CHAPTER 1

Introduction

The rich and dynamic change of people's subjective feelings is a fascinating topic yet less being touched from the cognitive science perspective. Overwhelming number of quantitative models and behavioral paradigms are formulated in the goal of maximizing some external rewards. My work in NYU instead focuses on exploring the internal rewards, often conceptualized in terms of subjective judgment or feelings, using computational models with a probabilistic perspective. Specifically, my work explored three different topics: the feeling of suspense in face of information to be revealed (Chapter 1); the satisfaction to an explanation given the statistical information of the causal structure (Chapter 2); the choice preference between snacks depending on the subjective value and active information collection process (Chapter 3).

1.1 modeling suspense as expected future learning

When does a sport match become most suspenseful, that the audience has to hold their breath, forget about eating pop corns or going to bathroom, paying full attention to the game so they do not miss anything? Usually this does not happen at the very beginning of the game because whichever team wins a point is not very consequential; when towards the end of the game, it may still not be suspenseful if one team already have a big advantage that the other side has no chance to flip the situation. But, if the game is towards the end and both sides have a fair chance of winning, then the game could be come really intense and suspenseful. Is this a universal intuition that people would generally agree with, regarding their feeling of suspense? If so, what would be a good way to explain this mechanism?

Empirically, previous studies showed that people do have general agreements to feel more suspense in certain conditions. For example, in a story-telling setting, people may feel more suspense if the chance of the protagonist fails is high and the possible solution for the protagonist has been removed (Comisky & Bryant, 1982; Gerrig & Bernardo, 1994); also, the presence of time pressure could also increases suspense (Alwitt, 2002). What could be an underlying principles behind these factors?

In Ely *et al.* (2015), they proposed a theory that suspense is in proportion to the expected belief update in the upcoming moment, where the belief refers to the estimated probability regarding a significant consequence (E.g. which team will win the game, which candidate will win the election, etc.). Ely et al. applied this framework to explain the suspense dynamics in different kinds of sports as well as mystery novels, political primaries, auctions etc, but no direct human experiment evidence has corroborated these ideas.

My work sought to develop an empirical paradigm that could test the predictions made by Ely et al. in a controlled but also engaging environment. This paradigm also allowed to compare the "expected future learning" model with other heuristics proposed by the previous literature that I quantified in this setting. In Chapter 1, I present the empirical data as well as the model comparison results in two studies.

1.2 modeling satisfying explanation as a combination of distinctive causes

In empirical research, scientists constantly face the problem of how to determine which explanation for the data is the best. Statistical methods for model comparison, such as AIC (Akaike, 1974), BIC (Schwarz *et al.*, 1978), Bayesian model selection (Stephan *et al.*, 2009), all aim to balance between the quality of description towards the data (often evaluated in terms of likelihood), and the complexity of the model (could be quantified by the number of parameter, the prior of the model, etc.). This is the "type" level explanation where a myriad of phenomena are summarized and explained by novel theories / models. Plenty of previous psychology studies on causal inference tasks also explored how people observe or even manipulate instances of causal events, then infer what is the underlying causal structure. Some studies also indicate that probability-based models do well account for people's behavior patterns (Griffiths & Tenenbaum, 2009; Lu *et al.*, 2008).

In contrast to the "type" level explanation, in the daily life, people also often seek for "token" level explanation where the general causal rules are already known, yet they need to find an explanation for a specific instance (E.g. what causes this specific student to be so successful? What disease causes this person to show such a symptom?). How do ordinary people determine the best explanation on the token level? From a computational perspective, do people perform some evaluation strategies similar to the computationally costly statistical algorithms, or do they use some kind of heuristics?

Many previous studies emphasize on heuristic explanation preferences. Specifically, the heuristics regarding preference towards simple or complex explanations have been discussed from different perspectives. People may prefer simple explanations to explain multiple phenomena all at once because they have a bias judging simpler explanations being more probable (Lombrozo, 2007). But if the explanations only probabilistically (not deterministically) causes the phenomena (Johnson *et al.*, 2019), or if the mechanisms behind complex explantion is provided (Zemla *et al.*, 2017, 2020), the explanation preference may shift towards complexity.

In my work, based on the previous research, the aim is to systematically investigate how the different probabilistic settings of the causal system will influence how people prefer a simpler or more complex explanation. By quantitatively manipulate how prevalent and strong each cause is, I can then also compare the Bayesian posterior model of explanation with the behavioral data, as well as developing heuristic models of explanation satisfaction. I will present the novel paradigm as well as analysis in Chapter 2.

1.3 modeling value-based choice as maximizing a posterior based utility function

When you enter a friend's party, standing in front of tables of snacks, how do you pick which one will you eat first? You may have a vague idea regarding generally how good a general category of snack to you (say, chips are always more attractive than hard candies); Then you may need to more carefully examine a few snacks, comparing between the different flavors, shapes, nutrition contents, etc., to further distinguish which one may be better for you. This is a process that combines internal preference with external information collection, although in the end still making decisions for one's subjective happiness. What do these two aspects influence the final decisions? How do people integrate their subjective values with sensory information?

Krajbich *et al.* (2010) studied this in a controlled environment and proposed an explanation for the underlying process. They used a paradigm where participants choose between two snack items on the screen, while the experimenters monitoring their eye movement sequences as an approximate of the information collection process. Before the selection phase participants also have seen all the snack images and rated each one, probing the subjective value of each item. It is not surprising that if people are more likely to choose the items they rated higher. The unique finding of this study is that when people look at one item for longer, they are more likely to choose them, on top of the rating difference. Previous studies also have shown that this is potentially causal, i.e. the extended fixation time on items increased the probability of choosing it, not the other way around. How to explain this phenomena?

The classic treatment from Krajbich et al. (and later extended in Krajbich & Rangel 2011; Krajbich *et al.* 2012 for other types of choice tasks) is the attentional drift-diffusion model where the decision variable is analogous to a drifting particle which goes towards either one of the decision-boundary for the choice options.

In this work, I explored a new, explicitly Bayesian model to explain the same data from Krajbich et al. Specifically, I postulated that when people are looking at one item, they are collecting pieces of information to update the posterior distribution of the item's value. The subjective value rating determines the mean value of each piece of evidence of that given item. Then the posterior distribution is fed into a utility function which includes the posterior mean and variance. The variance term represents people's tendency of being either uncertainty-seeking or uncertainty-averse. Thus the fixation process plays two roles: more evidence collection makes the posterior mean closer to the original subjective value, also makes the variance smaller thus the posterior estimation more certain. In Chapter 3, I will present the details of this novel posterior-based model with technical details of model fitting for choice and fixation data at the same time. I also performed rigorous model comparison with the original and extended version of attentional drift-diffusion model.