EVA 2.0: Emotional and Rational Multimodal Argumentation between Virtual Agents

Niklas Rach, Klaus Weber, Yuchi Yang, Stefan Ultes, Elisabeth André, and Wolfgang Minker

Abstract: Persuasive argumentation depends on multiple aspects, which include not only the content of the individual arguments, but also the way they are presented. The presentation of arguments is crucial - in particular in the context of dialogical argumentation. However, the effects of different discussion styles on the listener are hard to isolate in human dialogues. In order to demonstrate and investigate various styles of argumentation, we propose a multi-agent system in which different aspects of persuasion can be modelled and investigated separately. Our system utilizes argument structures extracted from text-based reviews for which a minimal bias of the user can be assumed. The persuasive dialogue is modelled as a dialogue game for argumentation that was motivated by the objective to enable both natural and flexible interactions between the agents. In order to support a comparison of factual against affective persuasion approaches, we implemented two fundamentally different strategies for both agents: The logical policy utilizes deep Reinforcement Learning in a multi-agent setup to optimize the strategy with respect to the game formalism and the available argument. In contrast, the emotional policy selects the next move in compliance with an agent emotion that is adapted to user feedback to persuade on an emotional level. The resulting interaction is presented to the user via virtual avatars and can be rated through an intuitive interface.

ACM CCS: Human-centered computing → Human computer interaction (HCI) →Interactive systems and tools; Human-centered computing → Human computer interaction (HCI) → HCI design and evaluation methods

Keywords: Persuasive Dialogue Systems, Multi-Agent Systems, Reinforcement Learning, Computational Argumentation

1 Introduction

How does persuasion work? This question was addressed in many different fields, including philosophy [12], psychology [19] and computational argumentation [37]. Despite the differences in all these approaches, it has become clear that the process of persuading a person includes an interplay of multiple different aspects. Especially in the case of dialogical persuasion, it involves not just the rational arrangement of suitable arguments, but also a presentation that is emotionally appealing to the interlocutor. However, isolating and investigating the contribution of these individual aspects is difficult in human dialogues, as humans usually act and react intuitively in a conversation. To address this issue, we introduce a new version EVA 2.0 of the multi-agent system proposed in [42] in which different persuasion strategies can be displayed and compared. To this end, each agent is represented by a virtual avatar that interacts with its counterpart through synthetic voice and multimodal emotions. In addition, each agent selects utterances in compliance with either an emotional or rational argumentation strategy.

The rational part of argumentation has recently gained a lot of interest in the field of computational argumentation and argumentative conversational systems in multiple domains have been introduced. Examples for areas of application range from full scale debates against a human debater[1] over persuasive dialogue [24; 29] up to customer support [8]. Besides, the field of argument mining [13] has shown remarkable progress in the task

[1] https://www.research.ibm.com/artificial-intelligence/project-debater/

of analyzing argument structures automatically. On the other hand, the role of different presentation styles was addressed in the field of human-robot interaction and included aspects like linguistic styles [28] and speech emotions [2].

Within this work, we include both aspects into a separate agent strategy, which enables us to compare their persuasive effectiveness and investigate subliminal biases. The presented version of our system extends the previous one [42] in multiple ways. First, we utilize a modified version of the interaction model that is motivated by enabling more natural and intuitive persuasive dialogues between the virtual agents [26]. Second, we extend the range of topics that can be discussed by the systems and include argument structures extracted from hotel and restaurant reviews. This choice is motivated by the goal to have a minimal bias of the user in evaluation studies. Moreover, the use of reviews ensures that arguments with different emotions are included, as they are based on subjective customer opinions. Finally, we include a conceptual extension to the decision making of the system: Whereas the original system separated the emotion the system is supposed to convey from the selection of the next dialogue utterance and treated them as individual problems, we herein combine both aspects into a new emotional policy that is adapted in real-time with respect to the individual user response. The next utterance is then selected in compliance with the adapted emotion and based on the emotional wording of the arguments. In addition, we include an updated version of the original rational strategy which is optimized prior to the interaction in self-play with respect to the new formal framework by means of deep Reinforcement Learning (RL). Consequently, the system allows for a direct comparison of the two argumentation strategies.

The remainder of this paper is as follows: Section 2 summarizes related work from the fields of argumentative dialogue systems and persuasion theory. The argument acquisition, including argument structure, discussion of the data, and extraction of arguments is discussed in Section 3. In Section 4, the interaction model is covered, followed by a description of the overall system in Section 5. Subsequently, we discuss the rational policy in Section 6 and the emotional policy in Section 7. A summary of the work, including discussion and outlook on future research, is given in Section 8.

2 Related Work

This section provides an overview of relevant related work from two perspectives: We start by discussing the role of emotions in the context of persuasion and subsequently summarize related systems and technologies from the field of computational argumentation.

2.1 Influence of Emotions on Persuasion

It is well-known that persuasion depends on far more than just the content of a persuasive message but also the emotions with which a message is conveyed. Psychological models distinguish between *central* and *peripheral* processing. While central processing focuses on the content of a message, peripheral processing focuses on non-verbal cues, own opinions, experiences, status, and overall expression of the persuader [19; 4]. An easy way to influence people is via (appropriate usage of) emotional subtones. We have seen a lot of examples in recent political history, such as Brexit and the American Election, both of which were driven by a specific emotional atmosphere. A theory, namely EASI theory (Emotions As Social Influence), developed by van Kleef [35] states that not only the emotions of the persuader have an impact on the persuasive outcome but also the emotions of the recipient itself. The strength of the impact is defined by the recipient's *epistemic motivation*, which describes the ratio between *inference* and *affective reaction*. Studies have proven this theory that people subconsciously use the source's emotion to form their own opinion [36]. In addition to that, DeSteno et al. [6] showed that persuasive messages are more successful when framed with the emotional state of the recipient. Therefore, it seems reasonable to generate dialogues conveying a specific emotional tone adapted to the recipient's emotional state to increase the persuasive effect.

2.2 Persuasive Dialogue Systems

As for argumentative systems, a variety of different approaches and related tasks were addressed in recent works. The IBM debater discusses controversial topics with a human in a debate setup with fixed speaking times and turn-taking and by means of natural language. The persuasive system in [29] uses arguments encoded in a weighted bipolar argumentation framework generated from an annotated corpus of human discussion. The strategy optimization is then formalized as a Partially Observable Markov Decision Process and addressed employing Monte-Carlo planning. Also, different approaches to argumentative chatbots were investigated: Rakshit et al. [27] proposed a system that retrieves counterarguments from a corpus by means of a semantic similarity measure, whereas Le et al. [14] compared a similar retrieval approach to a generative model. Moreover, Chalaguine and Hunter [5] introduced a persuasive chatbot that utilizes a crowd-sourced argument graph and recognizes user concerns to increase its persuasive effectiveness. The idea of a user-adaptive persuasive dialogue system was also explored in [38], where the authors collected a corpus of human-human dialogues and annotated it with different persuasion strategies. Moreover, approaches to estimate the annotated strategies and the interplay between the psychological background of human users and the different persuasion strategies were investigated. The role of emotions in persuasive dialogue systems was addressed by Asai et al. [2] through the collection of a corpus with emotional speech for the use in persuasive robots. The emotional text of the utterances was collected in a crowd-sourcing setup and the emotional speech was recorded by a voice actor. Whereas all the approaches discussed so far investigate human-machine persuasion, we propose a multi-agent setup to model and investigate different aspects of persuasion. Also along the line of multi-agent argumentation, Alahmari et al. [1] introduced an approach to optimize agent strategies in a dialogue game for argumentation using Reinforcement Learning. The learning agent in the proposed setup optimizes its argumentation strategy against two pre-defined baseline agents. In contrast, we utilize RL approaches that are independent of pre-defined opponent strategies.

3 Argument Acquisition from Reviews

The first component of our system is the knowledge base of arguments that is accessible for both agents. It comprises argument components and relations between them that are encoded according to the argument annotation scheme introduced in [33]. The annotation scheme distinguishes three different component types (*Major Claim*, *Claim* and *Premise*) and two directed relations between them (*support*, *attack*). The *Major Claim* is the overall topic of the discussion and the only component in the structure that does not target another component with a relation. A *Claim* is a general statement regarding the *Major Claim* and can therefore only target the *Major Claim* with a relation. Finally, a *Premise* provides evidence or additional information regarding another component and can hence target every other component type with a relation. Throughout this work, argument components are denoted as φ and if a component φ_i targets another component φ_j with a relation, φ_j is called the *target* of φ_i. Each component has exactly one target but can be the target of multiple other components. Since the formal difference between the components is only their allowed target component types and due to the unique target of each component, the resulting structure can be represented as a tree with components as nodes and relations as edges.

As in [41], we focus on argument components extracted from hotel and restaurant reviews since we can assume a minimal bias of the user in this domain which is necessary for evaluation studies. At the same time, reviews also include a wide range of emotional text, which is required for the emotional policy. The argument components were extracted from annotated reviews in the *SemEval-2015 Task 12* Test Datasets [20]. The data includes sentences from reviews annotated with the following labels:

- An aspect category consisting of an entity (e.g. Location, Service) and an attribute (e.g. Price, Quality).
- A polarity (positive, negative, neutral).
- An opinion target expression within the annotated sentence that explicitly refers to the entity.

We utilize the semi-automatic procedure introduced in [41] to generate argument structures based on these annotations. We generate one structure per hotel/restaurant and start with the definition of a *Major Claim* component with the textual representation *This hotel/restaurant is worth a visit.* Afterwards, a *Claim* is included into the structure for each entity in the annotation scheme. Its relation towards the *Major Claim* is determined based on the ratio of positive and negative sentences concerned with the corresponding hotel/restaurant and annotated with this entity. The textual representation of each *Claim* is of the form *The <ENTITY> is/are good/bad.* Afterwards, each sentence is included that has consistent polarity annotations, a target expression, and is not marked as 'OutOfScope'. We assume that sentences with the same entity label within the same review build on each other and therefore connect them in order of appearance. The first sentence is connected to the corresponding *Claim*. Afterwards, standalone sentences are connected to the *Claim* with the same entity, unless their target expression matches a previous component. In this case, the new component is connected to this one. All relations are determined based on the polarity annotation, meaning that a component *supports* its target if they have the same polarity and *attacks* it otherwise. So far, the procedure is completely automatic. However, as we want to utilize the annotated sentences directly in the interaction, we manually correct language errors in the sentences, merge components with the same content and separate sentences with multiple different entity labels into standalone components (one per annotated entity). We extracted four argument structures and use the one utilized in [41] with 43 components for examples and testing.

4 Dialogue Game for Argumentation

Next, we discuss the formal model of the agent-agent interaction that regulates aspects like allowed replies and turn-taking. We focus on dialogue games for argumentation as an approach to formally model persuasive dialogue and ensure a logically coherent discussion. The herein applied model is a modification of the dialogue game for argumentation introduced in [21; 22] that is motivated by enabling natural interactions while at the same time preserving the logical consistency [26]. We will discuss the formal notion of dialogue games for argumentation alongside the original and the modified framework in detail throughout the following subsections.

4.1 Original Framework

Within this work, we follow the formalism introduced in [22] that defines a dialogue game for argumentation as tuple $(\mathcal{L}, D)$. The logic for defeasible argumentation $\mathcal{L}$ includes arguments in the form of AND-trees. Each argument is comprised of elements out of a logical language L_t (nodes) that are connected by instantiations of inference rules R defined over L_t (links). The root of an argument Φ_i is called *conclusion* and denoted with $conc(\Phi_i)$ whereas the set of leaves is called *premises* and denoted with $prem(\Phi_i)$. For example, an argument $\Phi_i = b, e \Rightarrow c$ has the premises $prem(\Phi_i) = \{b, e\}$ and the conclusion $conc(\Phi_i) = c$ with $b, e, c \in L_t$. An argument Φ_i is called an extension of another argument Φ_j, if $conc(\Phi_i) \in prem(\Phi_j)$. Finally, binary defeat relations $\rightarrow$ are defined over the set of all arguments *Args*. Consequently, the formal notation of the logic for defeasible argumentation is $\mathcal{L} = (L_t, R, Args, \rightarrow)$.

Within this work, $\mathcal{L}$ is derived from the argument (tree) structure described in Section 3. Therefore, each argument has a single premise and the form $\Phi_j = \varphi_j \Rightarrow \varphi_i$ (if φ_j supports φ_i) or $\Phi_j = \varphi_j \Rightarrow \neg\varphi_i$ (if φ_j attacks φ_i). An argument Φ_i defeats another argument Φ_j, if the conclusion of Φ_i contradicts the premise of Φ_j. Conversely, an argument Φ_i extends another argument Φ_j, if the conclusion of Φ_i equals the premise of Φ_j. The generation of arguments and the implications in $\mathcal{L}$ regarding defeat and extensions are summarized in Table 1.

Arg Structure	*Args*	$\mathcal{L}$
φ_j supports φ_i	$\Phi_j = \varphi_j \Rightarrow \varphi_i$	-
φ_l attacks φ_j	$\Phi_l = \varphi_l \Rightarrow \neg\varphi_j$	Φ_l defeats Φ_j
φ_h supports φ_j	$\Phi_h = \varphi_h \Rightarrow \varphi_j$	Φ_h extends Φ_l

Table 1: Arguments in *Args* generated from the argument structure (Arg. Structure) and implications in the logic for defeasible argumentation $\mathcal{L}$.

The dialogue system proper D encodes the communication language L_c, the game protocol P and commitment rules C and is hence given as $D = (L_c, P, C)$. The communication language L_c includes the speech acts and an explicit reply structure between them whereas the game protocol P regulates turn-taking, allowed replies, and the outcome of the game. The commitment rules C regulate implications and obligations that arise for a player of the game from his or her previous utterances. The communication language for the herein discussed framework is shown in Table 2. A game is played turnwise and each turn can include one or more game moves. Each move m_k consists of a temporal identifier, a single speech act out of L_c, an identifier of the corresponding player and a target, which is the temporal identifier of the previous move the current one replies to. As a direct consequence, each move (except for the opening move) replies to exactly one previous move. The set of all moves is denoted with $\mathcal{M}$ and a temporarily ordered sequence of moves is called a dialogue $d = m_0, m_1, ..., m_n$.

The protocol P determines the set of legal moves $\mathcal{M}_d \subset \mathcal{M}$ for each dialogue based on a relevance criterion that determines if a move in d can be addressed by the current player to move. The relevance criterion is based on a binary status of each move that defines it as either *in* or *out*. A move m_k is *out*, if d includes an attacking reply to it that is *in*. If no such attack is included in d, m_k is *in*. If an attacking reply on a move m_k would change the status of the initial move m_0, m_k is called a relevant target and only relevant targets can be addressed by the player to move. A player has to play additional moves until the status of m_0 is switched in his or her favour, meaning that each turn consists of a varying number of surrendering replies, followed by a single attack. The game ends if the player to move is not able to play another move and he or she loses the game.

Speech Act	Attacks	Surrenders
claim(φ_i)	*why*(φ_i)	*concede*(φ_i)
why(φ_i)	*argue*$(\varphi_j \Rightarrow \varphi_i)$	*retract*(φ_i)
concede(φ_i)	-	-
retract(φ_i)	-	-
argue$(\varphi_j \Rightarrow \varphi_i)$	*why*(φ_j), *argue*$(\varphi_l \Rightarrow \neg\varphi_j)$	*concede*(φ_j)

Table 2: Communication language L_c of the original framework [22] for arguments of the herein considered form.

4.2 Preliminary Evaluation and Natural Language Generation

We utilized this formalism in [25] to generate artificial discussions between two virtual agents (Alice and Bob) and assessed the interplay between the argument structure discussed in Section 3 and the dialogue game formalism. To this end, we generated Natural Language Generation (NLG) templates for all speech act types that do not include an argument (*why*, *concede*, *retract*) and used the sentences in the employed argument structure as formulation for the remaining ones (*argue* and *claim*). In case an *argue* move referred directly to its predecessor, the conclusion of the corresponding argument was left implicit. If the current move did not refer to its immediate predecessor, a formulation to indicate this topic switch was included and the referenced component was explicitly repeated in the utterance. In the case of an *argue* move this reference component was the argument's conclusion. The explicit formulation was selected from the NLG templates randomly. Moreover, the agent policies within the dialogue game framework were based on probabilistic rules for the first evaluation.

The artificial dialogues were compared to human-generated ones in a user survey [25]. Each participant assessed the transcript of one dialogue by answering several questions regarding logical consistency and naturalness of the shown dialogue as well as the argumentation strategies of the transcribed speakers. The results showed that the artificial dialogues are logically consistent but on the other hand not perceived as natural. To understand the reason behind this perception, we look

Player	Utterance	Speech Acts
Alice	*You should visit this hotel.*	claim(φ_0)
Bob	*Why do you think that?*	why(φ_0)
Alice	*From my perspective the hotel is very good in general.*	argue($\varphi_4 \Rightarrow \varphi_0$)
Bob	*Could you please elaborate?*	why(φ_4)
Alice	*This property has really improved since our last stay.*	argue($\varphi_5 \Rightarrow \varphi_4$)

Table 3: Artificial dialogue between the agents Alice and Bob generated with the original framework.

Player	Utterance	Speech Acts
Alice	*You should visit this hotel.*	claim(φ_0)

Bob	*In my opinion the facilities are bad.*
	They promised a lot of services which was not provided.
		argue_ex($\varphi_1 \Rightarrow \neg\varphi_0$)
		argue($\varphi_9 \Rightarrow \varphi_1$)

Alice	*I think the Restaurant was great.*
	The restaurant downstairs is the best kept secret in the area!
		argue_ex($\varphi_2 \Rightarrow \neg\varphi_1$)
		argue($\varphi_3 \Rightarrow \varphi_2$)

Table 4: Artificial dialogue between the agents Alice and Bob generated with the new framework.

at the example of an artificial discussion in Table 3. We see that Bob assumes a passive role and is merely asking for further information, whereas Alice is providing her arguments step by step. This structure is enforced by the dialogue game protocol since the explicit reply structure demands an explicit request for further information (*why* reply) before supporting arguments can be used. To address this issue, we discuss an extension of the dialogue game framework that allows players to chain multiple arguments in an utterance in Section 4.3.

4.3 Extended Framework

To include chained arguments into the game we proposed an extension of the communication language L_c that introduces an additional speech act *argue_extend(Φ_i)* [26]. In terms of attacking and surrendering replies, it has the same properties as an *argue(Φ_i)* speech act (see Table 2) but does not end the turn of the current player. A move with an *argue_extend*(Φ_i) speech act can only be played if *Args* includes and argument Φ_j that extends Φ_i. Further, an extended attack has to include another argument, meaning that an *argue_extend* move has to be succeeded by another *argue(_extend)* move. A series of *argue(_extend)* moves is called an *argument chain* and anticipates *why* replies by addressing them in advance.

The protocol determines the relevance of moves in a chain again through the binary status (*in/out*) of the original framework discussed in the previous subsection. When introduced, all moves in the chain have the status *in*, as there is no attacking reply on them in *d*. However, only the first move in a chain is a relevant target, since an attack on it switches the status of the initial move. Hence, the freedom of choices for the players is increased whereas the logical consistency of the dialogue ensured by the relevance criterion is preserved. An example dialogue for the new framework is shown in Table 4.

To evaluate the modification and verify the assumed increase in the naturalness of the dialogues, we conducted an additional survey in [26]. To compare the results with our previous work, the setup was as similar as possible to the one described in [25]. We used the same questionnaire, the same policy, and the same argument structure and compared the ratings to the ones from the original study for the unmodified framework. Regarding the NLG, additional templates for the new combinations of moves within a turn were included. We found a significant improvement in the naturalness of the dialogues generated with the modified framework whereas their logical consistency did not change significantly [26]. Consequently, we herein use the modified framework to model the interaction between the agents.

4.4 Winning Criteria

The formal winning criterion defines that the player who first runs out of moves loses the game. Although this is reasonable from a purely logical point of view, it encourages strategies that are not considered to be optimal by humans. This is best understood in view of Bob's strategy in the example in Table 3. If Bob challenges the previous argument of Alice in each turn and does not present own arguments, Alice always runs out of moves eventually. Although this is clearly not an optimal strategy, we used this winning criterion as proof of principle setup to test the RL approach.

In order to learn a more natural strategy, we introduce a modification of the winning criterion based on a scoring system that assigns points for playing certain moves. At the end of the game, these points are summed up and the player with the most points wins. The game is still terminated by the player who is to move and has no move left. The possibilities of gaining points including a short motivation are as follows: Providing an argument (playing an *argue(_extend)* or *claim* move) always increases the score by two points to encourage the exchange of arguments. Conceding to a move increases the score by one point as reasonable acceptance of arguments is a cornerstone of argumentation. The player who drives the opponent into a situation where no move is left gains 4 points. This reflects the idea that the final agreement on the discussed issue should be considered in the game. Nevertheless, it is now possible to agree to the stance of the opponent and still win the game.

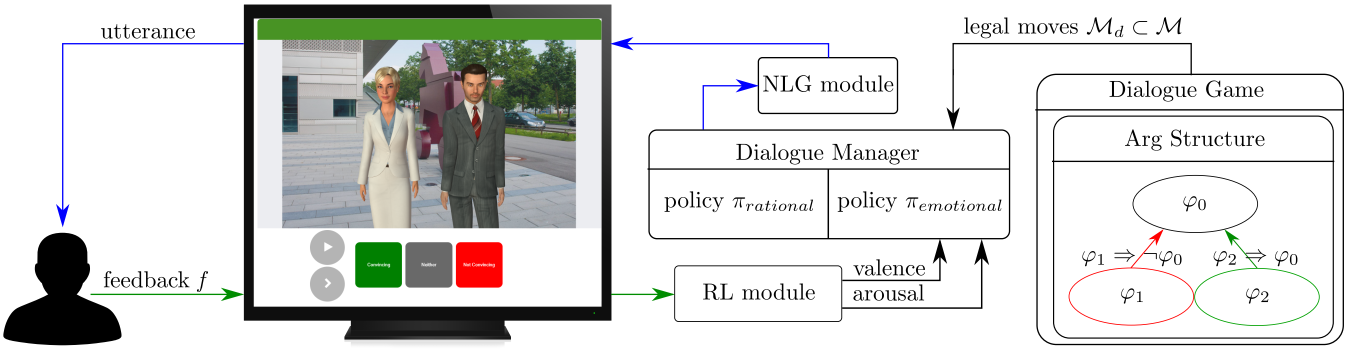

Figure 1: Conceptual overview of the EVA 2.0 system with two arguing virtual agents Alice (left) and Bob (right) allowing the user to give feedback using the three feedback buttons (*convincing*, *neither*, *not convincing*) at the bottom.

5 System

Throughout this section, we discuss the complete EVA 2.0 system which is comprised of 5 different modules as can be seen in Figure 1. The *dialogue game* module regulates the interaction and keeps track of the dialogue history, following the formalism introduced in Section 4. At each stage of the interaction, it provides a set of available game moves from which the *dialogue manager* selects one in compliance with the currently utilized policy. All selected moves that correspond to a turn in the dialogue game are transformed by the *NLG module* (described in Section 4.2) into a system utterance. Each utterance is presented by the avatar of the corresponding agent with synthetic speech and emotion in the interface. As in the prototype version [42], the interface is based on the Charamel™ avatar[2] and utilizes Nuance TTS in combination with Amazon Polly Voices[3]. In addition to both avatars, the interface also includes buttons that enable users to assess the presented turn if it includes an argument. This feedback is then used by the *RL module* to update the emotion conveyed by the system which is then used by the emotional policy in the next turn. Regarding the avatars, we selected one male and one female avatar for demonstration purposes. In practice, different avatars (male and female) can be chosen in compliance with the desired setup. This also allows for an investigation of gender effects on the perceived persuasive effectiveness [32].

The selection of the next game move in the *dialogue manager* and the update of the emotion conveyed by the system in the *RL module* are both decision making problems, meaning that the agents have to select the option in each state of the interaction that supports their goal best. This can be addressed as Reinforcement Learning [34] task, where the agents receive feedback in the form of a Reward for each action and optimize their strategies to maximize the overall return (sum of all Rewards). The problem is formally described as a Markov Game [3] with two players $(\mathcal{I}, \mathcal{S}, \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{T}})$. Here, $\mathcal{I}$ denotes the set of players (the two agents), $\mathcal{S}$ defines the state space, $\boldsymbol{\mathcal{A}} := \times_{p \in \mathcal{I}} \mathcal{A}_p$ the joint action space and $\boldsymbol{\mathcal{R}} := \times_{p \in \mathcal{I}} \mathcal{R}_p$ the joint reward function with $\mathcal{R}_p : \mathcal{S} \times \boldsymbol{\mathcal{A}} \times \mathcal{S} \to \mathbb{R}$ the reward function for player p. Finally, $\boldsymbol{\mathcal{T}} : \mathcal{S} \times \boldsymbol{\mathcal{A}} \times \mathcal{S} \to [0,1]$ denotes the transition function that determines the next state of an agent. The strategy of each agent is encoded in a policy function π which is a mapping from the state space to either the action space or a probability distribution over the action space (depending on the utilized learning method). The two different policies addressed throughout this work alongside the exact definition of the underlying Markov Games are discussed separately in the following sections.

6 Rational Dialogue: Policy $\pi_{rational}$

In the following, we address the optimization of the rational policy, i.e. a policy based solely on the objective winning criteria within the dialogue game formalism. As we discussed in [23], every dialogue game for argumentation that adheres to the general structure in Section 4.1 can be formulated as Markov Game. We herein build on these results and apply the introduced formalization to the modified dialogue game.

6.1 Formalization

To reduce the state space $\mathcal{S}$ to a reasonable size, we only include information about the played moves (and not their order in the dialogue). This is possible as the protocol determines the legal moves based on the relevance of moves and not on their temporal order.

Definition 1 (Rational State Space) *$\forall$ speech acts $\beta_j \in L_c$ let $\sigma^t_{\beta_j} \in \{0,1\}$ be a binary integer that is 1 if there is a move in d_t that includes β_j and 0 otherwise. Then the state of an agent p at time t is given as $s_{t_p} = (\sigma^t_{\beta_1}, ..., \sigma^t_{\beta_N})$.*

[2] https://www.charamel.com/competence/avatare

[3] https://docs.aws.amazon.com/polly/latest/dg/voicelist

Since the dialogue game defines legal moves on the basis of the current dialogue d, the set of actions available to an agent p in a state s_t is defined as follows:

Definition 2 (Rational Action Space) *Let S be the state space as defined above, $\mathcal{M}^{\leq\infty}$ the set of all finite-length dialogues in the dialogue game, $\Delta : \mathcal{S} \rightarrow \mathcal{M}^{\leq\infty}$ a function that maps each state to a corresponding legal dialogue and $P : \mathcal{M}^{\leq\infty} \rightarrow 2^{\mathcal{M}}$ the protocol of the dialogue game. The set of available actions for an agent p in state s_{t_p} is then defined as $\mathcal{A}(s_{t_p}) = P(\Delta(s_{t_p}))$.*

Consequently, each action corresponds to a legal move in the dialogue game. The transition function in this case is deterministic and updates the state based on the speech act in the selected action. As for the reward, we assign a $+20$ reward to the winning agent and a -20 reward to the losing agent at the end of the game based on the utilized winning criterion of the dialogue game. Although the reformulation is formally analogue to the referenced work, the addition of a new speech act in the modified framework increases the state and the action space and thus requires an advanced learning algorithm.

6.2 Deep Actor-Critic Reinforcement Learning

Actor-Critic methods [34] aim at combining the strong points of both value-based and policy-based RL methods by utilizing a value or Q-function as a Critic to update a parametrized policy (Actor). In the present case, the agent encodes the Q-function and the policy function within neural networks. The Critic estimates the parameters $\boldsymbol{\omega}$ of the Q-function network to minimize the mean-squared error

$$L(\boldsymbol{\omega}) = \mathbb{E}_{\pi_\theta}[(Q_{\pi_\theta}(s,a) - Q(s,a,\boldsymbol{\omega}))^2] \tag{1}$$

where Q_{π_θ} is the Q-function under policy π_θ which has to be estimated with observed rewards. The Actor updates the parameters $\boldsymbol{\theta}$ of the policy network in the direction suggested by the Critic with the policy gradient [34]

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = \mathbb{E}_{\pi_\theta}[\nabla_{\boldsymbol{\theta}} log\pi_\theta(a|s,\boldsymbol{\theta})Q(s,a,\boldsymbol{\omega})] \tag{2}$$

The training of the rational policy faces two major problems: First, since no data is available to initialize the networks it has to be collected solely from interactions between the agents. Second, winning dialogues for an agent become rare at the later stage of the training and the frequent negative reward therefore outweighs the positive. To mitigate the first problem, we use experience replay (ER) [15] to learn from samples collected in past interactions. To do so, we save each past dialogue experience into a replay memory. At each training step, a mini-batch of past dialogue experiences is randomly sampled from the replay memory to update the parameters. To address the rare occurrence of successful dialogues in the later stage of training, we use a prioritized ER approach which keeps track of dialogues in pools of different priorities. Rare transitions that provide positive rewards are assigned with higher priority and have higher chances to be sampled by the learning agent.

Throughout this work, we used the Actor-Critic with Experience Replay (ACER) algorithm proposed by Wang et al. [39] due to its high sample-efficiency and stability. In addition to the techniques discussed above, ACER also utilizes several optimizations, such as the Retrace algorithm for Q-function estimation [17], importance weight truncation to increase learning stability, and Trust Region Policy Optimization [31].

6.3 Experimental Results

During training, the agents are divided into a training-agent and a reference-agent. A game *stage* is defined as a period during which the training-agent keeps optimizing its policy until it has a superior policy against its opponent while the policy of the reference-agent remains unchanged. At the beginning of each stage, the policy of the reference-agent is updated with the training-agent's policy. This ensures that for each stage, the training-agent learns a policy that can beat its previous one. In case both policies are optimal against each other, this is called a Nash equilibrium. Although the existence of such an equilibrium is guaranteed for games of the herein discussed kind [16], it should be noted that convergence to this equilibrium can not be guaranteed in our setup.

We evaluated the ACER algorithm in different testing scenarios with the original dialogue game in Section 4 and the original winning criterion for which the optimal policy is known in [44]. The performance of the algorithm was tested on randomized argument structures with different sizes and on an annotated argument structure with 72 components on the topic *Marriage is an outdated institution* [25]. The results showed that an optimal policy could be found for argument structures of all investigated sizes, although convergence to local optima occurred for argument structures with more than 30 arguments and more frequently with an increasing number of arguments in the structure. For the sake of illustration, the training process has been divided into *super-iterations*, each consisting of 400 dialogue episodes. Every 100 episodes we calculated the overall average reward and the average reward in stance B (Bob). Figure 2 shows the average reward as a function of super-iterations for an example argument structure with 20 arguments. It can be seen that from the 8^{th} super-iteration, the overall average reward dropped to 0, meanwhile, the average reward of stance B remained $+20$, indicating that any agent with stance B won the game. This convergence indicates a Nash equilibrium where both policies are optimal against each other, which was confirmed in test games against the (known) optimal rule-based strategy. In addition, we applied the ACER algorithm to the modified dialogue game with

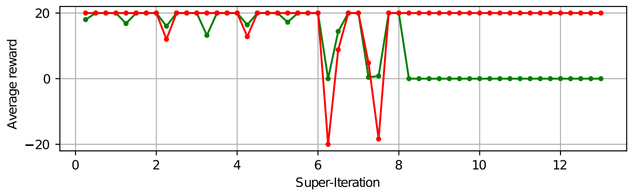

Figure 2: Average reward for the training agent as a function of super-iterations. The average over both stances is shown green, the average when assigned stance B is shown red.

Player	Utterance	Speech Acts
Alice	*You should visit this hotel.*	claim(φ_0)
Bob	*From my perspective the facilities are bad.*	argue($\varphi_1 \Rightarrow \neg\varphi_0$)

Alice	*It seems to me that the Restaurant was great.*
	The restaurant downstairs is the best-kept secret in the area!
		argue_ex($\varphi_2 \Rightarrow \neg\varphi_1$)
		argue($\varphi_3 \Rightarrow \varphi_2$)

Bob	*Hm, pardon the topic switch but I'm still thinking about one of our earlier points. You said that you should visit this hotel. What makes you so sure?*	why(φ_0)

Alice	*From my perspective the hotel is very good in general.*
	This property has really improved since our last stay.
	All in all it is a nice and affordable spot for sightseeing in the area.
		argue_ex($\varphi_4 \Rightarrow \varphi_0$)
		argue_ex($\varphi_5 \Rightarrow \varphi_4$)
		argue($\varphi_6 \Rightarrow \varphi_5$)

Bob	*It's hard to disagree with that. I see your point there.*
	I'm still thinking about your general claim. From my perspective the rooms are bad.
	All 4 lower rooms were 80 degrees and above.
		concede(φ_4)
		argue_ex($\varphi_7 \Rightarrow \neg\varphi_0$)
		argue($\varphi_8 \Rightarrow \varphi_7$)

Table 5: Artificial dialogue between the agents Alice and Bob with the rational policy trained with the ACER algorithm on the modified winning criterion.

both the original as well as the modified winning criterion. We used 5 randomized argument structures with 20 arguments and the original winning criterion again as a proof-of-principle setup and trained the final rational policy for the hotel structure on the modified one. For the proof-of-principle setup we report that in all investigated instances, the optimal policy could be found and was confirmed in test games against the rule-based strategy. An excerpt of a dialogue generated with the optimized rational policy and the new winning criterion is shown in Table 5.

7 Emotional Dialogue: Policy $\pi_{emotional}$

In the following section, we describe the (learning of the) emotional policy $\pi_{emotional}$ in detail. We recently proposed an approach, in which the emotional policy was adapted without taking the emotional content of the argument into account but following the rational policy $\pi_{rational}$ and only considering argument-related information, such as stance and relation, to determine the most-effective emotional tone [41; 42]. However, there is evidence that non-verbal inconsistencies lead to poor first impressions [43], which is in line with [35] who showed the importance of appropriateness of emotions.

Therefore, we focus on generating an emotion-based dialogue $d = m_0, m_1, ..., m_n$ conveying the most-influencing emotional tone by learning a policy $\pi_{emotional}$ during interaction with the user. While the rational policy $\pi_{rational}$ is trained beforehand, the emotional policy is adapted to the user during interaction without pretraining. This approach is motivated by several reasons: i) The effectiveness of emotions is highly subjective [18; 10], ii) adaptation of the strategy should be the main focus at an individual level [7] and iii) there is evidence that persuasive messages are more successful when framed with the emotional state of the recipient [6]. As per DeSteno et al. [6] this is mediated by biases induced by the own emotions of the recipient. Thus, we focus on learning the most effective affective state by modifying the agents' pleasure (valence) and arousal dimensions [30], which are used for selecting the next move $m_k \in \mathcal{M}_d$ of all available moves at dialog time step k. To allow for adaptation in real-time, RL is used since it is an effective tool for real-time adaptation [28; 40].

7.1 Formalization

In the following we give a formal description of the emotion-based strategy. As in [42], the decision making is described as Markov Game as follows:

Definition 3 (Emotional State Space) *Let $s_{t_p} \in \mathcal{S}$ be a state at time step t for player $p \in \mathcal{I}$. Further let $\rho_{t_p}(\varphi_0) \in [0,1]$ be a prediction that defines how persuasive agent p is compared to the opponent (see below), then the state is defined as: $s_{t_p} := (arousal, valence, \rho_{t_p}(\varphi_0))$, where arousal and valence are discrete values $\in [-1, 1]$.*

The prediction model of the persuasive effectiveness has been proposed and evaluated in [41] showing high accuracy and F1-Score for the herein used argument dataset. The persuasive effectiveness is based on *bipolar-weighted argument graphs* (BWAG) that assigns a user feedback $f_i \in [0,1]$ to each $\varphi_i \in L_t$, which is used for computing the component's effectiveness $\rho(\varphi_i)$ considering the effectiveness of its child nodes in the argument structure. If no feedback exists for a component φ_i, then $f_i = 0.5$ by default. In practice, we assign user feedback given through the interface described in Section 5 to *argue(_extend)* moves to the premise of the corresponding argument. Including the persuasive effectiveness in the state space allows for obtaining performance information with respect to the opponent's performance and allows for switching to an alternative strategy if necessary. Moreover, it is an intuitive metric to reward one agent taking its performance (change) into account [42].

Definition 4 (Emotional Action Space) *The action space $\mathcal{A}_p$ for player $p \in \mathcal{I}$ consists of an* INCREASE *and* DECREASE *action both for valence and arousal and an action* NONE *that leaves the state s_{t_p} unchanged. Consequently, there are five different actions.*

Definition 5 (Emotional Reward Function) *Let ρ_{t_p} be the prediction at time step t for $p \in \mathcal{I}$. As in [41; 42] the reward $\mathcal{R}_{t_p}$ is defined as the change of the persuasive prediction, i.e. $\mathcal{R}_{t_p} := \rho_{t_p}$ - ρ_{t_p-1}.*

7.2 Real-Time Adaptation of Policy $\pi_{emotional}$

To enable the agents to learn the emotional policy, we employ a linear function approximator along with a Fourier Basis transformation [11], which is an effective way to adapt to the user quickly [42]. At every learning step t, the agent $p \in \mathcal{I}$ selects one of the available actions $a_{t_p} \in \mathcal{A}$ according to the current player's state $s_{t_p} \in \mathcal{S}$ and policy $\pi_{emotional}$ ($\epsilon-$greedy with $\epsilon = 0.05$), modifies its current emotional state s_{t_p+1} and selects the next move(s) as follows: Let $\mathbf{f} : \mathcal{M} \to [-1,1] \times [-1,1]$ be the function that maps any component $m_k \in \mathcal{M}$ into the 2d valence-arousal space. For that, we employ DEVA, a text analysis tool designed for mapping any given sentence into the VA space (precision 82%, recall 78%, [9]). For any move m_k that includes an *argue_extend(Φ_i)* speech act, $\mathbf{f}(m_k) := {}^1\!/_2(\mathbf{f}(m_k)+\mathbf{f}(m_l))$ where m_k includes *argue(Φ_i)*, m_l includes *argue(Φ_j)* and Φ_j extends Φ_i. Further, let $\mathbf{g} : \mathcal{S} \to [-1,1] \times [-1,1]$ be the respective function for the state space. The agent uses the emotional state s_{t_p+1} to select the next move m_k that is closest to the agent's state using the L_2 norm:

$$m_k = \min_{m \in \mathcal{M}_d} \|\mathbf{f}(m) - \mathbf{g}(s_{t_p+1})\|_2 \qquad (3)$$

The obtained feedback signal f is used to compute the current effectiveness level ρ_{t_p} and with that the reward signal $\mathcal{R}_{t_p}$, which is used to update the policy. However, relying on the distance metric only leads to several issues, one of which is sketched in Table 6. Because an affect state of (-0.5, 0.5) is used, Alice concedes immediately after Bob's first argument. This seems odd at first glance, but has some good reasons:

1. The second quadrant of the VA space only contains two arguments, but only one argument ($\Phi_1 = \varphi_1 \Rightarrow \neg\varphi_0$) is allowed to be played by Bob.
2. Since $\Phi_1 = \varphi_1 \Rightarrow \neg\varphi_0$ is the closest one to (-0.5, 0.5), it is selected by Bob following the distance metric. However, Alice does not have any argument within quadrant two and the only arguments that she can make use of are within quadrant one ($\Phi_3, ..., \Phi_6$).
3. Computing the distance between all available arguments $\Phi_3, ..., \Phi_6$ and the *concede* move, inevitably leads to the concede move as the closest one.

Player	Utterance	Speech Acts
Alice	*You should visit this hotel.*	claim(φ_0)
Bob	*I think the facilities are bad.*	argue($\varphi_1 \Rightarrow \neg\varphi_0$)
Alice	*I concede.*	concede(φ_0)

Table 6: Artificial dialogue between the agents Alice and Bob with affect state (-0.5, 0.5).

Conceding right away seems irrational. However, rationality should not be completely excluded from the emotional policy but should support the emotional policy with regard to some general rational decisions during argumentation, such as: When is the right time...

1. ...to *concede* or *retract* an argument?
2. ...to introduce a new argument (*argue)*?
3. ...to request additional information (*why*)?

These questions can be handled by the rational policy $\pi_{rational}$, which decides which move type comes next, and whenever an *argue(_extend)* move is chosen, the agent follows the metrics of the emotional policy (Equ. 3). The full adaptation algorithm is sketched in the following and an example dialogue is shown in Table 7.

Algorithm 1: Emotional Dialogue Generation

```
Init: $t_p = 0, a_{t_p}, s_{t_p+1}, \forall p \in \mathcal{I}$
foreach $k = 1, ..., n$ do
    $p \leftarrow$ active player
    $s_{t_p+1} \leftarrow$ observe $s_{t_p+1} \in \mathcal{S}$
    $m_k \leftarrow \pi_{rational}$
    if $m_k$ isType(argue(_extend)) then
        $m_k \leftarrow \min_{m \in \mathcal{M}_{d,argue}} \|\mathbf{f}(m) - \mathbf{g}(s_{t_p+1})\|_2$
    apply $m_k$
    if $f_p$ then
        $\pi_{emotional} \leftarrow$ update policy using $f$
        $a_{t_p+1} \leftarrow$ select next action.
        $s_{t_p+2} \leftarrow$ modify emotional state.
        $t_p \leftarrow t_p + 1$
```

It should be noted that between any state transition $s_{t_p} \to s_{t_p+1}$ multiple moves $m_k, m_{k+1}, ..., m_{k+m}$ can be selected before the agent's emotional state changes again. This is a direct consequence of the herein employed framework and the communication language L_c since only speech acts of type *argue(_extend)* require feedback regarding the overall perceived persuasiveness. Again, the logical consistency of the dialogue is preserved by the relevance criterion of the game protocol.

Player	Utterance	Speech Acts
Alice	*You should visit this hotel.*	claim(φ_0)
Bob	*In my opinion the rooms are bad.*	argue($\varphi_7 \Rightarrow \neg\varphi_0$)
Alice	*I am not sure I understand what you are getting at.*	why(φ_7)

Bob	*I think that's enough for the moment. I would rather focus on another aspect of the topic. It seems to me that the facilities are bad.*
	An elevator was broken during our last stay and it was most annoying, but did not greatly impact the overall experience.
		argue_ex($\varphi_1 \Rightarrow \neg\varphi_0$)
		argue($\varphi_9 \Rightarrow \varphi_1$)

Alice	*I think the Restaurant was great.*	argue($\varphi_2 \Rightarrow \neg\varphi_1$)

Table 7: Artificial dialogue between the agents Alice and Bob with affect state (-0.5, -0.5) using algorithm 1.

7.3 Experimental Results

We have run multiple simulations to evaluate the adaptive feasibility of our proposed prototype. We considered two scenarios:

- First, an adaptation of one agent only to verify that it is able to increase its performance within the RL task (**S1**).
- Second, an adaptation of both agents to verify that they are able to optimize their policy with respect to each other (**S2**).

We simulated 150 users (see Fig. 3) with randomly assigned affective states $(x, y) \in$ VA space, i.e., the affective state that the agents had to learn, and ran multiple dialogues with an overall minimum length of 40 RL time steps for each agent. For every time step t_p the simulated user feedback f_i is defined as the normalized distance between the agent's affective state and the user's affective state:

$$f_i = 1 - \frac{\|(x, y) - \mathbf{g}(s_{t_p+1})\|_2}{2\sqrt{2}} \qquad (4)$$

Figure 3 shows initial results of the simulation: **S1**) Agent 1 is able to increase its performance with respect to its opponent, and **S2**) both are able to keep the balance of their individual persuasive effectiveness.

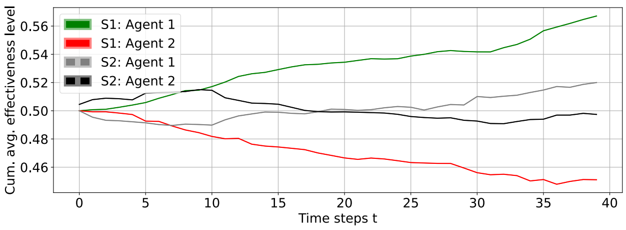

Figure 3: Initial simulation results of **S1** and **S2**

8 Summary and Future Work

We have introduced a fully integrated version of our persuasive multi-agent system EVA 2.0 in which two agents engage in a multimodal discussion. The system utilizes argument structures extracted from reviews and a dialogue game for argumentation to structure the interaction. Besides, we have discussed two approaches to policy optimization within the dialogue game framework that focus on different aspects of persuasion. The rational strategy is optimized on the objective winning criterion of the dialogue game and before the interaction by means of deep Actor-Critic RL. The emotional policy on the other hand utilizes a mapping of the available arguments into the valence-arousal space to select arguments that are close to the current emotion the system is supposed to convey. This emotion is adapted to user feedback to enable an individual and adaptive strategy. The adaptation is again approached by means of RL, although in this case the learning is done during interaction and in real-time. Both approaches were formally introduced and we discussed first results that indicate the feasibility of the proposed techniques.

In future work, we will focus on an extensive evaluation of the individual policies within a user study. Further, we will investigate more detailed aspects of the proposed approaches like the effect of different winning criteria and argument structures. Finally, we aim at combining the rational and the emotional decision making to find a hybrid policy that can consider and switch between both aspects, depending on the individual user and the current dialogue.

Acknowledgement

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) within the project "How to Win Arguments - Empowering Virtual Agents to Improve their Persuasiveness", Grant Number 376696351, as part of the Priority Program "Robust Argumentation Machines (RATIO)" (SPP-1999).

Niklas Rach studied Physics at Ulm University, Germany and received his M.Sc. in 2015. He is currently pursuing a joint Ph.D. in the Dialogue Systems group at Ulm University and the Ubiquitous Computing Systems laboratory at Nara Institute of Science and Technology, Japan. His research interests are centered around dialogue management with an emphasis on computational argumentation in dialogue systems and machine learning applications.

Address: Ulm University, Institute of Communications Engineering - E-Mail: niklas.rach@uni-ulm.de

Klaus Weber studied Computer Science at Augsburg University, Germany and received his M.Sc. in Computer Science in 2017, and his specialised M.Sc. in Computer Science and Multimedia in 2019. He is currently doing a Doctoral Degree (rer. nat.) in Computer Science in the Human-Centered AI group at Augsburg University. His research interests focus on human-agent interactions and real-time adaptation of agents to humans with an emphasis on investigating biases caused by subliminal argumentation.

Address: Augsburg University, Human-Centered Artificial Intelligence - E-Mail: klaus.weber@uni-a.de

Yuchi Yang received his M.Sc. in Computer Engineering at Ulm University in 2019. He is currently employed as Data Scientist at AXA Konzern AG.

Address: Ulm University, Institute of Communications Engineering - E-Mail: yangyuchi0617@gmail.com

Dr. Stefan Ultes received his doctorate in engineering (Ph.D.) from Ulm University (Germany) in 2015. Afterwards, he was a Research Associate at the Spoken Dialogue Systems Group at the University of Cambridge working with Prof. Steve Young and Prof. Milica Gasic. He is currently employed as Dialogue Research Lead at Mercedes Benz Research & Development.

Address: Mercedes-Benz AG, Sindelfingen, Germany

Prof. Elisabeth André received her Doctoral Degree in 1995 at Saarland University. She is a full professor at Augsburg University, Germany, since 2001. Her group currently consists of 20+ members, most of whom work on topics related to multimodal human-computer interaction, virtual agents and social robots. She is a very well-known researcher at the intersection of Human-Computer Interaction and Artificial Intelligence. She is both an elected member of the Sigchi Academy and a EurAI fellow. She is the Editor-in-Chief of IEEE Trans. on Affective Computing. In 2019, she was named by the German Society of Informatics (GI) as one of the most influential personalities in the history of German AI. In 2021, she was awarded with the Leibniz Prize for establishing the research field of conversational emotional agents in the field of artificial intelligence.

Address: Augsburg University, Human-Centered Artificial Intelligence - E-Mail: andre@informatik.uni-augsburg.de

Prof. Wolfgang Minker received his Ph.D. in Engineering Science at the University of Karlsruhe, Germany in 1997 and a Ph.D. in Computer Science from Université Paris-Sud, France in 1998. Since 2003 he is a full professor at Ulm University, Germany and also became a Co-Director of the International Research Laboratory *Multimodal Biometric and Speech Systems* at ITMO University St. Petersburg, Russia in 2017. The research at his group is focused on dialogue systems with special interest in adaptive and proactive spoken language dialogue interaction and argumentative dialogue systems.

Address: Ulm University, Institute of Communications Engineering - E-Mail: wolfgang.minker@uni-ulm.de

Bibliography

[1] S. Alahmari, T. Yuan, and D. Kudenko. Reinforcement learning for dialogue game based argumentation. In *CMNA@ PERSUASIVE*, pages 29–37, 2019.

[2] S. Asai, K. Yoshino, S. Shinagawa, S. Sakti, and S. Nakamura. Emotional speech corpus for persuasive dialogue system. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 491–497, 2020.

[3] M. Barlier, J. Perolat, R. Laroche, and O. Pietquin. Human-Machine Dialogue as a Stochastic Game. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 2–11, Prague, Czech Republic, 2015. ACL.

[4] S. Chaiken, A. Liberman, and A. Eagly. *Heuristic and Systematic Information Processing within and beyond the Persuasion Context*, pages 212–252. Guilford, 1989.

[5] L. A. Chalaguine and A. Hunter. A persuasive chatbot using a crowd-sourced argument graph and concerns. *Computational Models of Argument: Proceedings of COMMA 2020*, 326:9–20, 2020.

[6] D. DeSteno, R. E. Petty, D. D. Rucker, D. T. Wegener, and J. Braverman. Discrete emotions and persuasion: the role of emotion-induced expectancies. *Journal of personality and social psychology*, 86(1):43, 2004.

[7] B. Fogg. *Mobile Persuasion: 20 Perspectives on the Future of Behavior Change*. Stanford Captology Media, 2007.

[8] B. Galitsky. Enabling a bot with understanding argumentation and providing arguments. In *Developing Enterprise Chatbots*, pages 465–532. Springer, 2019.

[9] M. R. Islam and M. F. Zibran. DEVA: sensing emotions in the valence arousal space in software engineering text. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing - SAC '18*, pages 1536–1543, Pau, France, 2018. ACM Press.

[10] M. Kaptein, J. Lacroix, and P. Saini. Individual differences in persuadability in the health promotion domain. In *International Conference on Persuasive Technology*, pages 94–105. Springer, 2010.

[11] G. Konidaris, S. Osentoski, and P. Thomas. Value function approximation in reinforcement learning using the fourier basis. In *Proceedings of the AAAI Conference*, volume 25, 2011.

[12] G. Krapinger. Aristoteles: Rhetorik. *Übersetzt und herausgegeben von Gernot Krapinger. Stuttgart: Reclam*, 1999.

[13] J. Lawrence and C. Reed. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818, 2020.

[14] D. T. Le, C.-T. Nguyen, and K. A. Nguyen. Dave the debater: a retrieval-based and generative argumentative dialogue agent. In *Proceedings of the 5th Workshop on Argument Mining*, pages 121–130, 2018.

[15] L.-J. Lin. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, 8(3):293–321, May 1992.

[16] M. L. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the eleventh international conference on machine learning*, volume 157, pages 157–163, 1994.

[17] R. Munos, T. Stepleton, A. Harutyunyan, and M. Bellemare. Safe and efficient off-policy reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 1054–1062, 2016.

[18] D. J. OḰeefe and S. Jackson. Argument quality and persuasive effects: A review of current approaches. In *Argumentation and values: Proceedings of the 9th Alta conference on argumentation*, pages 88–92, 1995.

[19] R. E. Petty and J. T. Cacioppo. The elaboration likelihood model of persuasion. In *Communication and persuasion*, pages 1–24. Springer, 1986.

[20] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, 2015.

[21] H. Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In *JELIA*, pages 224–238. Springer, 2000.

[22] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of logic and computation*, 15(6):1009–1040, 2005.

[23] N. Rach, W. Minker, and S. Ultes. Markov games for persuasive dialogue. In *Computational Models of Argument: Proceedings of COMMA 2018*, pages 213–220, 2018.

[24] N. Rach, K. Weber, L. Pragst, E. André, W. Minker, and S. Ultes. Eva: A multimodal argumentative dialogue system. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pages 551–552. ACM, 2018.

[25] N. Rach, S. Langhammer, W. Minker, and S. Ultes. Utilizing argument mining techniques for argumentative dialogue systems. In *9th International Workshop on Spoken Dialogue System Technology*, pages 131–142. Springer, 2019.

[26] N. Rach, W. Minker, and S. Ultes. Increasing the naturalness of an argumentative dialogue system through argument chains. *Computational Models of Argument: Proceedings of COMMA 2020*, 326:331–338, 2020.

[27] G. Rakshit, K. K. Bowden, L. Reed, A. Misra, and M. Walker. Debbie, the debate bot of the future. In *Advanced Social Interaction with Agents*, pages 45–52. Springer, 2019.

[28] H. Ritschel, T. Baur, and E. André. Adapting a robot's linguistic style based on socially-aware reinforcement learning. In *26th International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 378–384. IEEE, 2017.

[29] A. Rosenfeld and S. Kraus. Strategical argumentative agent for human persuasion. In *ECAI*, volume 16, pages 320–329. IOS Press, 2016.

[30] J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[31] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz. Trust region policy optimization. In *International conference on machine learning*, pages 1889–1897, 2015.

[32] M. Siegel, C. Breazeal, and M. I. Norton. Persuasive Robotics: The influence of robot gender on human behavior. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2563–2568, St. Louis, MO, USA, Oct. 2009. IEEE.

[33] C. Stab and I. Gurevych. Annotating argument components and relations in persuasive essays. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1501–1510, Dublin, Ireland, Aug. 2014. Dublin City University and ACL.

[34] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction.* A Bradford Book, Cambridge, MA, USA, 2018.

[35] G. van Kleef. Emotions as Agents of Social Influence. In *The Oxford Handbook of Social Influence*, volume 1. Oxford University Press, 2014.

[36] G. A. Van Kleef, H. van den Berg, and M. W. Heerdink. The persuasive power of emotions: Effects of emotional expressions on attitude formation and change. *Journal of Applied Psychology*, 100(4): 1124, 2015.

[37] H. Wachsmuth, N. Naderi, Y. Hou, Y. Bilu, V. Prabhakaran, T. A. Thijm, G. Hirst, and B. Stein. Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the ACL: Volume 1, Long Papers*, pages 176–187. ACL, 2017.

[38] X. Wang, W. Shi, R. Kim, Y. Oh, S. Yang, J. Zhang, and Z. Yu. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Annual Meeting of the ACL*, pages 5635–5649, 2019.

[39] Z. Wang, V. Bapst, N. Heess, V. Mnih, R. Munos, K. Kavukcuoglu, and N. de Freitas. Sample efficient actor-critic with experience replay. *arXiv preprint arXiv:1611.01224*, 2016.

[40] K. Weber, H. Ritschel, I. Aslan, F. Lingenfelser, and E. André. How to Shape the Humor of a Robot - Social Behavior Adaptation Based on Reinforcement Learning. In *Proceedings of the International Conference on Multimodal Interaction*, pages 154–162, Boulder, CO, USA, 2018. ACM Press.

[41] K. Weber, K. Janowski, N. Rach, K. Weitz, W. Minker, S. Ultes, and E. André. Predicting Persuasive Effectiveness for Multimodal Behavior Adaptation using Bipolar Weighted Argument Graphs. In *19th International Conference on Autonomous Agents and Multiagent Systems*, Auckland, New Zealand, 2020. ACM, New York.

[42] K. Weber, N. Rach, W. Minker, and E. André. How to Win Arguments. *Datenbank-Spektrum*, 2020.

[43] M. Weisbuch, N. Ambady, A. L. Clarke, S. Achor, and J. V.-V. Weele. On being consistent: The role of verbal–nonverbal consistency in first impressions. *Basic and Applied Social Psychology*, 32(3):261–268, 2010.

[44] Y. Yang. Multi-agent actor-critic reinforcement learning for argumentative dialogue systems. Master's thesis, Ulm University, 2019.