The Presence of Large Language Model Biases in Human Surveying

Michelle Ho (mtho@princeton.edu)
Department of Mechanical and Aerospace Engineering, Princeton, NJ 08544 USA

Sam Liang (saml@princeton.edu)
Department of Computer Science, Princeton, NJ 08544 USA

Abstract

This study explores the presence of biases present in large language models when surveying humans. These biases include majority label, recency, and common token biases. We designed a survey using inspiration from prompts given to the GPT-3 language model in Zhao et al.'s 2021 paper "Calibrate Before Use: Improving Few-Shot Performance of Language Models." The questions were modified for human subjects by using questions with multiple correct answers and randomization of symbols to avoid biases associated with certain answers. The survey found that common token and recency biases were more prevalent than majority label bias in humans, though all were present to a degree. The survey had certain limitations such as prompt format length and repetition, question order, biases in the sample tested, and sub-optimal prompts for GPT-3. Still, this study presents preliminary results that shed light on how surveys of this nature can be improved and how large language models can better match human behaviors.

Keywords: Majority label bias, recency bias, common token bias, GPT-3

Introduction

Our project investigates whether three biases are present in humans: majority label bias, recency bias, and common token bias. These biases were discovered to be present in large language models (LMMs) like GPT-3 during in-context learning (Brown et al., 2020) and caused LMMs to have high accuracy variance because it was sensitive to the prompt structure (Zhao, Wallace, Feng, Klein, & Singh, 2021). In-context learning is appending a task-specific prompt to the input of the LMM as an alternative to fine-tuning without needing to update any model weights while still achieving great performance on tasks. For example, in sentiment classification of movies, we can append *(sentence, label)* pairs like *(the movie is bad, negative)* and *(the movie was amazing, positive)* prior to the test input sentence which the model needs to predict the sentiment of. There has been recent research in 2021 and 2022 that has sought to further calibrate GPT-3 to mitigate the effects of these biases (Han, Hao, Dong, Sun, & Wei, 2022)(Sun, Zheng, Hao, & Qiu, 2021). However, because LMMs are pre-trained on large amounts of online data, most of it generated by humans, we want to explore whether humans themselves also express these three biases.

Majority label bias refers to the phenomenon in which GPT-3 was prone toward answers that appeared in the prompt itself. For instance, if the language model is presented with a text classification prompt that has more positive than negative examples, then GPT-3's predictions are more likely to come from the positive class (Zhao et al., 2021). There is not a lot of previous research done with observing majority label bias in humans, rather computational research that cites majority label bias as a limitation to their work such as in research on words that act as polarity shifters(Schulder, Wiegand, & Ruppenhofer, 2020) or named entity recognition that used GPT-3(Vythoulkas, 2022).

Recency bias refers to the tendency of the language model to predict answers that appeared towards the end of a prompt. For example, if a prompt includes more negative examples towards the end, the model is more likely to predict from the negative class (Zhao et al., 2021). Previous research on recency bias in humans has examined its effect on food consumption (Garbinsky, Morewedge, & Shiv, 2014) and task completion when coupled with experience level (Arnold, Collier, Leech, & Sutton, 2000) Other research has explored how recency bias affects other computational, but non-language, models (Fudenberg & Levine, 2014). However, research on humans exhibiting recency bias when being surveyed or asked questions could not be found.

Common token bias refers to the model being more likely to predict tokens that it has seen during pre-training during downstream tasks. For example, it is more likely to predict entities like "America" than rare entities in the LAMA dataset (Zhao et al., 2021). Again, little research exists on how this bias appears in human behavior, rather research focuses on limitations due to common token bias in other computational research such as in models with knowledge graph completion (Xie et al., 2022).

Should humans be found to also be prone to these biases, it could bring more attention to the way in which surveys, screenings, and other examinations are written and presented. Additionally, it will shed light on the gap or similarities between human and GPT-3 probability distributions. Based on our preliminary background research regarding biases in prompting of large language models like GPT-3, we propose the following research question: are humans also prone to the biases (majority label bias, recency bias, and common token bias) that large language models suffer from?

Experimental Setup

We designed our experiment to be similar to Zhao et al., so that we could compare observations and make similar comparisons. We made a survey with questions designed to test each of the 3 biases.

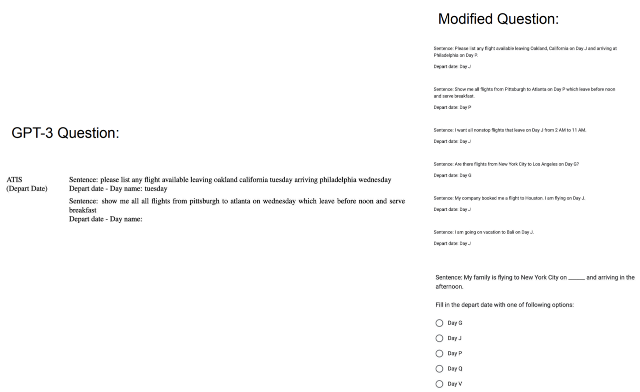

Figure 1: Example of Majority Label Bias Questions for GPT (a) and this experiment (b)

We began by examining the prompts given to GPT-3 in the sample paper and modified them to account for human tendencies. For instance, for an example majority label bias question, as depicted in Figure 1, we kept a similar structure for how the question was asked, but changed the options from days of the week, to randomly labeled days to remove any bias those days may perpetuate.

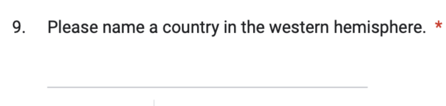

Figure 2: Example of Common Token Bias Question with Multiple Possible Answers

We also chose to rewrite some questions such that multiple correct answers were possible because, for simple questions, humans are often able to select correct answers over exhibiting bias. For instance, as seen in Figure 2, there is not just one answer to this question intended to show common token bias, but we predict that a person's background and upbringing might influence their answer.

We have included samples of our questions in the Appendix for reference. To test majority label bias, questions of the format of Figure 1b and Appendix A Figure 8 were used. To test recency bias, questions of the format of Appendix A Figure 9 were used. To test common token bias, questions of the format of Figure 2 and Appendix A Figure 10 were used. To gather additional data to attempt to make claims for common token bias, additional questions involving class year, hometown, ethnicity, and parents' birthplace were also asked. We were concerned with our participants losing interest and randomly answering if the survey was too long. Therefore, due to the lengths of the questions, we selected four questions for majority label bias, four for recency bias, and six for common token bias. We created our survey in Google Forms (see Appendix B for all the questions) and sent it to 35 participants who were a mix of students from our college, Princeton University, with some outliers from other colleges and relatives. All were given the same questions and arranged in the same order. Questions testing for the same bias were not adjacent. Since it is difficult to know a person's true probability distribution, we are exploring this statement: if humans have a similar probability distribution to GPT-3, then they should also exhibit these biases. We will feed the questions (in the same prompt format as in the survey) to OpenAI's GPT-3 API and compile GPT-3's output distributions for comparison with the human results. Importantly, we also designed the questions such that they allowed for intrinsic qualitative analysis of the biases without comparing them to GPT-3 results. Unless otherwise stated, the GPT-3 data was obtained from text-davinci-002.

Results

Survey:

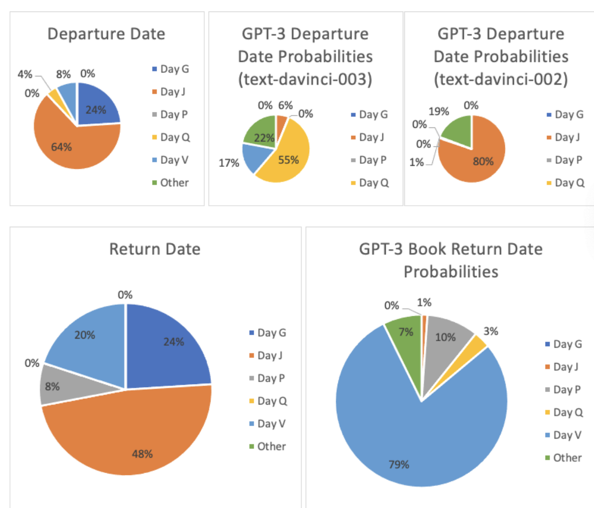

Figure 3: Results for Majority Label Bias Type 1

First, we will begin by discussing our observations from the survey results independent of the GPT-3 results. Of the 35 who received our survey, we received 25 responses. The charted results for the first type of majority label bias questions can be seen in Figure 3. The bias was seen most prominently in the Flight date question, with the most common label, Day J being selected as the answer 64% of the time. It is unclear though if this is also due to recency bias since it was the most recent label used as well. In the library question which had the same format, Day J was also picked most often at 48%, despite not being the majority label. We think that this occurred perhaps because the question may have been confusing. Therefore people might have picked the same answer as the previous question of similar format. We could avoid this in the future by randomizing the days with different random letter labels. Interestingly, the second most common answer was the most recent label, day G, at 24%, so this may have been a case of recency bias. Still, from the first question, there is promise that majority label bias is present.

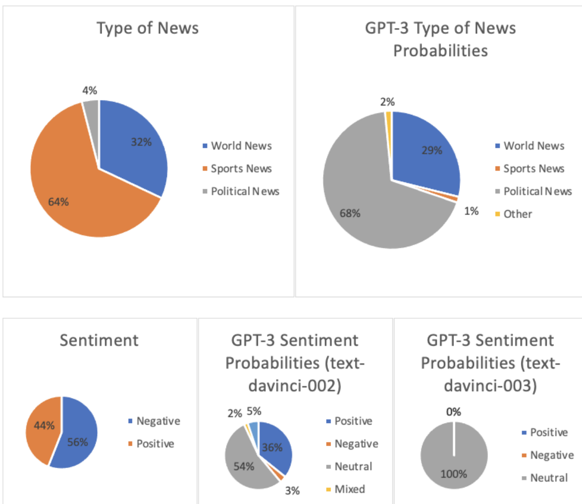

Figure 4: Results for Majority Label Bias Type 2

The results for the other form of the majority label bias questions can be seen in Figure 4. For the first one, we attempted to write a prompt that was equally political, sports-based, and worldly, though the success of that is up to interpretation. We expected the majority label, political news, to be the most commonly selected but it was only selected by one person, with sports news at 64% being the most common. This may be because we had selected a topic, the World Cup, that has been most recently associated with sports. As for the similarly formatted question involving sentiment, we expected negative to be more commonly selected, when positive was at 56%, which is interestingly close to what we expect the results to be when there is no majority label (50/50). This was nearly similar to a prompt given to GPT-3. However, a possible explanation is that since humans are skilled at identifying sentiment, the term ”ok” might already have certain connotations for certain settings rather than being completely neutral. Therefore, in future iterations, the questions could be written with these considerations in mind.

For the recency bias questions, with results pictured in Figure 5, the first one involving volcanoes nearly displayed this bias. The last listed type, shield, and second to last listed type, composite, tied for the most common answer at 40%. For the second similarly formatted question involving microscopes, the most common answer was actually the first, simple microscope at 32% with the most recently listed one, scanning probe as the third most common answer at 16%. This re-

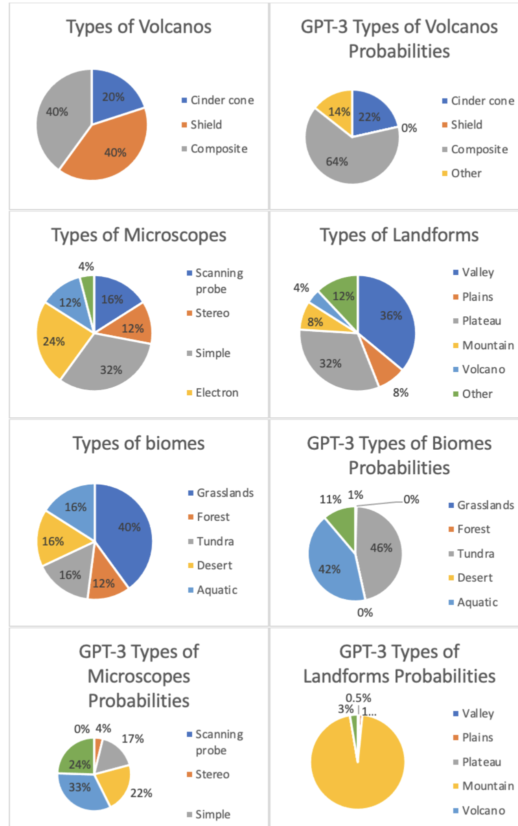

Figure 5: Results for Recency Bias Questions

sult could be portraying a primacy bias, or the type ”simple” might have just been easier to remember than the other more complicated names. For the third similarly formatted question involving landforms, recency bias was present, with the most recently listed, valleys, selected most often at 36%. Unlike in the other examples, the name did not appear in the same part of the sentence, whereas for instance in the volcano example, the type of volcano always came at the beginning of the sentence. This could have made a difference in the resulting answer selected. In the final recency bias question involving biomes, the most common answer, grasslands at 40%, appeared second. Then the first and last two mentioned biomes tied at 16%. It is possible at this point that participants picked up on the type of question being asked and did not read the whole prompt. From these results, it is possible that recency bias is present but the question format was not ideal for showing that; this needs to be considered in a future iteration.

For the common token bias questions, regarding the questions meant to test participants’ instinctual answers to major brands (results shown in Figure 6), both the coffee chain and burger questions’ results were as expected. In fact, the top two answers for burgers, McDonald’s at 44% and Burger King at 36% are the largest chains in the US (Restaurant Business, 2020) and the top answer for coffee chains, Starbucks at

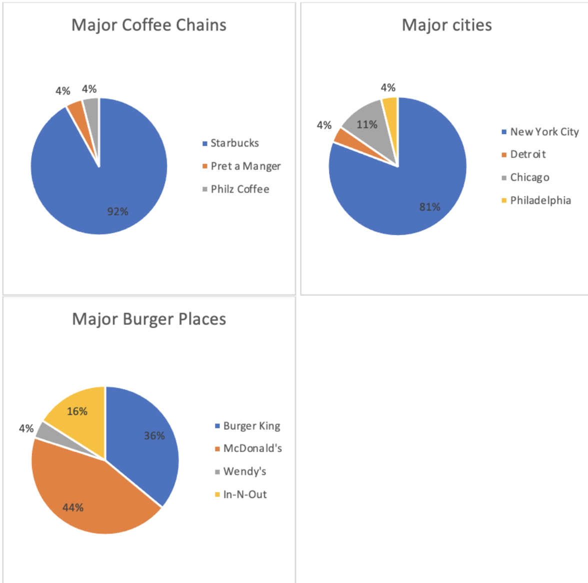

Figure 6: Results for Common Token Bias Questions Type 1

92% is the most common coffee chain worldwide (Brennan, 2022). Additionally, for the question asking for the city, participants associated with the phrase "a city" most with New York City. This correlates to most participants being from Princeton University. Moreover, the two people who grew up in Chicago answer "Chicago," further supporting the presence of common token bias.

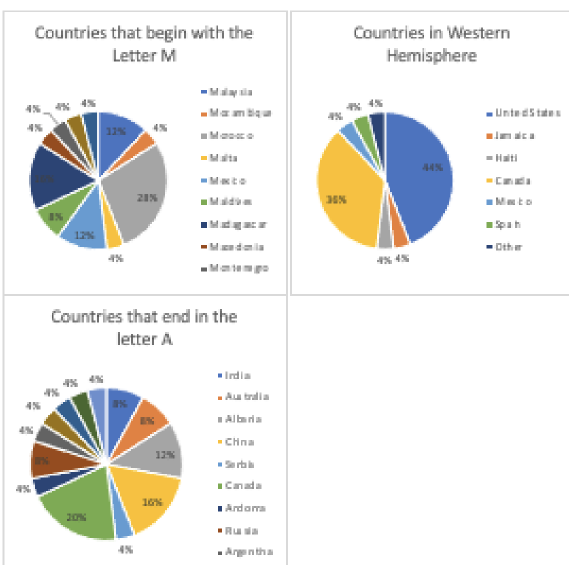

Figure 7: Results for Common Token Bias Questions Type 2

The results for common token bias questions about countries are seen in Figure 7. For the question involving naming a country that begins with the letter "M," there were several different answers, with Morocco being the most common over Mexico. This was slightly unexpected considering most participants grew up in the US. However, this result may be due to the Morocco match in the World Cup that occurred on the day this survey was distributed. Some answers were expected, like a participant from Korea answering "Malaysia" or a participant from Brazil answering "Mexico." In fact, the second most common answer was Madagascar, which might be due to the widely popular children's movie, "Madagascar", but that cannot be assumed without further exploration. Regarding the question about countries in the Western Hemisphere, the most common answer was the United States at 44% and then Canada at 36% which was expected. Additionally, the participants from outside the US answered Canada, which also was not surprising. Finally, regarding the question about countries ending in the letter "a," the most common answers were Canada and China at 20% and 16% which was expected. Additionally, those of Indian descent answered India, and two people of Chinese descent answered China, which was also expected. It is also possible that participants misread the question because several answers also started with the letter "A," and this question came after one asking them to name a country that began with a different letter. Still, these results shed light on the existence of common token bias appearing in surveying.

GPT-3:

The GPT-3 results presented in the figures are from text-davinci-002 unless otherwise stated. We ran additional experiments on the prompts given to humans to further confirm whether GPT-3 text-davinci-002 exhibited the three biases as presented in (Zhao et al., 2021).

The first set of majority label results in Figure 3 shows that text-davinci-002 exhibits majority label bias. In the departure date question, Day J is the majority answer in the prompt which text-davinci-002's predicts with a probability of 0.8. Similarly, for the book return date question, Day V is the majority answer and the most probable prediction at a probability of 0.79. These results are expected. However, we ran an additional experiment with text-davinci-003 which produced interesting results. We see that text-davinci-003 does not exhibit majority label bias for the departure date question. It predicts Day Q with a probability of 0.55 and a range of probabilities for the other days. This is most likely the result of the changes they made to training text-davinci-003 from text-davinci-002.

However, in the second set of majority label results, shown in Figure 4, we see conflicting results from GPT-3. When asked about the genre of news of the test input, GPT-3 assigns the highest probability of 0.68 to the majority label answer, political news. However, when asked about the sentiment of the test input, GPT-3 is unsure about the answer and actually assigns the highest probability of 0.5 to "Neutral". This is interesting because only "Positive" and "Negative" were answers provided in the prompt examples and the test input, "The movie was ok," has a neutral sentiment. In this case, GPT-3 exhibited an understanding of the test input.

However, we believe that if the test input were of positive or negative sentiment, then majority label bias would come into play as shown in (Zhao et al., 2021). Again, we tested the sentiment question on text-davinci-003 which assigned a probability of 100% to "Neutral", showing that it is smarter than text-davinci-002 which could be a reason why it exhibited less bias on the departure date question.

We see mostly negative results for recency bias in Figure 5. The question on the types of biomes had five listed biomes. Aquatic was the first listed type and tundra was the last listed type. GPT-3 gave 42% and 46% probability to aquatic and tundra, respectively, exhibiting some recency bias. The question on the types of microscopes had five listed microscopes. GPT-3 assigned 17%, 22%, and 33% to simple, electron, and compound microscopes, respectively. They are the first three listed microscopes. It assigned zero probability to the last listed microscope, the scanning probe. For the types of landforms question, we listed five landforms. GPT-3 assigned 96% probability to mountains which was the first listed landform. On the last question about the types of volcanoes, we listed three volcanoes. It assigned 64% and 22% to composite and cinder cone volcanoes in the types of volcanoes question. These last three experiments all show negative results for GPT-3 exhibiting recency bias. This happened because the prompts were not designed specifically solely to induce recency bias in GPT-3. They were designed for humans. The prompts were also long with a lot of content and each type only appeared once which may have made it difficult to examine recency bias in GPT-3.

Conclusion and Future Work

The purpose of this project was to explore the presence of large language model biases in humans, specifically majority label, recency, and common token biases. Without access to a person's true probability distribution, we used GPT-3's probability distribution as a baseline for comparison. If the estimated human probability distribution matches GPT-3, then that is one way to confirm that humans also exhibit these biases. We did this by giving a survey to a group of participants and engineered the prompts to allow for qualitative evaluation of these biases as well. We see from our results that these biases are most likely present in humans, particularly recency and common token biases, but not to the same degree as with GPT-3. Interestingly, our human-designed prompts also exhibited unexpected and interesting behaviors from GPT-3 that could provide more insight into how it works. This could be an avenue for future work.

We noticed a few flaws in our experimental design that should be addressed in future iterations. For instance, the order of the questions combined with reusing similar question formats could allow participants to predict upcoming questions, such that they skip over the reading that precedes the question. This can be addressed by creating more varieties of questions for testing these biases. Additionally, it is possible that participants were not answering with their first instinctual answer, rather than thinking more about it before typing in an answer. This can be addressed by making the survey verbal. The sample size could have been better varied to see how results change with different generations, upbringings, backgrounds, and education, among other things. Moreover, many of the recency bias questions did not allow us to investigate recency bias in GPT-3 well, and our common token questions would produce no meaningful insights into common token bias in GPT-3. Thus, designing prompts that work well for investigating biases in both humans and GPT-3 is paramount. Moving forward, we can use these insights and future related studies when considering how we engineer surveys for testing. A natural step from here would be to redesign the survey from the flaws we noticed to obtain higher fidelity results and move towards bridging the gap in the differences between the probability distributions for humans and GPT-3.

Acknowledgments

We would like to thank Prof. Griffiths, Dr. Tom McCoy, and Dr. Jake Snell for their insightful advice throughout this project.

References

Arnold, V., Collier, P. A., Leech, S. A., & Sutton, S. G. (2000, July). The effect of experience and complexity on order and recency bias in decision making by professional accountants. *Accounting & Finance*, *40*(2), 109–134. Retrieved 2022-10-06, from `https://onlinelibrary.wiley.com/doi/10.1111/1467-629X.000` doi: 10.1111/1467-629X.00039

Brennan, S. (2022). *10 most famous coffee chains in the world (2022 update).* Retrieved from `https://coffeeaffection.com/most-famous-coffee-chains/`

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. *CoRR*, *abs/2005.14165*. Retrieved from `https://arxiv.org/abs/2005.14165`

Fudenberg, D., & Levine, D. K. (2014). Learning with recency bias. *Proceedings of the National Academy of Sciences*, *111*, 10826-10829.

Garbinsky, E. N., Morewedge, C. K., & Shiv, B. (2014, July). Interference of the End: Why Recency Bias in Memory Determines When a Food Is Consumed Again. *Psychological Science*, *25*(7), 1466–1474. Retrieved 2022-10-06, from `http://journals.sagepub.com/doi/10.1177/0956797614534268` doi: 10.1177/0956797614534268

Han, Z., Hao, Y., Dong, L., Sun, Y., & Wei, F. (2022, May). Prototypical Calibration for Few-shot Learning of Language Models. Retrieved 2022-10-06, from `https://arxiv.org/abs/2205.10183v2` doi: 10.48550/arXiv.2205.10183

Restaurant Business. (2020). *The largest burger chains in the U.S.* Retrieved from `https://www.restaurantbusinessonline.com/financing/larges`

Schulder, M., Wiegand, M., & Ruppenhofer, J. (2020, May). Enhancing a lexicon of polarity shifters through the supervised classification of shifting directions. In *Proceedings of the 12th language resources and evaluation conference* (pp. 5010–5016). Marseille, France: European Language Resources Association. Retrieved from https://www.aclweb.org/anthology/2020.lrec-1.616

Sun, Y., Zheng, Y., Hao, C., & Qiu, H. (2021, September). NSP-BERT: A Prompt-based Zero-Shot Learner Through an Original Pre-training Task–Next Sentence Prediction . Retrieved 2022-10-06, from https://arxiv.org/abs/2109.03564 doi: arXiv:2109.03564

Vythoulkas, V. (2022). *Reformulating named entity recognition*. Unpublished doctoral dissertation, Athens University of Economics and Business.

Xie, X., Zhang, N., Li, Z., Deng, S., Chen, H., Xiong, F., … Chen, H. (2022). From discrimination to generation: Knowledge graph completion with generative transformer. *CoRR*, *abs/2202.02113*. Retrieved from https://arxiv.org/abs/2202.02113

Zhao, T. Z., Wallace, E., Feng, S., Klein, D., & Singh, S. (2021). Calibrate before use: Improving few-shot performance of language models. *CoRR*, *abs/2102.09690*. Retrieved from https://arxiv.org/abs/2102.09690

Appendix

A. Survey Question Examples

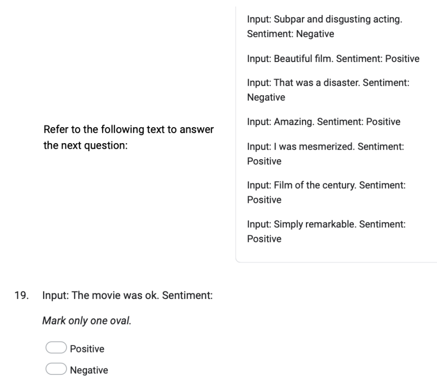

Figure 8: Example of Majority Label Bias Questions Type 2

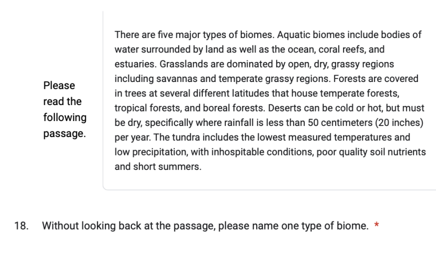

Figure 9: Example of Recency Bias Questions

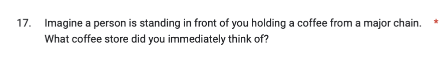

Figure 10: Example of Common Token Bias Question Type 2

B. Survey

This appendix includes the exact questionnaire given and answered by our 25 participants in this order. The survey was given with Google Forms and was live for a 3-day period.

454 Study Questionaire

Hey guys! Thanks for filling out our survey for our class, we really appreciate it (and just go with your gut instinct if you're unsure) ~Sam and Michelle

* Required

Data Collection Questions

1. What is your class year *

   Mark only one oval.

   - 2023
   - 2024
   - 2025
   - 2026
   - Other

2. Where did you grow up? *

   ____

3. What kind of school did you attend before college and where was this school? *

   ____

4. Where did your parents grow up? *

   ____

5. What is your ethnicity? *

   ____

Please read the following passage.

The most well-known types of volcanoes are cinder cones, composite volcanoes (stratovolcanoes), and shield volcanoes. Cinder cones are the most common type of volcano in the world. They may look like an idealized depiction of a volcano as they are steep, conical hills that usually have a prominent crater at the top. Composite volcanoes can be the most picturesque of all volcanoes. A classic composite volcano is conical with a concave shape that is steeper near the top, with a snow-covered peak. Shield volcanoes, the largest volcanoes on Earth, are broad volcanoes with gentle slopes and are shaped somewhat like a warrior’s shield lying flat on the Earth.

6. Without looking back at the passage, what is one type of volcano? *

7. Please name a country that starts with the letter M. *

Sentence: Please list any flight available leaving Oakland, California on Day J and arriving at Philadelphia on Day P.

Depart date: Day J

Sentence: Show me all flights from Pittsburgh to Atlanta on Day P which leave before noon and serve breakfast.

Depart date: Day P

Sentence: I want all nonstop flights that leave on Day J from 2 AM to 11 AM.

Depart date: Day J

Refer to the text below to answer the question:

Sentence: Are there flights from New York City to Los Angeles on Day G?

Depart date: Day G

Sentence: My company booked me a flight to Houston. I am flying on Day J.

Depart date: Day J

Sentence: I am going on vacation to Bali on Day J.

Depart date: Day J

8. Sentence: My family is flying to New York City on ______ and arriving in the afternoon.

Fill in the depart date with one of following options:

Mark only one oval.

- Day G
- Day J
- Day P
- Day Q
- Day V

9. Please name a country in the western hemisphere. *

Please read the following passage.	There are different types of microscopes and each of these has different purposes of use. A simple microscope is defined as the type of microscope that uses a single lens for the magnification of the sample. A compound microscope is defined as the type of microscope that has more than one lens. It has a combination of lenses and two optical parts known as an objective lens and an eyepiece or ocular lens. An electron microscope is defined as the type of microscope in which the source of illumination is the beam of accelerated electrons. A stereo microscope is defined as a type of microscope that provides a three-dimensional view of a specimen. The scanning probe microscope is defined as the type of microscope that finds applications in industries where the examination of the specimen is done at the nanoscale levels.

10. Without looking back at the passage, please name one type of microscope. *

11. Imagine a person standing in front of you with a burger from a major fast food place. Which chain did you immediately think of?

Refer to the following text to answer the next question:

Input: Thousands evacuated after Indonesia's Mount Semeru erupts. Casualties are unknown.

Label: World News

Input: Republican hopes fade as Warnock momentum picks up in Georgia ahead of Tuesday's runoff.

Label: Political News

Input: Lost remains of last Tasmanian tiger found hiding in plain sight at the Tasmanian Museum and Art Gallery (TMAG) in the Australian island state, Tasmania.

Label: World News

Input: Los Angeles Lakers' Anthony Davis is dominating at the position he's never really wanted to play. He scored 44 points against the Milwaukee Bucks.

Label: Sports News

Input: Nancy Pelosi announces that she is stepping down from House leadership.

Label: Political News

Input: The Iranian government has signaled that it could abolish its morality police after months of protests that erupted after the death of a young woman in the custody of the morality police.

Label: Political News

Input: Putin signs Law Banning Expressions of L.G.B.T.Q. Identity in Public.

Label: Political News

12. Input: FIFA brought disciplinary charges against Serbia on Monday for alleged misconduct by players and fans including offensive chants directed at Swiss players with Kosovonian roots.

Label:

Mark only one oval.

- World News
- Sports News
- Political News

13. Please name a country that ends in the letter "a". *

Please read the following passage.	Many different types of landforms make up Earth’s topography. The most common type of mountains arise where the Earth’s crust experienced folding or faulting, which results in fault-block mountains and volcanos. Most of the Earth's surface consists of low and high plains, defined by a mostly level profile that ranges from gently rolling to completely flat. Plateaus can be thought of as elevated plains – that is, elevated flattish areas. Valleys, either U or V shaped, are created from the river and glacier movements.

14. Without looking back at the passage, please name one type of landform. *

15. Your friend says to you, "'I'm going to a city." Which city did you immediately think of? *

Refer to the following text to answer the next question:

Sentence: I borrowed the book last week and need to return it to the library on Day V.

Return date: Day V

Sentence: The deadline to return the book is Day J otherwise I will receive a fine.

Return date: Day J

Sentence: John called me today to return the book to him on Day J.

Return date: Day J

Sentence: Since you're going out on Day V, can you return the book while you're on the way?

Return date: Day V

Sentence: My boss wants me to return the book on Day V, because he will be on vacation after that.

Return date: Day V

Sentence: I haven't finished reading the book, but I need to return it by Day G.

Return date: Day G

16. Sentence: I lost the book that I borrowed, but I need to find it before ______ which is when I must return it.

Fill in the return date with one of following options:

Mark only one oval.

- Day G
- Day J
- Day P
- Day Q
- Day V

17. Imagine a person is standing in front of you holding a coffee from a major chain. * What coffee store did you immediately think of?

____

Please read the following passage.	There are five major types of biomes. Aquatic biomes include bodies of water surrounded by land as well as the ocean, coral reefs, and estuaries. Grasslands are dominated by open, dry, grassy regions including savannas and temperate grassy regions. Forests are covered in trees at several different latitudes that house temperate forests, tropical forests, and boreal forests. Deserts can be cold or hot, but must be dry, specifically where rainfall is less than 50 centimeters (20 inches) per year. The tundra includes the lowest measured temperatures and low precipitation, with inhospitable conditions, poor quality soil nutrients and short summers.

18. Without looking back at the passage, please name one type of biome. *

____

Refer to the following text to answer the next question:

Input: Subpar and disgusting acting. Sentiment: Negative

Input: Beautiful film. Sentiment: Positive

Input: That was a disaster. Sentiment: Negative

Input: Amazing. Sentiment: Positive

Input: I was mesmerized. Sentiment: Positive

Input: Film of the century. Sentiment: Positive

Input: Simply remarkable. Sentiment: Positive

19. Input: The movie was ok. Sentiment:

Mark only one oval.

- Positive
- Negative