A Review of Bayesian Inferences about the Self (and Others)

Egemen Gülkılık

Yeditepe University

COGS 598: Cognitive Science Seminar

June 2022

Abstract

The free energy principle proposed by (Friston et al., 2006), inspired by the concept of free energy measure in thermodynamics, states that organisms minimize their free energy, which is the energy available to do work. (Friston, 2010) extended the theory to describe all brain functions. Both perceptions and inferences of people are the outcomes of just free energy minimizing processes governed by Bayesian Inferences. In this article, self-perception and social relations will be demonstrated under this model, the results of some previous research on this subject will be shown, and an example will be given for future research on this theory.

Keywords: free energy, bayesian inference, free energy, active inference, pain perception, self, interpersonal

A Review of Bayesian Inferences about the Self (and Others)

Introduction

In this section, some terms and theories, proposed mainly by Karl Friston, will be introduced to explain brain function before going into Bayesian inferences about self and others.

Bayesian Inference

Based on Bayes' Theorem, Bayesian Inference is a well-known statistical approach. In Bayesian Inference, the probability for a hypothesis is updated by combining prior beliefs with new evidence. Bayesian inference is commonly used in machine learning and data analysis, where it can be used to update beliefs about the state of the world as new data is observed.

Free Energy Principle

The free-energy principle is a theoretical framework that explains how living systems maintain themselves in a state of homeostasis. The principle states that organisms minimize their free energy, which is the energy available to do work.

(Friston, 2010) extended this principle to explain brain functions. The behavior and cognition of living can be expressed as energy minimization strategies. He proposes that living systems are motivated to reduce their free energy, which is the amount of energy available to do work. The principle has been used to explain various phenomena, including decision-making, sensing and motor control.

Active Inference

Active Inference is a framework, which was also developed by (Friston et al., 2009), for making inferences that are closely related to Bayesian inference. Active inference has been proposed to understand how the brain makes inferences and how perceptual systems might work.

The brain symbolizes the self to forecast and optimize the probable results of social interactions. (Friston et al., 2009) characterize it as active inference, which may be thought of as reducing surprise while expecting to end up in situations with high utility. Interpersonal representations, therefore, make interactions more predictable, and interpersonal inference's emotional outcome makes self-perception evaluative. In this mechanism, two phenomena form the basis of our actions and perceptions:

- Surprise: the difference between an agent's predictions about the world and the actual outcomes observed.
- Learning and motivation: a signal indicating that the beliefs about the environment are wrong and require revision.

$$P\left(s \mid o\right)=\frac{P\left(o \mid s\right) P\left(s\right)}{P(o)}=posterior=\frac{\text{likelihood } \times \text{ prior}}{\text{observation(evidence)}}$$

So, learning happens in a Bayesian way ,as in the previous equation. The brain learns and records likelihoods $P(o \mid s)$ and prior knowledge $P(s)$ for states and observations. Later, the brain uses these likelihoods and prior knowledge to update its beliefs about the environment again.

Literature Review

The Bayesian Model of Pain Perception

(Moutoussis et al., 2014) state that we should approach pain as perception, not just a sensation. Pain perception is a complex phenomenon, and it is not just affected by current stimuli but also by past experiences. They use the term pain expectation as one's beliefs about forthcoming pain. Ultimately, the pain experience results from the pain expectation and the pain stimuli.

There exist many studies ((Friston et al., 2014), (Anchisi et al., 2015), (Hechler et al., 2016), (Tabor & Burr, 2019), (Hoskin et al., 2019)) which have proved that Bayesian models of pain perception are plausible in explaining the pain experience. In these researches, the Bayesian approaches to measuring pain expectation showed great proximity to the actual pain experience.

For example, the study (Hechler et al., 2016) investigates individuals with chronic pain. They formulate the perception as follows:

$$p\left(pain \mid sensations\right) = \frac{p\left(sensations \mid pain\right) p\left(pain\right)}{p(sensations)}$$

In their model, $p(sensations)$ represents the prior probability of pain sensation, $p(pain)$ represents the pain experience prediction, $p(sensations \mid pain)$ represents the likelihood of pain expectation, and $p(pain \mid sensations)$ is the posterior probability of pain expectation given sensation. This model indicates a positive correlation between pain experience expectation and pain experience; in other words, the pain experience might result from previous events. In addition, the likelihood of pain expectation positively correlates with the pain experience; previously learned pain expectation likelihood might also cause pain experience in the future. They exemplify the validity of this model through a case study of a young child with chronic visceral pain. They assume this model can explain the girl's pain experience even without painful stimuli.

According to (Friston et al., 2014), the Bayesian inference model can also explain irrational behavior produced by pain beliefs. Although this model may simulate the pain experience, it is restricted because it does not explain the functional role of inference about pain. There might be unforeseen roles of the painful experiences which are not a function of the previous pain experience.

Active Inference to Reach Goals

$$state \rightarrow action \rightarrow ... \rightarrow state \rightarrow action \rightarrow goal$$

(Friston et al., 2009) describe active inference as a way to reach goals similar to computational reinforcement learning in which the agents learn their policies gradually by Bayesian inference—beliefs influence agents' policies to minimize surprise and update policy about hidden states of the environment. The active inference works, as shown in the simple schema above. In reinforcement, learning and actions depend on environmental signals and the policy for that state ($a = \pi(\tilde{s})$). Here a represents action, π represents the policy, and $\tilde{s}$ represents the hidden state of the environment.

Beliefs about the Self (and Others) and the Impacts on Social Interactions

Like in the active inference model to reach goals, people tend to form beliefs about themselves and others based on Bayesian inferences. (Friston et al., 2014) claims that psychological studies showed that people rate 'good traits' and 'bad attributes' of both 'self' and 'other.' (Friston et al., 2014) use the term 'type' as a person's traits suitable to the current context. People use these types during social interactions, and like in reinforcement learning, previous beliefs and new evidence update these types.

A 'Trust Task' is a classical game in which people try to figure out whether or not their partner is cooperative. They will give them more money if they believe their partner is cooperative. They will give their partner less money if they believe they are not cooperating.

People try to deduce the 'type' of their partners based on their beliefs and behavior in a game-theoretical approach. (Friston et al., 2014) have developed a Bayesian Attribution Self Representation (BASR) model to help them understand it better. This model aims to predict what people will do based on what they think about others and how confident they feel. In the model, people have beliefs about how cooperative each player is, and they use these beliefs to estimate the likely evolution of the game. The player whose turn it is when selects an action that is consistent with their goals and beliefs about the other player. (Friston et al., 2013), (Moutoussis et al., 2014) explain the details of the decision-making and interpersonal inference model in their previous research and claim that making inferences about hidden states via BASR can be the most efficient method like computational reinforcement learning.

Clinical Applications

Self-perception issues are common in people with psychiatric disorders. They may perceive themselves differently than others, or they may be unable to anticipate how others will react to their actions. In addition, many mental disorders can be caused by deficiencies in perceiving the environment.

(Mokady & Reggev, 2022) credit active inference may help explain how people process self-relevant knowledge and maintain a stable and positive self-concept. Understanding how rewards help us believe in ourselves can help us understand how to better treat people with mental disorders.

Future Studies

Even though Bayesian inference models may present approximate predictions of the self and the others, the model has some limitations. It assumes that people are trying to optimize their decisions, which may not always be accurate. Another disadvantage of this model is that it only applies to people whose type is known.

Bayesian Causal Inference to Predict the Body Perception Illusions

According to (Shams & Beierholm, 2022), the Bayesian Causal Inference model is an extension of the classic Bayesian Inference model that explains how humans process information using multiple senses. It is based on the idea of "competitive priors," which means that there is competition among different hypotheses to explain the self and the others. In some studies, the Bayesian Causal Inference model outperformed other inference models.

(Samad et al., 2015) states that the perception of body ownership is highly flexible in the human mind. A famous example of this flexibility is the rubber hand illusion, in which people believe an artificial plastic hand is their own. (Schürmann et al., 2019) have conducted some experiments to gain more insight into the deviated perceptions of the body. They found that the Bayesian Causal Inference model outperforms uniform models, predicting these body perception illusions.

Understanding how bodily illusions work could help create new machines with that people can interact. These machines can trick people's bodies into feeling like they are moving when they are not and make the illusion work better. (Schürmann et al., 2019) also states that future research should study the effect of visual precision to enhance the performance of this illusion. Using such models in robotic assistive systems for online user adaptation may improve user acceptance and body scheme integration.

References

[1] Karl Friston, James Kilner, and Lee Harrison. A free energy principle for the brain. *Journal of physiology-Paris*, 100(1-3):70–87, 2006. doi:10.1016/j.jphysparis.2006.10.001.

[2] Karl J Friston, Jean Daunizeau, and Stefan J Kiebel. Reinforcement learning or active inference? *PloS one*, 4(7):e6421, 2009. doi:10.1371/journal.pone.0006421.

[3] Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, Feb 2010. ISSN 1471-0048. doi:10.1038/nrn2787.

[4] Karl Friston, Philipp Schwartenbeck, Thomas Fitzgerald, Michael Moutoussis, Tim Behrens, and Raymond Dolan. The anatomy of choice: active inference and agency. *Frontiers in Human Neuroscience*, 7, 2013. ISSN 1662-5161. doi:10.3389/fnhum.2013.00598.

[5] Michael Moutoussis, Nelson Trujillo-Barreto, Wael El-Deredy, Raymond Dolan, and Karl Friston. A formal model of interpersonal inference. *Frontiers in Human Neuroscience*, 8, 2014. ISSN 1662-5161. doi:10.3389/fnhum.2014.00160.

[6] Michael Moutoussis, Pasco Fearon, Wael El-Deredy, Raymond J. Dolan, and Karl J. Friston. Bayesian inferences about the self (and others): A review. *Consciousness and Cognition*, 25:67–76, 2014. ISSN 1053-8100. doi:10.1016/j.concog.2014.01.009.

[7] Davide Anchisi and Marco Zanon. A bayesian perspective on sensory and cognitive integration in pain perception and placebo analgesia. *PLOS ONE*, 10(2):1–20, 02 2015. doi:10.1371/journal.pone.0117270.

[8] Majed Samad, Albert Jin Chung, and Ladan Shams. Perception of body ownership is driven by bayesian sensory inference. *PLOS ONE*, 10(2):1–23, 02 2015. doi:10.1371/journal.pone.0117178.

[9] Tanja Hechler, Dominik Endres, and Anna Thorwart. Why harmless sensations might hurt in individuals with chronic pain: About heightened prediction and perception of pain in the mind. *Frontiers in Psychology*, 7, 2016. ISSN 1664-1078. doi:10.3389/fpsyg.2016.01638.

[10] Abby Tabor & Christopher Burr. Bayesian learning models of pain: A call to action. *Current Opinion in Behavioral Sciences*, 26:54–61, 2019. ISSN 2352-1546. doi:10.1016/j.cobeha.2018.10.006. Pain and Aversive Motivation.

[11] R. Hoskin, C. Berzuini, D. Acosta-Kane, W. El-Deredy, H. Guo, and D. Talmi. Sensitivity to pain expectations: A bayesian model of individual differences. *Cognition*, 182:127–139, 2019. ISSN 0010-0277. doi:10.1016/j.cognition.2018.08.022.

[12] Tim Schürmann, Joachim Vogt, Oliver Christ, and Philipp Beckerle. The bayesian causal inference model benefits from an informed prior to predict proprioceptive drift in the rubber foot illusion. *Cognitive Processing*, 20(4):447–457, Nov 2019. ISSN 1612-4790. doi:10.1007/s10339-019-00928-9.

[13] Aviv Mokady and Niv Reggev. The role of predictions, their confirmation, and reward in maintaining the self-concept. *Frontiers in Human Neuroscience*, 16, 2022. ISSN 1662-5161. doi:10.3389/fnhum.2022.824085.

[14] Ladan Shams and Ulrik Beierholm. Bayesian causal inference: A unifying neuroscience theory. *Neuroscience & Biobehavioral Reviews*, 137:104619, 2022. ISSN 0149-7634. doi:https://doi.org/10.1016/j.neubiorev.2022.104619.