Modeling Emotional Functions in Shared Decision-Making: A Bayesian Approach

Abtin Mogharabin (abtin.mogharabin@metu.edu.tr)
Department of Statistics, Middle East Technical University

Abstract

Decision-making, particularly in high-stakes environments such as healthcare and emergency response situations, is heavily influenced by emotional and pragmatic factors. However, past work in Bayesian Theory of Mind mainly focused on the pragmatic aspect of rational reasoning, often neglecting the critical role that emotional and contextual influences play in shaping decisions under pressure. As a result, they are unable to suitably represent phenomena such as life changing medical decisions under pressure. In this paper, I propose a model for the dynamics of high-stake shared decision-making (SDM) based on the rational speech acts (RSA) framework. For this purpose, I use various strategies to incorporate the four main effects of emotions in human decision-making, as theorized by Pfister and Böhm (2008). Particularly, I expand this framework by introducing the *multi-agent listener/observer* model for scenarios consisting of more than two agents. Then, I provide a case study of medical SDM for a cancer patient about the post-surgery treatment options. I conduct detailed comparisons of the model's results and the past known literature on the possible effects of emotions on the patient's decision.

Keywords: Theory of Emotions, Rational Speech Acts, Shared Decision Making, Multi-Agent models

I. Introduction

> "The heart has its reasons which reason knows nothing of" — Blaise Pascal, Pens ́ees

Introducing prospect theory, Kahneman and Tversky (1979) criticized the expected utility theory proposed by Neumann and Morgenstern (1947) by analyzing human decision-making under risk. Through experimental methods, they challenged the notion that rational agents must always maximize utility in all scenarios. They demonstrated that individuals evaluate potential losses and gains relative to a reference point, i.e., their perception, rather than in absolute terms. Similarly, Richard Lazarus, in his book *Psychological Stress and the Coping Process* (1966), defined emotions as a relationship between the person and the environment, shaped by how individuals subjectively assess the personal significance of a situation, goal, or action. This concept led to the introduction of the cognitive appraisal theory.

Building on previous literature, Pfister and Böhm (2008) analyzed the past theories of emotion. They argued that much of the earlier research on the effect of emotion in decision-making suffers from three main problematic assumptions: (1) in the irrational view of emotions, decision making is considered a purely rational mental process that is jeopardized by emotions. However, evidence is accumulating that this conception might be false (Bechara & Damasio, 2005; Bechara, Damasio, Tranel, & Damasio, 1997). In fact, without emotional involvement, decision making might be far from optimal (Damasio, 1994). (2) The positive-negative dichotomy of emotions assumes that all emotional states can be mapped onto a one-dimensional valence scale characterized by contrasting labels, such as positive versus negative (Barrett, 2006b; Russell, 2003). Pfister and Böhm (2008), similar to Solomon and Stone (2002), point out that classifying all emotions into pure positive/negative groups might be impossible. For example, what is good or beneficial need not be pleasurable, and what is harmful might nevertheless be satisfying. (3) The assumption of emotions as a homogeneous category treats emotions as a single, unified category of real and natural phenomena rather than as conceptual constructions (Charland, 2002; Boyd, 1999; Griffiths, 2004). More recent work provides evidence that generalizations about emotions are based on similarity by analogy rather than homology, meaning that no universal physiological, neurological, or behavioral regularities reliably identify all emotions (as discussed in Griffiths, 1997, 2004). In addition, Barrett (2006a) demonstrates that particular emotions, such as anger and sadness, do not typically exhibit a homeostatic property cluster.

Taking the above evidence into account, Pfister and Böhm (2008) introduced a four-way theory of emotion in decision making that not only follows many of the strengths from previous theories, but also avoids the abovementioned three key criticism. Pfister and Böhm's approach distinguished the effect of emotional phenomena on decisions in four distinct functions: the information function, associated with joys and (dis)likes, provides insights into the pleasure and pain linked to specific actions or outcomes; the speed function, amplified by fear and disgust, facilitates quick decisions under time constraints; the relevance function, influenced by regret and disappointment, helps focus attention on the most pertinent aspects of a decision; and the commitment function, strengthened by guilt, love, and anger, ensures adherence to decisions that align with social and moral values. In the computational modeling section, I will further elaborate on these four functions of emotion and discuss my interpretations of them from a modeling perspective.

Stivers, T., & Timmermans, S. (2021). Arriving at no: patient pressure to prescribe antibiotics and physicians' responses. Social Science & Medicine, 290, 114007.

On the computational modeling side, Bayesian models of cognition, have shown considerable success in modeling learning and reasoning as inference in complex probabilistic models (Griffiths and Tenenbaum, 2007; Griffiths, Kemp and Tenenbaum, 2008). The "planning as inference" paradigm treats action selection as an inference problem (Botvinick and Toussaint, 2012), while frameworks like Markov Decision Processes (MDPs) and Partially Observable Markov Decision Processes (POMDPs) model sequential decision-making, where actions depend on anticipated future decisions and states (Evans et al., 2017). Structural Estimation utilizes sequential data to infer preferences and beliefs, contributing to the understanding of decision-making processes (Aguirregabiria and Mira, 2010; Darden et al., 2017). These frameworks have been applied in Inverse Reinforcement Learning (IRL), where agents learn utility functions through observed actions, with successful implementations in tasks such as parking cars, flying helicopters, and playing table tennis (Abbeel et al., 2008, 2010; Muelling et al., 2014; Evans et al. 2017; Shamsi et al. 2020), as well as in computational social science (Ermon et al., 2014), to describe security (Barthe et al. 2013), in quantum mechanisms (Ying, 2011), and as the models of perfect metareasoning (Zhang and Amin, 2022).

One of the most influential models on pragmatic reasoning and communicative decision making is Rational Speech Acts (RSA), a model defined based on Bayesian decision theory to accurately replicate a wide variety of pragmatic phenomena using utility maximization methods. (Frank and Goodman, 2012; also see Degen 2023). As will be demonstrated in the subsequent sections, RSA provides one of the most intuitive frameworks to represent the four-way theory of emotions. Moreover, recent research has seen a renewed interest in Bayesian models of cognition, fueled by various case studies demonstrating their potential to enhance the reasoning capabilities and interpretability of modern large language models (LLMs) through novel integrations (e.g., White et al., 2024; Wang and Demberg, 2024; Zhi-Xuan et al., 2024; Ying et al., 2023, 2024; Wong et al., 2023).

In this paper, I extend upon the traditional methods and introduce a model of emotional functions in medical shared decision making (SDM). SDM typically takes place in medical decision making between a physician and a patient. In SDM, an advisor and a patient (decision maker) collaborate towards a joint decision (Elwyn et al., 2012; Elwyn et al., 2015; Elwyn et al., 2021). The advisor has more knowledge about the uncertain medical condition (e.g. a disease) and its outcomes but they don't completely know the patient's mind (e.g. interests, costs, and emotions). Wrong decision might result in significant health issues for the patient in the future, even risking their lives in cases. Thus, SDM scenarios often involve time constraints and risks, making it challenging to model such scenarios without constructing a detailed framework of human emotions, appraisals, and communication. While my model focuses primarily on emotional dimensions rather than the pragmatic aspects of medical conversations, its foundation in RSA allows it to complement various pragmatic phenomena and analyze language games and strategies. In SDM, for example, past literature has highlighted various strategies employed by patients and doctors during medical interactions (e.g., Stivers, 2007, 2021; Heritage et al., 2010; Stivers and Timmermans, 2021), and my aim is for this model to provide an optimal framework for capturing such strategies, especially under pressure and in stressful scenarios.

II. Relevant Works

Extending the previous applications of Bayesianism in cognitive sciences, a new interest in apply Bayesian models to human emotions has begun. Unfortunately, the vast majority of these models focus on third-person models, following the cognitive appraisal theory (e.g., Saxe & Houlihan, 2017; Ong et al., 2019), instead of focusing on the effects of one's own emotions on actions. Perhaps, the most relevant work to my model is the Bayesian Drift-Diffusion Model introduced by Ying et al. (2022), where they designed a model of Schachter-Singer's Two Factor Theory (1962).

Utilizing Drift-Diffusion Models, Ying et al. (2022) introduced a method to model emotions as the outcome of cognitive labeling, or attribution of a diffuse pattern of autonomic arousal. Their method, utilizing random walks, modeled a decision-maker who accumulates evidence until a relative decision value meets one of the two decision boundaries. A choice is made corresponding to the boundary being crossed. Then, the corresponding choice is selected to be the resulting decision (as discussed in Fudenberg et al., 2020). They investigated two model scenarios for each of the two aspects of Two Factor Theory. Unfortunately, their method separated the two components of the Two-Factor Theory of Emotion in two separate studies, which made it unideal for more realistic models where arousal and context interact.

III. Computational Models

Rational Speech Acts

The RSA framework, as defined by Goodman and Frank (2012), views communication as a process of recursive reasoning between a speaker and a listener. The listener interprets the speaker's utterance by reasoning about a cooperative speaker who is trying to inform a naive listener about some state of affairs. Using Bayesian inference, the listener reasons about what the state of the world is likely to be, given that the speaker produced an utterance, and knowing that the speaker is reasoning about how the listener is most likely to interpret that utterance. There are multiple levels of inference: at the top, the sophisticated, pragmatic listener reasons about the pragmatic speaker, who selects the best utterance by maximizing a utility function over the literal listener.

The literal listener, L_0, interprets an utterance u based on its literal meaning. The probability that a literal listener L_0 infers a state s given an utterance u is given by

$$P_{L0}(s|u) \propto [\![u]\!](s) \cdot P(s) \qquad (1)$$

where $[\![u]\!](s)$ denotes the truth conditions of the utterance u in state s, and $P(s)$ is the prior probability of state s.

The pragmatic speaker, S_1, chooses (makes a decision) an utterance u to convey a state s by maximizing the utility of u. The probability that a pragmatic speaker S_1 chooses an utterance u given a state s is defined as

$$P_{S1}(u|s) \propto \exp(\alpha \cdot U_{S1}(u; s)) \qquad (2)$$

where α is the rationality parameter that controls the speaker's sensitivity to utility differences, and $U_{S1}(u;s)$ is the utility function representing the speaker's preference for utterance u given state s. Where the utility allows speaker to decide on the utterances are preferred (liked by the speaker) at communicating, while taking a cost for each choice (unlikability of the choice for the speaker). In other words:

$$U_{S1}(u; s) = \log L_0(s|u) - C(u) \qquad (3)$$

The pragmatic listener, L_1, then interprets an utterance u by updating their beliefs about the possible states s based on the likelihood of u being chosen by a rational speaker.

$$P_{L1}(s|u) \propto P_{S1}(u|s) \cdot P(s) \qquad (4)$$

Common Ground in RSA

It is worth noting that, despite its promise as a model of conversational reasoning, RSA has rarely been used to model more than a single conversation turn. As clear from the abovementioned definition, RSA, in its most basic form, is defined as a model of a single agent's recursive reasoning. One of the few efforts of adapting RSA for multi-agent conversations was done by Anderson (2021). This work demonstrated a system that iteratively updates conversational states, including the common ground and participants' beliefs, as they interact over multiple conversational turns. In each iteration, the agents' roles switches between being the speaker and the listener. In Anderson's framework (2021), the posterior output of each speaker model over the utterances was used to linearly update the prior distribution of the other agent's listener model over the world states (i.e., the common ground update), based on a learning rate. Mathematically, this update is represented as follows:

$$CG_{L1(i+1)}(s)=(1 - lr) \cdot CG_{L1(i)}(s) + lr \cdot P_{Li}(s \mid u) \qquad (5)$$

where $CG_{L1(i+1)}$ represents the common ground of the $(i+1)^{th}$ iteration's listener model for the i^{th} iteration's speaker agent, $CG_{L1(i)}$ represents the i^{th} iteration's prior common ground of the speaker agent, and P_{Li} represents the posterior of the listener model for the current iteration's speaker agent. P_{Li} contains the probability distribution of the world states after processing the speaker's utterance. The learning rate determines the updated posterior's weight compared to the prior.

Information Function

The information function of emotions (Pfister and Böhm, 2008) provides essential evaluative information for decision-making by indicating pleasure or pain associated with different options. Emotions, such as joy or distress, inform individuals about the desirability or undesirability of potential outcomes, guiding their preferences and choices. I consider an SDM scenario where an advisor (S_1) provides suggestions on the best treatment options to the patient (the listener), taking into account both the likability and unlikability of each option. In this case, I denote the treatment option as t. By representing the advisor's undesirability as $D_{S1}(t)$, I redefine the model to better represent the information function.

$$P_{L0}(s| t) \propto [\![t]\!](s) \cdot P(s) \qquad (6)$$

$$P_{S1}(t| s) \propto \exp(\alpha \cdot U_{S1}(t; s)) \qquad (7)$$

$$U_{S1}(t; s) = \log L_0(s| t) - D_{S1}(t) \qquad (8)$$

$$P_{L1}(s| t) \propto P_{S1}(t| s) \cdot P(s) \qquad (9)$$

$$CG_{L1(i+1)}(s)=(1 - lr) \cdot CG_{L1(i)}(s) + lr \cdot P_{L1(i)}(s \mid t) \qquad (10)$$

Here, $P_{L0}(s| t)$ represents the likability of each treatment for the advisor, i.e., the probability of a treatment to positively improve the patient's state. The unlikability of each treatment, $D_{S1}(t)$, is evaluated as a linear combination of the weights assigned by the advisor to each potential side effect of the treatment.

Speed Function

The speed function (Pfister and Böhm, 2008) of emotions emphasizes the role of affective responses in enabling rapid decision-making under time constraints. It suggests that certain emotions, like fear and disgust, act as rapid response mechanisms to stimuli that require immediate action. Research on bounded rationality has similarly highlighted that humans often rely on heuristic methods and simplified decision processes when faced with limited time (Herbert, 1955, 1990, 1991). To replicate this effect, the assumption is made that under pressure, agents assign greater weights to the most significant outcomes and side effects of each treatment, while disregarding less critical factors. Consequently, they seek a satisfactory option rather than the optimal one. In mathematical terms, the decision-maker under pressure is assumed to have a utility of $W_{S1}(t) * U_{S1}(t; s)$ for treatments.

Relevance Function

The relevance function (Pfister and Böhm, 2008) is integral to how emotions influence decision making by focusing the decision maker's attention to the most significant aspects of a situation based on their personal appraisals and concerns. In this context, the relevance function influences how both the advisor and the patient perceive the situation. The advisor might deem the situation irrelevant to themselves, leaving them unaffected by this function. Conversely, the patient would find this decision to be highly relevant. Depending on the patient's values, their decision-making process would be impacted. For example, if choosing the wrong treatment could endanger the patient's life, the relevance function might lead the patient to focus more on safer options, knowing they would regret selecting a riskier one, even if the risk is minimal. To quantify this function, I introduce the *relevance weight*. This weight, directly adjusts the side effects' cost terms for each agent based on what they consider most important and relevant.

Commitment Function

The commitment function (Pfister and Böhm, 2008) pertains to the role of emotions in sustaining long-term decisions and fostering social and moral behavior, even when these actions contradict their original self-interest. Emotions like guilt, shame, love, and anger act as commitment devices, offering internal motivation to behave ethically, even when short-term selfish actions might seem more advantageous.

To avoid overcomplication, I mainly incorporate some aspects of love and its associated social and moral norms. As discussed by Ho (2008), there have been observations of patients being influenced by their family members and changing their minds under the influence of their loved ones. Previous work has examined whether such effects are detrimental to patients (Breslin, 2005).

To model these effects, I introduce a third agent into my system, representing a family member of the patient. By generating a mind model of their family members, the patient assigns a weight to their family member in the decision-making process. The patient's mind model of the family member acts as a second pragmatic speaker model (S_2), making suggestions after considering the doctor's advice. To accommodate this, I extend the traditional RSA model with a new variation of pragmatic listener models, allowing each listener to consider the suggestions of multiple sources simultaneously. The family agent is defined as follows:

$$P_{S2}(s| t) \propto P_{S1}(t| s) \cdot P(s) \quad (11)$$

$$P_{S2}(t| s) \propto \exp(\alpha \cdot W_{S2}(t) * U_{S2}(t; s)) \quad (12)$$

$$U_{S2}(t; s) = (1-\phi) \cdot \log L_0(s| t) + \phi \cdot \log P_{S1}(s| t) - D_{S2}(t) \quad (13)$$

where ϕ is the weight family member puts on the advisor's suggestions.

The patient (DM) then incorporates their mind model of the doctor, the family member, as well as their own self-interests, to make the final decision by utility maximization.

$$P_{DM}(s| t) \propto P(s) \cdot P(t| s) \quad (14)$$

$$P_{DM}(t| s) \propto \exp(\alpha \cdot W_{DM}(t) * U_{DM}(t; s)) \quad (15)$$

$$U_{DM}(t; s) = (1-\delta) \cdot \log L_0(s| t) + \delta \cdot \log P_{S2}(s| t) - D_{DM}(t) \quad (16)$$

Setting the Speed Weight

As described, emotions such as fear affect the patient's actions under time constraints. Considering that these emotions are highly dependent on the severity of the patient's situation, I define a model to select a prior for the patient's stress level.

First, the advisor explains the surgery results to the patient but prefers to avoid lengthy descriptions, especially if they believe the patient has sufficient medical knowledge. Therefore, the advisor decides on the amount of detail (utterance's detailedness – UD) based on the patient's medical understanding (PMU). If the patient demonstrates a high level of medical understanding, the advisor provides a concise summary. Conversely, if the patient has a lower level of medical understanding, the advisor offers a more detailed explanation to ensure comprehension. This model follows the basic RSA format:

$$P_{L0}(UD| PMU) \propto [\![UD]\!](PMU) \cdot P(PMU) \quad (17)$$

$$P_{S1}(UD| PMU) \propto \exp(\alpha \cdot U_{S1}(UD; PMU)) \quad (18)$$

$$U_{S1}(t; s) = \log L_0(s| t) - D_{S1}(t) \quad (19)$$

```
# Initializations
initialize  relevance_func_doc, relevance_func_fam, relevance_func_pat, commitment_func, speed_func,
worldStates, learning_rate, true_state
initialize  initial_CG_doc, initial_CG_fam, initial_CG_pat = uniform(worldStates)
define models  patient_cure, doctor_speaker, family_listener, family_speaker, patient_listener, decision_model

# Agent 1: Doctor's Suggestion
doctorSpeaker_post = doctor_speaker(true_state, relevance_func_doc, initial_CG_doc)
treatment1 = sample(doctorSpeaker_post)
doctor_mind_posterior = post(patient_cure(treatment1))
updated_CG_pat = (1- learning_rate) * initial_CG_pat + learning_rate * doctor_mind_posterior
updated_CG_fam = (1- learning_rate) * initial_CG_fam + learning_rate * doctor_mind_posterior

# Agent 2: Family Member's Suggestion
familySpeaker_post = family_speaker(true_state, relevance_func_fam, initial_CG_fam, updated_CG_fam)
treatment2 = sample(familySpeaker)
family_mind_posterior = post(family_listener(treatment2))
updated_CG_pat = (1- learning_rate) * updated_CG_pat + learning_rate * family_mind_posterior

# Agent 3: Patient's Decision Model
patientDM = decision_model(updated_CG_patient, commitment_func, speed_func, relevance_func_pat,
updated_CG_pat) # The patient's decision posterior
final_decision = sample(patientDM)
```

Figure 1. The complete decision model.

After deciding on UD, the advisor then decides on the utterance itself, depending on the patient's state. The patient then appraises the situation and would get affected depending on their inference on the doctor's description.

$$P_{\text{Patient}}(SR \mid PMU, DU1, UD) \propto P(SR) \cdot P_{DS1}(DU1, UD \mid PMU, SR) \quad (20)$$

The patient's speed weight then follows a truncated beta distribution (bounded between 0 and 1), decided based on the patient's inference of the state.

The pseudocode of the algorithm is available in Figure 1.

IV. Experiments

For a better demonstration of the model, I work on a medical SDM scenario for a rectal cancer patient about the post-surgery treatment options, a decision complicated by emotional weight and potential life-changing outcomes. Colorectal cancer (CRC), prevalent in Western societies and increasingly in Asia, typically affects the rectum. The primary treatment for localized rectal cancer is surgical resection, removing the cancerous part of the rectum along with a margin of healthy tissue. Post surgery, multiple treatment options are available depending on the surgery's outcome. I investigated various aspects of three-agent SDM following the surgery to demonstrate the model's applicability. While I could not access a complete dataset due to patient privacy constraints, prior distributions and parameters were selected to closely align with established literature (Shaukat et al., 2021; El-Shami et al., 2015; Edwards et al., 2012; Kosmider & Lipton, 2007; Desch et al., 2005; Compton et al., 2000).

V. Results

The result posterior probabilities of these agents are available in Figures 2-4. For the results of the doctor and patient's model, I can observe a consistent output with the information in the literature (Shaukat et al., 2021; El-Shami et al., 2015; Edwards et al., 2012; Kosmider & Lipton, 2007; Desch et al., 2005; Compton et al., 2000). i.e., for the doctor, when the state is complete resection with clean margins, the most likely suggested treatment is to only monitor the patient. But the doctor also takes some amount of probability for the rest of the options because taking an extra treatment like chemotherapy ensures that any possible

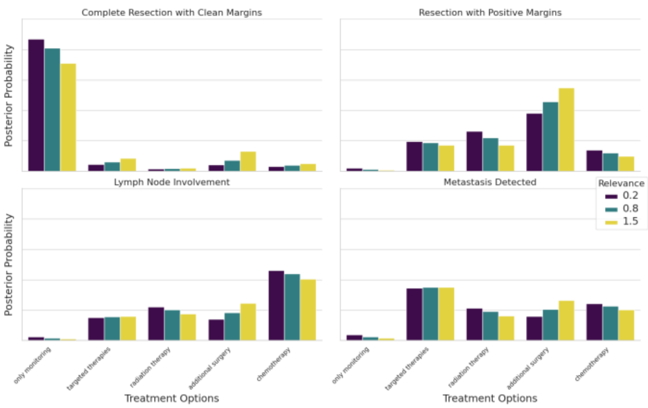


Figure 2. The resulting posterior on different treatment options for the doctor, demonstrated for different relevance weights and different states.

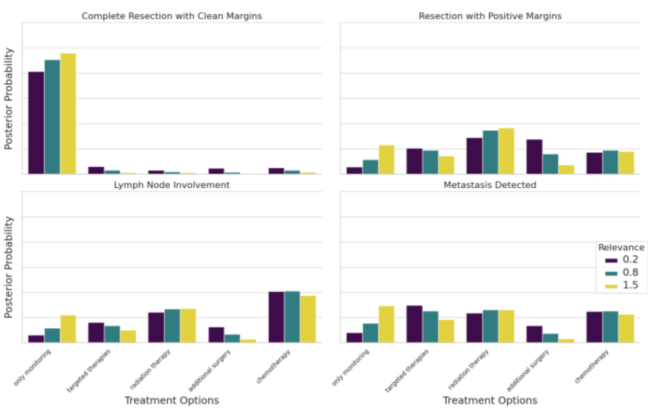


Figure 3. The resulting posterior on different treatment options for the family member, demonstrated for different relevance weights and different states.

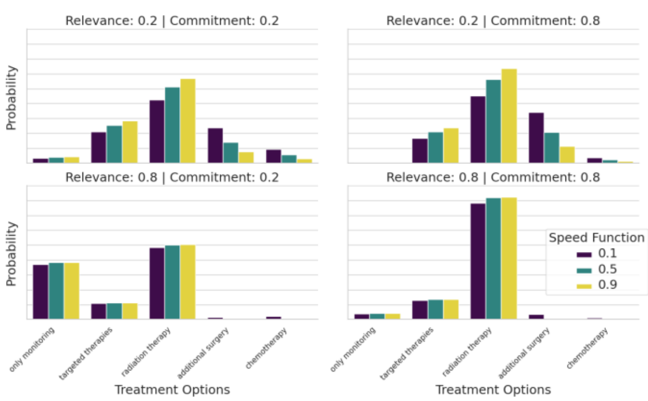


Figure 4. The resulting posterior on different treatment decision for the patient, demonstrated for different relevance, speed, and commitment weights and state 'Resection with Positive Margins'.

issues are solved, and the patient is guaranteed to avoid any relapse later on. Similarly, for when resection leaves positive margins, the most likely options, as described in the past literature, often involves either a second surgery to achieve clear margin, or radiation therapy may be employed to target residual microscopic disease. In addition, the standard care for node-positive colorectal cancer, i.e., when lymph node is involved, is using adjuvant chemotherapy to target the micro metastatic disease to reduce recurrence risk. While Radiation Therapy and Targeted Therapies are also possible options for local control, their choice depends on whether biomarker-driven options are available. Lastly, for when metastasis is detected, decisions have some degree of uncertainty and depend on the extent of metastasis, resectability, and patient condition.

For the emotion aspect, there are 3 main observations: For the relevance weight, as the relevance weight increases, the weight of side effects also increases for all 3 agents. Each treatment has its own associated side effects, some of which are less extreme than others. For example, chemotherapy is commonly associated with nausea, hair loss, fatigue, and infection. Radiation therapy, on the other hand, is linked to bladder issues, skin problems, and fatigue. It could be seen that each agent suitably incorporates these effects.

In addition, we observe that as the effects of the commitment function increase, the patient's selections get increasingly similar to the family member's possible suggestions, rather than the doctor's, which is as expected from the commitment function. The speed function's effect also correctly incorporates the possible fears of the patient from specific side effects (e.g., hair loss in this case). Depending on the speed function's weight, the patient's actions might deviate from the patient's mind model of the doctor and the family member, and their suggestions.

VI. Conclusion

In this paper, I introduced a model of medical shared decision making based on the rational speech acts framework of communication. I demonstrated how the four aspects of emotional functions in Pfister and Böhm's four factor theory of emotion (2008) could be integrated into the available framework, and how these modifications suitably demonstrate the assumed effects of information, speed, commitment, and relevance functions.

I explored how each emotional function is mapped differently onto the decision-making process, and the posterior probability. I confirmed that the finding was in line with multiple other past work on the theories of emotion, and Colorectal cancer. Unfortunately, due to the lack of individual experimental data, I was not able to fully assess the goodness of fit of my simulations. Nevertheless, this study proposes a Bayesian model for various emotional functions of Pfister and Böhm's (2008) proposed framework, and the experiments show some encouraging baseline results. I hope that, by further experiments, future studies could further validate these models and improve the results. In addition, although I adapted the traditional RSA model of pragmatic reasoning for a scenario more following the theory of mind and emotions rather than pragmatics, it should be noted that as my method fully supports the integration of pragmatic aspects. In SDM alone, past literature has shown many interesting communication strategies between doctors and patients. For example, for antibiotics prescription, patients frequently use various language strategies, such as using intensifiers, discourse markers, and metaphor, in order to lead the doctor to prescribe antibiotics for them. The doctors, on the other hand, are under pressure to prescribe antibiotics, and they need to align their medical observations, with their understanding of the patient, to decide on the best treatment (Stivers, 2007, 2021; Heritage et al., 2010). I hope that this model creates an optimal framework to model such language strategies under pressure and stressful scenarios.

References

Barrett, L. F. (2006a). Are emotions natural kinds? Perspectives on Psychological Science, 1, 28-58.

Barrett, L. F. (2006b). Valence as a basic building block of emotional life. Journal of Research in Personality, 40, 35-55.

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. Games and Economic Behavior, 52, 336-372

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. Science, 275, 1293-1295.

Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), Species: new interdisciplinary essays (pp. 141-185). Cambridge, MA: MIT Press.

Breslin J. Autonomy and the role of the family in making decisions at the end of life. J Clin Ethics 2005; 16: 11–19.

Charland, L. C. (2002). The natural kind status of emotion. British Journal for the Philosophy of Science, 53, 511-537.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. Trends in cognitive sciences, 10(7), 335–344.

Compton, C. C., Fielding, L. P., Burgart, L. J., Conley, B., Cooper, H. S., Hamilton, S. R., ... & Willett, C. (2000). Prognostic factors in colorectal cancer: College of American Pathologists consensus statement 1999. Archives of pathology & laboratory medicine, 124(7), 979-994.

Damasio, A. R. (1994). Descartes' error: Emotion, rea son and the human brain. New York: Putnam.

Degen, Judith (2023). "The rational speech act framework". In: Annual Review of Linguistics 9, pp. 519–540.

Desch, C. E., Benson III, A. B., Somerfield, M. R., Flynn, P. J., Krause, C., Loprinzi, C. L., ... & Petrelli, N. J. (2005). Colorectal cancer surveillance: 2005 update of an American Society of Clinical Oncology practice guideline. Journal of Clinical Oncology, 23(33), 8512-8519.

Edwards, M. S., Chadda, S. D., Zhao, Z., Barber, B. L., & Sykes, D. P. (2012). A systematic review of treatment guidelines for metastatic colorectal cancer. *Colorectal Disease*, *14*(2), e31-e47.

Elwyn, G. (2021). Shared decision making: what is the work?. Patient Education and Counseling, 104(7), 1591-1595.

Elwyn, G., Frosch, D. L., & Kobrin, S. (2015). Implementing shared decision-making: consider all the consequences. Implementation science, 11, 1-10.

Elwyn, G., Frosch, D., Thomson, R., Joseph-Williams, N., Lloyd, A., Kinnersley, P., ... & Barry, M. (2012). Shared decision making: a model for clinical practice. Journal of general internal medicine, 27, 1361-1367.

Frank, Michael C. and Noah D. Goodman (2012). "Predicting Pragmatic Reasoning in Language Games". In: Science 336.6084, pp. 998–998.

Fudenberg, D., Newey, W., Strack, P., & Strzalecki, T. (2020). Testing the drift-diffusion model. Proceedings of the National Academy of Sciences, 117(52), 33141–33148.

Griffiths, P. E. (1997). What emotions really are: The problem of psychological categories. Chicago: University of Chicago Press.

Griffiths, P. E. (2004). Emotions as natural and normative kinds. Philosophy of Science, 71, 901-911.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), The Cambridge handbook of computational psychology (p. 59–100). Cambridge University Press.

Griffiths, T., & Tenenbaum, J. B. (2007). Two proposals for causal grammars. Causal learning: Psychology, philosophy, and computation, 323–345.

Heritage, J., Elliott, M. N., Stivers, T., Richardson, A., & Mangione-Smith, R. (2010). Reducing inappropriate antibiotics prescribing: the role of online commentary on physical examination findings. Patient education and counseling, 81(1), 119-125.

Ho, A. (2008). Relational autonomy or undue pressure? Family's role in medical decision-making.

Scandinavian journal of caring sciences, 22(1), 128-135.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica 47.2: 363-391.
Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. Annu. Rev. Psychol., 55, 271–304.
Kosmider, S., & Lipton, L. (2007). Adjuvant therapies for colorectal cancer. World Journal of Gastroenterology: WJG, 13(28), 3799.
Lazarus, R. S. (1966). Psychological stress and the coping process.
Ong, D. C., Zaki, J., & Goodman, N. D. (2019). Computational models of emotion inference in theory of mind: A review and roadmap. Topics in cognitive science, 11(2), 338–357.
Pfister, H. R., & Böhm, G. (2008). The multiplicity of emotions: A framework of emotional functions in decision making. Judgment and decision making, 3(1), 5-17.
Russell, J. A. (2003). Core affect and the psychological construction of emotion. Psychological Review, 110, 145-172
Saxe, R., & Houlihan, S. D. (2017). Formalizing emotion concepts within a Bayesian model of theory of mind. Current opinion in Psychology, 17, 15–21.
Simon, H. A. (1955). A behavioral model of rational choice. The quarterly journal of economics, 99-118.
Simon, H. A. (1990). A mechanism for social selection and successful altruism. Science, 250(4988), 1665-1668.
Simon, H. A. (1991). Bounded rationality and organizational learning. Organization science, 2(1), 125-134.
Solomon, R. C., & Stone, L. D. (2002). On "positive" and "negative" emotions. Journal for the Theory of Social Behavior, 32, 417-435.
Stivers, T. (2007). Prescribing under pressure: Parent-physician conversations and antibiotics. Oxford University Press.
Stivers, T. (2021). Managing patient pressure to prescribe antibiotics in the clinic. Pediatric Drugs, 23(5), 437-443.
Von Neumann, J., & Morgenstern, O. (1947). Theory of games and economic behavior, 2nd rev.
Ying, L., Michal, A., & Zhang, J. (2022). A Bayesian Drift-Diffusion Model of Schachter-Singer's Two Factor Theory of Emotion. Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44).
White, I., Pandey, S., & Pan, M. (2024). Communicate to Play: Pragmatic Reasoning for Efficient Cross-Cultural Communication in Codenames. arXiv preprint arXiv:2408.04900.
Wong, L., Grand, G., Lew, A. K., Goodman, N. D., Mansinghka, V. K., Andreas, J., & Tenenbaum, J. B. (2023). From word models to world models: Translating from natural language to the probabilistic language of thought. arXiv preprint arXiv:2306.12672.
Ying, L., Collins, K. M., Wei, M., Zhang, C. E., Zhi-Xuan, T., Weller, A., ... & Wong, L. (2023). The neuro-symbolic inverse planning engine (nipe): Modeling probabilistic social inferences from linguistic inputs. arXiv preprint arXiv:2306.14325.
Zhi-Xuan, T., Ying, L., Mansinghka, V., & Tenenbaum, J. B. (2024). Pragmatic Instruction Following and Goal Assistance via Cooperative Language-Guided Inverse Planning. arXiv preprint arXiv:2402.17930.
Ying, L., Zhi-Xuan, T., Wong, L., Mansinghka, V., & Tenenbaum, J. (2024). Grounding Language about Belief in a Bayesian Theory-of-Mind. arXiv preprint arXiv:2402.10416.
Wang, Y., & Demberg, V. (2024). RSA-Control: A Pragmatics-Grounded Lightweight Controllable Text Generation Framework. arXiv preprint arXiv:2410.19109.
Botvinick, M., & Toussaint, M. (2012). Planning as inference. Trends in cognitive sciences, 16(10), 485-488.
Aguirregabiria, V., & Mira, P. (2010). Dynamic discrete choice structural models: A survey. Journal of Econometrics, 156(1), 38-67.
Darden, M. (2017). Smoking, expectations, and health: a dynamic stochastic model of lifetime smoking behavior. Journal of Political Economy, 125(5), 1465-1522.
Abbeel, P., Dolgov, D., Ng, A. Y., & Thrun, S. (2008, September). Apprenticeship learning for motion planning with application to parking lot navigation. In 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 1083-1090). IEEE.
Abbeel, P., Coates, A., & Ng, A. Y. (2010). Autonomous helicopter aerobatics through apprenticeship learning. The International Journal of Robotics Research, 29(13), 1608-1639.
Muelling, K., Boularias, A., Mohler, B., Schölkopf, B., & Peters, J. (2014). Learning strategies in table tennis

using inverse reinforcement learning. Biological cybernetics, 108, 603-619.

Ermon, S., Xue, Y., Toth, R., Dilkina, B., Bernstein, R., Damoulas, T., ... & Gomes, C. (2015). Learning large-scale dynamic discrete choice models of spatio-temporal preferences with application to migratory pastoralism in East Africa. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 29, No. 1).

Stivers, T. (2007). Prescribing under pressure: Parent-physician conversations and antibiotics. Oxford University Press.

Stivers, T. (2021). Managing patient pressure to prescribe antibiotics in the clinic. Pediatric Drugs, 23(5), 437-443.

Heritage, J., Elliott, M. N., Stivers, T., Richardson, A., & Mangione-Smith, R. (2010). Reducing inappropriate antibiotics prescribing: the role of online commentary on physical examination findings. Patient education and counseling, 81(1), 119-125.

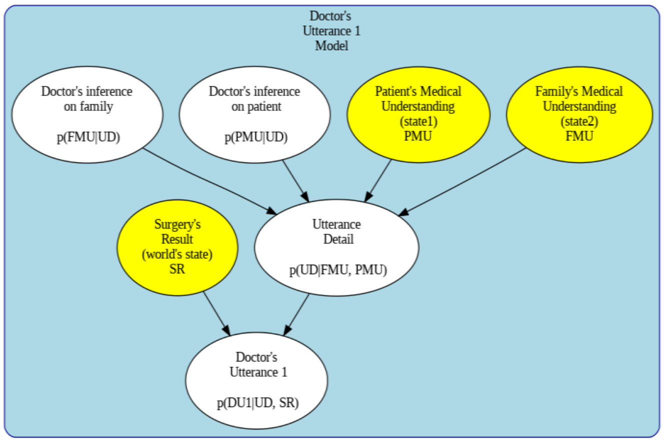
Barthe, G., Köpf, B., Olmedo, F., & Zanella-Béguelin, S. (2013). Probabilistic relational reasoning for differential privacy. ACM Transactions on Programming Languages and Systems (TOPLAS), 35(3), 1-49.

Ying, M. (2012). Floyd--hoare logic for quantum programs. ACM Transactions on Programming Languages and Systems (TOPLAS), 33(6), 1-49.

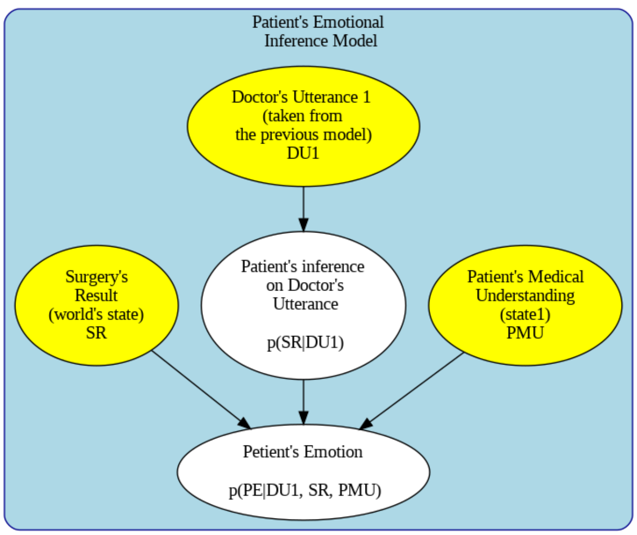
Zhang, Y., & Amin, N. (2022). Reasoning about “reasoning about reasoning”: semantics and contextual equivalence for probabilistic programs with nested queries and recursion. Proceedings of the ACM on Programming Languages, 6(POPL), 1-28.

VII. Appendix: Directed acyclic graphs

Directed acyclic graph for the doctor's detailedness:



Directed acyclic graph for the patient's emotion:



Directed acyclic graph for the patient's decision: