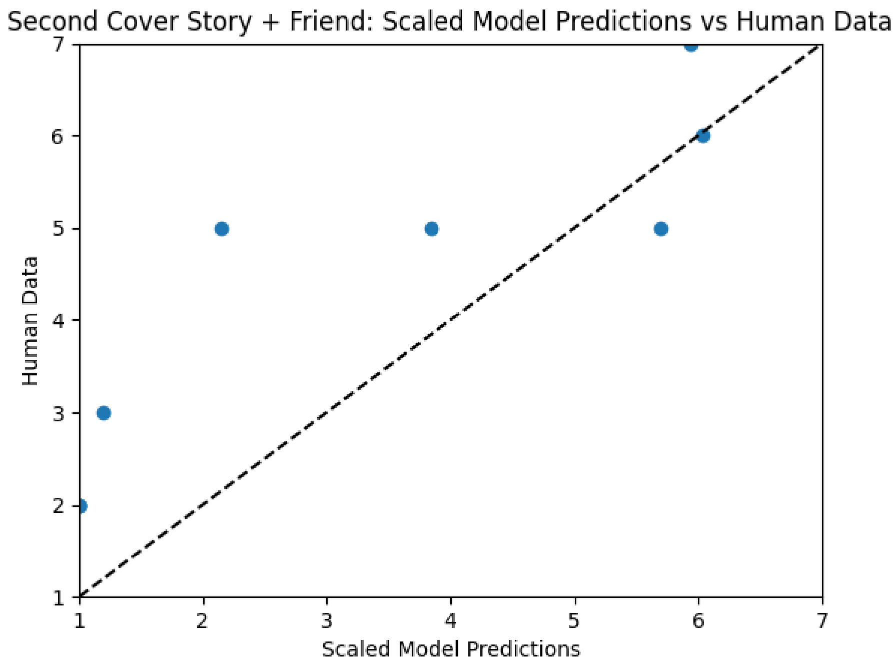
Correlation Coefficient: 0.899912660169252

C. Second Cover Story + Myself - 4, 5, 5, 2, 5, 6, 1, 5, 7



Correlation Coefficient: 0.8478927123898757

D. Second Cover Story + Friend - 3, 3, 5, 2, 3, 6, 1, 6, 6



Correlation Coefficient: 0.901393733013654

II. What settings of a and b seem to work best (you don't need to explicitly search for the best a and b; just try a few)?
   A. For my case, a = 2, b =1 looks good.

III. How did you assess the goodness of fit for a and b?
   A. Quantitatively, I calculated the correlation coefficient between the model predictions and observed values to assess the strength of their linear relationship.
   B. Besides, from a qualitative perspective, I also visually inspected the scatter plot, observing the distance of the data points from the middle line (y=x), to intuitively evaluate the degree of fit of the model.

IV. How well does the model qualitatively capture people's judgments in each condition? Are there any systematic differences between people and the model?
   A. In the second cover story, even though the model's predictions remained unchanged, there were significant differences in people's judgments. People were more inclined to select answers closer to 7 and less inclined to choose answers closer to 1, indicating a greater willingness to believe that the coin is fair. The model's predictions did not reflect this preference in this context.
   B. Besides, regarding the first question, we can observe a horizontal distribution in the graph. This suggests that as the likelihood of the coin being fair decreases, people are more inclined to make significant shifts in their decision-making within the context of the second story. The previous cover story may have made them more certain about this choice.
   C. In the second question, humans tended to provide more conservative answers and were less likely to choose 1 or 7 readily, while the model had a greater tendency to select 1 or 7. This demonstrates that in this scenario, there is a noticeable difference

between humans and the model, with humans being more cautious and the model being more inclined towards extreme choices.

Part C

I. What would the effect be of varying P(H1) on the model predictions (remember that) and why?
   A. Varying P(H1) would have a significant effect on the model predictions. When P(H1) increases, it would imply a stronger belief in hypothesis H1, which represents the fairness of the coin. Because P(H2) = 1−P(H1), as a result, if P(H1) is higher, the model would be more inclined to predict outcomes that align with the belief that the coin is fair (H1), and it would be less likely to predict coin is unfair (H2).
   B. Conversely, when P(H1) decreases, indicating a weaker belief in the coin's fairness, the model would be more willing to predict outcomes consistent with hypothesis H2, which implies bias or unfairness in the coin.

II. (ii) Can you draw any conclusions about which values of P(H1) fit your participants' judgments best in each condition?
   A. First Cover Story + Myself - 3, 5, 7, 2, 6, 7, 2, 6, 7
      1. P(H1) = 0.4
   B. First Cover Story + Friend - 3, 4, 5, 4, 6, 7, 3, 6, 7
      1. P(H1) = 0.35
   C. Second Cover Story + Myself - 4, 5, 5, 2, 5, 6, 1, 5, 7
      1. P(H1) = 0.35 (which is strange)
   D. Second Cover Story + Friend - 3, 3, 5, 2, 3, 6, 1, 6, 6
      1. P(H1) = 0.8

Part D

I. What does the hypothesis space {H1,H2} not capture about people's intuitions?
   A. The hypothesis space {H1, H2} does not capture people's perception of patterns within sequences.

II. Give two examples of coin flip sequences where the hypothesis test above will fail to predict human judgments.
   A. In both cases, the hypothesis test may fail to predict human judgments because it relies solely on statistical probabilities while human participants often consider patterns, sequences, and their own expectations when making judgments about coin fairness
      1. Streaks of Heads and Tails - H H H H H T T T T T
      2. Alternating Pattern - H T H T H T H T H T

2.The Number Game

Part A

I. Attach your modified cover story and all four data sets. What effect did you anticipate from the two cover stories you used?

A. Manually compute the posterior probabilities of the hypotheses “all multiples of 10” and “all even numbers” given the data 10 70 30 (assuming those two are the only hypotheses). Show your work.

h1: all multiples of 10
h2: all even numbers
D: {10, 70, 30}

In this problem set, we’ll use a simple hypothesis space that includes two equally likely types of hypotheses: intervals and multiples-of-k (for integer k) + assuming those two are the only hypotheses
-> P(h1)=P(h2)=0.5

P(h1 | D) = P(D | h1) * P(h1) / Sigma(P(X| h’)* p(h’))
= (1/10)^3 * 0.5 / ((1/10)^3 * 0.5 + (1/50)^3 * 0.5)
= 0.99206349206

P(h2 | D) = P(D | h2) * P(h2) / Sigma(P(X| h’)* p(h’))
= (1/50)^3 * 0.5 / ((1/10)^3 * 0.5 + (1/50)^3 * 0.5)
= 0.00793650793

Part B

A. Manually compute the probability the concept contains the number 40 given the data 10 70 30. (Hint: This should be a simple calculation, the results from part (a) only trivially affect calculation here.)

h1: all multiples of 10
h2: all even numbers
D: {10, 70, 30}

y = 40
P(y ∈ C | h1) = 1
P(y ∈ C | h2) = 1

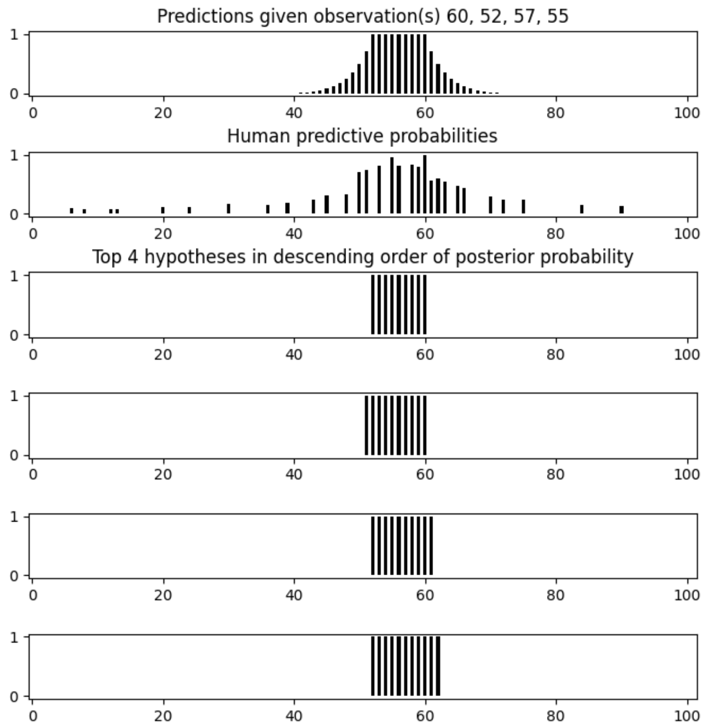
P(y ∈ C | D) = Sigma(P(y ∈ C | h) * P(h | D))
= 1*P(h1 | D) + 1*P(h2 | D)
= 1

Part C

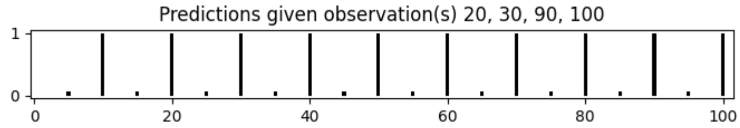
```
def number_game_likelihood(hypothesis, data):

    # Check if data is consistent with the given hypothesis
    data_b = np.array(data, dtype=bool)
    hypothesis_b = np.array(hypothesis, dtype=bool)
    if np.any(data_b & ~hypothesis_b):
        return np.log(0)  # log(0) represents -inf, as P(D|H) = 0 if data is not consistent with hypothesis

    # Count the number of options in the hypothesis
    num_options = np.sum(hypothesis)

    # If hypothesis has no options, it is not consistent with any data
    if num_options == 0:
        return np.log(0)  # log(0) represents -inf

    # Compute the log likelihood
    log_likelihood = np.sum(data) * np.log(1.0 / num_options)

    return log_likelihood
```

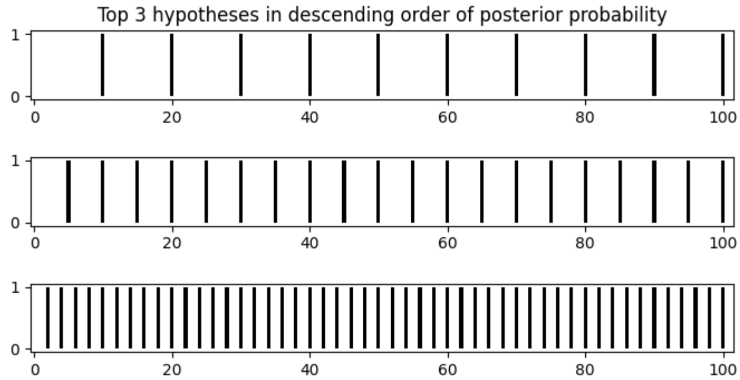
Part D

A. Generate plots showing the predictive distribution for the dataset [60, 52, 57, 55] and one of your own choosing.
   a. dataset [60, 52, 57, 55]



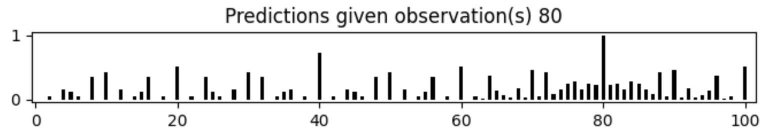
b. dataset [20, 30, 90, 100]

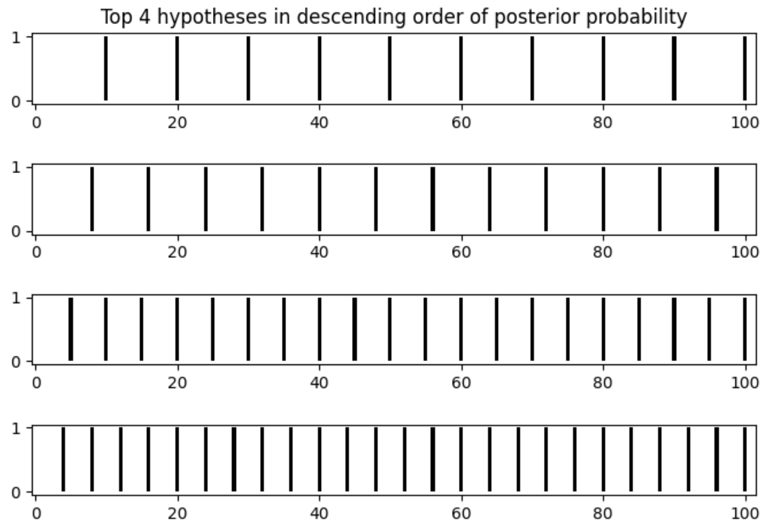




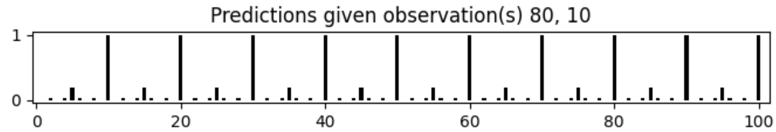
B. Generate plots in sequence for [80], [80, 10], [80, 10, 60], and [80, 10, 60, 30] demonstrating how new data changes the predictive distribution.

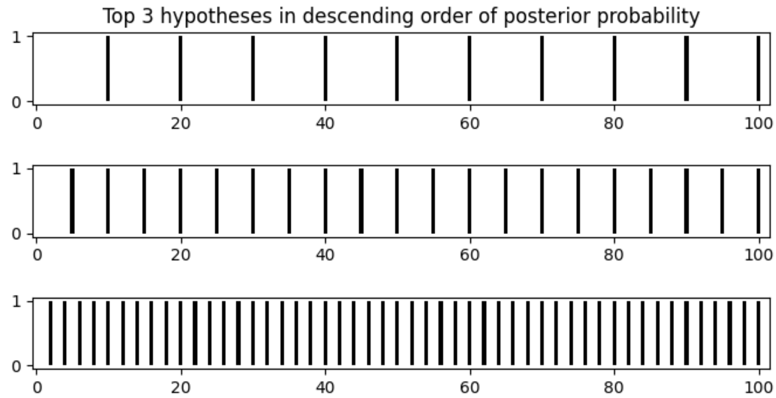
a. dataset [80]



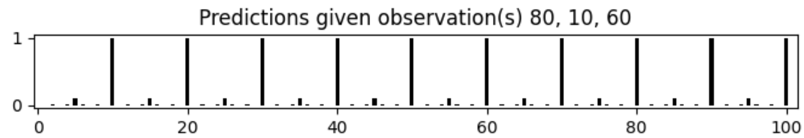


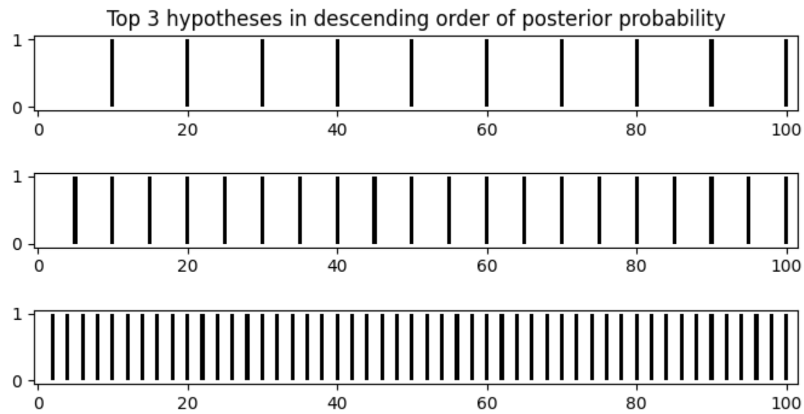
b. dataset [80, 10]



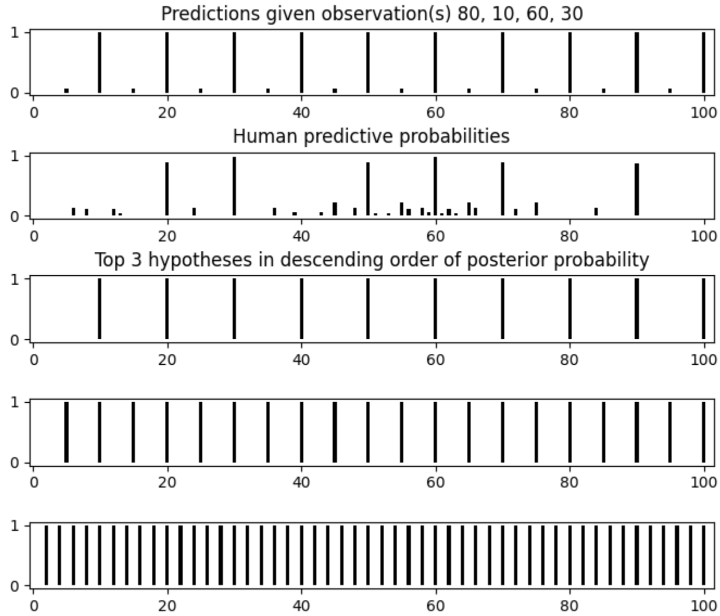


c. dataset [80, 10, 60]



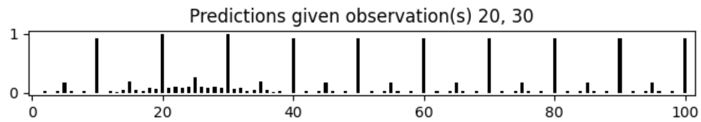


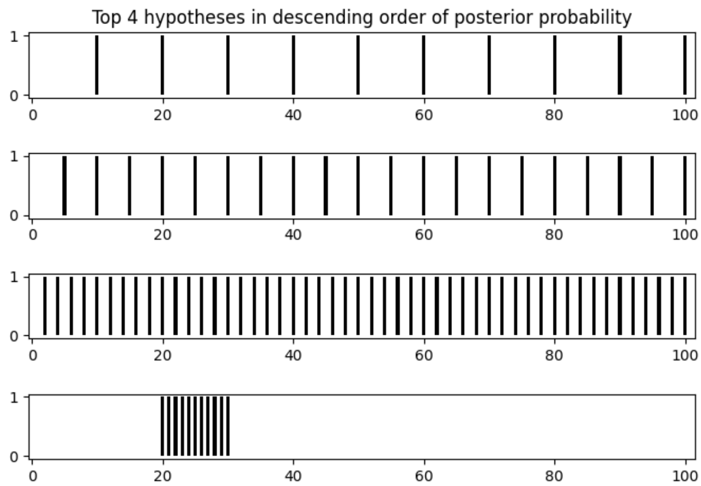
d. dataset [80, 10, 60, 30]



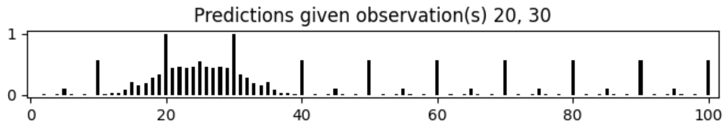
C. Experiment with three alternative settings of the prior, and show how the patterns of generalization change. Discuss and explain the effect of varying the prior.

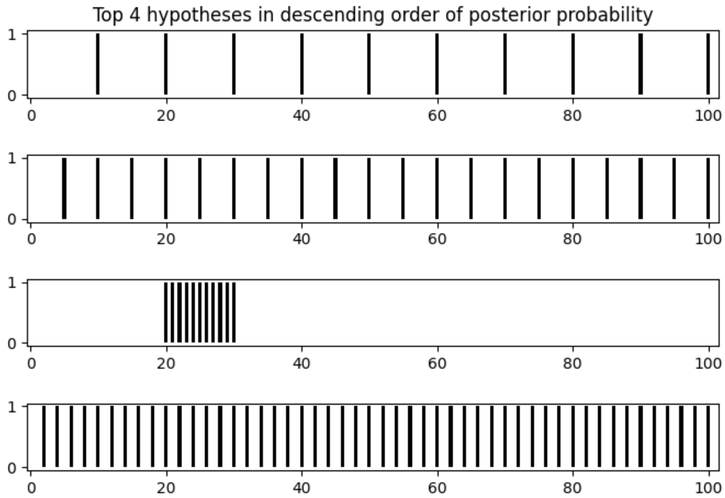
a. P(math)=0.5, P(interval)=0.5, dataset [20, 30]



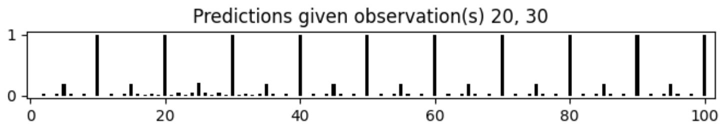


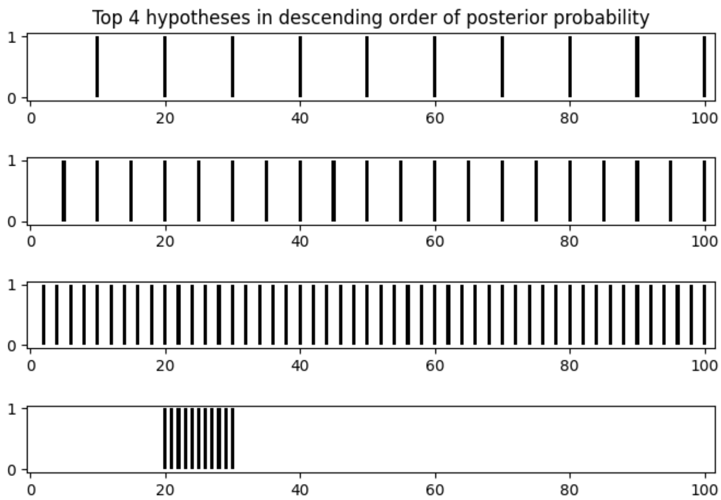
b. P(math)=0.1, P(interval)=0.9, dataset [20, 30]



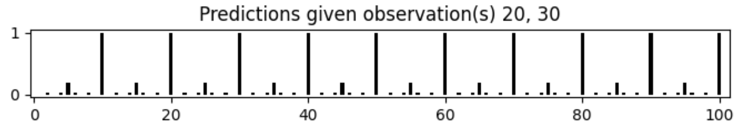


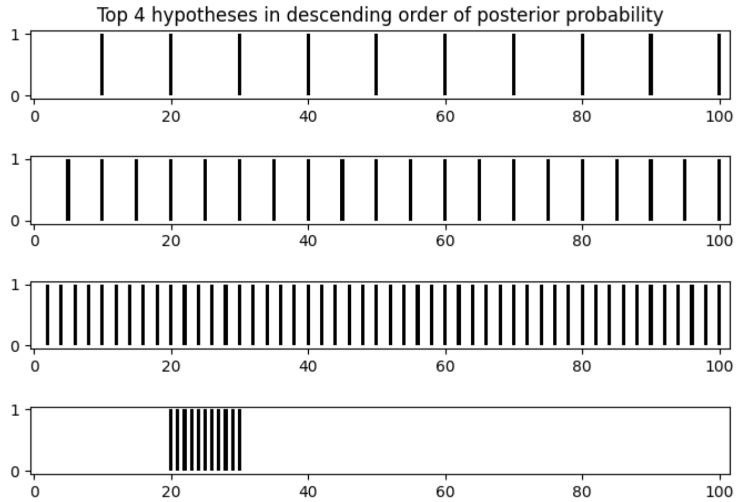
c. P(math)=0.9, P(interval)=0.1, dataset [20, 30]



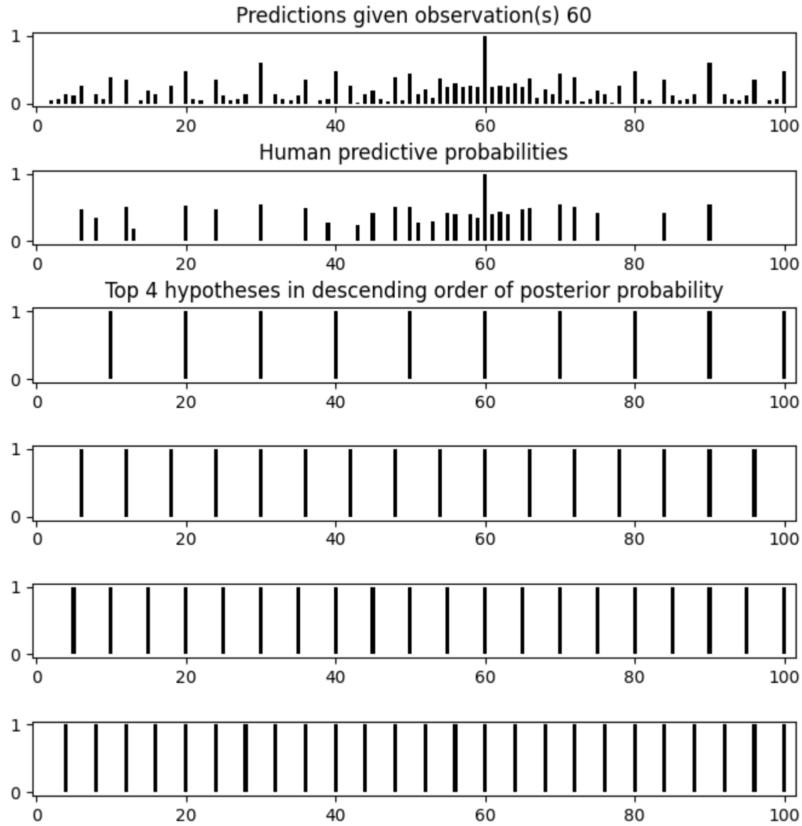


d. P(math)=0.99, P(interval)=0.01, dataset [20, 30]





D. Which settings for the prior best capture the human data? Explain what this implies. (Human data – average subject ratings – will appear as the second, labeled plot whenever the dataset corresponds to one which people were asked about.)



P(math)=0.55, P(interval)=0.45, dataset [60]

Part E

A. How do Marr’s levels apply to the number game? For instance, what level of explanation does it aim for? What aspects of human concept learning does or doesn’t it capture?
    a. The number game primarily operates at Marr's computational level, aiming to understand the abstract principles and goals underlying human concept learning, but it doesn’t fully capture the algorithmic and implementational intricacies of human cognition.
    b. It provides insights into hypothesis formation and testing aspects of concept learning but may not capture the richness and variability of real-world human concept learning experiences.
B. Do you think the number game is an ecologically relevant task to study in cognitive psychology (i.e. Does it give us intuitions about how human cognition works outside of the lab?)?
    a. The number game is somewhat ecologically relevant as it models aspects of how humans generalize and form concepts, offering insights into cognitive processes; however, its abstract and simplified nature may limit the direct applicability of its findings to more complex, real-world cognitive tasks.
C. If people play the number game by considering a hypothesis space similar to this one, where might this hypothesis space come from? How might it differ from the one above?
    a. The hypothesis space in the number game might originate from innate cognitive structures or learned experiences, and it might differ from the one above by incorporating more complex, nuanced, or diverse hypotheses, reflecting individual differences, cultural variations, and environmental influences.(e.g. May includes missing values)