Emergent Communication Through Negotation:

https://openreview.net/pdf?id=Hk6WhagRW

Paper Summary Notes

1 Main Ideas

Central Question of the Paper: *Study the emergence of effective communication in a negotiation environment, to see how different communication methods affect performance in pro-social and self-interested agents*

1.1 Background

- human interactions are not fully cooperative, yet we can still successfully use language to communicate our private information and thoughts, discuss plans and ask questions, in order to be effective
- in the negotiation environment, effective communication is crucial, as the agents needs to exchange strategic information about their desires, and infer their opponent's desires through communication

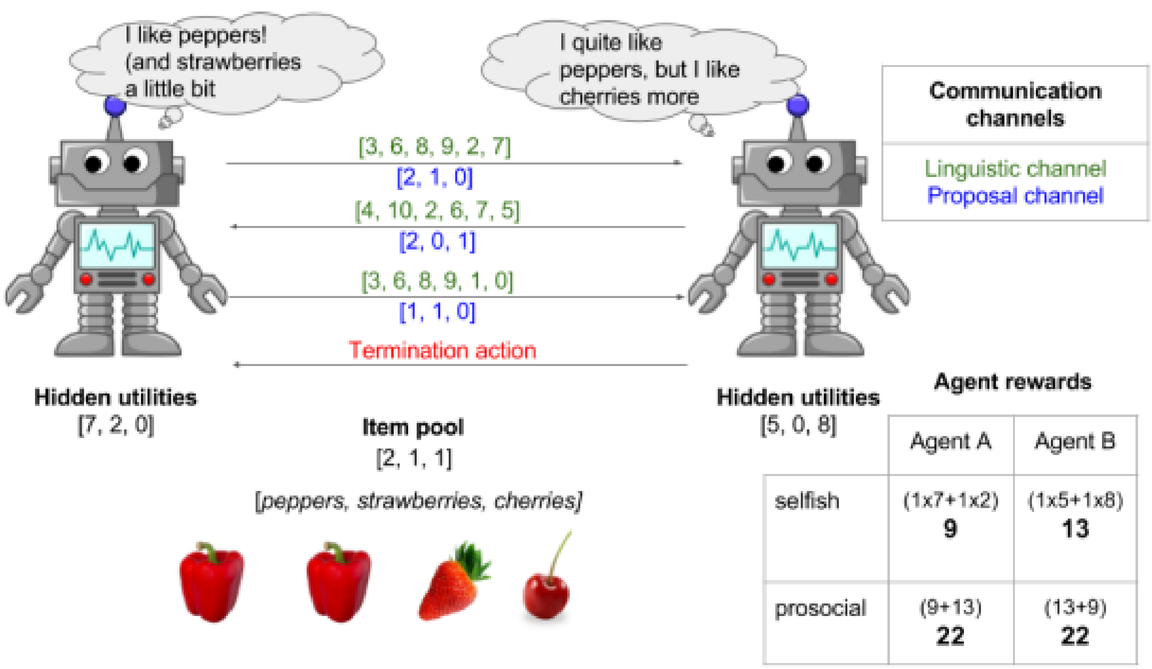
Cheap Talk

Definition 1. *Cheap Talk* is a communication protocol between agents that *does not* explicitly affect the payoffs of the game. In other words, the speaker can choose strategically what to say and what not to say, based on it's internal state.

- costless to transmit and receive
- does not limit strategic choices by either party
- unverifiable

2 Game Setting

2.1 Negotiation Environment



- Agents are presents with three items: *peppers, cherries, and strawberries*
- each round i, the item pool is sampled uniformly $I \sim \mathcal{U}[0,5]^3$ and each agent j receives a hidden utility function sampled uniformly, $u_j \sim \mathcal{U}[0,10]^3$ denoting how rewarding a given item is for agent j
- The game plays $N \sim \mathcal{P}[7]$ rounds, agents alternating steps, where for each time-step t the two agents exchange *messages* $m_t^j \in \mathbb{R}^n$ and *proposals* $p_t^j \in [0,5]^3$
- Either agent can terminate the negotiation at some time-step t with signifies agreement with the most recent proposal made by the other agent, resulting in terminal rewards

$$R_{term} = u_{term}^T(i - p_{t-1}^{no-term}) \tag{1}$$

$$R_{no_term} = u_{no-term}^T(p_{t-1}^{no_term}) \tag{2}$$

$$\tag{3}$$

  where *term* denotes the agent who signalled the termination, and $no - term$ is the other agent
- invalid proposals that are accepted result in no rewards for either agent

2.2 Communication Channels

Proposal Channel

- directly communicate via the transmission of proposed division p_j
- since action space if finite and discrete, this upper bounds the amount of information that can be passed between agents at ≈ 10 bits

Linguistic Channel

- *Non-bindingness*: messages sent via this channel do not commit the sender to any course of action, unlike directly transmitting a proposal
- *Unverifiability*: no direct link between the linguistic utterance and private proposal made, so agent can lie
- hence it is a instantiation of **cheap talk** communication
- interested to see if agents can make use of this channel to facilitate negotiation and establish a common ground for the symbol usage

Both channels

- agents have access to both proposal and linguistic channel

No communication

- agents have no communication. In the case of the proposal channel, the opponent doesn't see the request, but the item division is still computed

2.3 Reward schemes

Reward: $R = \alpha \cdot R_A + \beta R_B$ the joint reward for both agents

- Selfish agent: $\alpha = 1$ $\beta = 0$
- Prosocial agent: $\alpha = \beta = 1$

2.4 Agent Architecture and learning

Proposer receives three inputs at step t:

1. item **context** $c^J = [i; u_j]]$ the item pool and proposer's utilities
2. m_{t-1} the **message** produced by the other agent in the previous time step; *zeros* if linguistic channel is off
3. p_{t-1} the **proposal** produced by the other agent in the previous time step; *zeros* if the proposal channel is off

Embed the discrete tables into dense vectors using an embedding $\mathcal{E}_1$ for the item context and proposal and $\mathcal{E}_2$ for the linguistic message

Input sequence fed through an LSTM producing 3 fixed size vectors: h_t^c, h_t^m, h_t^p feed through MLP to produce:

$$\pi_{term} \in [0,1] : \textit{probability of termination action} \tag{4}$$

$$\pi_{utt} \in \mathbb{R}^n : \textit{the linguistic message to be sent} \tag{5}$$

$$\pi_{pro} \in [0,5]^3 : \textit{the proposal to be sent} \tag{6}$$

The negotiation result in a sequence of tuples:

$$\tau = \langle (e_t, u_t, p_t) \rangle_{t=1,\ldots,N} \tag{7}$$

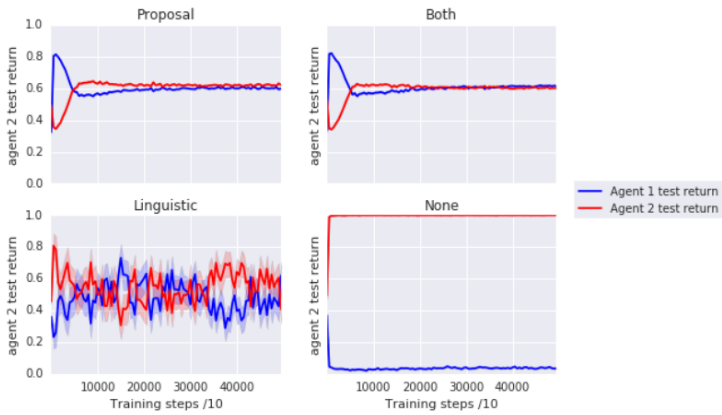
During training each agent maximizes the objective:

$$\pi_i^* = argmax_{\pi_i, \tau \sim (\pi_A, \pi_B)} \mathbb{E}[R_i(\tau) + \lambda \mathcal{H}(\pi_i)] \tag{8}$$

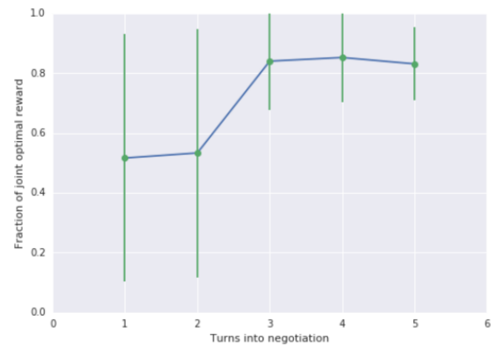
using REINFORCE style updates.

3 Experiments

3.1 Experiment 1: Can self-interested agents learn to negotiate?



(a)



(b)

- self-interested agents learn to divide up the items fairly when they exchange proposals directly
- when using the linguistic channel, agents do not negotiate optimally. Instead the agents randomly alternate taking all the items

Item pool: [5, 5, 1]	A utilities: [8, 7, 1]	B utilities: [8, 3, 2]
Turn	Agent	Proposal
1	A	[3, 4, 4]
2	B	[4, 2, 0]
3	A	[3, 4, 0]
4	B	[4, 1, 0]

3.2 Experiment 2: Can pro-social agents learn to coordinate?

Agent sociality		Proposal	Linguistic	Both	None
Self-interested	Fraction of joint reward	0.87 ± 0.11	0.75 ± 0.22	0.87 ± 0.12	0.77 ± 0.22
	25^{th} & 75^{th} percentiles	[0.81, 0.95]	[0.61, 0.94]	[0.81, 0.95]	[0.62, 0.97]
	Turns taken	3.55 ± 1.12	5.36 ± 1.20	3.43 ± 1.10	3.00 ± 0.13
Prosocial	Fraction of joint reward	0.93 ± 0.10	0.99 ± 0.02	0.92 ± 0.11	0.95 ± 0.11
	25^{th} & 75^{th} percentiles	[0.89, 1.0]	[1.0, 1.0]	[0.88, 1.0]	[0.93, 1.0]
	Turns taken	3.10 ± 0.99	3.71 ± 0.58	2.98 ± 0.97	2.27 ± 0.69

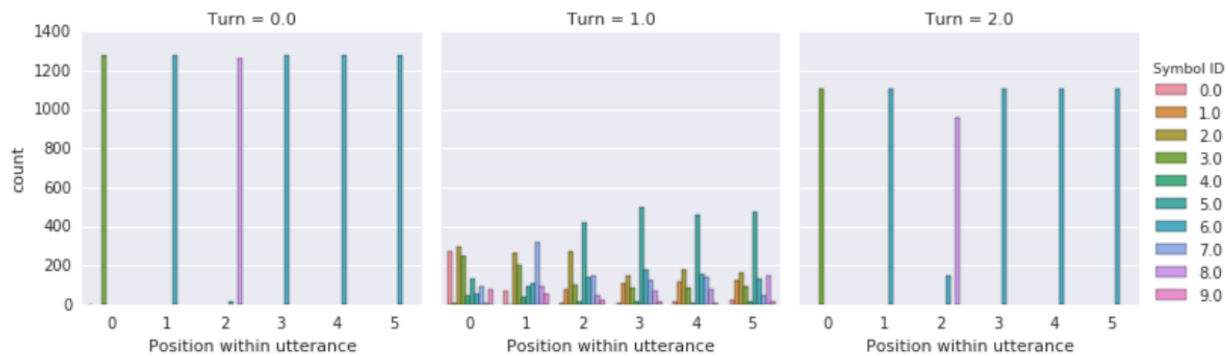
- looking at the joint utility, linguistic channel results in much better task success than any other communication scheme.
- in line with theory that says communication using cheap talk is Nash equilibrium
- both channels don't do well, likely due to optimization difficulty: proposal channel is pre-grounded making it a strong local optima

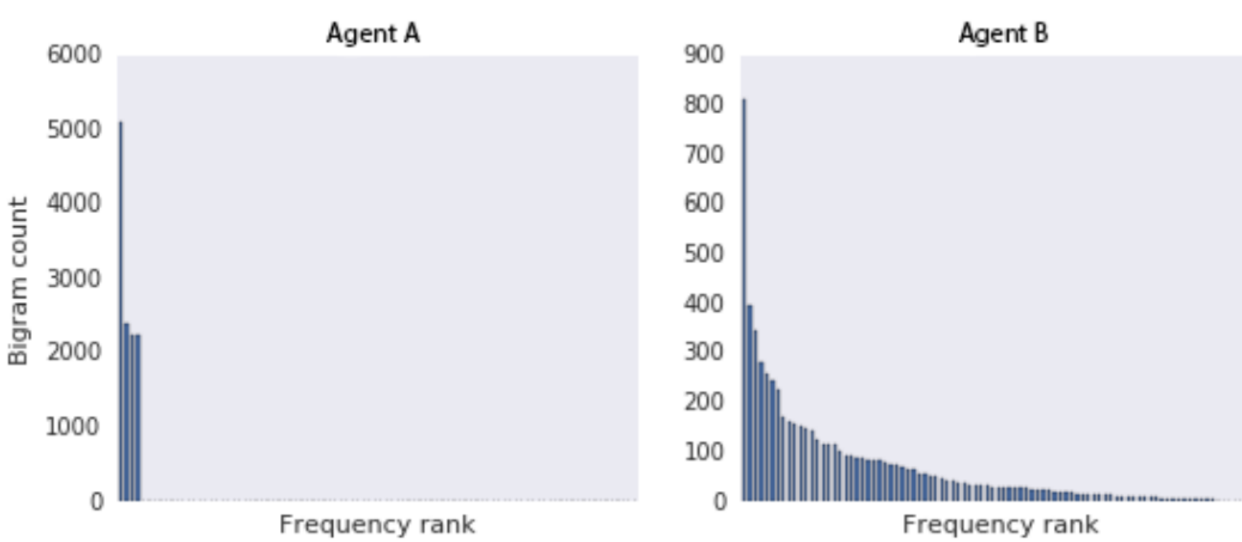
Item pool: [4, 5, 1]		A utilities: [1, 7, 9]	B utilities: [7, 0, 10]
Turn	Agent	Linguistic utterance	Proposal (hidden)
1	A	[3, 3, 3, 3, 3, 3]	[0, 5, 1]
2	B	[6, 6, 6, 4, 7, 4]	[4, 0, 1]
3	A	[3, 3, 3, 3, 3, 3]	[0, 5, 0]

- proposal pro-social agents seems to be using the proposal channel to learn about each others preferences. And hence, when we manually terminate early, results are bad. Intuitively, the proposal channel is not being used for actual proposals until later in the game.

3.2.1 Analysis of linguistic Communication

Symbol Usage





- interestingly, the agent A that starts the negotiation, is not communicating any information using the linguistic channel.
- agent B does use a diversity of symbols, resulting in long-tailed bigram symbols distributed like *Zipf* law would predict for informative languages.

- thus, even though the tasks are symmetric, the agents take on nature roles, where the agent A shares its utilites, while the other agent compares the shared utility with their own, and generates an optimal counter-proposal

Content of the Messages

Agent types	% agent A value correct	% agent B value correct	% proposal correct
Selfish-Selfish	21	25	94 (94)
Prosocial-Prosocial	26	81	80 (57)
Random baseline	20	20	17

- training probe classifiers, based on the information content being passed, we see that the messages also have semantic content:

$$\textit{they reduce the entropy about the hidden utilities of the other agent} \tag{9}$$

- purely selfish agent's do not seem to transmit meaningful symbols in contrast

3.3 Experiment 3: A society of Agents

			Fixed agent A			Fixed agent B		
Fixed agent type	# proso-cial agents	IDs given	Proposal	Linguistic	Both	Proposal	Linguistic	Both
Self	1	False	0.81 ± 0.17	$\mathbf{0.97 \pm 0.05}$	0.87 ± 0.11	0.76 ± 0.18	$\mathbf{1.0 \pm 0.0}$	0.76 ± 0.16
		True	$\mathbf{0.96 \pm 0.06}$	$\mathbf{0.97 \pm 0.10}$	$\mathbf{0.97 \pm 0.04}$	$\mathbf{0.95 \pm 0.08}$	$\mathbf{1.0 \pm 0.0}$	$\mathbf{1.0 \pm 0.0}$
	5	False	0.82 ± 0.17	$\mathbf{1.0 \pm 0.0}$	0.98 ± 0.07	0.88 ± 0.14	$\mathbf{1.0 \pm 0.0}$	0.88 ± 0.14
		True	$\mathbf{1.0 \pm 0.01}$	$\mathbf{1.0 \pm 0.0}$	$\mathbf{1.0 \pm 0.01}$	$\mathbf{1.0 \pm 0.0}$	$\mathbf{1.0 \pm 0.0}$	$\mathbf{1.0 \pm 0.02}$
Pros.	1	False	0.91 ± 0.11	0.95 ± 0.09	0.91 ± 0.12	0.93 ± 0.11	0.95 ± 0.10	0.92 ± 0.11
		True	0.95 ± 0.10	0.93 ± 0.11	0.95 ± 0.09	0.92 ± 0.01	0.96 ± 0.08	0.92 ± 0.10
	5	False	0.90 ± 0.13	$0.98^{*} \pm 0.06$	0.93 ± 0.10	0.92 ± 0.12	0.95 ± 0.08	0.90 ± 0.16
		True	0.94 ± 0.09	0.97 ± 0.06	0.94 ± 0.01	0.93 ± 0.10	0.95 ± 0.08	0.92 ± 0.10

A more realistic scenarios we negotiate amongst a diverse population of agent, with varying selfishness. In this settings, maximizing ones rewards requires finding which agents are most prosocial: they can be exploited best by selfish agents, and cooperate best with pro-social agents.

- Training a community of 10 agents, a submit of them being prosocial, we see if selfish agents can exploit others, and prosocial agents can coordinate with others.
- agent's are optionally labelled with one-hot-vectors, which improves performance uniformly, allowing agents to build models of which agents are best to interact with