Social Influence Dialogue Systems: A Scoping Survey of the Efforts Towards Influence Capabilities of Dialogue Systems

Kushal Chawla∗ Weiyan Shi2∗ Jingwen Zhang3∗∗
Gale Lucas1∗∗ Zhou Yu2∗∗ Jonathan Gratch1∗∗

1University of Southern California  2Columbia University
3University of California Davis

1{chawla, lucas, gratch}@ict.usc.edu  2{ws2634, zy2461}@columbia.edu
3jwzzhang@ucdavis.edu

Abstract

Dialogue systems capable of social influence such as persuasion, negotiation, and therapy, are essential for extending the use of technology to numerous realistic scenarios. However, existing research primarily focuses on either task-oriented or open-domain scenarios, a categorization that has been inadequate for capturing influence skills systematically. There exists no formal definition or category for dialogue systems with these skills and data-driven efforts in this direction are highly limited. In this work, we formally define and introduce the category of social influence dialogue systems that influence users’ cognitive and emotional responses, leading to changes in thoughts, opinions, and behaviors through natural conversations. We present a survey of various tasks, datasets, and methods, compiling the progress across seven diverse domains. We discuss the commonalities and differences between the examined systems, identify limitations, and recommend future directions. This study serves as a comprehensive reference for social influence dialogue systems to inspire more dedicated research and discussion in this emerging area.

1 Introduction

Dialogue research has traditionally focused on either task-oriented systems, that aim for task completion, or open-domain systems, that target social companionship. Besides these two broad categories, efforts have also been made for conversational systems that revolve around another core function in human communication, that is, social influence (Perloff, 1993; Cialdini and Goldstein, 2004; Cialdini, 2009; Dillard and Wilson, 2014). Such systems attempt to influence the users’ behaviors, feelings, thoughts, or opinions, such as persuasive dialogue agents (Schulman and Bickmore, 2009), negotiation chatbots (Kraus, 1997), and recommendation assistants (Miao et al., 2007). Developing these systems holds importance in AI research for multiple reasons. Social influence is at the core of human communication in everyday scenarios, from games (Peskov et al., 2020) to political or marketing campaigns (Connolly, 2002), and from social platforms (Tan et al., 2016) to therapeutic interactions (Tanana et al., 2016). Therefore, tackling these tasks not only involves AI but also aspects of game theory, communication, linguistics, and social psychology, making them an ideal testbed for interdisciplinary AI research. Most importantly, they reflect AI’s general ability to consider their partners’ inputs, tailor the communication strategies, personalize the responses, and lead the conversation actively.

Despite existing efforts in identifying and analyzing various social influence scenarios, data-driven efforts for dialogue systems in this space are limited. Further, there exists no formal definition or category that appropriately unifies these efforts across diverse domains. Hence, in this work, we introduce the concept of social influence dialogue systems that encompass a broad range of domains but share the unique goal of achieving social influence. This distinguishes them from purely task-oriented or open-domain interactions, that target task completion and social companionship.

Various social influence dialogue tasks have been proposed throughout the years. We categorize them by their application: games (Peskov et al., 2020), multi-issue bargaining (Lewis et al., 2017), social good (Wang et al., 2019), e-commerce (He et al., 2018), therapy and support (Tanana et al., 2016), argumentation (Thomas et al., 2006), conversational recommendations (Dodge et al., 2016), and miscellaneous tasks (Tang et al., 2019). We further organize the developed methods along four axes: strategy representation, language realization, partner modeling, and training paradigms.

Our findings reveal four foundational challenges
that must be addressed by future work. First, realistic social influence settings can be complex, involving diverse skills in the same interaction, making it essential to unify the existing research into a common framework. Second, it remains unclear how to effectively incorporate fundamentals from social influence theories (Cialdini, 2009; Lewicki et al., 2016) in dialogue models and how to realize them in language. Third, accurate evaluation of social influence requires a comprehensive protocol measuring not only the system’s linguistic sophistication and task performance but also the subjective perception of the human partners. And fourth, social influence occurs through all viable modalities, which calls for an integration of dialogue research with other means of communication.

Our contributions are three-fold. First, we formally define the concept of social influence dialogue systems and discuss key measures of success (Section 2). Second, we summarize the available corpora (Section 3) and categorize existing approaches in this area (Section 4). And third, we identify key limitations in data collection and modeling practices, providing recommendations for future work (Section 5). Over the years, research in task-oriented and open-domain dialogues has benefited from a myriad of survey efforts (Huang et al., 2020; Zhang et al., 2020c; Ni et al., 2021). We instead focus on dialogue systems with social influence capabilities and present a thorough review, unifying efforts across various domains. We sincerely hope that this work serves as a timely entry point for interested researchers to take this area further, inspiring dedicated effort and discussion on social influence in the dialogue community.

2 Social Influence Dialogue Systems

Quoting Robert Gass (2015), “Social influence is a fact of everyday life”. Formally, it is the change in an individual’s thoughts, feelings, attitudes, or behaviors that results from interaction with another individual or a group (Rashotte, 2007). AI systems have the ability to act interactively and thus, potentially influence their partners in decision-making and behavioral contexts (Zhang et al., 2020a; Lee et al., 2020). Social influence dialogue systems influence their partners through natural conversations. This calls for an active role by the system in the conversation, distinguishing them from other well-studied scenarios, such as purely task-oriented, where systems passively assist their partners to complete tasks, and open-domain, that target social companionship. Key social influence tasks include persuasion (Wang et al., 2019), aiming to change users’ attitudes or behaviors, and negotiation, aiming to change the users’ perspective to achieve a common ground (Lewis et al., 2017). We provide more examples in Section 3. The conceptual overview in Figure 1 distinguishes between the kinds of conversational content in social influence interactions. The task-oriented content focuses on the task of influencing for a domain-specific goal, like persuading for donation, bargaining with trade-offs, or encouraging healthier eating habits. These interactions may also contain optional social content, such as small talk, empathy, or self-disclosure. Task-oriented content relates to the system’s goal, and provides a context for social interactions. Depending on the task, social content is optional, but if present, can in turn build rapport and enhance user-system relationship for improved task outcomes (Liao et al., 2021).

We do not see social influence dialogue systems as exclusive to other more familiar categories in dialogue research. Simply put, if a scenario involves the system’s goal to influence its partner, we consider it under social influence. For instance, He et al. (2018) studied the task of buyer-seller price negotiations. The task of the buyer is to negotiate for a reasonable price (arguably making it task-oriented), but achieving it requires social influence skills of engaging in trade-offs and building a rapport with the seller so as to reach an agreement.

Measures of Success: The above discussion makes it clear that a comprehensive evaluation of social influence dialogue systems must draw inspiration from both task-oriented and open-domain dialogue research. Since there already exist extensive surveys that discuss the evaluation in these
settings (Deriu et al., 2021; Li et al., 2021), we will not cover them here in detail. However, we will define three essential axes for evaluation: 1) **Linguistic Performance**, focusing on the system’s linguistic sophistication based on automatic metrics (e.g. perplexity, BLEU) and human evaluation (e.g. fluency, consistency, coherency). 2) **Influence Outcome**, covering the ability to influence, as defined by objective task-specific goals like the negotiated price or weight loss after interacting with the system. 3) **Partner Perception**, covering the subjective evaluation of the user, for instance, the person’s satisfaction, likeness towards the system, and interest in interacting with the same system again. In a buyer-seller negotiation, if the seller hates the buyer in the end, no matter how favorable the deal is for the buyer, one might argue that this is still a failed negotiation for the buyer. Hence, as recent work also argued (Chawla et al., 2021b), we encourage future work to take all three dimensions into account collectively.

3 Social Influence Across Diverse Application Areas

We now illustrate social influence across numerous domains and application areas. Alongside, we briefly discuss the available datasets, providing a clearer distinction between tasks that constitute social influence and those that don’t. The literature surveyed in this work was compiled from various sources, including conference proceedings, web repositories, and search engines. We refer the readers to Appendix A for more details about this compilation process, which ultimately led us to 22 datasets, spanning 12 publication venues, 4 languages, and 7 application domains. In general, one sample in these datasets involves the non-conversational context for the participants (e.g. negotiation preferences or other role-specific information), the conversation between them, and outcome assessment. Optionally, some datasets also gather participant demographics and personality traits, utterance-level annotations, and subjective evaluations via post-surveys. We provide a list of the datasets in Appendix B, along with key statistics, and information about the available metadata that can be useful for developing more practical and personalized systems. We categorize these datasets based on their domains and discuss them below.

**Games**: Multiplayer strategy games provide a remarkable platform to study various aspects of social influence like dynamics of trust and deception, based on complex interactions between linguistic and non-linguistic game contexts. For instance, players in Diplomacy roleplay as European powers while attempting to forge and break alliances to dominate the map. This involves extensive tactics in a complex environment that captures deception in long-lasting relationships (Peskov et al., 2020). The game of Catan (Asher et al., 2016; Boritchev and Amblard, 2021) revolves around the trade of resources (ore, wood, wheat, clay, and sheep) for the acquisition of roads, settlements, and cities. Often, the players have access to only a subset of resources that they need, which naturally encourages strategic influential conversations and trade.

**Multi-Issue Bargaining Tasks (MIBT)**: Numerous other efforts rely on a tractable closed-domain abstraction from the negotiation literature, known as the Multi-Issue Bargaining Task, or MIBT (Fershtman, 1990). MIBT negotiations involve a fixed set of issues each with a predefined priority for each player, which governs what the players want out of the negotiation. MIBT can lead to diverse mixed-motive interactions, depending on how these priorities of the players are aligned. If they match (known as a distributive scenario), this leads to more competitive negotiations, where each party attempts to convince their partner with trade-offs and persuasive arguments. If the priorities do not match (known as an integrative scenario), this paves the way for more cooperative interactions where the negotiators try to find optimal divisions that benefit everyone. MIBT has been popular in the negotiation literature and the industry, and thus, has been adopted by recent dialogue datasets as well. DealOrNoDeal (Lewis et al., 2017) involves negotiations over three arbitrarily-defined issues: books, balls, and hats. This provides a controlled environment to study trade-offs but lacks the semantic context necessary for persuasive arguments. Targeting a more realistic scenario, CaSiNo (Chawla et al., 2021b) involves two participants who role-play as campsite neighbors to negotiate over food, water, and firewood. Along with the trade-offs, the authors found the usage of multiple persuasive strategies in the dialogues. The JobInterview dataset (Yamaguchi et al., 2021) instead focuses on a more complex setup with a greater number of possible negotiation outcomes. The dataset involves recruiter-applicant interactions over salary, day off, position, company, and workplace.
Social Good: Social influence is critical for systems that aim to change people’s opinions for social good. For effectiveness, the system must be personalized based on knowledge that is both relevant and appealing. PersuasionForGood (Wang et al., 2019) involves two participants, where one attempts to persuade the other to donate to a charity. Unlike negotiations where both players actively engage in social influence, this design is more asymmetric, where the persuader tries to convince the persuadee to donate to a social cause using a variety of influence tactics. For instance, Logical Appeal involves providing reason and evidence to support the argument, while Emotional Appeal involves the elicitation of emotion. In total, the authors annotated 7 influence strategies that the persuaders use at different stages in the conversation.

E-commerce: Participant goals in E-commerce are usually conflicting, and are governed by their predefined roles. For example, a buyer influences the seller to agree for a reasonable price, while the seller tries to maximize their own profit. The influence tactics rely on back-and-forth offers, often associated with strategies such as side offers and appealing to sympathy. An effective system in this space must be able to combine price-related reasoning along with language realization. CraigslistBargain (He et al., 2018) involves buyer-seller price negotiations over products listed on Craigslist. The interactions in this case are more open-ended as compared to the MIBT setup with a fixed set of issues. The dialogues are rich with diverse influence strategies such as embellishments, side offers, emotional appeals, and leveraging product-specific world knowledge. On the other hand, AntiScam (Li et al., 2020) focuses on a customer service scenario where users try to defend themselves against attackers who try to steal sensitive personal information with convincing arguments.

Therapy & Support: Being able to influence is fundamental for providing effective therapy and counseling, which aids in the treatment of mental disorders, and substance use disorders, along with changing undesirable behaviors like unhealthy diets. A counselor needs to be adaptive, personalized, should understand the core issues, and should facilitate a change in perspective of the patient, which has been linked to a higher likelihood of success (Althoff et al., 2016). In SMS counseling, Althoff et al. (2016) found that linguistic influence like pushing the conversation in the desired direction is associated with this desired perspective change. Similar scenarios were captured in other datasets as well (Demasi et al., 2019; Liang et al., 2021). Tanana et al. (2016) collected the Motivational Interviewing (MI) dataset from psychotherapy logs. MI is an evidence-based approach for behavior change where the goal is to elicit and explore the patient’s own motivations for change. Lastly, EmpatheticDialogues (Rashkin et al., 2019) contains support conversations where the listener provides support by expressing empathy, which has been associated with rapport and better task outcomes in many domains (Kim et al., 2004; Norfolk et al., 2007; Fraser et al., 2018).

Argumentation: Scenarios such as debates or court hearings involve interaction between parties that hold opposing views and use persuasive arguments (often based on factual information) to convince the other party and the audience. Multiple factors govern how persuasive an argument will be. In addition to factuality and social proof, how the argument is presented in terms of intensity, valence, authoritativeness, and framing can be crucial as well (Chaiken, 1987; Althoff et al., 2014). Tan et al. (2016) released a dataset of ChangeMyView logs from Reddit, involving discussions on numerous controversial topics. Other datasets include DDO Debates on diverse topics (Durmus and Cardie, 2019), congressional proceedings (Thomas et al., 2006), and court hearings (Fornaciari and Poesio, 2012; D.-N.-M. et al., 2012; Ji et al., 2020).

Conversational Recommendation: Conversational recommender systems provide contextual recommendations through dialogue. These scenarios are ubiquitous such as a discussion with a friend about a movie or with a librarian about a book. Naturally, these dialogues hold potential for influence through recommendations that may be biased towards a specific movie or a book: for instance, a movie fan persuading their friends to watch a movie that they adore, instead of simply providing an opinion. Li et al. (2018) and Dodge et al. (2016) collected conversational recommendation datasets around movies. The conversations in these datasets are not necessarily guided towards a specific movie - the goal is simply to provide recommendations based on facts and personal experiences. Hence, they do not explicitly involve social influence but still provide interesting examples of scenarios that can involve social influence.

Miscellaneous: Finally, we describe the Target-
Guided dataset (Tang et al., 2019) which was constructed from the open-domain PersonaChat corpus (Zhang et al., 2018) where participants engage in natural conversations while being consistent with their predefined personas. Instead of open-ended conversations, in Target-Guided scenario, the system is tasked with a more concrete goal of naturally guiding the conversation to a designated target subject, thereby, making it a social influence setting.

3.1 Discussion

Differences in Task Design: We reviewed how various forms of social influence appear across domains. In some scenarios like negotiations and debates, the tasks are symmetric with multiple active players. In others such as persuasion for social good or strategic recommendations, the design is asymmetric with one active party influencing their partner. Tasks also differ in terms of possibilities for cooperation. Argumentation and scamming scenarios are inherently adversarial involving conflicting goals, but negotiations are generally seen as mixed-motive. Depending on the context, a negotiation can be non-collaborative, where the participants employ selfish strategies to maximize their own rewards but it can also be collaborative, where creative solutions to maximize the joint value can be explored. In realistic contexts, conversations involving social influence can include aspects of both negotiation and persuasion. As previous studies have found, specific behaviors of the participants also show variance across demographics, cultural backgrounds, and individual personalities (Bogaert et al., 2008; Xu et al., 2017; Wang et al., 2019).

Potential for Social Influence: The potential for social influence depends more on the task definition than the domain in consideration. For example, a minor modification to the design inculcates social influence in the Target-Guided dataset (Tang et al., 2019). Restaurant booking, which has been traditionally defined as assistive in nature where the system only passively assists the users, can involve social influence if both parties are instead trying to fit their own schedules for maximizing convenience and profit. In essence, we argue that real-world conversations can be more complex than the typical ways dialogue tasks are defined, and incorporating social influence is crucial for bridging this gap.

Limitations of Existing Datasets: The datasets discussed above have been very successful in fueling interest among the researchers and several recent modeling approaches in this space. However, several limitations exist that can hinder the progress in the future. Most datasets do not contain user attributes like demographics or personality which hinders the research in personalized dialogue systems. Almost all of them focus only on objective outcomes like the points scored or the final agreed price in a negotiation, with only a few that assess partner perceptions. We further observed that crowdsourced datasets provide more structure and metadata for modeling as opposed to crawled datasets. However, as is the challenge with dialogue research in general, the former remain expensive to collect, resulting in relatively smaller dataset sizes. Finally, except for a few cases, most datasets that were surveyed are in English, indicating a lack of datasets that capture diverse cultures and demographics for research in this area.

4 Methodological Progress

We now summarize the modeling approaches developed for social influence dialogue systems. Most domains have seen efforts in analyzing human dialogue behaviors and their impact on task outcomes. Examples include analyzing deception in games (Peskov et al., 2020), the impact of persuasive strategies and dialogue acts on charity donations (Wang et al., 2019), cooperative and non-cooperative strategies in MIBT (Chawla et al., 2021b), the use of emotion expression for predicting partner perceptions (Chawla et al., 2021a), and studying semantic categories of persuasive arguments on web forums (Egawa et al., 2019).

In addition, researchers have targeted various domain-specific subtasks that can be crucial for the eventual development of dialogue systems in this space. This involves research in lie detection methods (Yeh and Ku, 2021; Yu et al., 2015), discourse parsing (Shi and Huang, 2019; Ouyang et al., 2021), strategy prediction (Chawla et al., 2021b; Wang et al., 2019), breakdown detection (Yamaguchi et al., 2021), outcome prediction (Sinha and Dasgupta, 2021; Chawla et al., 2020; Dutt et al., 2020), and argument mining (Dutta et al., 2022).

Research that directly targets the development of dialogue systems in this space is still nascent. Among other challenges like limited cross-cultural diversity and relatively smaller dataset size, social influence dialogue settings pose a unique challenge: an average human often exhibits sub-optimal strategic behaviors in social influence tasks (Wunderle,
Figure 2: A categorization for modeling efforts in social influence dialogue systems. These categories are not exclusive and many approaches use them in conjunction.

2007; Babcock and Laschever, 2009). This means that standard seq2seq approaches trained on these collected datasets using supervised learning are fundamentally insufficient for developing effective dialogue systems with influence capabilities. Hence, prior work gives a special focus to the system strategy, employing different ways to model the strategy and language together. To study the existing methods systematically, we organize them along four axes: Strategy Representation, Language Realization, Partner Modeling, and Training (Figure 2). We discuss them individually below.

4.1 Strategy Representation

Implicit: The most obvious way to represent the system strategy is implicitly, without any intended decoupling between system strategy and response realization. This corresponds to the usual sequence-to-sequence framework that has been a standard baseline for the methods developed in this space. An important example is the work by Lewis et al. (2017), who were one of the first works to train end-to-end dialogue models that exhibit social influence. The authors employed a neural network based on GRUs, one for encoding the negotiation context, one to encode the dialogue utterances, and two recurrent units to generate the output agreement in a bidirectional manner.

Latent vectors: Yarats and Lewis (2018) explored latent vectors to decouple utterance semantics from its linguistic aspects. Their hierarchical approach first constructs a latent vector from the input message, which is then used for response generation and planning. These latent vectors are trained to maximize the likelihood of future dialogue messages and actions, which enables the decoupling between semantics and realization.

Dialogue Acts (DAs): DAs, such as offer propose, accept, or reject, are effective at capturing a high-level structure of the dialogue flow in social influence settings, reducing the model strategy to first predicting the dialogue act for the next response. This makes it convenient to apply reinforcement learning approaches (Zhang et al., 2020b; Yang et al., 2021) and also helps in developing a modular dialogue system design (He et al., 2018).

Annotated Strategies: The structural properties expressed by DAs are insufficient for capturing semantics like emotion, small talk, and appeal. To better incorporate them, researchers have relied on additional utterance-level annotations grounded in prior theories in social influence contexts (Wang et al., 2019; Chawla et al., 2021b). These strategies have been used in conjunction with DAs (Zhou et al., 2019; Joshi et al., 2020).

4.2 Language Realization

An important aspect of the system design is an effective way to realize the language, that is, to generate the next response so that it portrays the desired strategic behaviors. Borrowing from task-oriented and open-domain research, existing dialogue models for social influence use a variety of methods to generate the final system response.

Templates and retrieval methods: Predefined templates and response retrieval from the training data simplify the generation pipeline, improving controllability and modularity. He et al. (2018) used templates in their generator which are later filled by retrieving similar responses from the data. This allowed the authors to explore supervised and reinforcement learning at the level of DAs for the influence strategy of the system.

Conditional Generation: Text generation methods result in more diverse responses, but negatively impact the controllability and interpretability. Prior work relies on autoregressive text generation conditioned on the dialogue history, non-conversational context, and additional annotations. These are either encoder-decoder networks (Lewis et al., 2017; Li et al., 2020; Joshi et al., 2020) or use a decoder-only design (Li et al., 2020). A useful future direction is to combine generation with retrieval for knowledge-grounded settings like argumentation. Similar methods have been explored for other NLP
tasks like open-domain question answering and question generation (Lewis et al., 2020).

4.3 Partner Modeling

Partner modeling refers to inferring the mental states of the partner based on the conversation. For example, understanding the cause that the persuadee cares about in the PersuasionForGood context, or inferring the priorities of the partner in DealOrNoDeal negotiations. Building an accurate partner model can be crucial in social influence settings for guiding the decision-making of the system (Baarslag et al., 2013; Zhang et al., 2020b). Hence, we discuss various ways in which prior work tackles partner modeling.

Implicit: A majority of the efforts do not explicitly model the behavior of the partner but instead, this behavior implicitly guides the next response of the sequence-to-sequence dialogue system pipeline.

Simulated User: Lewis et al. (2017) trained a simulated user on the available data in a supervised manner. This was then used to further train the dialogue system. Instead of inferring mental states explicitly, this takes a more behavioral approach of estimating the future actions of the partner and using these for training via reinforcement learning.

Dialogue Act Look-Ahead: With a similar idea, Zhang et al. (2020b) proposed OPPA model with a look-ahead based partner modeling strategy at the level of DAs. At each step, OPPA first estimates the user’s future DA, which is then used to select the next DA of the system. The authors found significant improvements on the DealOrNoDeal task. Yang et al. (2021) used a similar method for buyer-seller negotiations.

Taking a different approach, Chawla et al. (2022) trained a ranking model to directly predict the hidden preferences of the partner in a multi-issue negotiation. Instead of predicting future actions, these methods assume that the partner’s behavior can be explained by their context and goals in the dialogue. However, this approach is yet to be used in an end-to-end dialogue system.

4.4 Training

We now discuss how the above components are tied together via various training paradigms.

Architecture Choices: One crucial aspect is the architecture design: End-to-end (Lewis et al., 2017; Radford et al., 2019) vs Modular (He et al., 2018). While end-to-end methods improve the diversity and need less manual effort, a modularized design enhances controllability and explainability. Perhaps, this is why modular methods are popular in large-scale industrial models like the IBM Project Debater1 and the design by Hadfi et al. (2021). Improving the control of desired variables such as topics, strategy, or emotion in the end-to-end methods is an open area of research and is yet to be explored for social influence dialogue systems.

Supervised Learning (SL) and Reinforcement Learning (RL): Zhou et al. (2019) used SL to train a hierarchical encoder-decoder for generating the next response and used Finite State Transducers (FSTs) to encode the historic sequence of DAs and persuasive strategies into the model, showing improvements in negotiation and persuasion tasks. The performance was later improved by Joshi et al. (2020), who replaced FSTs with Graph Neural Networks to better model the interdependencies. Others have relied on RL to explicitly optimize the model on task-specific objective outcomes. While SL trains the model to mimic the average human behavior, RL techniques based on REINFORCE (Williams, 1992) allow the system to explore its own strategies in the wild while being guided by one or more overall reward metrics. Lewis et al. (2017) used RL in negotiations, with the final points scored in the agreed deal as the reward. More recent work employed RL to incorporate simplistic partner models into the decision-making process of the dialogue system, showing improvements in negotiation tasks (Zhang et al., 2020b; Yang et al., 2021).

Multi-tasking and Pretraining: Limited efforts have also explored multi-tasking and pretrained language models for social influence dialogue systems, which provide promising ways to deal with the challenge of insufficient training data. Liu (2021) trained a sequence-to-sequence transformer on a mix of Cornell Movie Dialogue corpus (Danescu-Niculescu-Mizil and Lee, 2011) and psychotherapy data. Li et al. (2020) fine-tuned the GPT model (Radford et al., 2018), while employing multi-tasking to incorporate intents and slots for both the human and the system. Wu et al. (2021) recently introduced ARDM which uses GPT2 (Radford et al., 2019) to separately encode the utterances of the human and the dialogue system, reducing the reliance on additional annotations.

---

1https://research.ibm.com/interactive/project-debater/
4.5 Discussion
Integrating dialogue systems with social influence skills is a challenging and an underdeveloped research direction. Although a variety of techniques have been employed, individually, they have only been explored in a limited manner, requiring a significant amount of dedicated investigation in the future to reach maturity. In order to guide this investigation, we identify some key limitations in prior work. The primary drawback is the limited exploration carried out till now. Existing methods are yet to capture user personality and demographics, have not explored long-term interactions or controllability, and have not considered partner perception in their modeling or evaluations, which is essential for an accurate reflection of model performance. Further, directly applying RL is expected to be insufficient to reliably learn complex influence capabilities, especially in the cases where effective influence strategies are underrepresented in the collected datasets. Integrating theories from psychology, communication, and behavior change literature about human-human social interactions can be one way to tackle this (Section 5).

5 The Path Forward
Building sophisticated social influence dialogue systems remains a challenging research endeavor. To guide future work, we summarize our insights in Appendix C, encouraging a holistic understanding of system attributes, target audience, underlying modeling techniques, and evaluation mechanisms. We lay out some pressing future directions below.

Task unification: Transfer learning approaches that exploit the commonalities among diverse social influence tasks are yet to be explored. Roller et al. (2020) proposed to blend various open-domain tasks and built Blenderbot that aims to address different challenges together (e.g., persona-based, knowledge-enriched, etc.). Future research should explore similar unified approaches for social influence settings as well, given a common conceptual foundation (Figure 1), with similar evaluation and influence principles (Cialdini, 2009).

Task Evaluation: A comprehensive evaluation in these settings is challenging since it must consider the partner perception along with objective task outcomes. Building user simulators (Li et al., 2016; Jain et al., 2018; Shi et al., 2019) could potentially alleviate the problem. Most existing simulators are developed for task-oriented systems which follow certain agenda. Future research should study how to utilize partner modeling to build social influence user simulators for more efficient task evaluation (He et al., 2018; Yang et al., 2020). For instance, we could potentially design different user personalities and simulate the change in user’s beliefs, opinions, and attitudes accordingly (Yang et al., 2021).

Theory integration: There has been a great amount of research effort on building theories for social influence (Cameron, 2009; Giles, 2016; Lewicki et al., 2016; Cialdini and Goldstein, 2004). Instead of solely relying on the collected data, future work should consider leveraging fundamentals from this research to guide the dialogue policy.

Multimodal systems: Social influence occurs through all possible modalities. In fact, Schulman and Bickmore (2009) show that embodied agents achieve better persuasion results than text-only agents. Other studies have recognized the importance of emotion in social influence tasks (Asai et al., 2020; Chawla et al., 2021a). Further, Nguyen et al. (2021) proposed a speech dataset in debates and study the influence of spoken tactics on persuasiveness across genders. Given these findings, we encourage future research to explore social influence across multimodal interactions.

Knowledge-enriched systems: Social influence tasks often involve constantly-changing world knowledge such as organizational facts and recent social news. Oftentimes, the system’s internal state (e.g., the change of task setting from one set of products to a different set) needs to be updated. Retraining the entire system is costly to maintain after the initial development. Recent work has proposed to augment the dialogue system with internet-search ability to generate more factual and updated responses in open-domain dialogues (Komeili et al., 2021). Future efforts in this direction will benefit social influence dialogue systems as well.

6 Conclusions
We introduced the category of social influence dialogue systems that aim to influence their partners through dialogue. We presented a survey of the recent prior work in this space, compiling datasets and methods across diverse application domains. We pointed out key limitations in existing methodologies and proposed promising directions for designing more sophisticated systems in the future. Our survey reveals that although substantial progress has been made, this is still an emerging re-
search area. We hope our work inspires more dedicated interdisciplinary effort and discussion, which is necessary for making progress in this space.

7 Broader Impact and Ethical Considerations

Social influence is ubiquitous in everyday life. Research on how we use influence in all aspects of our lives spans a number of fields, including social psychology, communication, consumer behavior, behavioral change, and behavioral economics. This research has led to crucial findings about the strategies of social influence and how they impact our decision-making. Over the past few decades, research has accumulated and demonstrated the effectiveness of using various strategies across contexts and domains. Prominent examples include core principles of social influence by Cialdini from social psychology: reciprocity, commitment and consistency, social proof, liking and attractiveness, authority, and scarcity (Cialdini, 2009). Further, communication strategies used in persuasion and general social influence contexts include credibility appeals, two-sided argumentation, emotional tactics, and appeals to social norms, among others (Cameron, 2009; O’keefe, 2015).

First, the well-studied principles in social influence research can guide the development of effective dialogue systems with influence capabilities. In fact, many of the strategies found in the datasets developed for social influence tasks (Section 3) directly map to the principles laid out by Cialdini, for instance, credibility and emotional appeal in PersuasionForGood dataset (Wang et al., 2019) and reciprocity observed in CaSiNo negotiation dataset (Chawla et al., 2021b). Second, research in social influence dialogue systems provides novel datasets on human-human and human-machine communication, and therefore, holds a great potential to advance theories of human cognition and influence processes (Gratch et al., 2015). The datasets and subsequent analyses can further contribute new theoretical insights to social influence research.

Although dialogue systems have already been used in a number of applications involving chatbots and AI assistants, advancements in social influence dialogue systems can help to bridge the gap between our existing task definitions and a number of other real-world applications. For instance, realistic customer support interactions often involve active behaviors from both the support agent and the user where the agent uses social cues for improved customer satisfaction and retention, while the user attempts to address their queries. These settings naturally involve aspects of social influence, unlike traditional task-oriented definitions where the dialogue system plays a passive role to assist the human users. As discussed earlier, social influence dialogue systems can positively help to advance other areas as well. In therapy domain, these systems can assist in various psychological treatments such as by increasing the willingness to disclose (Lucas et al., 2014). In pedagogy, they can help to make social skills training more accessible (Johnson et al., 2019).

While we think about these applications, it is crucial to also lay out proper ethical guidelines to avoid any misuse of these systems. Primary concerns are around the use of deception (e.g. in Diplomacy and other negotiation tasks), emotional appeals (e.g. in persuasion), and behavior change (e.g. in conversational recommendations).

To mitigate possible misuse scenarios or unintended harms, we now lay out a few ethical guidelines which also apply to dialogue research in general. First, rigorous attempts must be made to ensure that the data collection, design processes, and evaluations, strictly abide by the guidelines and regulations laid out by the relevant Institutional Review Board (IRB). Second, the research team needs to develop a thorough plan to monitor and understand the behaviors of the developed systems before deployment. This includes identifying the goals of the dialogue system, identifying potential toxic language use, and any discriminatory behaviors. Third, investment into improved data collection practices, along with explainable and controllable dialogue systems can help identify these issues early on and allow manipulation to avoid them. Fourth, we argue that transparency is the key. All stakeholders must be made aware of the goals and design objectives of the system, along with any known misbehaviors or potential risks. The users must also be informed of any data collected during the deployment phase. Lastly, we believe that continuous monitoring of dialogue systems is necessary to ensure that the system performs consistently and does not diverge to unexpected conditions that may incur offensive or discriminative actions. We hope that our work promotes a more systematic study of social influence dialogue systems, which
in turn will help to tackle the ethical concerns in a more principled way.

8 Limitations

Literature Search: We presented a survey of efforts in social influence dialogue systems. Although every attempt was made to provide the readers with a comprehensive overview of the research in this space, our work does not claim exhaustiveness in the covered literature and it is likely that we missed out on other relevant research in this space.

Need for more investigation: Further, our objective in this paper was to only provide a high-level view of the research that serves as an entry-point and encourages more discussion in this nascent area. We believe that a more detailed analysis of just the corpora or only the methods in this space can lead to more insights for future research. For instance, it is still unclear how the similarities of different tasks within social influence can be leveraged to enable transfer learning approaches in this area. Regardless, we hope that our current work will provide a strong grounding for these next steps.

References

Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. Transactions of the Association for Computational Linguistics, 4:463–476.

Tim Althoff, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2014. How to ask for a favor: A case study on the success of altruistic requests. In Proceedings of the International AAAI Conference on Web and Social Media, volume 8, pages 12–21.

Sara Asai, Koichiro Yoshino, Seitaro Shinagawa, Sakrani Sakti, and Satoshi Nakamura. 2020. Emotional speech corpus for persuasive dialogue system. In Proceedings of The 12th Language Resources and Evaluation Conference, pages 491–497.

Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afentenos. 2016. Discourse structure and dialogue acts in multiparty dialogue: the stac corpus. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16), pages 2721–2727.

Tim Baarslag, Mark Hendrikx, Koen Hindriks, and Catholijn Jonker. 2013. Predicting the performance of opponent models in automated negotiation. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), volume 2, pages 59–66. IEEE.

Linda Babcock and Sara Laschever. 2009. Women don’t ask: Negotiation and the gender divide. Princeton University Press.

Sandy Bogaert, Christophe Boone, and Carolyn Declerck. 2008. Social value orientation and cooperation in social dilemmas: A review and conceptual model. British Journal of Social Psychology, 47(3):453–480.

Maria Boritchev and Maxime Amblard. 2021. Ding–a corpus of transcriptions of real-life, oral, spontaneous multi-party dialogues between french-speaking players of catan. In Journées du GdR LIFT.

Kenzie A Cameron. 2009. A practitioner’s guide to persuasion: An overview of 15 selected persuasion theories, models and frameworks. Patient education and counseling, 74(3):309–317.

Shelly Chaiken. 1987. The heuristic model of persuasion. In Social influence: the ontario symposium, volume 5, pages 3–39.

Kushal Chawla, Rene Clever, Jaysa Ramirez, Gale Lucas, and Jonathan Gratch. 2021a. Towards emotion-aware agents for negotiation dialogues. In 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII), pages 1–8. IEEE.

Kushal Chawla, Gale Lucas, Jonathan May, and Jonathan Gratch. 2020. Exploring early prediction of buyer-seller negotiation outcomes. arXiv preprint arXiv:2004.02363.

Kushal Chawla, Gale M Lucas, Jonathan May, and Jonathan Gratch. 2022. Opponent modeling in negotiation dialogues by related data adaptation. arXiv preprint arXiv:2205.00344.

Kushal Chawla, Jaysa Ramirez, Rene Clever, Gale Lucas, Jonathan May, and Jonathan Gratch. 2021b. Casino: A corpus of campsite negotiation dialogues for automatic negotiation systems. arXiv preprint arXiv:2103.15721.

R. B. Cialdini. 2009. Influence: Science and Practice. fifth ed. Pearson/Allyn & Bacon, Boston, MA.

Robert B Cialdini and Noah J Goldstein. 2004. Social influence: Compliance and conformity. Annual review of psychology, 55(1):591–621.

William E Connolly. 2002. Identity, difference: Democratic negotiations of political paradox. U of Minnesota Press.

Cristian D.-N.-M., Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In Proceedings of the 21st international conference on World Wide Web, pages 699–708.
Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics, pages 76–87.

Orianna Demasi, Marti A Hearst, and Benjamin Recht. 2019. Towards augmenting crisis counselor training by improving message retrieval. In Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, pages 1–11.

Jan Deriu, Alvaro Rodrigo, Arantxa Otegi, Guillermo Ecchegoyen, Sophie Rosset, Eneko Agirre, and Mark Cieliebak. 2021. Survey on evaluation methods for dialogue systems. Artificial Intelligence Review, 54(1):755–810.

David DeVault, Johnathan Mell, and Jonathan Gratch. 2015. Toward natural turn-taking in a virtual human negotiation agent. In AAAI Spring Symposia. Citeseer.

James Price Dillard and Steven R Wilson. 2014. Interpersonal influence. Interpersonal communication, 6:155.

Jesse Dodge, Andreea Gane, Xiang Zhang, Antoine Bordes, Sumit Chopra, Alexander H Miller, Arthur Szlam, and Jason Weston. 2016. Evaluating prerequisite qualities for learning end-to-end dialog systems. In 4th International Conference on Learning Representations, ICLR 2016.

Sebastian Duerr and Peter A Gloor. 2021. Persuasive natural language generation–a literature review. arXiv preprint arXiv:2101.05786.

Esin Durmus and Claire Cardie. 2019. A corpus for modeling user and language effects in argumentation on online debating. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.

Ritam Dutt, Rishabh Joshi, and Carolyn Rose. 2020. Keeping up appearances: Computational modeling of face acts in persuasion oriented discussions. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 7473–7485.

Subhabrata Dutta, Jeevesh Juneja, Dipankar Das, and Tam moy Chakraborty. 2022. Can unsupervised knowledge transfer from social discussions help argument mining? arXiv preprint arXiv:2203.12881.

Ryo Egawa, Gaku Morio, and Katsuhide Fujita. 2019. Annotating and analyzing semantic role of elementary units and relations in online persuasive arguments. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 422–428.

Chaim Fershtman. 1990. The importance of the agenda in bargaining. Games and Economic Behavior, 2(3):224–238.

Tommaso Fornaciari and Massimo Poesio. 2012. Decour: a corpus of deceptive statements in italian courts. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12), pages 1585–1590.

Jamie Fraser, Ioannis Papaioannou, and Oliver Lemon. 2018. Spoken conversational ai in video games: Emotional dialogue management increases user engagement. In Proceedings of the 18th International Conference on Intelligent Virtual Agents, pages 179–184.

Robert H Gass. 2015. Social influence, sociology of.

Howard Giles. 2016. Communication accommodation theory: Negotiating personal relationships and social identities across contexts. Cambridge University Press.

Jonathan Gratch, David DeVault, Gale M Lucas, and Stacy Marsella. 2015. Negotiation as a challenge problem for virtual humans. In International Conference on Intelligent Virtual Agents, pages 201–215. Springer.

Rafik Hadli, Jawad Haqbeen, Soña Sahab, and Takayuki Ito. 2021. Argumentative conversational agents for online discussions. Journal of Systems Science and Systems Engineering, 30(4):450–464.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 2333–2343.

Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in building intelligent open-domain dialog systems. ACM Transactions on Information Systems (TOIS), 38(3):1–32.

Alankar Jain, Florian Pecune, Yoichi Matsuyama, and Justine Cassell. 2018. A user simulator architecture for socially-aware conversational agents. In Proceedings of the 18th International Conference on Intelligent Virtual Agents, pages 133–140.

Changzhen Ji, Xin Zhou, Yating Zhang, Changlong Sun, Conghui Zhu, and Tiejun Zhao. 2020. Cross copy network for dialogue generation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1900–1910.

Emmanuel Johnson, Gale Lucas, Peter Kim, and Jonathan Gratch. 2019. Intelligent tutoring system for negotiation skills training. In International Conference on Artificial Intelligence in Education, pages 122–127. Springer.

Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan Black, and Yulia Tsvetkov. 2020. Dialograph: Incorporating interpretable strategy-graph networks into negotiation dialogues. In International Conference on Learning Representations.
Sung Soo Kim, Stan Kaplowitz, and Mark V Johnston. 2004. The effects of physician empathy on patient satisfaction and compliance. *Evaluation & the health professions*, 27(3):237–251.

Mojtaba Komeili, Kurt Shuster, and Jason Weston. 2021. Internet-augmented dialogue generation. *arXiv preprint arXiv:2107.07566*.

Sarat Kraus. 1997. Negotiation and cooperation in multi-agent environments. *Artificial intelligence*, 94(1-2):79–97.

Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–27.

Roy J Lewicki, Bruce Barry, and David M Saunders. 2016. *Essentials of negotiation*. McGraw-Hill.

Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Drhuv Batra. 2017. Deal or no deal? end-to-end learning of negotiation dialogues. In *EMNLP*.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards deep conversational recommendations. *Advances in neural information processing systems*, 31.

Xinneng Li, Wansen Wu, Long Qin, and Quanjuan Yin. 2021. How to evaluate your dialogue models: A review of approaches. *arXiv preprint arXiv:2108.01369*.

Xiujuan Li, Zachary C Lipton, Bhuvan Dhingra, Lihong Li, Jianfeng Gao, and Yun-Nung Chen. 2016. A user simulator for task-completion dialogues. *arXiv preprint arXiv:1612.05688*.

Yu Li, Kun Qian, Weiyan Shi, and Zhou Yu. 2020. End-to-end trainable non-collaborative dialog system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 883–8302.

Kai-Hui Liang, Patrick L Lange, Yoo Jung Oh, Jingwen Zhang, Yoshimi Fukuoka, and Zhou Yu. 2021. Evaluation of in-person counseling strategies to develop physical activity chatbot for women. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 32–44.

Wang Liao, Jingwen Zhang, Yoo Jung Oh, and Nicholas A Palomares. 2021. Linguistic accommodation enhances compliance to charity donation: The role of interpersonal communication processes in mediated compliance-gaining conversations. *Journal of Computer-Mediated Communication*, 26(3):167–185.

Houjun Liu. 2021. Towards automated psychotherapy via language modeling. *arXiv preprint arXiv:2104.10661*.

Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It’s only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37:94–100.

Chunyan Miao, Qiang Yang, Haijing Fang, and Angela Goh. 2007. A cognitive approach for agent-based personalized recommendation. *Knowledge-Based Systems*, 20(4):397–405.

Huyen Nguyen, Ralph Vente, David Lupea, Sarah Ita Levitan, and Julia Hirschberg. 2021. Acoustic-prosodic, lexical and demographic cues to persuasiveness in competitive debate speeches. *Proc. Interspeech 2021*, pages 1034–1038.

Jinjie Ni, Tom Young, Vlad Pandelea, Fuzhao Xue, Vinay Adiga, and Erik Cambria. 2021. Recent advances in deep learning based dialogue systems: A systematic survey. *arXiv preprint arXiv:2105.04387*.

Tim Norfolk, Kamal Birdi, and Deirdre Walsh. 2007. The role of empathy in establishing rapport in the consultation: a new model. *Medical education*, 41(7):690–697.

Daniel J O’keefe. 2015. *Persuasion: Theory and research*. Sage Publications.

Siri Ouyang, Zhuosheng Zhang, and Hai Zhao. 2021. Dialogue graph modeling for conversational machine reading. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3158–3169.

Richard M Perloff. 1993. *The dynamics of persuasion: Communication and attitudes in the 21st century*. Routledge.

Denis Peskov, Benny Cheng, Ahmed Elgohary, Joe Barrow, Cristian Danescu-Niculescu-Mizil, and Jordan Boyd-Graber. 2020. It takes two to lie: One to lie, and one to listen. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 3811–3854.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381.
Lisa Rashotte. 2007. Social influence. *The Blackwell encyclopedia of sociology.*

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M Smith, et al. 2020. Recipes for building an open-domain chatbot. *arXiv preprint arXiv:2004.13637.*

Daniel Schulman and Timothy Bickmore. 2009. Persuading users through counseling dialogue with a conversational agent. In *Proceedings of the 4th international conference on persuasive technology,* pages 1–8.

Weiyan Shi, Kun Qian, Xuewei Wang, and Zhou Yu. 2019. How to build user simulators to train rl-based dialog systems. *arXiv preprint arXiv:1909.01388.*

Zhouxing Shi and Minlie Huang. 2019. A deep sequential model for discourse parsing on multi-party dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence,* volume 33, pages 7007–7014.

Manjira Sinha and Tirthankar Dasgupta. 2021. Predicting success of a persuasion through joint modeling of utterance categorization. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management,* pages 3423–3427.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 23th international conference on world wide web,* pages 613–624.

Michael Tanana, Kevin A Hallgren, Zac E Imel, David C Atkins, and Vivek Srikumar. 2016. A comparison of natural language processing methods for automated coding of motivational interviewing. *Journal of substance abuse treatment,* 45:43–50.

Jianheng Tang, Tiancheng Zhao, Chenyan Xiong, Xiaodan Liang, Eric Xing, and Zhiting Hu. 2019. Target-guided open-domain conversation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,* pages 5624–5634.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing,* pages 327–335.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,* pages 5635–5649.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning,* 8(3):229–256.

Qingyang Wu, Yichi Zhang, Yu Li, and Zhou Yu. 2021. Alternating recurrent dialog model with large-scale pre-trained language models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume,* pages 1292–1301.

William Wunderle. 2007. How to negotiate in the middle east. *Military review,* 87(2):33.

Yuyu Xu, Pedro Sequeira, and Stacy Marsella. 2017. Towards modeling agent negotiators by analyzing human negotiation behavior. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII),* pages 58–64. IEEE.

Atsuki Yamaguchi, Kosui Iwasa, and Katsuhide Fujita. 2021. Dialogue act-based breakdown detection in negotiation dialogues. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume,* pages 745–757.

Runze Yang, Jingxiao Chen, and Karthik Narasimhan. 2020. Improving dialog systems for negotiation with personality modeling. *arXiv preprint arXiv:2010.09954.*

Runze Yang, Jingxiao Chen, and Karthik Narasimhan. 2021. Improving dialog systems for negotiation with personality modeling. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers),* pages 681–693.

Denis Yarats and Mike Lewis. 2018. Hierarchical text generation and planning for strategic dialogue. In *International Conference on Machine Learning,* pages 5591–5599. PMLR.

Min-Hsuan Yeh and Lun-Wei Ku. 2021. Lying through one’s teeth: A study on verbal leakage cues. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing,* pages 4504–4510.

Dian Yu, Yulia Tsyshchuk, Heng Ji, and William Wallace. 2015. Detecting deceptive groups using conversations and network analysis. In *53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL-IJCNLP 2015,* pages 857–866. Association for Computational Linguistics (ACL).

Jingwen Zhang, Yoo Jung Oh, Patrick Lange, Zhou Yu, Yoshimi Fukuoka, et al. 2020a. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet. *Journal of medical Internet research,* 22(9):e22845.
Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213.

Zheng Zhang, Lizi Liao, Xiaoyan Zhu, Tat-Seng Chua, Zitao Liu, Yan Huang, and Minlie Huang. 2020b. Learning goal-oriented dialogue policy with opposite agent awareness. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 122–132.

Zheng Zhang, Ryuichi Takanobu, Qi Zhu, MinLie Huang, and XiaoYan Zhu. 2020c. Recent advances and challenges in task-oriented dialog systems. *Science China Technological Sciences*, pages 1–17.

Yiheng Zhou, Yulia Tsvetkov, Alan W Black, and Zhou Yu. 2019. Augmenting non-collaborative dialog systems with explicit semantic and strategic dialog history. In *International Conference on Learning Representations*.

### A Literature Compilation

In this section, we provide details about how the literature was curated for our survey. We hope this helps the overall reproducibility and also guides similar studies in the future. We followed a simple two-stage process. First, we compiled the relevant datasets that capture various forms of social influence across diverse domains (presented in Section 3) and then, we compiled the techniques developed on these datasets (presented in Section 4).

**Step I - Datasets:** Our objective was to gather datasets that (by design) capture forms of social influence. We primarily focused on dialogue interactions but include the datasets based on transcripts from multimodal interactions as well. Given the large breadth of research in this space across a number of domains, our collection is not exhaustive but is rather restricted to the following sources.

We surveyed the past 6 years of *ACL* conference proceedings. We then covered several online repositories of dialogue data to capture datasets published at other venues. This includes ParlAI\(^2\), Huggingface\(^3\), NLP-Progress\(^4\), and Convokit\(^5\). Further, we revisited several recent surveys in dialogue systems and Natural Language Generation (NLG) research (Huang et al., 2020; Zhang et al., 2020c; Ni et al., 2021; Duerr and Gloor, 2021). Datasets that were categorized as task-oriented or open-domain in these surveys but also contain some aspects of social influence have been included in our discussion. As discussed in Section 4, we also include the datasets that have not been directly used for designing dialogue systems but rather for various Natural Language Understanding (NLU) sub-tasks that can be crucial for the eventual development of dialogue systems in this space. Finally, we also reviewed the citation graphs of the collected papers from Google Scholar. Overall, we ended up with 22 dataset papers, spanning 12 publication venues, 4 languages, and 7 application domains.

**Step II - Methods:** Compiling the methodological progress was based on the models developed on the curated datasets. For this purpose, we simply reviewed the citations of all the dataset papers using Google Scholar.

\(^2\)https://github.com/facebookresearch/ParlAI

\(^3\)https://huggingface.co/docs/datasets/index

\(^4\)http://nlpprogress.com/english/dialogue.html

\(^5\)https://convokit.cornell.edu/documentation/datasets.html
B Datasets

A comprehensive list of the available datasets for investigating social influence in dialogues is provided in Table 1. For each dataset, we mention the application domain, source, key statistics, as well as the available metadata and annotations apart from the conversation logs.

C Five Stages for Designing Social Influence Dialogue Systems

We develop a five-stage framework to summarize our recommendations for future work. These stages cover key decisions in the design of a dialogue system in this space, encouraging a holistic understanding of the system characteristics, target audience, underlying modeling techniques, and evaluation mechanisms. These steps are inspired by a behavior change model in healthcare research (Zhang et al., 2020a). We adapt this model to make it suitable for general social influence tasks in NLP. We present these steps in Figure 3.
| Name (Citation)          | Domain                  | Source                   | Key Statistics                  | Metadata & Annotations                                                                 |
|-------------------------|-------------------------|--------------------------|---------------------------------|----------------------------------------------------------------------------------------|
| STAC (Asher et al., 2016) | Games                   | Crowdsource              | Dialogues: 1081                  | Dialogue Acts; Discourse Structures                                                     |
| Diplomacy (Peskov et al., 2020) | Games                   | Crowdsource              | Games: 12                        | Intended and perceived truthfulness; Participant demographics and self-assessment of lying abilities; Ground-truth betrayals |
| DinG (Boritchev and Amblard, 2021) | Games                   | University game night logs | Games: 10                       | Annotated question-answer pairs                                                          |
| Tabletop (DeVash et al., 2015) | MIBT                    | Face-to-face             | Face-to-face Dialogues: 89       | Participant demographics; Subjective questionnaire-based assessment                      |
| DealOrNoDeal (Lewis et al., 2017) | MIBT                    |                          | Dialogues: 5808                  |                                                                                         |
| CaSiNo (Chawla et al., 2021b) | MIBT                    | Crowdsource              | Dialogues: 1030                  | Participant demographics and personality traits; Outcome satisfaction; Partner perception; Strategy Annotations |
| JobInterview (Yamaguchi et al., 2021) | MIBT                    |                          | Dialogues: 2639                  | Dialogue acts                                                                          |
| PersuasionforGood (Wang et al., 2019) | Social Good             | Crowdsource              | Dialogues: 1017                  | Participant sociodemographics, personality, and engagement in the conversation; Strategy annotations; Dialogue Acts |
| CraigslistBargain (He et al., 2018) | E-commerce              | Crowdsource              | Dialogues: 6682                  | Dialogue Acts                                                                          |
| AntiScam (Li et al., 2020) | E-commerce              | Crowdsource              | Dialogues: 220                    | Dialogue Acts; Semantic Slots                                                            |
| Motivational Interviewing (Tanana et al., 2016) | Therapy & Support       | Psychotherapy session logs | Sessions: 341                    | Behavior annotations                                                                   |
| SMS Counseling (Althoff et al., 2016) | Therapy & Support       | SMS chat logs            | Dialogues: 80,885                 | Post-conversation assessment for both the counselor (e.g. suicide risk, main issue etc.) and user (how they feel afterwards) |
| EmpatheticDialogues (Redkin et al., 2019) | Therapy & Support       | Crowdsource              | Dialogues: 24,850                 |                                                                                         |
| Hotline Counseling (Demasi et al., 2019) | Therapy & Support       | Synthetic Transcripts    | Dialogues: 254                    | Demographics; Physical activity related pre and post surveys; Strategy annotations         |
| mPED (Liang et al., 2021) | Therapy & Support       | Physical activity clinical trials | Sessions: 107                   |                                                                                         |
| Congressional Debates (Thomas et al., 2006) | Argumentation           | U.S. Congressional transcripts | Debates: 53                     | Ground-truth label with each speech segment for support/oppose of the proposed bill       |
| Supreme Court (D.-N.-M. et al., 2012) | Argumentation           | Oyez.org transcripts     | Cases: 7700                       | Case-related metadata such as key dates, citation, parties involved, and voting results   |
| DeCour (Fornaciari and Poesio, 2012) | Argumentation           | Italian court hearings   | Hearings: 35                     | Metadata for testimonies like place, date, demographics; Hearing related details; Truthfulness annotations |
| ChangeMyView (Tan et al., 2016) | Argumentation           | Reddit                   | Discussion Trees: 20,626          | The original post with mutual arguments and explicitly recognized successful persuasive arguments from the opposing side |
| DDD Debates (Durmus and Cardie, 2019) | Argumentation           | debate.org logs          | Debates: 78,376                   | User demographics; Debate metadata like dates and category; Audience votes and comments   |
| Court Debates (Ji et al., 2020) | Argumentation           | China Court transcripts  | Dialogues: 260,190                |                                                                                         |
| Target-Guided (Tang et al., 2015) | Miscellaneous           | Crowdsource              | Dialogues: 9939                  |                                                                                         |

Table 1: Social Influence Dialogue Corpora, grouped by task domains and sorted by publishing year within a domain. All statistics of the form X/Y denote average numbers. MIBT: Multi-Issue Bargaining Task. *Only computed for dialogues with additional survey responses, + Only computed for the training data, # Only for Speaker utterances in front of the judge (doesn’t include other members of the court). Note that not all datasets listed above have been directly used for designing end-to-end dialogue systems, but instead, these have enabled research into various sub-tasks and analyses that can eventually be useful for dialogue systems in this area. Please refer to Section 3 in the main paper for a detailed discussion about these datasets and to Section 4 for information about various methods developed using them.
Figure 3: A theoretical model for the development of dialogue systems for social influence tasks. Curved arrows represent forward relations and the straight arrow represents the feedback.