MindDial: Belief Dynamics Tracking with Theory-of-Mind Modeling for Neural Dialogue Generation

Shuwen Qiu [1] **Song-Chun Zhu** [1 2 3 4] **Zilong Zheng** [2]

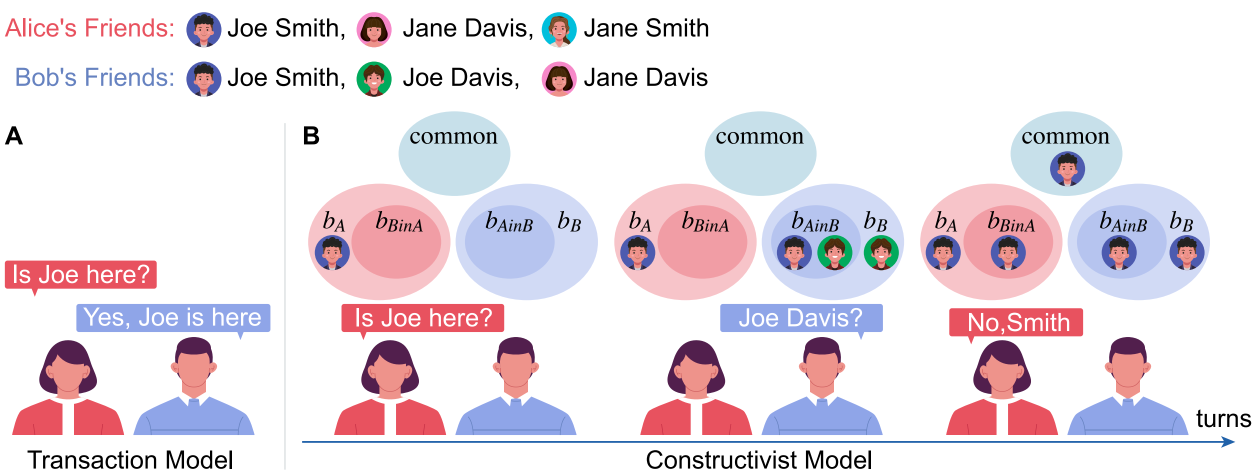


Figure 1: **Illustrative comparison between our Construstivist model (B) and conventional Transaction model (A).**

Abstract

Humans talk in free-form while negotiating the expressed meanings or common ground. Despite the impressive conversational abilities of the large generative language models, they do not consider the individual differences in contextual understanding in a shared situated environment. In this work, we propose MindDial, a novel conversational framework that can generate situated free-form responses to negotiate common ground. We design an explicit mind module that can track three-level beliefs – the speaker's belief, the speaker's prediction of the listener's belief, and the common belief based on the gap between the first two. Then the speaking act classification head will decide to continue to talk, end this turn, or take task-related action. We augment a common ground alignment dataset MutualFriend (He et al., 2017b) with belief dynamics annotation, of which the goal is to find a single mutual friend based on the free chat between two agents. Experiments show that our model with mental state modeling can resemble human responses when aligning common ground meanwhile mimic the natural human conversation flow. The ablation study further validates the third-level common belief can aggregate information of the first and second-order beliefs and align common ground more efficiently.

[1]UCLA Center for Vision, Cognition, Learning, and Autonomy (VCLA) [2]National Key Laboratory of General Artificial Intelligence, BIGAI [3]Department of Automation, Tsinghua University [4]Institute for Artificial Intelligence, Peking University. Correspondence to: Zilong Zheng <zlzheng@bigai.ai>.

Proceedings of the First Workshop on Theory of Mind in Communicating Agents, Honolulu, Hawaii, USA. PMLR 202, 2023.

1. Introduction

Large generative language models (LGLMs), such as GPT (Radford et al., 2019; Brown et al., 2020) and T5 (Raffel et al., 2020), have dominated the natural language processing (NLP) community over recent years. By providing carefully designed prompts with zero or few-shot examples, LGLMs are capable of adapting to various downstream NLP tasks in the form of natural language generation. ChatGPT, one of the most prevalent models, has demonstrated "human-like" multi-turn conversational generation abilities by learning from massive natural language datasets and human feedbacks (RLHF) (Ouyang et al., 2022). Despite the inspiring performance demonstrated by these large language chat models, the objective of these models is to provide *general*, *helpful* and *objective* information based on users' queries (Glaese et al., 2022; OpenAI, 2023) rather than daily chit-chats. Therefore, they can hardly be applied to daily conversational agents that deliver *situated*, *free-form* and *subjective* responses within a shared conversational context. Critically, two vital components are not carefully addressed in previous dialogue models:

I. Shared Situated Context. Conventional neural dialogue agents model dialogues as multi-turn question answering, where they treat conversational agents as sender and receiver: the sender initiates the topic by sending out a query and the receiver takes in the query and produces the next sentence with the highest probability by learning from a massive dialogue corpus. These models are considered *Transaction models*. Refer to Figure 1A for an example, Bob responds according to his individual knowledge and assumes there is no ambiguity in Alice's question. However, these models differ from the nature of human communication — we communicate based on a shared understanding of contextual meanings (also known as negotiated meanings or common ground) (Burleson, 2007; Delia & O'Keefe, 1982). Such perspective is particularly essential when the dialogue agents can only partially perceive the environment and effective communication can only occur when they negotiate to obtain a common ground. These models are in line with the *Constructivist model*. As shown in the talking pairs from Figure 1B, Alice and Bob have to negotiate to confirm the common understanding of "`Joe`" is "`Joe Smith`".

II. Individual Mental Dynamics. One prerequisite for the Constructivist framework is that agents have to explicitly model individual differences in understanding, intention and goals, *i.e.*, mental states. Compared with the Transaction framework that only models the individual's state, a proper inference of the conversational partner's state can result in faster convergence of the common ground. Moreover, based on the individual mental state modeling, one can easily produce free-form dialogues, *i.e.*, the goal is not to be forced to generate a single response based on the context but to keep speaking or stop based on individual differences.

In order to step towards real-world situated conversation, in this work, we introduce **MindDial**, a new dialogue framework built upon the theory-of-mind (ToM) modeling in cognitive theory, aiming at modeling free-form neural dialogue generation with the Constructivist model's point of view. Specifically, a mind module on top of the dialogue context encoder will start by predicting the first and second-order belief dynamics of the current speaker. Then a common belief distribution about which entities will be aligned is estimated based on the mind gap between the first and second-order beliefs. The neural dialogue generator will generate the next response taking the dialogue context, world knowledge, and the common belief distribution into account. We adopt a cooperation communication game dataset with additional belief annotations to demonstrate the belief update process and how agents negotiate common ground through conversation. It is worth noting that some works simulate human values through their feedback to enable dialogue systems to generate responses aligning better with human expectations (Bai et al., 2022; Yuan et al., 2022). Different variants of Sally-Anne test are also proposed to test ToM of large language models (Kosinski, 2023; Ullman, 2023; Sileo & Lernould, 2023). In comparison, we consider mind modeling in situated daily dialogue scenarios and differing from value alignment and ToM question answering from three perspectives:

- **A structured "five mind" representation** (Fan et al., 2021): as the Constructivist model example shown in Figure 1B, we model two first-order beliefs, two second-order beliefs of each other's mind, and the third-level common belief;
- **Belief dynamics prediction**: we explicitly model how each utterance in the context results in the occurrence, disappearance, or no change of the entities in the agents' beliefs as the belief dynamics. Then, the final belief distribution will be the summation of all belief dynamics estimation over utterances in the dialogue history. It avoids the problem of losing track of some entities if we directly model the final belief given a long dialogue input;
- **Common belief modeling**: we define the common belief as how probably the current speaker thinks about an entity that will be aligned to the common ground. We hypothesize that the common belief is based on the gap between the speaker's belief of the physical world and her belief estimation of the listener.

Contribution We consider our contributions as four-fold:

i. We build a speaking act classifier to model the free-form conversation. Experimental results show that the model can accurately predict whether the current speaker wants to talk more or finishes the current turn.
ii. We design an explicit mind module that can track the first and second-order beliefs over long contexts by aggregating belief dynamics. A third-level common belief based on the gap between the two will support the next response generation.
iii. We augment a collaborative dialogue dataset MutualFriend with belief dynamics annotations for each utterance. The dataset can serve as a new benchmark for ToM in situated dialogue tasks.
iv. The evaluation results demonstrate the efficacy of each

component in our mind module. The responses generated with the three-level beliefs are shown to be more accurate and efficient for negotiating common ground.

2. Related Work

Theory-of-Mind (ToM) ToM is a crucial capability for human social interactions developed in early life (Kovács et al., 2010; Richardson et al., 2018). In literature, early works model belief update through time in sequential games with partially observable Markov decision process (POMDP) (Baker et al., 2011; De Weerd et al., 2013; Doshi et al., 2010; Han & Gmytrasiewicz, 2018). One agent's belief update is based on the estimate of others' current beliefs, resulting in an infinite recursion. However, in real life, studies have shown that humans could go no deeper than two levels of recursion (Camerer et al., 2004). Therefore, works (Fan et al., 2021) began the efforts to end the recursion when their beliefs merge into the "common mind".

Modeling the belief of others has been extensively studied in symbolic-like environments (Wunder et al., 2011; Rabinowitz et al., 2018; Kleiman-Weiner et al., 2016; Ho et al., 2016), where agents need to incorporate or compete for a goal. Efforts to measure models' ability to recognize false beliefs and perspective-taking also emerge in robotics (Yuan et al., 2020; Milliez et al., 2014), computer vision (Eysenbach et al., 2016; Fan et al., 2021), and natural language processing (Qiu et al., 2022; Nematzadeh et al., 2018) using the Sally-Anne test (Baron-Cohen et al., 1985). It is also shown that augmenting the model with external mind modules can help improve the performance of tasks involving intensive belief exchange and rich social interaction scenarios (Fan et al., 2021; Qiu et al., 2022). In this work, we explore how the common belief modeling with first and second-order belief difference can enhance the quality and efficiency of the response generation in dialogue tasks.

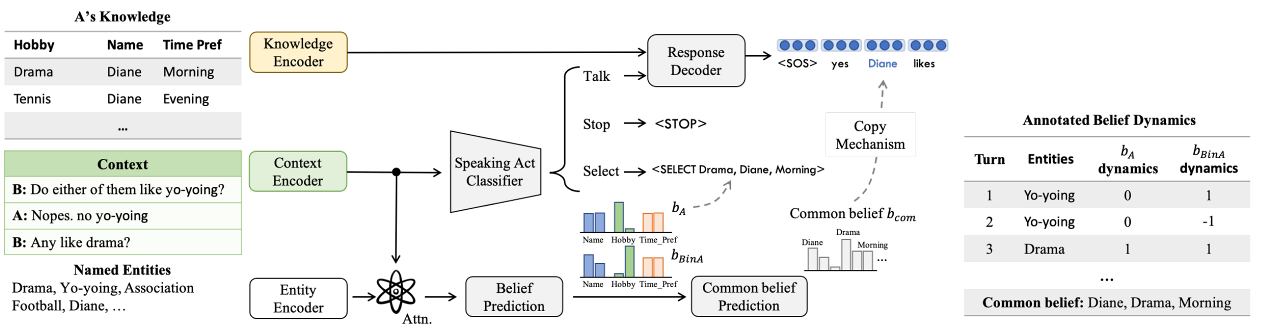
Neural Dialogue Generation Neural dialogue generation has made impressive progress after various datasets and advanced model architectures are proposed. Both SEQ2SEQ and decoder-based models (Lewis et al., 2019; Zhang et al., 2020) are introduced into the open-domain dialogue systems for style and personality-controlled generation (Hu et al., 2022; Cho et al., 2022), with knowledge and emotion-aware abilities (Varshney et al., 2022; Liu et al., 2022), etc. In addition, researchers explore reinforcement learning-based methods to enable agents to learn from human feedback (Bai et al., 2022), coordinate, and compete with each other in task-oriented dialogues (Verma et al., 2022; Jang et al., 2022).

Cooperative Communication For a cooperative task, efficient communication could be essential, especially in a situation when each agent can only have a partial observation of the environment. To guarantee that the communication takes the least cost meanwhile provides the most informative messages, previous works proposed multiple methods to align the common ground between agents (Bohn et al., 2019; Anderson, 2021). Specially for dialogue tasks, datasets have been collected to provide golden utterances when people try to align the common ground with each other based on structured knowledge (He et al., 2017a), in situated tasks (Bara et al., 2021; Kim et al., 2019), multimodal and continuous environment (Haber et al., 2019; Udagawa & Aizawa, 2021). Frameworks have been adopted to model the belief dynamics using GNN, RNN, and transformers (He et al., 2017a; Udagawa & Aizawa, 2021; Qiu et al., 2022). Fan et al. (2021) introduce low-level visual cues that may possibly indicate mind transition. However, most of the models only focus on the first-order belief (the current speaker's belief of the world). In this work, we track the speaker's both first and second-order beliefs (the current speaker's belief of others) and demonstrate how the jointly modeling between the two can help align the common ground.

3. The MindDial Framework

The dialogue corpus can be denoted as $\mathcal{D} = \{(U_n, KB_n^p, E_n^p, C_n, Y_n)\}_{n=1}^N$, where $U_n = (u_{n,1}, ..., u_{n,K})$ represents the dialogue history and K is the number of turns. $KB_n^p = (kb_{n,1}, ...kb_{n,I})$ where I is the number of knowledge passages. $p \in A, B$ represents the two agents. We assume that the current speaker is A, and p will be dropped for the following formulation. The knowledge base contains entities under different attributes. E_n is a union set of entities visible for the current speaker. Each entity in $E_n = \{E_n^U, E_n^{KB}\}$ is either shown in the dialogue history or in the speaker's knowledge base. $Y_n = \{y_n^1, .., y_n^L\}$ is A's next response consisting of several utterances. $C_n = \{c_n^1, ..., c_n^L\}$ is A's corresponding speaking act sequence, with $c_n^1, ..., c_n^{L-1}$ is "continue to talk" and c_n^L belongs to stop talking or make task related action. Given the dialogue contexts and a partial response $\{y_n^1, .., y_n^{l-1}\}$, speaker A will first decide the next speaking act c_l. If she decides to talk, the next utterance y_n^l will be generated.

Utterance Encoder The utterance and the structured knowledge encoder are built upon sequential models like recurrent neural networks (Cho et al., 2014) or a transformer encoder (Vaswani et al., 2017). We define the output states at all time steps of one utterance and knowledge passage as o_{u_k}, o_{kb_i}. We take the last hidden state as the turn-level representation s_{u_k}, $s_{kb_i} \in \mathbb{R}^{1\times d_h}$. Similarly, when we flatten the turns and passages into a single sequence, the encoded output is written as o_U and o_{KB}. The sentence-level representation is s_U and s_{KB}. The entity representation s_e is a linear transformation of the corresponding word embedding



(a) Overview of MindDial framework. (b) Example of belief dynamics annotation.

Figure 2: (a) Illustration of the MindDial framework. The colored barplots denote the belief prediction's outputs b_A and b_{BinA} for each attribute; the barplot in lightgray denotes the normalized probability of the next entity to be mentioned w.r.t. the common belief prediction module. (b) We annotate belief dynamics of each turn for the current context; refer to Section 4.1 for details.

of the entity for a given $e \in E$.

Belief prediction We define the first and second-order belief as the confidence distribution over entities of each attribute. For a knowledge base with M attributes, the belief is $b = \{b^m = (b^{m_1}, ..., b^{m_{J_m}})\}_{m=1}^{M}, \sum_j b^{m_j} = 1$. J_m is the number of entities for attribute m. We further define belief dynamics as the state change of each entity at each time step k as Δb_k with each entry value ranging from -1 to 1. -1 indicates the disappearance of an entity in the belief and 1 indicates an occurrence of an entity. Then the current belief is the accumulated prediction of the dynamics over all dialogue turns: $b^m = \text{Softmax}(b_0^m + \sum_k \Delta b_k^m)$. b_0 is initialized to all zeros. The belief dynamics of speaker A are obtained by calculating the attention score between the current utterance and the entity shown in her own knowledge base

$$
\begin{aligned}
\Delta b_{k,A}^m &= tanh(s_{E^{KB}}^m \cdot (s_{u_k})^T) \\
d_A^m &= \sum_k (b_{k,A}^m)^T \cdot s_{E^{KB}}^m,
\end{aligned}
\tag{1}
$$

where $s_{E^{KB}}^m = [s_e^{m_1}; ...; s_e^{m_{J'_m}}] \in \mathbb{R}^{J'_m \times d_z}$ concatenates all entity representations for the attribute m shown in E^{KB}. $d_A^m \in \mathbb{R}^{1 \times d_z}$ is the belief representation of attribute m for the speaker. Since A's current estimate of B's knowledge base is based on her own knowledge base along with the entities mentioned by B in the context, using the same attention module applied to the context and the whole entity set of A, we can get d_{BinA}, b_{BinA}:

$$
\begin{aligned}
\Delta b_{k,BinA}^m &= tanh(s_E^m \cdot (s_{u_k})^T) \\
d_{BinA}^m &= \sum_k (\Delta b_{k,BinA}^m)^T \cdot s_E^m.
\end{aligned}
\tag{2}
$$

We further define common belief b_{com} as how likely each entity is agents' next talking focus over all possible entities based on the gap $bdiff$ between b_A and b_{BinA}. The output range is between 0 and 1. 1 indicates that this entity will be mentioned in the next response and 0 otherwise. The common belief will be learned through another attention layer $b_{com} = \text{Sigmoid}(s_E \cdot d_{bdiff}^T)$, where $s_E = [s_E^0; ..., s_E^M] \in \mathbb{R}^{|Ent| \times d_z}$. $|Ent| = \sum_m J_m$ is the total number of entities over all attributes.

To get the gap representation d_{bdiff}, we first aggregate d_A and d_{BinA} over entities by a weighted summation of p^m: $d_{bdiff}^m = (p^m d_A^m + (1 - p^m) d_{BinA}^m)$, where $p^m = \sigma(W_{ent}[b_A^m, b_{BinA}^m]^T)$, and $W_{ent} = \mathbb{R}^{1 \times 2d_z}$. d_{bdiff}^m is then aggregated over attributes: $d_{bdiff} = g(W_{Att}[d_{bdiff}^0; ...; d_{bdiff}^M])$, where $W_{Att} = \text{Softmax}(W_0, ..., W_M)$. W_m is computed by the Jensen–Shannon divergence (Lin, 1991) over b_A^m and b_{BinA}^m in that we hope to pay more attention to attributes if they differ a lot between b_A and b_{BinA}. g is a linear transformation.

Speaking act Classifier We divide the speaking act into 3 categories: {continue to talk, end the current turn, take task-related action}. Based on the current context and the partial response, the action is predicted using $p_{c^l} = \text{MLP}(s_U)$, where $\text{MLP}(\cdot)$ denotes a multi-layer perceptron network.

Response Decoder We take another recurrent neural network or a transformer decoder as our response generator. For each word prediction, it receives the embedding vector y_{t-1} of the word predicted at time-step $t-1$ and outputs the last hidden state $h_t \in \mathbb{R}^{1 \times d_{z'}}$ and $P_{vocab}(w_t)$ over the fixed vocabulary obtained from the training set.

Multi-source Copy Mechanism. We follow Bai et al. (2021); He et al. (2017a) to adopt the copy mechanism so that the final word distribution depends on both the decoder output and a copy probability of words shown in dialogue history, speaker's knowledge base and the common belief. The context representation of utterance and structured knowledge at each time step is obtained through the Attention module:

$$d_t^{\phi} = \text{Softmax}(h_t \cdot o_{\phi}^T) o_{\phi} \tag{3}$$

where $\phi \in \{U, KB\}$. Then the decoder state h_t attends over the dialogue history representation d_t^U and the knowledge representation d_t^{KB} by

$$\begin{aligned} \alpha_t &= \text{Softmax}(h_t \cdot [d_t^U; d_t^{KB}]^T) \\ d_t &= \alpha_t [d_t^U; d_t^{KB}] \end{aligned} \tag{4}$$

where $\alpha_t = (\alpha_t^U, \alpha_t^{KB}) \in \mathbb{R}^{1\times 2}$ is used to combine the distributions of the two inputs as shown in Equation (5). We also use a generation probability $p_t^{gen} \in [0, 1]$ to balance the distribution between input sources and the fixed vocabulary, where $p_t^{gen} = \sigma(W_{gen}[y_{t-1}, h_t, d_t]^T)$, and $W_{gen} \in \mathbb{R}^{1\times(d_{emb}+d_z+d_{z'})}$. Besides, we set a mind weight p_{com} to leverage the common belief distribution into the final prediction. The overall distribution is obtained by

$$\begin{aligned} P(w_t) &= (1 - p_{com})\big[p_t^{gen} P_{vocab}(w_t) \\ &+ (1 - p_t^{gen}) \sum\nolimits_{\phi:\phi\{U,KB\}} \alpha_t^{\phi} P_{\phi}(w_t)\big] \\ &+ p_{com}\, \text{Softmax}(b_{com}) \end{aligned} \tag{5}$$

Objective Mean squared error (MSE) loss will be used to measure the difference between the predicted and ground truth belief dynamics. The common belief prediction loss will be measured by Binary Cross Entropy (BCE). The action classification head is updated by Cross-Entropy loss. Apart from belief and act training loss, we use the NLL loss to capture the word order information:

$$\mathcal{L}_{NLL} = -\frac{1}{|y^l|} \sum_{t=1}^{|y^l|} \log(P(y_t^l | y_{1:t-1}^l, U, KB)) \tag{6}$$

The final loss is composed of three parts:

$$\mathcal{L} = \mathcal{L}_{NLL} + \mathcal{L}_{belief} + \mathcal{L}_{act} \tag{7}$$

4. Experiments

4.1. Settings

Dataset To provide a reasonable quantitative measure of belief dynamics in the dialogue, the expected dataset should contain rich belief exchanges. Meanwhile, the belief exchange and the final common ground can be easily labeled. Therefore, we choose MutualFriend (He et al., 2017b) to evaluate our dialogue generation framework for its clear definition of belief (distribution over structured knowledge) and common ground (the mutual friend). In the MutualFriend task, each agent has a private knowledge base including a list of friends and their attributes like name, school, *etc*. There is a shared friend that both agents have and they need to chat with each other to find this mutual friend. We only keep the successful dialogues and the final data split for train/val/test is 4922/608/581. Each dialogue in the training set contains a maximum of 53 turns and each turn with a maximum length of 29.

To get the supervision signal for belief dynamics, we manually label each entity after one turn of utterance as occur (mentioned by the speaker), no change (not mentioned in the last turn), or disappear (negated by the speaker). Figure 2 illustrate one annotation process. For example, when B is asking about "yo-yoing", this entity is marked as 1 for b_{BinA} dynamics. However, since it does not belong to A's knowledge, for the first-order belief of speaker A, we annotate it as no change. Then, when "yo-yoing" is negated by A, it will be marked as a "disappear" in b_{BinA} dynamics. One entity is labeled as the common belief to be aligned next if it is shown in the response utterance.

Implementation To serve as a baseline in this task, the model is expected to encode current contexts and predict the belief dynamics. Then it will further generate the next response based on both the dialogue history and the belief prediction. Therefore, we select dialogue baselines from three categories: 1) We use the Gated Recurrent Unit (GRU) (Cho et al., 2014) among the recurrent neural networks for its memory efficiency of modeling sequential data; 2) We combine the powerful encoder Transformer (Vaswani et al., 2017) with the decoder Transformer (Radford et al., 2019) for its strong conversation abilities; 3) We employ pre-trained encoder-decoder Transformer architectures such as BART (Lewis et al., 2020) which can be flexibly adopted to sequence-to-sequence tasks.

For all transformer models, we finetune the pretrained model on the MuturalFriend dataset. For context encoding, we prepend the BOS token at the beginning of the context and use its corresponding hidden representation as the turn and sentence-level representation for the following attention layers and speaking act prediction. The entity encoding will be a linear transformation of the corresponding word embedding. Meanwhile, the decoder's predicted vocabulary distribution will be mediated by the copy mechanism (Bai et al., 2021). The model is trained on a single A6000 GPU for 30k steps with an initial learning rate of 1e-4. The batch size is set to 32. Results are gathered over 3 random seeds.

4.2. Evaluation and Results

Mind prediction We first determine whether the models can accurately track both the first and second-order beliefs. The dynamics prediction performance is evaluated using

Table 1: **Belief dynamics prediction**. Δb specifies belief dynamics for b_A and b_{BinA}.

Models	Δb	Precision	Recall	F1
GRU	b_A	69.00±0.02	83.33±0.02	74.67±0.02
Transformer		65.67±0.01	70.67±0.09	67.67±0.04
BART		70.67±0.02	62.33±0.07	64.67±0.03
GRU	b_{BinA}	73.33±0.02	83.33±0.02	77.33±0.01
Transformer		70.00±0.01	69.33±0.11	68.33±0.06
BART		73.33±0.02	60.00±0.07	62.33±0.04

Table 2: **Next response generation and speaking act classification**. +mind indicates the generator copies from common belief distribution.

Models	METEOR	ROUGE-L	BLEU-1	BLEU-2	Action acc
GRU	7.64±0.33	9.26±0.90	12.65±1.06	4.70±0.42	76.49±0.48
GRU+mind	8.56±0.03	9.89±0.27	12.70±0.56	5.17±0.11	77.64±0.54
Transformer	9.93±0.70	12.44±0.15	13.45±0.21	3.95±0.21	77.24±0.50
Transformer+mind	10.45±0.11	13.14±0.23	14.15±0.07	4.90±0.14	77.66±0.42
BART	10.70±0.28	12.54±1.90	15.10±1.13	4.75±0.35	75.77±2.62
BART+mind	11.72±0.26	14.01±1.10	16.95±1.34	6.25±1.06	76.90±0.69

the macro-average of Precision, Recall, and F1-score. We can see from Table 1 that all three types of encoders can predict the belief dynamics in mind A and $BinA$ fairly well compared with the random guess (0.33).

For the next common entity prediction, we compare the performance using both information from b_A and b_{BinA} as shown in Section 3 with a method computing d_{bdiff} by d_A or d_{BinA} only. In d_A only method, p^m will be set to 1 and the Jensen-Shannon divergence is set to 0 for all attributes. In d_{BinA}, p^m will be 0. We also report the same metrics as the belief dynamics. However, since the common belief label is pretty sparse (selecting one/two entities from over 20 entities), all models get similar results. Therefore, we further treat it as a ranking task and use MRR (Mean Reciprocal Rank) to measure how well the target entity can be returned among all available entities. As shown in the first three columns of Table 4, combining b_A and b_{BinA} can achieve higher ranking score, which suggests that people consider both their self belief and the belief estimation of others when choosing the next entity to align with.

Next Response Generation We evaluate the response generation by both speaking act prediction accuracy and commonly used textual generation metrics (BLEU, METEOR, and ROUGE (Papineni et al., 2002; Lin, 2004; Lavie & Agarwal, 2007)). For the textual evaluation metrics, we align the generated texts with the ground truth utterances only when both speaking acts are "continue to talk". If the ground truth action is "continue to talk" while the predicted action is "end the current turn", the score will be set to 0. To study the contributions of our mind module, we compare the performance of our full model with generations not copying from common belief distribution ($p_{com} = 0$ in Equation (5)).

From the results shown in Table 2, we can see that generators combining with the external mind module achieve better performances for all three categories. Without considering the mind prediction, the performances drop for both response generation and speaking act prediction. This indicates that reasoning about the belief dynamics can help the model resemble human responses when aligning common ground meanwhile form human-like speaking flows.

Case Study Figure 3 demonstrates how the response differs with and without mind/first/second-order belief modeling in Transformer category. The example on the left shows that the model without belief modeling cannot effectively pull the conversation towards the mutual friend based on the context history. Model b_A and b_{BinA} respond with entities in their corresponding order of beliefs but do not address the false belief of B. For the right example, though all models can capture the correct knowledge and generate reasonable responses, only b_A+b_{BinA} provides additional information of other unknown attributes.

4.3. Ablation study

To assess the contributions of each component in the proposed method, we derive the following variants as an ablation study:

- *No dynamics*: We predict the current beliefs b_A and b_{BinA} directly from the given K turns of dialogues instead of summing over belief dynamics across turns in contexts.
- b_A *only* / b_{BinA} *only*: The common belief distribution is learned only based on the first or second-order beliefs (the same in Section 4.2).
- $p_{comm} = X$: We apply different values of $X \in [0, 1]$ for the copy weight p_{com} of the common belief distribution.

Will the belief prediction with dynamics be more accurate? We contrast belief prediction with and without belief dynamics in the GRU category. We can observe from Table 3 that predictions by summing belief dynamics over turns can reach higher precision and F1. We reckon that modeling the state change of entities per utterance can help better track beliefs in longer contexts.

Does each order of belief benefit the response generation? From the pair-wise comparison between +b_A only and +b_{BinA} of columns 4-5 in Table 4, we can see that +b_A performs better in BART while +b_{BinA} is better in the other two models. And jointly modeling both two levels of beliefs can achieve the best results. This suggests that there is no clear dominance between the first and second-order beliefs. Our full model b_A+b_{BinA} can weight the information of the two beliefs accordingly based on the current given contexts.

A's Knowledge Base

Hobby	School	Name
Roller Skating	Boise State Univ.	Randy
Mahjong	Mount Holyoke	Lawrence
Mahjong	Claremount	Julia

B's Knowledge Base

Hobby	School	Name
Mahjong	Mount Holyoke	Lawrence
3D Printing	UIUC	Patricia
3D Printing	Rider Univ.	Eric

Context

B: Hello
B: I know 2 people who are into 3D printing
A: Have Julia and Randy
A: No 3D printing

B's next Response

GT: I don't know any Julia or Randy. Do you know a Lawrence or Eric?
w/o mind: I have a friend into Mahjong
b_A only: I have a friend named Lawrence
b_{BinA} only: I have a friend named Randy
$b_A + b_{BinA}$: I don't know a Julia

A's Knowledge Base

Hobby	Major	Location
Astronomy	Veterinary Sciences	Indoor
Astronomy	Vocational Education	Indoor

B's Knowledge Base

Hobby	Major	Location
Knapping	Physical Education	Outdoor
Astronomy	Veterinary Sciences	Indoor

Context

A: Hey
B: Hi do you know anyone in Astronomy?

A's next Response

GT: 2 of my friends
w/o mind: I have a friend that went to Astronomy.
b_A only: I have 2 that work at Astronomy
b_{BinA} only: I have a friend that work at Astronomy
$b_A + b_{BinA}$: I have two that at indoor

Figure 3: **Responses generated by model variants**. For each case, the current contexts are shown in the top-right; bottom-right visualizes the ground truth responses and the responses generated by models without mind (w/o mind), with common belief estimation using the first-order belief (b_A only), the second-order belief (b_{BinA} only) and with both the first and second-order beliefs (b_A + b_{BinA}).

Table 3: **Belief estimation with/without dynamics prediction for each turn**. w/o Δb represents models trained directly with b_A and b_{BinA}. w/ Δb denotes models trained with belief dynamics. The belief estimation sums over dynamics over all turns in the context.

Models	Belief	Precision	F1
w/o Δb	b_A	45.00±0.00	46.00±0.00
	b_{BinA}	45.33±0.01	46.33±0.01
w/ Δb	b_A	53.67±0.01	56.00±0.01
	b_{BinA}	55.00±0.00	57.00±0.01

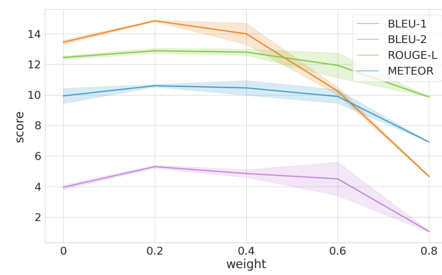


Figure 4: The textual generation score with different copy weights of p_{com}.

Table 4: **Common belief prediction, next response generation and self-talk simulation by ablating the first and second-order belief**. Rows with model+b_A specify results with common belief predicted by the first-order belief only, with model+b_{BinA} by the second-order belief only and with +b_A+b_{BinA} by both the first and second-order beliefs.

Models	F1	MRR	ROUGE-L	Action acc	Success rate	# of Turn	# of Entity
GRU+b_A	60.00±0.00	26.75±0.31	8.89±1.12	77.54±0.66	6.58±1.25	37.74±0.13	12.80±2.90
+b_{BinA}	60.00±0.00	26.65±0.51	9.88±0.26	77.63±0.55	6.33±1.01	37.77±0.25	12.78±2.86
+b_A+b_{BinA}	60.00±0.00	26.87±0.43	9.89±0.27	77.64±0.54	6.92±1.60	37.71±0.17	12.78±2.94
Transformer+b_A	58.67±0.01	24.70±1.74	12.63±0.09	77.43±0.04	9.99±4.82	19.20±0.94	12.23±2.72
+b_{BinA}	59.00±0.01	25.12±1.77	12.75±0.55	77.77±0.11	8.83±2.18	17.79±1.28	10.55±0.99
+b_A+b_{BinA}	58.67±0.01	25.17±2.06	13.14±0.23	77.66±0.42	10.76±4.55	17.65±0.29	10.91±2.73
BART+b_A	60.00±0.00	26.46±0.71	14.41±1.29	76.84±0.55	5.77±0.04	31.77±0.21	15.31±2.84
+b_{BinA}	59.67±0.01	25.93±1.03	12.83±1.43	76.25±1.58	7.37±3.00	28.93±1.84	13.71±4.92
+b_A+b_{BinA}	60.00±0.00	26.51±0.65	14.10±1.10	76.90±0.69	7.65±2.60	30.68±4.57	14.87±4.52

How much does the mind modeling contribute to response generation? We investigate the effects of changing the copy weight of the common belief distribution when generating the next response. The line plot of Figure 4 gives an illustration of the variation of the performance when p_{com} in $\{0, 0.2, 0.4, 0.6, 0.8\}$. We observe that the generation score first increases and then decreases as the weights vary. This indicates that the model not only needs to learn what should be mentioned next but also how to express it in the response. It is supposed to trade-off between entity name prediction accuracy and grammar fluency.

Common ground alignment We further use self-talk to evaluate how well the agents align their common ground with and without modeling the gap between their beliefs and the belief of others. For each round, the model will play as both the speaker and the listener to generate the next response. The game will end when the two agents exceed the maximum of turns, or both of the agents decide to make the choice. When the model predicts the next action is "take task-related action", we will choose the friend with

ChatGPT1's Knowledge Base

School	Company	Location
Pittsburg State	American Motors Corporation	Outdoor
West Virginia	Apache Software Foundation	Indoor
Lake Superior State	L&L Hawaiian Barbecue	Outdoor

ChatGPT2's Knowledge Base

School	Company	Location
Pittsburg State	American Motors Corporation	Outdoor
West Virginia	L&L Hawaiian Barbecue	Indoor
Lake Superior State	Pelican Products	Outdoor

(b_1: American Motors Corporation; $b_{2\ in\ 1}$: None) Do you have a friend who works for the Apache Software Foundation?

Note: ChatGPT1 fails to mention entities w.r.t b_1 .

(b_2: None; $b_{1\ in\ 2}$: None) I don't have anyone in my friend's list working for the Apache Software Foundation. Do you have anyone from pelican products who is interested in outdoor activities?

(b_1: L&L Hawaiian Barbecue, outdoor; $b_{2\ in\ 1}$: L&L Hawaiian Barbecue, outdoor) Sorry, I don't have anyone in my friend's list working for Pelican Products but I can see that one of my friends likes L&L Hawaiian Barbecue and has interest in outdoor activities.

(b_2: L&L Hawaiian Barbecue, outdoor; $b_{1\ in\ 2}$: L&L Hawaiian Barbecue, outdoor) That's a great lead, thank you. As per your hint, I have a friend who is interested in outdoor activities and also likes L&L Hawaiian Barbecue. He could be the mutual friend we are looking for.

Note: ChatGPT2 provides incorrect information w.r.t its KB.

(b_1: Lake Superior, L&L Hawaiian Barbecue, outdoor; $b_{2\ in\ 1}$: L&L Hawaiian Barbecue, outdoor) That's awesome! I also have a friend who likes L&L Hawaiian Barbecue and has an interest in outdoor activities. Let me check his profile quickly to see if he works at Pelican Products.

Note: "Pelican Products" has been checked in turn 2.

Figure 5: **Self-talk simulation using ChatGPT**. The left table contains the knowledge bases of two ChatGPT agents. The right box shows the conversation flow. We additionally prompt ChatGPT to obtain their belief estimation as shown in parenthesis.

the maximum probability based on b_A as its choice of friend. Since models without a mind cannot make this choice, we only compare models with b_A or b_{BinA} only and our full model. We report the average success rate when the agents choose the same person in their knowledge bases as their mutual friend at the end of each game round. Meanwhile, we also calculate the average number of turns they use to end the game and the average number of entities they mention when they reach the common ground.

The right three columns of Table 4 show the self-talk results. Compared with b_A and b_{BinA} only, our full model can achieve higher success while take less turns in the game and mention fewer number of entities to achieve the alignment. This indicates that attending both the first and second-order beliefs helps the model align the common ground more efficiently.

4.4. How well do the large language models (LLMs) perform this task?

LLMs have gained great attention for their impressive conversational abilities. Instead of directly asking ChatGPT questions regarding beliefs and false beliefs like the Sally-Anne test (Kosinski, 2023; Ullman, 2023; Sileo & Lernould, 2023), we investigate its capability of solving cooperative communication tasks involving intense belief exchanges. As shown in Figure 5, we initialize two ChatGPT models as two chat agents. Instructions are given to both of them about the goal and rules of the game. Additional prompts are given to help circulate the conversation meanwhile probe their belief estimation and action prediction. As marked in Figure 5, we observe several potential questions of current LLMs: 1) **Inappropriate belief estimation:** when ChatGPT1 proposes friends interested in L&L Hawaiian Barbecue and outdoor activities, ChatGPT2's first-order belief over possible entities is also outdoor which is not consistent with his knowledge base; 2) **Mind inconsistent utterance**: ChatGPT1 is asking Apache Software Foundation while she believes more in people working for American Motors Corporation as the mutual friend; 3) **No belief tracking:** when the two agents confirm with L&L Hawaiian barbecue and outdoor, ChatGPT1 returns back to Pelican Products which it already negated before.

5. Conclusion

In this study, we present MindDial, a novel framework for generating human-like dialogues. Our approach incorporates an external mind module, which predicts the first and second-order beliefs of the speaker. The response generation takes into account a third-level common belief, which is determined based on the disparity between the first two levels. Through extensive experiments, we demonstrate that responses that consider belief estimation can enhance common ground negotiation between agents. Our ablation studies further validates the effectiveness of our design in capturing belief dynamics and modeling common beliefs by aggregating the first and second-order beliefs.

Acknowledgement

The authors would like to thank Prof. Yixin Zhu (Peking University) for helpful discussions, Miss Chen Zhen (BIGAI) for designing the introductory figure. This work herein is supported by National Key R&D Program of China (2022ZD0114900).

References

Anderson, C. J. Tell me everything you know: a conversation update system for the rational speech acts framework. In *Proceedings of the Society for Computation in Linguistics 2021*, pp. 244–253, 2021.

Bai, J., Yang, Z., Liang, X., Wang, W., and Li, Z. Learning to copy coherent knowledge for response generation. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pp. 12535–12543, 2021.

Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Drain, D., Fort, S., Ganguli, D., Henighan, T., et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

Baker, C., Saxe, R., and Tenenbaum, J. Bayesian theory of mind: Modeling joint belief-desire attribution. In *Annual Meeting of the Cognitive Science Society (CogSci)*, volume 33, 2011.

Bara, C.-P., CH-Wang, S., and Chai, J. MindCraft: Theory of mind modeling for situated dialogue in collaborative tasks. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 1112–1125, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.85. URL https://aclanthology.org/2021.emnlp-main.85.

Baron-Cohen, S., Leslie, A. M., and Frith, U. Does the autistic child have a "theory of mind"? *Cognition*, 21(1): 37–46, 1985.

Bohn, M., Tessler, M. H., and Frank, M. C. Integrating common ground and informativeness in pragmatic word learning. 2019.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:1877–1901, 2020.

Burleson, B. R. Constructivism: A general theory of communication skill. *Explaining communication: Contemporary theories and exemplars*, pp. 105–128, 2007.

Camerer, C. F., Ho, T.-H., and Chong, J.-K. A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3):861–898, 2004.

Cho, I., Wang, D., Takahashi, R., and Saito, H. A personalized dialogue generator with implicit user persona detection. In *Proceedings of the 29th International Conference on Computational Linguistics*, pp. 367–377, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics. URL https://aclanthology.org/2022.coling-1.29.

Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. On the properties of neural machine translation: Encoder–decoder approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pp. 103–111, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/W14-4012. URL https://aclanthology.org/W14-4012.

De Weerd, H., Verbrugge, R., and Verheij, B. How much does it help to know what she knows you know? an agent-based simulation study. *Artificial Intelligence*, 199:67–92, 2013.

Delia, J. G. and O'Keefe, B. J. The constructivist approach to communication. In *Human communication theory: Comparative essays*, pp. 147–191. Harper and Row, 1982.

Doshi, P., Qu, X., Goodie, A., and Young, D. Modeling recursive reasoning by humans using empirically informed interactive pomdps. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 1223–1230, 2010.

Eysenbach, B., Vondrick, C., and Torralba, A. Who is mistaken? *arXiv preprint arXiv:1612.01175*, 2016.

Fan, L., Qiu, S., Zheng, Z., Gao, T., Zhu, S.-C., and Zhu, Y. Learning triadic belief dynamics in nonverbal communication from videos. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7312–7321, 2021.

Glaese, A., McAleese, N., Trębacz, M., Aslanides, J., Firoiu, V., Ewalds, T., Rauh, M., Weidinger, L., Chadwick, M., Thacker, P., et al. Improving alignment of dialogue agents via targeted human judgements. *arXiv preprint arXiv:2209.14375*, 2022.

Haber, J., Baumgärtner, T., Takmaz, E., Gelderloos, L., Bruni, E., and Fernández, R. The PhotoBook dataset: Building common ground through visually-grounded dialogue. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1895–1910, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1184. URL https://aclanthology.org/P19-1184.

Han, Y. and Gmytrasiewicz, P. Learning others' intentional models in multi-agent settings using interactive pomdps. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018.

He, H., Balakrishnan, A., Eric, M., and Liang, P. 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)*, pp. 1766–1776, Vancouver, Canada, July 2017a. Association for Computational Linguistics. doi: 10.18653/v1/P17-1162. URL https://aclanthology.org/P17-1162.

He, H., Balakrishnan, A., Eric, M., and Liang, P. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017b.

Ho, M. K., MacGlashan, J., Greenwald, A., Littman, M. L., Hilliard, E., Trimbach, C., Brawner, S., Tenenbaum, J., Kleiman-Weiner, M., and Austerweil, J. L. Feature-based joint planning and norm learning in collaborative games. In *Annual Meeting of the Cognitive Science Society (CogSci)*, 2016.

Hu, Z., Cao, Z., Chan, H. P., Liu, J., Xiao, X., Su, J., and Wu, H. Controllable dialogue generation with disentangled multi-grained style specification and attribute consistency reward. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:188–199, 2022.

Jang, Y., Lee, J., and Kim, K.-E. Gpt-critic: Offline reinforcement learning for end-to-end task-oriented dialogue systems. In *International Conference on Learning Representations (ICLR)*, 2022.

Kim, J.-H., Kitaev, N., Chen, X., Rohrbach, M., Zhang, B.-T., Tian, Y., Batra, D., and Parikh, D. CoDraw: Collaborative drawing as a testbed for grounded goal-driven communication. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 6495–6513, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1651. URL https://aclanthology.org/P19-1651.

Kleiman-Weiner, M., Ho, M. K., Austerweil, J. L., Littman, M. L., and Tenenbaum, J. B. Coordinate to cooperate or compete: abstract goals and joint intentions in social interaction. In *Annual Meeting of the Cognitive Science Society (CogSci)*, 2016.

Kosinski, M. Theory of mind may have spontaneously emerged in large language models. *arXiv preprint arXiv:2302.02083*, 2023.

Kovács, Á. M., Téglás, E., and Endress, A. D. The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012):1830–1834, 2010.

Lavie, A. and Agarwal, A. METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pp. 228–231, Prague, Czech Republic, June 2007. Association for Computational Linguistics. URL https://aclanthology.org/W07-0734.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*, 2019.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7871–7880, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.703. URL https://aclanthology.org/2020.acl-main.703.

Lin, C.-Y. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL https://aclanthology.org/W04-1013.

Lin, J. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.

Liu, Y., Maier, W., Minker, W., and Ultes, S. Empathetic dialogue generation with pre-trained roberta-gpt2 and external knowledge. In *Conversational AI for Natural Human-Centric Interaction: 12th International Workshop on Spoken Dialogue System Technology, IWSDS 2021, Singapore*, pp. 67–81. Springer, 2022.

Milliez, G., Warnier, M., Clodic, A., and Alami, R. A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. In *The 23rd IEEE international symposium on* *robot and human interactive communication*, pp. 1103–1109. IEEE, 2014.

Nematzadeh, A., Burns, K., Grant, E., Gopnik, A., and Griffiths, T. Evaluating theory of mind in question answering. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2392–2400, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1261. URL https://aclanthology.org/D18-1261.

OpenAI. Gpt-4 technical report, 2023.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems (NeurIPS)*, 35:27730–27744, 2022.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics. doi: 10.3115/1073083.1073135. URL https://aclanthology.org/P02-1040.

Qiu, L., Zhao, Y., Liang, Y., Lu, P., Shi, W., Yu, Z., and Zhu, S.-C. Towards socially intelligent agents with mental state transition and human value. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 146–158, Edinburgh, UK, September 2022. Association for Computational Linguistics. URL https://aclanthology.org/2022.sigdial-1.16.

Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S. A., and Botvinick, M. Machine theory of mind. In *International Conference on Machine Learning (ICML)*, pp. 4218–4227. PMLR, 2018.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020. URL http://jmlr.org/papers/v21/20-074.html.

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., and Saxe, R. Development of the social brain from age three to twelve years. *Nature communications*, 9(1):1027, 2018.

Sileo, D. and Lernould, A. Mindgames: Targeting theory of mind in large language models with dynamic epistemic modal logic. *arXiv preprint arXiv:2305.03353*, 2023.

Udagawa, T. and Aizawa, A. Maintaining common ground in dynamic environments. *Transactions of the Association for Computational Linguistics*, 9:995–1011, 2021. doi: 10.1162/tacl_a_00409. URL https://aclanthology.org/2021.tacl-1.59.

Ullman, T. Large language models fail on trivial alterations to theory-of-mind tasks. *arXiv preprint arXiv:2302.08399*, 2023.

Varshney, D., Prabhakar, A., and Ekbal, A. Commonsense and named entity aware knowledge grounded dialogue generation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1322–1335, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.95. URL https://aclanthology.org/2022.naacl-main.95.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 2017.

Verma, S., Fu, J., Yang, S., and Levine, S. CHAI: A CHatbot AI for task-oriented dialogue with offline reinforcement learning. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4471–4491, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.332. URL https://aclanthology.org/2022.naacl-main.332.

Wunder, M., Kaisers, M., Yaros, J. R., and Littman, M. L. Using iterated reasoning to predict opponent strategies. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pp. 593–600, 2011.

Yuan, L., Gao, X., Zheng, Z., Edmonds, M., Wu, Y. N., Rossano, F., Lu, H., Zhu, Y., and Zhu, S.-C. In situ bidirectional human-robot value alignment. *Science Robotics*, 7(68):eabm4183, 2022. doi: 10.1126/scirobotics.abm4183. URL https://www.science.org/doi/abs/10.1126/scirobotics.abm4183.

Yuan, T., Liu, H., Fan, L., Zheng, Z., Gao, T., Zhu, Y., and Zhu, S.-C. Joint inference of states, robot knowledge, and human (false-) beliefs. In *International Conference on Robotics and Automation (ICRA)*, pp. 5972–5978. IEEE, 2020.

Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Gao, J., Liu, J., and Dolan, B. DIALOGPT : Large-scale generative pre-training for conversational response generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 270–278, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-demos.30. URL https://aclanthology.org/2020.acl-demos.30.