Question Asking as Model Updating

Introduction: Questions are a key part of human communication and learning. An important aspect of asking effective questions is to tailor them to our information goals. For example, learning how cars work you might ask, "What is the difference between a car's brake pads and brake shoes?" or "How does the engine work?" You would probably ask a different set of questions if you were trying to buy a new car - "What is the warranty on the car?" and "What is the fuel efficiency of the car?" Previous work has been done studying if humans ask good questions in goal-oriented situations and how to engineer models that generate questions for specific goals [1]. However, there is a lot more richness in human question asking than just seeking information for a specific goal. When learning, humans may ask questions because there are gaps in our knowledge, and we seek some understanding. This requires a sense of the knowledge that one lacks and the kind of information that would bring clarity. But what are the formal mechanisms of this process? **Can a computational model flexibility look at its own knowledge state and ask questions about its lack of understanding?**

Background and Rationale: Building off previous work, this proposal views human thinking as building generative probabilistic models of the world [2]. A probabilistic generative model is a model that approximately represents the causal processes which generate data, while allowing for uncertainty. For example, a generative model of how plants grow would represent how plants need sunlight, carbon dioxide, and water. A question is a linguistic statement that is used to elicit information. Human language, including questions and explanations, can often be used to build those probabilistic generative models of the world [2]. For example, your third-grade science teacher explaining how plants grow. But, what if one day I overwater my plant and it dies? I may question why it died - what did I do wrong. It seems like my generative model of how plants grow did not consider this overwatering phenomenon. But it also seems like my model had a lot of things right (e.g., that water is needed for plants to grow). I do not want to completely erase my model just to build a new one with a few differences. Thus, given new data, how do I update my model? How does the mind find a new concept or structural property to add to its world model to explain the new data I have just seen (or been told)? Even though much work has been done in how people build generative probabilistic models of their world [2], not much work has studied how to incrementally change the structure of a generative model - such as adding a new concept or causal relationship. What role does question asking play in how humans change their generative models?

Aims: **This proposal claims that through investigating what questions people ask, we can infer what edits people make to their generative models.** There are two main aims: 1) When do people ask questions? The hypothesis is that people ask questions when they detect that their model is incomplete. In computational terms, when they receive data that is inconsistent with their generative model. Data that is inconsistent with the generative model is data that has zero probability of occurring within the generative model. 2) Why do people ask certain questions and what kind of questions do people ask? This hypothesis is that people will ask questions based on a theory they have of how the world works. This theory is formalized in a probabilistic generative model. **People ask questions to help guide search for how to update their generative model that can explain the new data that they have just seen. The kinds of questions people ask help them fix errors in the code of their generative model.**

Experimental Methods: Participants are randomly assigned to one of two conditions. In the first condition, participants are presented with a story [Fig. 1]. They will then be presented with pieces of data relating to the story. The data will either be inconsistent or consistent with the generative model that the story elicits. Anytime new data is presented, participants have the option to press a button to ask a question. They can then type their question in the text box. The purpose of this first condition is to investigate if participants will ask questions when presented with inconsistent data, even when there is no direct goal that participants are tasked with. It is predicted that when presented with inconsistent data more questions will be asked than when presented with consistent data.

In the second condition, participants are presented with a story (like in the first condition) but are then told a goal-question that they must answer at the end of the trial. Then participants are presented with data that will either be inconsistent or consistent with the generative model that the story elicits. Once

again, anytime new data is presented, participants have the option to press a button to type in a question. This goal-question serves as a goal that participants are trying to find the answer to through the course of the trial. However, this goal-question will ask about data that is inconsistent with the generative model that is elicited by the story. Participants must use a different or updated model to answer the goal-question. Thus, the purpose of this condition is to investigate if participants ask questions that are directly aimed at updating their generative model.

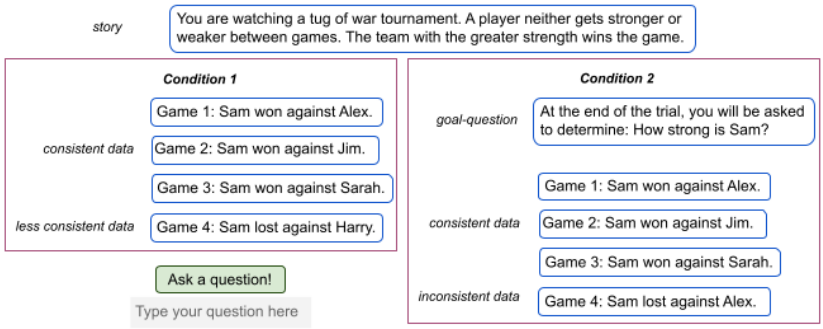

Figure 1: An example of the experimental set up. The story is a tug of war tournament. In Condition 1, no goal-question is asked. In Condition 2, a goal-question is asked.

Computational Modeling: This proposal seeks to also formalize its aims computationally. To correlate model predictions with human predictions I will create a set of probabilistic generative models that correspond to unique linguistic stories that participants will be presented. Every story and subsequent piece of data corresponds to a specific probabilistic generative model. In the second experimental condition, to investigate if participants ask questions with the goal of updating their generative model, a qualitative evaluation of the questions people ask in the experiments will be done (this kind of evaluation has been done in similar studies, see [1]). Then I will manually make expert updates to the generative model to reflect the kind of updates humans do through asking questions. A correlation between the non-updated model and the participant response of the goal-question will be calculated, as well as a correlation between the updated model and participant responses to the goal-question (it is predicted that this correlation will be higher).

The main computational question is, if there is an already existing generative model, how does the model propose edits? In other words, how does the model change its code (e.g., write new code)? The hypothesis is that humans have a sense of how to update their code based on the kinds of questions they ask. Could computers do something similar? In previous work, a model was automatically revised based on a grammar of potential modifications, but this was explored only in the domain of Gaussian processes [3]. This proposal will extend these ideas to a broad range of probabilistic generative programs. Additionally, future studies may include training large language models, such as OpenAI's Codex, to automatically update edits in a generative model's code.

Intellectual Merit: The relationship between how humans learn general knowledge from limited observations and why they ask certain questions is unclear. By bridging this gap, this proposal takes a step toward understanding how humans update their knowledge and, similarly, how to build models that can learn more effectively through updating its knowledge state.

Broader Impacts: As AI increasingly becomes more a part of human life, it is important that humans have a sense of trust towards these technologies and can seek formal interpretable explanations when these technologies produce undesired output. Understanding how humans ask questions, why they ask certain questions over others, and how these processes can be represented computationally will help to build AI models that know when its model of the world is incomplete and how to update it. Because question-asking is such an everyday behavior, this project offers a concrete and accessible example for public and youth engagement in the computational and cognitive sciences. As a graduate student, I will mentor and support undergraduate students on research to further foster an inclusive research program.

References: [1] A. Rothe et al. (2018). *Computational Brian & Behavior.* [2] N. Goodman (2014). Concepts in a probabilistic language of thought. In E. Margolis & S. Laurence (Ed.), *The Conceptual Mind.* [3] D. Duvenaud (2013). *International Conference on Machine Learning*.