Collaborative Rational Speech Act: Pragmatic Reasoning for Multi-Turn Dialog

Anonymous ACL submission

Abstract

As AI systems take on collaborative roles, they must reason about shared goals and beliefs—not just generate fluent language. The Rational Speech Act (RSA) framework offers a principled approach to pragmatic reasoning, but existing extensions face challenges in scaling to multi-turn, collaborative scenarios. In this paper, we introduce Collaborative Rational Speech Act (CRSA), an information-theoretic (IT) extension of RSA that models multi-turn dialog by optimizing a gain function adapted from rate-distortion theory. This gain is an extension of the gain model that is maximized in the original RSA model but takes into account the scenario in which both agents in a conversation have private information and produce utterances conditioned on the dialog. We demonstrate the effectiveness of CRSA on referential games and template-based doctor–patient dialogs in the medical domain. Empirical results show that CRSA yields more consistent, interpretable, and collaborative behavior than existing baselines—paving the way for more pragmatic and socially aware language agents.

1 Introduction

Modeling conversations is central to the development of grounded and useful agentic AI systems, which are increasingly characterized by collaborative interactions between humans and machines. Several applications benefit from dialog systems capable of natural and pragmatic interactions with users. For instance, in the medical domain, conversational agents could support diagnostic interviews (Tu et al., 2025) or serve as tools for physician training in controlled environments (Karunanayake, 2025). In enterprise settings, dialog agents could autonomously handle routine tasks—such as scheduling, data entry, or report generation—freeing human effort for higher-level decision-making (Tupe and Thube, 2025; Satav, 2025). In education, they offer the potential to personalize content delivery, adapting to learners' styles and paces (Nabhani et al., 2025; Vorobyeva et al., 2025). While such applications are still emerging, a key enabler is the development of models that can manage collaborative, goal-oriented interactions in a robust and interpretable manner.

To succeed in real-world settings, dialog generative-based models must do more than generate fluent language—they must track shared tasks to communicate meaningfully and contextually (Lin et al., 2024). For example, a physician in a diagnostic exchange refines hypotheses as the conversation evolves, requiring interpretable and scalable frameworks for reliable interaction.

Yet, many existing models prioritize task-specific response generation (He et al., 2017; Jiang et al., 2019; Meta Fundamental AI Research Diplomacy Team (FAIR) et al., 2022), or optimize for superficial conversation properties using narrowly defined objectives (Khani et al., 2018; Dafoe et al., 2020; Lin et al., 2024; Jeon et al., 2020). While these methods often yield strong performance, they typically lack principled foundations and depend on ad hoc design choices.

The Rational Speech Act (RSA) framework (Frank and Goodman, 2012) offers a principled foundation for modeling pragmatic reasoning as recursive social inference between speakers and listeners. Viewed through an information-theoretic (IT) lens, RSA approximates a Rate-Distortion solution (Cover and Thomas, 2001), where the listener reconstructs intended meaning from observed utterances (Zaslavsky et al., 2021). RSA has successfully captured phenomena such as reference (Degen et al., 2020), implicature (Bergen et al., 2016), and vagueness (Herbstritt and Franke, 2019), and powered applications from grounded captioning (Cohn-Gordon et al., 2018) to controlled generation (Wang and Demberg, 2024). Yet, despite this promise, existing RSA extensions remain limited in multi-turn, task-oriented dialog: they

struggle to model evolving beliefs or integrate dialog history (Carenini et al., 2024; Degen, 2023). We argue this shortfall stems from the absence of a unified, theoretically grounded mechanism for belief and task tracking in collaborative interaction.

Contributions

Our main contributions are as follows:

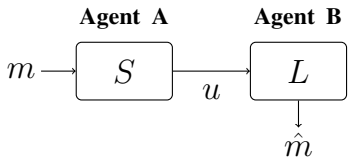
- We introduce **Collaborative RSA (CRSA)**, a novel, information-theoretically grounded extension of the RSA framework tailored for multi-turn, goal-driven dialog.
- **A generalized multi-turn gain function:** We extend the rate-distortion to model multi-turn collaborative settings of RSA, capturing both task progression and evolving partner beliefs. CRSA jointly models the agent's belief about (i) the shared task target and (ii) the interlocutor's private knowledge—enabling socially aware and context-sensitive communication.
- **Empirical validation:** We evaluate CRSA on referential games and semi-automatically generated doctor-patient dialogs, showing that it improves consistency, interpretability, and collaborative alignment compared to existing baselines.

2 Related work

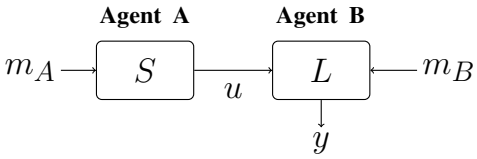
RSA model and pragmatics. The Rational Speech Act (RSA) framework (Frank and Goodman, 2012) serves as a model for pragmatic communication designed to emulate human behavior in linguistic tasks (Degen et al., 2020; Bergen et al., 2016; Herbstritt and Franke, 2019). This framework is both conceptually intuitive and computationally versatile, making it readily adaptable for integration with neural language models to tackle more intricate challenges, including machine translation (Cohn-Gordon and Goodman, 2019), image captioning (Cohn-Gordon et al., 2018), controllable text generation (Shen et al., 2019; Wang and Demberg, 2024). Extensions to the original RSA framework have been proposed to accommodate more complex scenarios. For instance, adaptations have addressed cases where agents lack shared vocabularies (Bergen et al., 2016) or where common ground evolves dynamically during interaction (Degen et al., 2015). A comprehensive overview of RSA's development and its numerous variants is provided by Degen (2023).

Information-theoretic results for interactive rate-distortion. Information theory offers a robust framework for analyzing communication as the exchange of information between agents. Within this domain, the rate-distortion problem (Shannon, 1993) offers a principled way to balance compression efficiency with the fidelity of reconstruction. This problem has been pivotal in exploring the trade-offs between fidelity and compression in message transmission. Kaspi (1985) investigated scenarios involving two agents engaging in iterative interactions to collaboratively infer each other's observations. Building on this foundation, Rey Vega et al. (2017) extended the analysis to multi-agent contexts, accommodating communication frameworks with three or more participants and significantly advancing the understanding of collective information exchange. Focusing on two-agent systems, Vera et al. (2019) explored a variation wherein each agent is tasked not merely with understanding one another but with predicting a target random variable representing a (possible stochastic) function of each other's observations. This approach highlights the promise of IT methods in supporting more efficient and collaborative communication among agents in complex environments, as shown by Zaslavsky et al. (2021), who reformulate the standard RSA framework as a rate-distortion optimization problem.

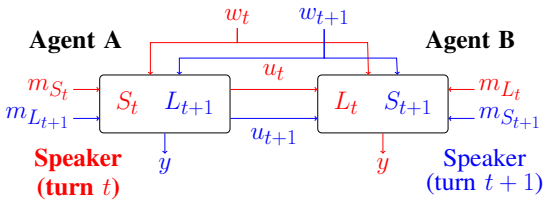
Collaborative dialog modeling. Multiple works frame a collaborative or task-oriented dialog as a Partially Observable Markov Decision Process (POMDP) (Williams and Young, 2007), which provide a suitable framework to model end-to-end networks on specific tasks (Wen et al., 2017; Jiang et al., 2019). Reinforcement Learning has been widely used in this context, in order to provide interpretable and trackable training procedures that incorporate the structure of the dialog in their policy training or decoding strategy (Lin et al., 2024; Li et al., 2016; Xu et al., 2025). Related to this, game-theoretic perspective has also been used in dialog modeling (Jeon et al., 2020; Lin et al., 2022). In this context, multiple tasks and datasets have been developed to evaluate dialog modeling (He et al., 2017; Khani et al., 2018; Macherla et al., 2023), usually by assessing the task performance and the similarity with human conversations. The RSA model has also found applications in dialog systems, often complementing neural models to enhance agent self-awareness (Kim et al., 2020) or to improve the interpretation of emotional subtext (Kim et al., 2021).



(a) RSA



(b) YRSA



(c) CRSA

Figure 1: RSA variants proposed in this work(1b, 1c) compared to the original one (1a).

3 Review of the RSA Model from the Lens of Information Theory

Fig. 1a presents a schematic view of the classic RSA model from an information-theoretic perspective. Here, a meaning $m \in \mathcal{M}$ is received by the speaker $S : \mathcal{M} \times \mathcal{U} \to [0, 1]$ who uses it to produce a posterior probability $S(u|m)$ for all possible utterances $u \in \mathcal{U}$. The utterance u is then transmitted to the listener $L : \mathcal{U} \times \mathcal{M} \to [0, 1]$ who produces a posterior $L(\hat{m}|u)$ for all possible reconstructions $\hat{m} \in \mathcal{M}$ of the meaning m that the speaker is trying to convey. Additionally, there is a distribution $P : \mathcal{M} \to [0, 1]$ that is known by the two agents and represents the prior of the meanings. Finally, the function $C : \mathcal{U} \to \mathbb{R}$ assigns a prior cost value to each utterance produced by the speaker.

In the classic RSA model, agents update their values based on the other's perspective. For simplicity, and without loss of generality, we adopt the listener's viewpoint—assuming the speaker updates first[1]:

$$S^{k+1}(u|\hat{m}) \propto \exp\left[\alpha(\log L^k(\hat{m}|u) - C(u)\right],$$
$$L^{k+1}(\hat{m}|u) \propto S^{k+1}(u|\hat{m})P(\hat{m}).$$

[1]In the classic RSA literature, the literal listener (speaker) is usually represented with L_0 (S_0) and the pragmatic with L_1 (S_1). Here, we will reserve the subindex notation for the turn number and denote the level of pragmatism of each agent by using the super index L^k (S^k) with $k = 0, 1, \ldots, K$.

In this case, the listener is initialized with a predefined lexicon function $\mathcal{L} : \mathcal{U} \times \mathcal{M} \to 0, 1$, which specifies the possible meanings associated with each utterance:

$$L^0(\hat{m}|u) \propto P(\hat{m})\mathcal{L}(u, \hat{m}).$$

Zaslavsky et al. (2021) show that this iteration process is equivalent to maximize the next objective:

$$\mathcal{G}^{\alpha}_{RSA}(L,S) = H_S(U|\hat{M}) + \alpha\mathbb{E}_S[V_L(U,\hat{M})], \quad (1)$$

where $H_S(U|\hat{M})$ is the conditional entropy between the estimated meanings and the utterances, $V_L(u, \hat{m}) \triangleq \log L(\hat{m}|u) - C(u)$ is called the "listener value", and $\mathbb{E}_S[V_L]$ is computed with respect to the distribution of the speaker. That is,

$$H_S(U|\hat{M}) = -\sum_{\forall\,(u,\hat{m})} P_S(u, \hat{m}) \log S(u|\hat{m}),$$
$$\mathbb{E}_S[V_L] = \sum_{\forall\,(u,\hat{m})} P_S(u, \hat{m}) V_L(u, \hat{m}),$$

where $P_S(u, \hat{m}) \triangleq S(u|\hat{m})P(\hat{m})$ represents the joint probability of the speaker.

4 Main Theoretical Results

4.1 Modeling private meanings (YRSA)

To extend the RSA model to bidirectional dialogue with explicit task modeling, we first distinguish between private meanings and shared task outcomes. In real conversations, each participant holds their own prior knowledge and worldview, which may differ from that of their interlocutor. In our example of a dialogue between a patient and a physician: the patient must describe their symptoms, which are not directly observable by the physician, while the physician brings medical expertise the patient lacks. Both types of knowledge are essential to determine the appropriate diagnosis or treatment plan. Notably, neither the patient's symptoms nor the physician's prior knowledge fully align with the shared goal of the conversation, i.e. the identification of a suitable medical outcome.

In this context, we identify the need of representing a private set of meanings $\mathcal{M}_A$ and $\mathcal{M}_B$ for each agent, which may or may not match. In addition, the result y of the shared task is going to be represented with a separate space $\mathcal{Y}$ that contains all the possible outcomes of it. For simplicity, we will assume that all these are discrete spaces. Fig. 1b represents a schematic of this model. We will refer to this extension as the YRSA model.

The YRSA model redefines the notion of prior from the classic RSA framework by conditioning the dialogue on the joint realization of the agents' private meanings (m_A, m_B) and the shared task target y, which together define the context in which the interaction unfolds. Importantly, we assume for the development of our model that both the realizations and the joint distribution of these three variables do not change over time during the conversation. This implies that the prior is completely defined by the joint distribution $P : \mathcal{M}_A \times \mathcal{M}_B \times \mathcal{Y} \to [0, 1]$ given to both agents.

We now turn to defining the updated agent posteriors. The new speaker $S : \mathcal{M}_A \times \mathcal{U} \to [0, 1]$ produces a posterior $S(u|m_A)$ that only depends on its the private meaning m_A. Similarly, the listener $L : \mathcal{M}_B \times \mathcal{U} \times \mathcal{Y} \to [0, 1]$ is represented by the posterior $L(y|m_B, u)$, which is conditional independent of the private meanings m_A. In this formulation, the representation of task performance is delegated to the listener, who updates their belief upon receiving the utterance.

We can now propose the corresponding gain function to be maximized by this model:

$$\mathcal{G}^{\alpha}_{YRSA}(L, S) = H_S(U|M_A) + \alpha\mathbb{E}_S[V_L(U, M_B, Y)] \quad (2)$$

with $V_L(u, m_B, y) = \log L(y|u, m_B) - C(u)$ and $H_S(U|M_A)$ defined as in the classic RSA. A detailed derivation of the equations used to maximize this function is provided in Appendix A.

4.2 The CRSA Model

Effective collaboration requires not only modeling agents' private meanings and the shared task, but also supporting multi-turn dialogue. In a medical consultation, for instance, the patient shares symptoms and background, while the physician asks questions, proposes diagnoses, and recommends treatments. To capture such interactions, we denote the speaker's utterance at turn t as U_t, and the dialogue history up to that point as $W_t = (U_1, \ldots, U_{t-1})$, representing the sequence of prior exchanges.

4.2.1 Modeling multi-turn dialog with explicit history (baseline)

The attempt of previous approaches to incorporate the history of the conversation to the RSA model rely on defining the lexicon (or directly the literal listener/speaker) as a function of each turn (Wang and Demberg, 2024; Kim et al., 2020; Lin et al., 2022). In many cases, this lexicon is given by the output of a neural language model and can be very robust to the evolving dialog. However, that variant of the RSA does not correspond to maximizing the gain of Eq. (1), but a modified version of it in which U_t is replaced by (U_t, W_t):

$$H_S(U_t,W_t|\hat{M})+\alpha\mathbb{E}_S[V_L(U_t,W_t,\hat{M},Y)]. \quad (3)$$

This is equivalent to applying an RSA model at each turn by initializing it with a lexicon $\mathcal{L}(u_t, \hat{m}, w_t)$ depending on w_t, the past utterances.

The issue with Eq. (3) is that the speaker's utterance U_t at turn t is modeled *jointly* with the dialogue history W_t, rather than being explicitly *conditioned* on it. To express the gain in terms of the conditional entropy of the current utterance alone, we condition it on both the dialogue history W_t and the speaker's intended meaning $\hat{M}$, rather than on $\hat{M}$ alone. In Section 4.2.2, we formally introduce the corresponding expressions of the CRSA model, which incorporates this notion of multi-turn conditioned to the past utterances, as well as private meanings and target task.

4.2.2 Equations of the CRSA model

Figure 1c illustrates our extension of the YRSA model to the collaborative setting. As in the original setup, agents alternate roles—one acting as the speaker, the other as the listener—to achieve a shared task. Each agent has access to a private meaning space, $\mathcal{M}_A$ or $\mathcal{M}_B$, which remains hidden from their counterpart. Then, at turn t, the private meanings of the speaker will correspond to the meanings of the agent playing the role of the speaker and vice-versa. We refer as $\mathcal{M}_{S_t}$ and $\mathcal{M}_{L_t}$ to the private meanings of the speaker and the listener at turn t, respectively. Both agents also have access to the conversation history, denoted as $w_t = (u_1, \ldots, u_{t-1}) \in \mathcal{W}_t \triangleq \mathcal{U}_1 \times \cdots \times \mathcal{U}_{t-1}$, where each $\mathcal{U}_i$ represents the space of possible utterances at turn i. The shared objective is to jointly predict a target class y from a finite discrete set $\mathcal{Y}$.

As discussed earlier, the joint distribution $P(m_A, m_B, y)$ serves as a fixed prior throughout the conversation. To maintain consistency as agents alternate roles, we define the prior at turn t over the active speaker and listener meanings, i.e., $P(m_{S_t}, m_{L_t}, y)$, as follows:

$$P_t(m_{S_t}, m_{L_t}, y) = \begin{cases} P(m_{S_t}, m_{L_t}, y) & \text{if } S_t = A \\ P^{\top}(m_{L_t}, m_{S_t}, y) & \text{if } S_t = B \end{cases}$$

where $P^\top : \mathcal{M}_B \times \mathcal{M}_A \times \mathcal{Y} \to [0,1]$ is such as $P^\top(b,a,y) = P(a,b,y)$.

Formally, we define the distribution of each agent at turn t. The speaker $S_t : \mathcal{M}_{S_t} \times \mathcal{U}_t \times \mathcal{W}_t \to [0,1]$ produces a posterior $S_t(u_t|m_{S_t}, w_t)$ that depends on its private meaning m_{S_t} and the past utterances w_t. On the other hand, the listener $L_t : \mathcal{M}_{L_t} \times \mathcal{U} \times \mathcal{W}_t \times \mathcal{Y} \to [0,1]$ is represented by the posterior $L_t(y|m_{L_t}, u_t, w_t)$ which is independent of the private meanings of the speaker.

Building on the gain function in Eq. (1), we extend the joint speaker distribution and listener utility to incorporate private meanings and multi-turn dialogue:

$$P_S(u_t, w_t, m_{S_t}, m_{L_t}, y) \triangleq S_t(u_t|m_{S_t}, w_t) \times P_S(w_t|m_{S_t}, m_{L_t}) P_t(m_{S_t}, m_{L_t}, y),$$

$$V_L(u_t, w_t, m_{L_t}, y) \triangleq \log L_t(y|u_t, m_{L_t}, w_t) - C(u_t).$$

Then, we define one gain function at each turn to be maximized:

$$\mathcal{G}^\alpha_{CRSA}(L_t, S_{S_t}) = H_{S_t}(U_t|M_{S_t}, W_t) + \alpha \mathbb{E}_{S_t}[V_L(U_t, W_t, M_{S_t}, M_{L_t}, Y)], \quad (4)$$

where the expectation of both terms is over P_S. In all cases, we will model $P_S(w_t|m_{S_t}, m_{L_t})$ with the past speakers' utterances:

$$P_S(w_t|m_{S_t}, m_{L_t}) = \underbrace{\prod_{\substack{i<t \\ S_i = S_t}} S_i(u_i|w_i, m_{S_t})}_{B_{L,t}(m_{S_t})} \underbrace{\prod_{\substack{i<t \\ S_i \neq S_t}} S_i(u_i|w_i, m_{L_t})}_{B_{S,t}(m_{L_t})}. \quad (5)$$

This formulation naturally leads to interpreting $B_{L,t}(m_{S_t})$ and $B_{S,t}(m_{L_t})$ as each agent's belief about their interlocutor's private meaning. In Section 5, we illustrate why this interpretation is reasonable with a concrete example.

Once modeled the gain, the equations that correspond to its maximization are the following:

$$S_t^{k+1}(u_t|w_t, m_{S_t}) \propto \exp\left[\alpha \sum_{\forall (m_{L_t}, y)} B'_t(m_{S_t}, m_{L_t}, y) V_L(u_t, w_t, m_{L_t}, y)\right],$$

$$L_t^{k+1}(y|u_t, w_t, m_{L_t}) \propto \sum_{\forall m_{S_t}} B_{S,t}(m_{S_t}) P_t(m_{S_t}, m_{L_t}, y) S_t^{k+1}(u_t|w_t, m_{S_t}),$$

where we replace $B'_t(m_S, m_L, y) =$

$$\frac{B_{S,t}(m_L) P(m_L|m_S)}{\sum_{\forall m'_L} B_{S,t}(m'_L) P(m'_L|m_S)} P(y|m_L, m_S). \quad (6)$$

A complete derivation of these equations is provided in Appendix B. Finally, there is no single prescribed method for initializing the iteration at each turn. In Section 5, we adopt the listener's perspective and explore two variants of the initial lexicon $\mathcal{L}$, initializing the literal listener as:

$$L^0(y|u_t, w_t, m_{L_t}) \propto \sum_{\forall m_S} P(m_S, m_{L_t}, y) \mathcal{L}_{u_t, w_t}(m_{S_t}) \quad (7)$$

with $\mathcal{L}_{u_t, w_t}(m_{S_t})$ depending on the variant of the RSA. In contrast, in Section 6 we initialize the literal speaker directly with a LLM:

$$S_t^0(u_t|m_{S_t}, w_t) \propto P_{LM}(u_t|w_t, \text{prompt}(m_{S_t})), \quad (8)$$

where $\text{prompt}(m_{S_t})$ is the text used to prompt the speaker at that turn. As shown, CRSA retains the flexibility of the original RSA framework in modeling both the listener's and speaker's perspectives.

5 CRSA for Reference Games

To evaluate CRSA, we adapt the reference game of Khani et al. (2018). In this setting, two agents are shown the same sequence of N cards, each labeled with one letter (A or B) and one number (1 or 2). Agent A sees only the letter on each card, while Agent B sees only the number. Their goal is to collaboratively identify the position of the card labeled A1. At each turn, an agent may utter a number from 1 to N, indicating a card position. For simplicity, we assume that each round contains at most one A1 card and that Agent A always initiates the exchange.

5.1 Experimental set-up

For this simulation we consider that the set $\mathcal{U}_t$ of possible utterances at turn t is always ($\forall t$) the set $\mathcal{U}_t = \{1, \ldots, N\}$ representing the messages of the form *"A1 card may be at position n"* with $n \in \mathcal{U}_t$. For the set $\mathcal{Y}$ of possible classes, the results can be as well *"A1 card may be at position n"*, with the addition that there is also the possibility of *"There is no A1 card"*. That is, $\mathcal{Y} = \{0, 1, \ldots, N\}$ with 0 representing the mentioned possibility. Regarding the meaning spaces,

they correspond to the possible sequences of length N that can be obtained combining without replacement the letters A and B (for agent A) and the numbers 1 and 2 (for agent B). That is, for instance if $N = 3$, $\mathcal{M}_A = \{\text{AAA}, \text{AAB}, \ldots, \text{BBB}\}$ and $\mathcal{M}_B = \{111, 112, \ldots, 222\}$. Finally, the prior distribution $P(m_A, m_B, y)$ can be defined as follows:

$$P(m_A, m_B, y) \propto \begin{cases} 1 & \text{if } m_A \text{ and } m_B \text{ form A1 at position } y \\ 0 & \text{otherwise} \end{cases}$$

Since this is a reference game, we adopt the listener's perspective. In all cases, the literal listener is initialized using Eq. (7), and different model variants are defined based on the update equations and the specification of the lexicon $\mathcal{L}_{u_t,w_t}(m_{S_t})$.

- **CRSA**: We apply the CRSA update equations and define a lexicon $\mathcal{L}_{u_t,w_t}(m_{S_t}) = \mathcal{L}(u_t, m_{S_t})$ that do not depend on w_t:

$$\mathcal{L}(u_t, m_{S_t}) = \begin{cases} 1 & \text{if } m_{S_t} \text{ contains A/1 at position } n \text{ and } u_t = n \\ 1 & \text{if there is no A/1 in } m_{S_t} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

- **CRSA-W_t**: We apply the CRSA update equations, but with the lexicon $\mathcal{L}_{u_t,w_t}(m_{S_t}) = \mathcal{L}(u_t, m_{S_t}, w_t)$ depending on the the past w_t. To define $\mathcal{L}(u_t, m_{S_t}, w_t)$, we follow the simple rule:

$$\mathcal{L}(u_t, m_{S_t}, w_t) = \begin{cases} 0 & \text{if } u_t \in w_{t-1} \\ & \wedge\ u_t \neq u_{t-1} \\ \mathcal{L}(u_t, m_{S_t}) & \text{otherwise} \end{cases} \quad (10)$$

  We expect efficient conversational behavior in this game to involve repeating an utterance only to confirm the correct A1 card position. If the correct position is identified, agents should repeat the utterance until the round ends; otherwise, repeating it would be inefficient. The rule in Eq. (10) explicitly encodes this behavior.
- **YRSA**: We initialize the listener using the YRSA iterative equations and the lexicon from Eq. (9), effectively applying the RSA iteration in the setting where each agent holds a private meaning—that is, the standard YRSA setup.
- **YRSA-W_t**: The same as the one above but using Eq. (10) as lexicon instead of Eq. (9).
- **Literal**: In this case, there is no iteration and we simply use Eq. (7) to predict the target. We use lexicon of Eq (9).

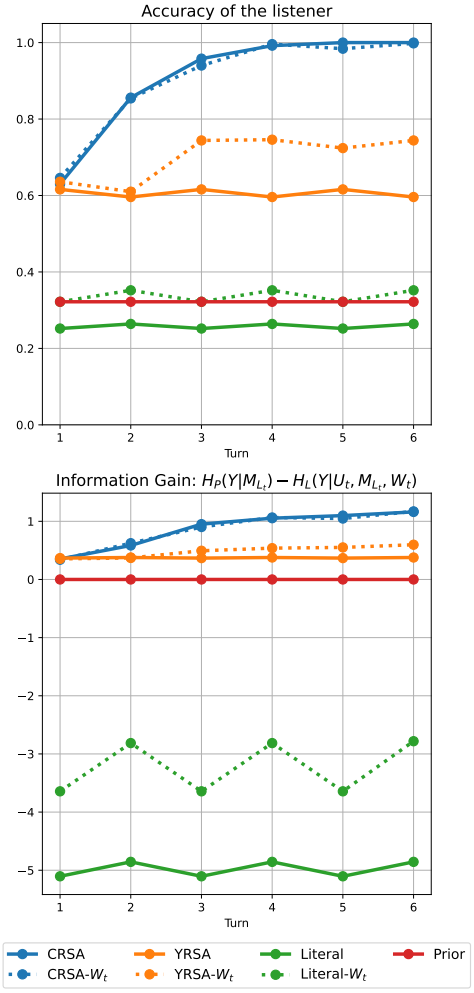


Figure 2: Average of correct predictions with the listener value (top) and information gain (bottom) for 500 rounds of the reference game.

- **Literal-W_t**: This is the same as above but using Eq. (10) as lexicon.
- **Prior**: In this case, we compute $P(y|m_{L_t})$ from $P(m_{S_t}, m_{L_t}, y)$ for all turns instead of $L_t(y|u_t, m_{L_t}, w_t)$. This case does not account for the dialog or the current utterance.

5.2 Numerical results and discussion

Figure 2 presents the performance of the CRSA model compared to baseline models for $\alpha = 2.5$. Each curve corresponds to a different model evaluated over 500 rounds of the game. The top plot displays task accuracy, measured as the proportion of correct guesses obtained by taking the argmax of the listener's posterior probability. As accuracy may not be fully representative of the confidence on the decision made by the listener, we also show the *Information Gain* (in the bottom plot) for each turn t, computed as the difference

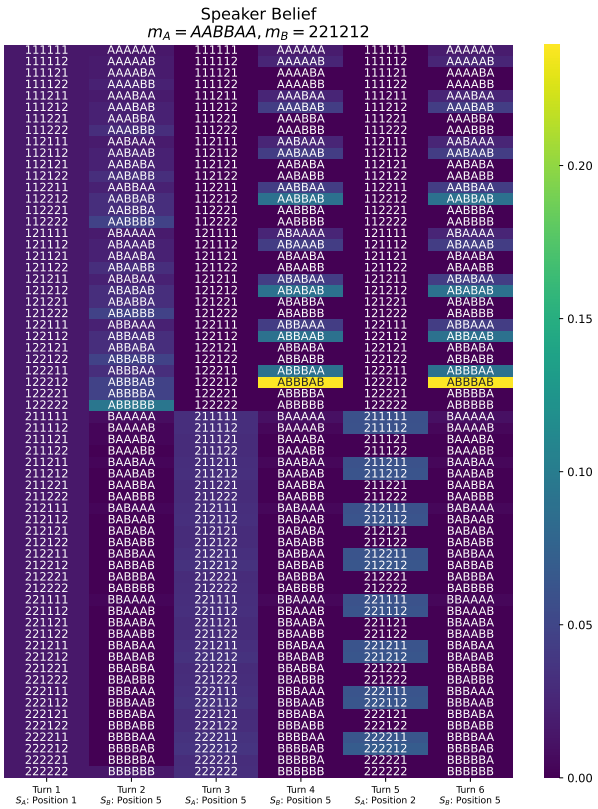


Figure 3: Internal belief of both agents.

$\text{IG}(L_t) = H_P(Y|M_{L_t}) - H_L(Y|U_t, M_{L_t}, W_t)$. That is, given a set of N rounds (all with the same number of turns), the listener's conditional entropy is defined as $H_L(Y|U_t, M_{L_t}, W_t) = -1/N \sum_{i=1}^{N} \log L(y^{(i)}|u_t^{(i)}, w_t^{(i)}, m_{L_t}^{(i)})$, and the conditional entropy of the prior is defined as $H_P(Y|M_{L_t}) = -1/N \sum_{i=1}^{N} \log P(y^{(i)}|m_{L_t}^{(i)})$, where the super-index (i) denotes the value at round i. As $P(y|m_{L_t})$ takes no account for the interchanged utterances, this metric could be interpreted as the amount of information gained by using the utterances of the dialog up to turn t. For all models where there is iteration, we run the model until the gain converged using a tolerance of $1e-3$, so the number of iterations may vary between each turn. We tried various values of $\alpha > 1$ and all values showed best performance of the CRSA model. For values $\alpha \leq 1$, all iterative algorithms always produced uniform distributions.

As shown in the plots, the CRSA model outperforms all baselines across both metrics. Moreover, incorporating a lexicon that depends on the past w_t neither improves nor diminishes performance, suggesting that the information encoded in Eq. (10) is already effectively captured by the CRSA model. In contrast, the information in Eq. 10 is not captured by the YRSA-W_t model, which appears to improve as the conversation progresses. As expected, models that do not incorporate dialog history maintain consistent performance across turns, with variations driven only by role changes. We also observed that the CRSA model's variance decreases over time, although this is not shown in the plots for clarity.

Figure 3 presents an example of a dialogue between the agents, along with their internal belief states at each turn. Each column displays the value of $B_{S,t}(m_{L_t})$ for each possible meaning m_{L_t} of the listener at turn t. Notably, as the conversation progresses, the meanings associated with previously uttered messages tend to gain higher belief values, reflecting a refinement in the speaker's inference about the listener's state. Here, the value that maximizes $B_{S,t}(m_{L_t})$ at turn 6 does not correspond exactly to the correct meaning, but it is a close approximation since the utterance "Position 6" never occurred during the round. This supports interpreting $B_{S,t}(m_{L_t})$ as the speaker's belief about the listener's meaning m_{L_t} at turn t.

6 Modeling Conversations Using Pragmatic LLMs

In this Section, we provide preliminary evidence that the CRSA model can produce reasonably good estimations of both the likelihood of each utterance u_t and the target y of the task in doctor–patient conversations, which is the disease corresponding to the symptoms described by the patient. To this end, we used the MDDial dataset (Macherla et al., 2023), which consists of template-based conversations between a doctor and a patient. In each dialog, the patient is assigned a subset of predefined symptoms, and the doctor must determine the correct disease from a set of possible pathologies.

As anticipated in Section 4.2.2, in order to apply the pragmatic models, we compute the literal speaker with equation 8 using a pre-trained LLaMA3.2–1B-Instruct language model. In this equation, $\text{prompt}(m_{S_t})$ is the text used to prompt the model with the relevant medical scenario. When S_t is the doctor, the prompt includes specific instructions to ask questions and produce a diagnosis, followed by two example doctor–patient conversations. When S_t is the patient, the prompt instructs the model to play the role of the patient. It uses the same conversation examples as in the doctor prompt but additionally includes the patient's

	H_S	H_L
CRSA	**8.18**	**2.19**
RSA	14.27	2.86
Literal	18.00	3.29

Table 1: Entropies for $\alpha = 2.5$ of the listener and the speaker for each model, computed for 65 samples of the MDDial dataset.

current symptoms at that turn. The full prompts used can be found in Appendix C. In order to save computation, we pre-computed the value of the literal speaker for each possible combination of utterance u_t and symptoms m_{S_t} for each turn t and then applied the update equations to that.

Finally, since each dialog ends with the doctor explicitly stating the disease, we compute the final listener distribution $L_{T^{(i)}}(y|u_{T^{(i)}}, m_{L_{T^{(i)}}}, w_{T^{(i)}})$ for the last turn $T^{(i)}$ of round i by replacing the disease name in the utterance with each candidate diagnosis $y \in \mathcal{Y}$. This probability is used at each round as the listener probability of the literal model.

6.1 Numerical results and discussion

To evaluate performance, we compute the speaker entropy as $H_S = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\log S_t\big(u_t^{(i)} \mid w_t^{(i)}, m_{S_t}^{(i)}\big)$, where N is the number of rounds and T_i is the number of turns in round i. For the Listener, we compute $H_L = -\frac{1}{N}\sum_{i=1}^{N} L_{T_i}\big(y \mid u_{T_i}, m_{L_{T_i}}, w_{T_i}\big)$, which is analogous to the method described in Section 5, but considers only the distribution at the final turn of each round.

The results are presented in Table 1 for 65 samples of the dataset and for a value of $\alpha = 2.5$, which is the same as used for Section 5. We observed the same trend mentioned in that Section when varying the value of α. The CRSA model achieves substantially lower entropy for both the speaker and the listener, outperforming both the RSA and Literal baselines. While the utterances in the MDDial dataset are not naturally produced—since they follow pre-defined templates—these results offer initial evidence that CRSA captures key aspects of pragmatic reasoning and task-oriented belief updates across turns.

7 Summary and Concluding Remarks

In this work, we introduced the Collaborative Rational Speech Act (CRSA) framework, an information-theoretic extension of RSA tailored for principled pragmatic reasoning in multi-turn, task-oriented dialogues. By integrating a novel multi-turn gain function grounded in interactive rate-distortion theory, CRSA effectively models the evolving belief dynamics of both interlocutors, overcoming key limitations of traditional RSA in collaborative contexts. Our preliminary results demonstrate that CRSA successfully captures the progression of shared understanding, partner beliefs, and utterance generation, providing the way for more natural and efficient communication in complex conversational settings.

CRSA lays the foundation for developing conversational agents driven by mathematically grounded principles of pragmatic reasoning. This principled formulation enhances both the interpretability and controllability of agent behavior, enabling the construction of language models that move beyond surface-level fluency to demonstrate structured, socially coherent, and contextually appropriate dialogue. In this way, CRSA represents a significant step toward building pragmatic agents whose interactions are not only effective but also firmly rooted in the formal theory of communication.

Limitations

This work focuses on simulated referential games and template-based doctor–patient dialogues, which, while controlled and insightful, do not capture the full variability and complexity of real-world conversations. Additionally, the CRSA framework relies on a fixed, predefined set of possible utterances at each turn, limiting its applicability to open-ended or generative dialogue scenarios involving variable-length token sequences. These factors currently restrict the scalability of our approach to more naturalistic domains. Future work will aim to overcome these limitations by extending CRSA to handle dynamically generated utterance spaces and by evaluating its effectiveness in less structured, real-world conversational settings.

Ethical considerations

This work presents a theoretically grounded framework for pragmatic reasoning in multi-turn dialogs. It is primarily methodological and does not involve direct deployment or interaction with real users. The datasets employed—simulated referential games and template-based medical dialogues—are synthetic and contain no personal or sensitive data.

However, since CRSA aims to inform the devel-

opment of more interpretable, goal-driven conversational agents, potential applications in sensitive domains like automatic medical diagnosis raise important ethical considerations. In such contexts, errors in belief tracking or task inference could result in incorrect recommendations, especially if users overestimate the system's understanding or authority. While our medical domain experiments are purely illustrative and not intended for clinical use, they underscore the critical need for caution when adapting theoretical models to real-world diagnostic settings. Future deployments must involve rigorous domain-specific validation, proper oversight, and human supervision to ensure safety and reliability.

References

Leon Bergen, Roger Levy, and Noah Goodman. 2016. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9:20:1–91.

Gaia Carenini, Luca Bischetti, Walter Schaeken, and Valentina Bambini. 2024. Towards a Fully Interpretable and More Scalable RSA Model for Metaphor Understanding. *arXiv preprint*. ArXiv:2404.02983 [cs].

Reuben Cohn-Gordon and Noah Goodman. 2019. Lost in Machine Translation: A Method to Reduce Meaning Loss. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 437–441, Minneapolis, Minnesota. Association for Computational Linguistics.

Reuben Cohn-Gordon, Noah Goodman, and Christopher Potts. 2018. Pragmatically Informative Image Captioning with Character-Level Inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 439–443, New Orleans, Louisiana. Association for Computational Linguistics.

Thomas M. Cover and Joy A. Thomas. 2001. *Elements of Information Theory*. Wiley.

Imre Csiszár and Paul Shields. 2004. *Information Theory and Statistics: A Tutorial*. Now Foundations and Trends.

Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R. McKee, Joel Z. Leibo, Kate Larson, and Thore Graepel. 2020. Open Problems in Cooperative AI. *arXiv preprint*. ArXiv:2012.08630 [cs].

Judith Degen. 2023. The Rational Speech Act Framework. *Annual Review of Linguistics*, 9(Volume 9, 2023):519–540. Publisher: Annual Reviews.

Judith Degen, Robert D. Hawkins, Caroline Graf, Elisa Kreiss, and Noah D. Goodman. 2020. When redundancy is useful: A Bayesian approach to "overinformative" referring expressions. *Psychological Review*, 127(4):591–621.

Judith Degen, Michael Henry Tessler, and Noah D. Goodman. 2015. Wonky worlds: Listeners revise world knowledge when utterances are odd. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 37(0).

Michael C. Frank and Noah D. Goodman. 2012. Predicting Pragmatic Reasoning in Language Games. *Science*, 336(6084):998–998.

He He, Anusha Balakrishnan, Mihail Eric, and Percy Liang. 2017. Learning Symmetric Collaborative Dialogue Agents with Dynamic Knowledge Graph Embeddings. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1766–1776, Vancouver, Canada. Association for Computational Linguistics.

Michele Herbstritt and Michael Franke. 2019. Complex probability expressions & higher-order uncertainty: Compositional semantics, probabilistic pragmatics & experimental data. *Cognition*, 186:50–71.

Hong Jun Jeon, Smitha Milli, and Anca Dragan. 2020. Reward-rational (implicit) choice: A unifying formalism for reward learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 4415–4426. Curran Associates, Inc.

Zhuoxuan Jiang, Xian-Ling Mao, Ziming Huang, Jie Ma, and Shaochun Li. 2019. Towards End-to-End Learning for Efficient Dialogue Agent by Modeling Looking-ahead Ability. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 133–142, Stockholm, Sweden. Association for Computational Linguistics.

Nalan Karunanayake. 2025. Next-generation agentic AI for transforming healthcare. *Informatics and Health*, 2(2):73–83.

A. Kaspi. 1985. Two-way source coding with a fidelity criterion. *IEEE Transactions on Information Theory*, 31(6):735–740.

Fereshte Khani, Noah D. Goodman, and Percy Liang. 2018. Planning, Inference and Pragmatics in Sequential Language Games. *Transactions of the Association for Computational Linguistics*, 6:543–555. Place: Cambridge, MA Publisher: MIT Press.

Hyunwoo Kim, Byeongchang Kim, and Gunhee Kim. 2020. Will I Sound Like Me? Improving Persona Consistency in Dialogues through Pragmatic Self-Consciousness. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language*

Processing (EMNLP), pages 904–916, Online. Association for Computational Linguistics.

Hyunwoo Kim, Byeongchang Kim, and Gunhee Kim. 2021. Perspective-taking and Pragmatics for Generating Empathetic Responses Focused on Emotion Causes. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2227–2240, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jiwei Li, Will Monroe, Alan Ritter, Dan Jurafsky, Michel Galley, and Jianfeng Gao. 2016. Deep Reinforcement Learning for Dialogue Generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202, Austin, Texas. Association for Computational Linguistics.

Jessy Lin, Daniel Fried, Dan Klein, and Anca Dragan. 2022. Inferring Rewards from Language in Context. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8546–8560, Dublin, Ireland. Association for Computational Linguistics.

Jessy Lin, Nicholas Tomlin, Jacob Andreas, and Jason Eisner. 2024. Decision-Oriented Dialogue for Human-AI Collaboration. *Transactions of the Association for Computational Linguistics*, 12:892–911. Place: Cambridge, MA Publisher: MIT Press.

Srija Macherla, Man Luo, Mihir Parmar, and Chitta Baral. 2023. MDDial: A Multi-turn Differential Diagnosis Dialogue Dataset with Reliability Evaluation. *arXiv preprint*. ArXiv:2308.08147 [cs].

Meta Fundamental AI Research Diplomacy Team (FAIR), Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, Athul Paul Jacob, Mojtaba Komeili, Karthik Konath, Minae Kwon, Adam Lerer, Mike Lewis, Alexander H. Miller, Sasha Mitts, Adithya Renduchintala, and 8 others. 2022. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074.

Fatema Al Nabhani, Mahizer Bin Hamzah, and Hassan Abuhassna. 2025. The role of artificial intelligence in personalizing educational content: Enhancing the learning experience and developing the teacher's role in an integrated educational environment. *Contemporary Educational Technology*, 17(2):ep573.

Leonardo Rey Vega, Pablo Piantanida, and Alfred O. Hero. 2017. The Three-Terminal Interactive Lossy Source Coding Problem. *IEEE Transactions on Information Theory*, 63(1):532–562.

Ashay Satav. 2025. Enterprise API & Platform Strategy in the era of Agentic AI. *Journal of Computer Science and Technology Studies*, 7(1):380–385.

Claude E. Shannon. 1993. *Coding Theorems for a Discrete Source With a Fidelity CriterionInstitute of Radio Engineers, International Convention Record, vol. 7, 1959.*, pages 325–350.

Sheng Shen, Daniel Fried, Jacob Andreas, and Dan Klein. 2019. Pragmatically Informative Text Generation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4060–4067, Minneapolis, Minnesota. Association for Computational Linguistics.

Tao Tu, Mike Schaekermann, Anil Palepu, Khaled Saab, Jan Freyberg, Ryutaro Tanno, Amy Wang, Brenna Li, Mohamed Amin, Yong Cheng, Elahe Vedadi, Nenad Tomasev, Shekoofeh Azizi, Karan Singhal, Le Hou, Albert Webson, Kavita Kulkarni, S. Sara Mahdavi, Christopher Semturs, and 7 others. 2025. Towards conversational diagnostic artificial intelligence. *Nature*, pages 1–9. Publisher: Nature Publishing Group.

Vaibhav Tupe and Shrinath Thube. 2025. AI Agentic workflows and Enterprise APIs: Adapting API architectures for the age of AI agents. *arXiv preprint*.

Matías Vera, Leonardo Rey Vega, and Pablo Piantanida. 2019. Collaborative Information Bottleneck. *IEEE Transactions on Information Theory*, 65(2):787–815.

Klarisa I. Vorobyeva, Svetlana Belous, Natalia V. Savchenko, Lyudmila M. Smirnova, Svetlana A. Nikitina, and Sergei P. Zhdanov. 2025. Personalized learning through AI: Pedagogical approaches and critical insights. *Contemporary Educational Technology*, 17(2):ep574.

Yifan Wang and Vera Demberg. 2024. RSA-Control: A Pragmatics-Grounded Lightweight Controllable Text Generation Framework. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5561–5582, Miami, Florida, USA. Association for Computational Linguistics.

Tsung-Hsien Wen, David Vandyke, Nikola Mrkšić, Milica Gašić, Lina M. Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. 2017. A Network-based End-to-End Trainable Task-oriented Dialogue System. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 438–449, Valencia, Spain. Association for Computational Linguistics.

Jason D. Williams and Steve Young. 2007. Partially observable Markov decision processes for spoken dialog systems. *Computer Speech & Language*, 21(2):393–422.

Kai Xu, Zhenyu Wang, Yangyang Zhao, and Bopeng Fang. 2025. An Efficient Dialogue Policy Agent with Model-Based Causal Reinforcement Learning. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7331–7343, Abu Dhabi, UAE. Association for Computational Linguistics.

Noga Zaslavsky, Jennifer Hu, and Roger P. Levy. 2021. A Rate–Distortion view of human pragmatic reasoning? In *Proceedings of the Society for Computation in Linguistics 2021*, pages 347–348, Online. Association for Computational Linguistics.

A Detailed Expressions of the YRSA Model

In Zaslavsky et al. (2021), the authors propose to use the alternation maximization (AM) algorithm (Csiszár and Shields, 2004) to maximize the gain function of expression 1:

$$S^{k+1} = \arg\max_S \mathcal{G}(S, L^k),$$
$$L^{k+1} = \arg\max_L \mathcal{G}(S^{k+1}, L).$$

If the same procedure is applied to the gain of equation 2 (the one corresponding to the YRSA model), then the following equations are obtained:

$$S^{k+1}(u|m_A) \propto \exp\big[\alpha(\sum_{\forall(m_B,y)} P(m_B, y|m_A) (\log(L^k(y|m_B, u)) - C(u))\big], \quad (11)$$

$$L^{k+1}(y|m_B, u) \propto \sum_{\forall m_A} P(m_A, m_B, y) \cdot S^{k+1}(u|m_A). \quad (12)$$

Additionally, if a lexicon $\mathcal{L}(u, m_A)$ is given, the listener is initialized as

$$L^0(y|m_B, u) \propto \sum_{\forall m_A} P(m_A, m_B, y) \cdot \mathcal{L}(u, m_A). \quad (13)$$

The proof of how to arrive to these equations is very similar to the ones to obtain the CRSA, which is presented in appendix B so we suggest to read that Section instead.

B Derivation of the CRSA Model Expressions

For the following derivation we have assumed that the speaker is agent A and the listener is agent B in order to simplify notation. In addition, since every variable depends on the turn, we will omit the subindex t for the same reason. We start by representing the speaker, the listener, the prior and the cost as matrices:

$$s_{awu} = S(u|m_A, w) = [\mathbf{S}]_{awu}$$
$$\mathbf{S} \in [0,1]^{\mathcal{M}_A \times \mathcal{W} \times \mathcal{U}}$$

$$l_{buwy} = L(y|m_B, u, w) = [\mathbf{L}]_{buwy}$$
$$\mathbf{L} \in [0,1]^{\mathcal{M}_B \times \mathcal{U} \times \mathcal{W} \times \mathcal{Y}}$$

$$P_{abyw} = P_S(m_A, m_b, y, w) = [\mathbf{P}]_{abyw}$$
$$\mathbf{P} \in [0,1]^{\mathcal{M}_A \times \mathcal{M}_B \times \mathcal{Y} \times \mathcal{W}}$$

$$c_u = C(u) = [\mathbf{C}]_u$$
$$\mathbf{C} \in \mathbb{R}^{\mathcal{U}}$$

with the restrictions

$$\sum_u s_{awu} = 1, \qquad \sum_y l_{buwy} = 1. \quad (14)$$

The gain function at the turn t as a function of the matrices $\mathbf{S}$ and $\mathbf{L}$ can be written as

$$\begin{aligned}\mathcal{G}(\mathbf{S}, \mathbf{L}) &= -\sum_{abywu} s_{awu} P_{abyw} (\log s_{awu} + \\ &\quad \alpha(\log l_{buwy} - c_u)) \\ &= -\sum_{awu} s_{awu} P_{aw} \log s_{awu} + \\ &\quad \alpha \sum_{abywu} s_{awu} P_{abyw} \log l_{buwy} - c_u \\ &= \sum_w \mathcal{G}_w(\mathbf{S}, \mathbf{L}), \end{aligned} \quad (15)$$

where

$$\begin{aligned}\mathcal{G}_w(\mathbf{S}, \mathbf{L}) &= -\sum_{au} s_{awu} P_{aw} \log s_{awu} \\ &\quad +\alpha \sum_{abyu} s_{awu} P_{abyw} \log l_{buwy} - c_u. \end{aligned} \quad (16)$$

Since the overall gain is a sum of the gain for a specific utterance history w, taking the derivative with respect to a different value of w cancels out the other terms in the sum, so we can abbreviate the notation by omitting the w subindex. Then, the problem reduces to maximize the following Lagrangian:

$$\begin{aligned}\mathcal{L}(\mathbf{S}, \mathbf{L}) &= -\sum_{au} s_{au} P_a \log s_{au} \\ &\quad + \alpha \left(\sum_{abyu} s_{au} P_{aby} \log l_{buy} - c_u \right) \\ &\quad - \sum_a \lambda_a g_a(\mathbf{S}) - \sum_{bu} \lambda_{bu} g_{bu}(\mathbf{L}) \end{aligned}$$

with

$$g_a(\mathbf{S}) = 1 - \sum_u s_{au} = 0,$$

$$g_{bu}(\mathbf{L}) = 1 - \sum_y l_{buy} = 0.$$

Taking the gradient w.r.t $s_{\hat{a}\hat{u}}$ and $l_{\hat{b}\hat{u}\hat{y}}$, we get

$$\frac{\partial \mathcal{L}}{\partial s_{\hat{a}\hat{u}}} = -P_a(\log s_{\hat{a}\hat{u}} + 1) + \alpha \sum_{by} P_{aby}(\log l_{b\hat{u}y} - c_{\hat{u}}) - \lambda_{\hat{a}} = 0,$$

$$\frac{\partial \mathcal{L}}{\partial l_{\hat{b}\hat{u}\hat{y}}} = \frac{\alpha}{l_{\hat{b}\hat{u}\hat{y}}} \sum_a s_{a\hat{u}} P_{a\hat{b}\hat{y}} - \lambda_{\hat{b}\hat{u}} = 0.$$

So it is straightforward to see that

$$l_{\hat{b}\hat{u}\hat{y}} \propto \sum_a s_{a\hat{u}} P_{a\hat{b}\hat{y}}$$

$$s_{\hat{a}\hat{u}} \propto \exp\left(\alpha \sum_{by} \frac{P_{\hat{a}by}}{P_{\hat{a}}} (\log l_{b\hat{u}y} - c_{\hat{u}})\right).$$

We can rewrite these equations in terms of the the original probabilities adding the past w and the turn t subindex:

$$L(y|m_B, u_t, w_t) \propto \sum_{\forall m_A} S(u_t|m_A, w_t) P_S(m_A, m_B, y, w_t)$$

$$S(u_t|m_A, w_t) \propto \exp(\alpha \sum_{\forall (m_B, y)} P_S(m_B, y|m_A, w_t) (\log L(y|m_B, u_t, w_t) - C(u_t))$$

Then, by applying equations 5 and 6 of Section 4.2 we can directly obtain

$$S_t(u_t|w_t, m_{S_t}) \propto \exp(\alpha \sum_{\forall (m_{L_t}, y)} B'_t(m_{S_t}, m_{L_t}, y) V_L(u_t, w_t, m_{L_t}, y), \quad (17)$$

$$L_t(y|u_t, w_t, m_{L_t}) \propto \sum_{\forall m_{S_t}} B_{S,t}(m_{S_t}) P_t(m_{S_t}, m_{L_t}, y) S_t(u_t|w_t, m_{S_t}). \quad (18)$$

These are the equations that maximize the gain $\mathcal{G}(\mathbf{S}, \mathbf{L})$ subject to the restrictions 14. Then, by applying again the alternation maximization algorithm we obtain the CRSA algorithm.

C Prompts used in the MDDial dataset

We prompt two different models for generating the lexicons defined in Section 6. The first one (the one for the patient) contained the following instructions:

```
You are an assistant that simulates to be a patient who has a disease and describes the symptoms to the user, which is a medical doctor.

Here is an example of a conversation between the assitant (i.e., the patient) and the user (i.e., the doctor). You are experiencing the following symptoms: Fear of cold, Poor sleep, Feel sick and vomit, Syncope, Increased stool frequency, Blood in the tears, Hoarse, Loss of appetite, Dizzy, Expectoration, Pharynx discomfort, Limb numbness, Acid reflux, Thirst, Diarrhea, Difficulty breathing, Stuffy nose, Vomiting, Bitter, Runny nose, Bloody stools, Bloating, Fatigue, Hard to swallow, Chest tightness, Shortness of breath, Palpitations, Fever, Headache, Constipation, Thin, Dizziness, Body aches, Diplopia, Hemoptysis, Burning sensation behind the breastbone, Sweating, Hiccup, Poor spirits, Frequent urination, Black stool, Consciousness disorder, Thin white moss, Edema, Anorexia, Cough, Stomach ache, Pain behind the breastbone, Sleep disorder, Hematemesis, Nausea, Twitch, Hiccough, Chest tightness and shortness of breath
  Assistant: I have been feeling Burning sensation behind the breastbone
  User: Oh, do you have any Stomach ache?
  Assistant: I am experiencing that sometimes
  User: Oh, do you have any Acid reflux?
  Assistant: Yes Doctor, I am feeling that as well
  User: This could probably be Esophagitis.

Here is an example of a conversation between the assitant (i.e., the patient) and the user (i.e., the doctor). You are experiencing the following symptoms: Pain in front of neck, Nose bleeding, Fear of cold, Syncope, Feel sick and vomit, Poor sleep, Headache and dizziness, Hazy, Loss of appetite, Hearing loss, Dizzy, Cry, Diarrhea, Difficulty breathing,
```

Stuffy nose, Vomiting, Runny nose, Bloating, Fatigue, Palpitations, Chest tightness, Headache, Fever, Dizziness, Incontinence, Poor spirits, Redness, Waist pain, Unconsciousness, Vertigo, Consciousness disorder, Head trauma pain, Poor physical activity, Anorexia, Cough, Stomach ache, Pain behind the breastbone, Hematemesis, Sleep disorder, Earache, Nausea, Tinnitus, Twitch, Limb numbness, Chest tightness and shortness of breath

Assistant: I have Nausea and Dizziness

User: I believe you are having from Traumatic brain injury.

Now, participate in a real conversation with the user. You are experiencing the following symptoms: {patient symptoms}

The prompt used for the doctor contained the following instructions:

You are an assistant that simulates to be a doctor who is diagnosing a patient based on the symptoms that he or she describes. You can ask questions to the patient, but ultimately, you have to provide a diagnosis based on the symptoms described by the patient.

Here is an example of a conversation between the assitant (i.e., the doctor) and the user (i.e., the patient):

User: I have been feeling Burning sensation behind the breastbone

Assistant: Oh, do you have any Stomach ache?

User: I am experiencing that sometimes

Assistant: Oh, do you have any Acid reflux?

User: Yes Doctor, I am feeling that as well

Assistant: This could probably be Esophagitis.

Here is an example of a conversation between the assitant (i.e., the doctor) and the user (i.e., the patient):

User: I have Nausea and Dizziness

Assistant: I believe you are having from Traumatic brain injury.

Now, participate in a real conversation with the user. You can ask questions to the patient, but ultimately, you have provide a diagnosis based on the symptoms described by the patient.

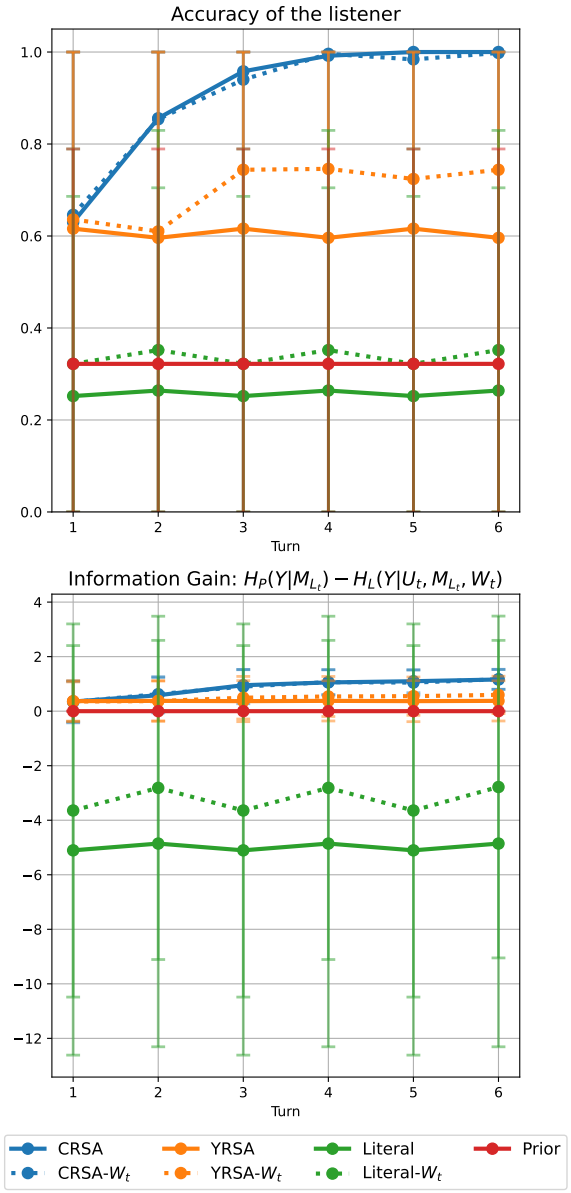


Figure 4: Performance of the CRSA compared to baselines, including error bars.

D Errors intervals in the reference game

Fig. 4 shows the same results as fig. 2 but with the standard deviation of each model. We did not included this plot in the main text for readability, but it can also be noted that the CRSA reduce the variance of the results in comparison with the other models.