Title

Reinforcement of Semantic Representations in Pragmatic Agents Leads to theEmergence of a Mutual Exclusivity Bias

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Ohmer, Xenia
Konig, Peter
Franke, Michael

Publication Date

2020

Peer reviewed

Reinforcement of Semantic Representations in Pragmatic Agents Leads to the Emergence of a Mutual Exclusivity Bias

Xenia Ohmer (xenia.ohmer@uni-osnabrueck.de)
Institute of Cognitive Science, University of Osnabrueck, Germany

Peter König* (peter.koenig@uni-osnabrueck.de)
Institute of Cognitive Science, University of Osnabrueck, Germany

Michael Franke* (michael.franke@uni-osnabrueck.de)
Institute of Cognitive Science, University of Osnabrueck, Germany

*Authors contributed equally.

Abstract

We present a novel framework for building pragmatic artificial agents with explicit and trainable semantic representations, using the Rational Speech Act model. We train our agents on supervised and unsupervised communication games and compare their behavior to literal agents lacking pragmatic abilities. For both types of games pragmatic but not literal agents evolve a mutual exclusivity bias. This provides a computational pragmatic account of mutual exclusivity and points out a possible direction for solving the *mutual exclusivity bias challenge* posed by Gandhi and Lake (2019). We find that pragmatic reasoning can cause the bias either by promoting lexical constraints during learning, or by affecting online inference. In addition we show that pragmatic abilities lead to faster learning and that this advantage is even stronger when meanings to be communicated follow a more natural distribution as described by Zipf's law.

Keywords: mutual exclusivity; reinforcement learning; Rational Speech Act model; gradient-based learning

Introduction

(Deep) neural networks have surpassed not only other learning architectures but also human performance in a great variety of tasks (e.g., Schmidhuber, 2015; Li, 2017). Still, in most domains humans learn more efficiently and apply their knowledge in more flexible ways (e.g., Lake, Ullman, Tenenbaum, & Gershman, 2017). This is also true for the domain of language. In terms of language learning, humans are equipped with useful biases, at least some of which are not exploited by neural networks (Gandhi & Lake, 2019). In terms of language use, humans easily produce and interpret utterances based on the context such that their meanings are much richer than their semantic content. This is an ability unmatched by neural networks (Hirschberg & Manning, 2015; Jacquet, Masson, Jamet, & Baratgin, 2019). In order to approach the design of artificial agents that are able to quickly acquire a grounded language and flexibly communicate with humans it seems promising to take inspiration from the cognitive mechanisms at play in human language use.

In this work we develop such a design and show how it induces an important word learning constraint, the mutual exclusivity bias. Mutual exclusivity (ME) describes the tendency to avoid assigning a second name to an object that already has a name. In a standard ME paradigm, when children are presented with two objects and know the label for one of them, they will tend to associate a new label with the other object (Markman & Wachtel, 1988). While the ME bias allows for fast language learning in humans as it helps to disambiguate how words map to referents, Gandhi and Lake (2019) showed that neural networks lack this bias and even have the reverse tendency of selecting a familiar class when presented with input from a class that was not part of the training. Therefore, capturing the ME bias is not only a major concern for agents learning word meanings but for any model performing categorization.

There are different theories about the mechanisms underlying the ME bias. The lexical constraint account, for example, proposes that due to an innate or early emerging constraint children are biased towards lexica which favor one-to-one mappings between states and messages (Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003). Our model focuses on an alternative explanation based on pragmatic reasoning. When producing and interpreting utterances humans reason about each other's intentions and take into account contextual influences on meaning (H. H. Clark, 1996). The field of pragmatics studies exactly this crucial aspect of human language use. Under a pragmatic account the bias arises from the assumption that the speaker follows cooperative principles of communication (e.g., E. V. Clark, 1988), leading to the following reasoning process in a standard ME paradigm: 'If the speaker wanted to talk about object x she would use its label l, which we both know. Given that she used a novel label k, this must refer to the other object y, for which I do not know a label.' Thus, among others, pragmatic inference offers a possible explanation for the ME bias.

There are various ways of modeling pragmatic behavior. We use a particularly prominent model of pragmatics, the Rational Speech Act (RSA) model (Frank & Goodman, 2012). It treats communication as a recursive process in which speaker and listener reason about each other's mental states to complement the literal meanings of utterances. The RSA framework has successfully modeled various pragmatic phenomena (e.g., Scontras, Tessler, & Franke, 2018), among others the ME bias. For example, Frank, Goodman, and Tenenbaum (2009) showed how RSA-like pragmatics leads to an ME bias in cross-situational word learning. Their architecture was ex-

tended to evaluate the pragmatic account and the lexical constraint account as two competing explanations of the bias, with the result that either mechanism is sufficient (M. Lewis & Frank, 2013). Smith, Goodman, and Frank (2013) applied the RSA model to iterated pragmatic inference games in a language learning setting, for which they demonstrate an ME bias, as well as a language emergence setting. In the aforementioned models, which we will group together as *probabilistic pragmatic word learning models*, the agents infer the most likely lexicon via Bayesian inference, given a history of observed state-message pairs. This line of work demonstrates the successful formalization of the ME bias in word learning with RSA-like models.

We present a new model of learning in pragmatic agents and show how the ME bias arises naturally within this framework. The model is a straightforward adaptation of the RSA model to gradient-based learning settings. We train pragmatic agents' mental lexica with reinforcement learning in a single agent word learning setting and in a two agent language emergence setting, using an explicit representation of the lexicon in the form of a matrix. Due to pragmatic reasoning about alternatives, agents utilize the entire lexicon when producing and interpreting an utterance, which leads to gradient-based updates of all state-message mappings for a single input example. These lateral effects can be inhibitory as well as facilitating, in sum leading to the emergence of an ME bias.

This setup allows us to make several contributions. 1) While the probabilistic pragmatic word learning models estimate the entire lexicon based on a history of collected evidence, using gradient-based learning as in our model makes it unnecessary to track past observations. The successive updates of particular associations based on individual samples are arguably more natural and can more easily be used in AI applications. 2) We show that the ME bias can arise from either, a *lexical* ME bias caused by pragmatic inference during training, or pragmatic inference at test time. The former suggests that a pragmatic approach can in principle accommodate the lexical constraint account. 3) We demonstrate that the ME bias leads to faster learning with an even stronger effect when the probability over meanings to be communicated is more naturally distributed. 4) We extend these analyses to a language emergence setting.

Model

Agents and training were implemented with Tensorflow 2.0 (Abadi et al., 2015). The project is entirely open-source and accessible via GitHub.[1]

Pragmatic Agents

The agents in our model feature two main components: 1) explicit lexical representations and 2) rules of pragmatic behavior telling them how to use these representations to produce and interpret messages. The lexicon is a matrix providing a mapping between states and messages. If there are N states and messages, the lexicon B_A of agent A is an $N \times N$ matrix. Each matrix entry $B(s_i, m_j) \in \mathbb{R}^+$ is an unnormalized value of how appropriate (in a semantic sense) message j is for state i. The matrix entries are the only trainable parameters of the model. For modeling the pragmatic rules the vanilla RSA model is used. In the RSA model conditional probabilities describe how speakers produce and listeners interpret utterances, recursively taking into account each other's reasoning process. It can be formalized as

$$LL(s \mid m) \propto [\![m]\!](s) \times P(s)\,, \tag{1a}$$

$$PS(m \mid s) \propto \exp\left(\alpha \times [\log LL(s \mid m) - C(m)]\right), \tag{2a}$$

$$PL(s \mid m) \propto PS(m \mid s) \times P(s)\,. \tag{3a}$$

At the basis of the recursive reasoning process is a literal listener (1a) who maps a message onto any state for which it is true, at the same time considering the prior probability of that state. In Equation (1a), $[\![m]\!](s)$ is the denotation function returning the truth value of message m for state s. A pragmatic speaker (2a) chooses its messages such that the probability of being understood correctly by a literal listener is maximized while production cost, $C(m)$, stays low. The parameter $\alpha \in \mathbb{R}^+$ regulates the speaker's optimality. The pragmatic listener (3a) in turn interprets a message as if coming from a pragmatic speaker, also taking into account the prior probability of states.

We adapt the vanilla model to our purpose in several ways. The standard RSA model assumes that all agents have access to the same lexicon which is a truth table of messages across states. In our case every agent learns the (real-valued) entries of its own lexicon which it also uses for reasoning about other agents. We assume a flat prior over states, zero costs for every utterance, and set $\alpha = 5$. This leads to the following formalization:

$$LL(s \mid m, B_{LL}) \propto B_{LL}(s, m)\,, \tag{1b}$$

$$PS(m \mid s, B_{PS}) \propto LL(s \mid m, B_{PS})^5\,, \tag{2b}$$

$$PL(s \mid m, B_{PL}) \propto PS(m \mid s, B_{PL})\,. \tag{3b}$$

In addition, we include a literal speaker defined analogously to the literal listener in (1b). Generic versions of listener and speaker with either reasoning ability are denoted by L and S.

Communication Game

We investigate the performance of pragmatic and literal agents in a signaling game. Specifically, we have chosen the Lewis game (D. Lewis, 1969) which is used to study the emergence of meaningful language from initially random messages. There are N states and N messages and two agents, a speaker and a listener. One round of the game proceeds as follows. The world state (target) is sampled from a state prior —which we assume to be uniform unless otherwise mentioned— and observed by the speaker, who then selects a message to convey this world state to the listener. The listener

[1] https://github.com/XeniaOhmer/pragmatic_agents_me_bias

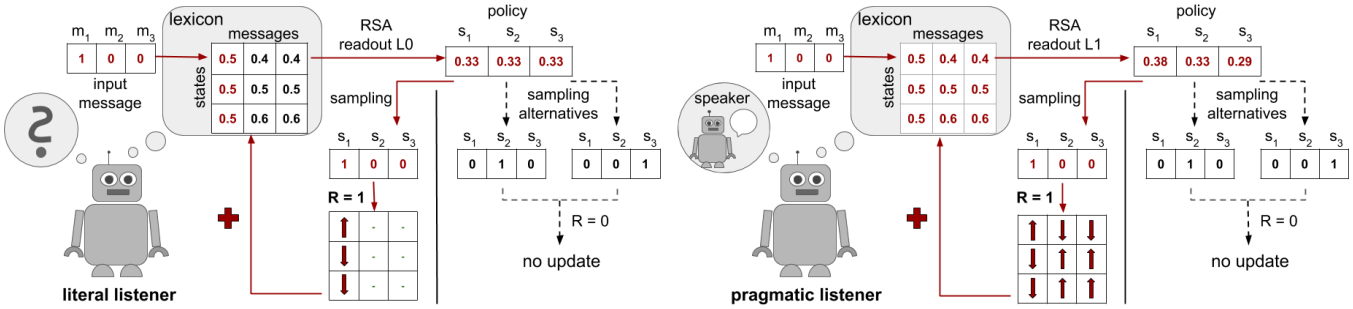

Figure 1: Example of one training step for a literal and a pragmatic listener. The agents receive an input message for which a policy is calculated dependent on their literal or pragmatic interpretation. In contrast to the policy of the literal listener the policy of the pragmatic listener depends on the entire lexicon, which is also why they are different despite identical input and lexicon. This difference also leads to different reinforcement effects (thick red arrows).

in turn selects a world state based on this message. If selection and target are the same both players receive a positive reward, otherwise they receive zero reward.

We look at two different versions of the game, a *single agent setting* and a *two agent setting*. The single agent setting is essentially a supervised learning setting. There is a predetermined state-message mapping which is known to the speaker. The listener is trained to map the speaker's message to the correct state. The two agent setting corresponds to the classical Lewis game where the state-message mapping is not predetermined but emerges during game play. The two settings allow us to verify the benefits of pragmatic reasoning in different language learning contexts, one where an existing language is learned and one where a new one is created.

Reinforcement Learning

Based on lexicon and reasoning abilities, the speaker maps an input state to a probability distribution over messages, $S(m \mid s, B_S)$, and the listener maps an input message onto a probability distribution over states, $L(s \mid m, B_L)$. These distributions define the agents' policies. The agents minimize a loss function defined as the negative expected reward (R):

$$\mathcal{L}(B_S, B_L) = -\mathbb{E}[R]\,.$$

The matrix entries are updated with REINFORCE (Williams, 1992), which belongs to the family of policy gradient algorithms. Figure 1 illustrates a training step for a literal (left) and a pragmatic (right) listener. Here, both agents receive the same input message and have the same lexicon. Given an input message, the literal listener only takes into account the semantic meaning and thus normalizes across the corresponding column to obtain a selection policy. Accordingly, if there is a positive feedback from the environment, the entries in that column are updated. The pragmatic listener, in contrast, calculates for each state which message the pragmatic speaker would have used to refer to that state, to determine its selection policy given the actual message. Due to this consideration of alternatives all lexicon entries are updated upon positive feedback. The thick red arrows show how updates based on the pragmatic policy calculation automatically reinforce a bijective mapping between the correctly identified state-message pair and at the same time strengthen all associations between the other states and messages. By this mechanism associations between unseen states and messages are reinforced, leading to the ME bias.

Experiment 1: Mutual Exclusivity Bias

In the first experiment we ran different simulations to establish whether literal or pragmatic agents evolve an ME bias, how this bias develops over the course of training, and whether the bias is due to pragmatic inference at training time (by changing the lexicon) or test time.

Methods

We looked at $N \in \{3, 10\}$ states and messages. Importantly, $K \in \{1, 2\}$ messages (single agent setting) or states (two agent setting) were withheld from training and reserved for testing the agents' behavior on novel inputs. For each of the four combinations the agents were trained on the remaining $N - K$ inputs. The lexica were initialized randomly with $b_{i,j} \sim \mathcal{N}(0.5, 0.01)$ and updated using the Adam optimizer (Kingma & Ba, 2015). In every epoch the agents saw the same 1000 random instances of the $N - K$ training inputs. The last or the last two states or messages (by index) were withheld from training. For every game and hyperparameter setup we trained 100 listeners or speaker-listener-combinations.

The emergence of an ME bias was evaluated by calculating the listener's average probability of selecting any novel state —that was not part of the training— when presenting the K test examples to the system. We formalize this idea in terms of an *ME index*:

$$I_{ME} = \frac{p(\text{new state selected} \mid \text{new input}) - p(\text{new input})}{p(\text{old input})}\,.$$

If the probability of selecting a novel state given a novel input is at chance level the ME index is equal to zero and if the entire conditional probability mass is on the new states it is

equal to one. With $M = N - K + 1$ the ME index for the single agent setting is

$$I_{ME}(B_L) = \frac{\frac{1}{K}\sum_{i=M}^{N}\sum_{j=M}^{N} L(s_i \mid m_j, B_L) - \frac{K}{N}}{\frac{N-K}{N}},$$

and for the two agent setting it is

$$I_{ME}(B_L, B_S) = \frac{\frac{1}{K}\sum_{i=M}^{N}\sum_{j=M}^{N}\sum_{q=1}^{N} L(s_i \mid m_q, B_L) S(m_q \mid s_j, B_S) - \frac{K}{N}}{\frac{N-K}{N}}.$$

Single Agent Setting In the single agent setting we compare literal and pragmatic listeners in a supervised learning setting. The correct state-message mappings are predetermined such that every state is associated with the message of the same index $\{(s_1, m_1), (s_2, m_2), \ldots, (s_N, m_N)\}$. For $N = 3$ we trained for 50 and for $N = 10$ for 100 epochs.

Two Agent Setting In the two agent setting we compare a literal speaker and listener with a pragmatic speaker and listener playing a full Lewis game together. Here, the state-message mapping is not predetermined but emerges during game play. During training the agents continuously adapt to each other's behavior which makes longer training necessary. For $N = 3$ we trained for 100 and for $N = 10$ for 1000 epochs.

Ablation Test We run an ablation test to disentangle the effects of pragmatic reasoning at training and test time on the ME bias. For the single agent setting we take the lexica acquired by the literal listeners and evaluate pragmatic listeners with these lexica, and vice versa. In analogy, in the two agent setting the pragmatic reasoning ability of both agents is switched.

Results

Single Agent Lewis Game In the single agent setting all agents converged to maximum performance. Figure 2 shows the average lexica of the agents as well as their average probability of selecting the different states when presented with the test data. First, we will analyze the literal listener (left). Looking at the average lexicon we see that the probability of matching a learned message to the correct state is close to one (diagonal) whereas all possible errors (off-diagonal) have values close to zero. When presented with a test message the selection probability is equally high for all states with almost no variance between individual agents. The flat probability distribution of states for new messages demonstrates the absence of an ME bias in this simple system. Let us now turn to the pragmatic listener (right). Also here the learned one-to-one mapping is clearly visible in the average lexicon. In contrast to the literal listeners, the pragmatic listeners have a high probability of selecting a new state when presented with a new input message. If two messages are excluded from training, the two states are selected with equal probability irrespective of the test message. This can be concluded from the equal means and small standard deviations. Table 1 (top) summarizes the results for the single agent setting in terms of ME indices. The literal listener's ME index is equal to zero in all cases which means that selection of new states is at chance level and there is no ME bias. The pragmatic listener's ME index is equal to one in all cases which means that the entire conditional selection probability mass is on the new states and the ME bias is maximal.

Table 1: ME indices ($\mu(\pm\sigma)$) for Experiment 1.

	N × K	Literal	Pragmatic
1 agent	3 × 1	0.00(±0.01)	1.00(±0.00)
	3 × 2	0.00(±0.01)	1.00(±0.00)
	10 × 1	0.00(±0.00)	1.00(±0.01)
	10 × 2	0.00(±0.00)	1.00(±0.00)
2 agent	3 × 1	−0.5(±0.00)	0.99(±0.00)
	3 × 2	−2.00(±0.00)	0.96(±0.12)
	10 × 1	−0.11(±0.00)	0.84(±0.26)
	10 × 2	−0.25(±0.00)	0.84(±0.18)

Two Agent Lewis Game In the two agent Lewis game, for $N = 3$ all agents converged to optimal performance while for $N = 10$, 84% and 88% of the literal agents and 88% and 91% of the pragmatic agents converged to optimal performance, for $K = 1$ and $K = 2$ respectively. Given that state-message mappings are arbitrary we do not plot the average lexica but look directly at the ME indices at the bottom part of Table 1. When a literal speaker plays with a literal listener, the ME index becomes negative. So when the speaker is presented with a previously unseen state and generates a message the listener will select one of the novel states with less than chance probability. This means that an anti-ME bias emerges exactly as in standard neural networks. This is different for the pragmatic speaker-listener combination. When presented with the test states the probability of selecting a novel state is clearly higher than chance with ME indices between 0.84 and 0.99. In conclusion the pragmatic, but not the literal agents, develop an ME bias in the supervised as well as the unsupervised communication game.

ME Bias Development Over Time Investigating how the ME bias develops over time reveals whether it can be exploited early during training or emerges first when the training examples are already consolidated. Figure 3 shows the average reward and ME index over the course of training for both game settings and system sizes, always for the case where one example was excluded from training ($K = 1$). In the single agent setting the literal agents have a constant ME bias of zero because the relevant lexicon entries are not updated during training. For the pragmatic agents the ME index starts to increase with training onset and converges to one at approximately the same time as the reward. In the two agent setting the ME index of the literal agents quickly converges to a negative value. The agents unlearn selecting states that are not in the training data as this never yields a reward. For the pragmatic agents the ME index starts increasing immediately with the rewards but does so more slowly than in the

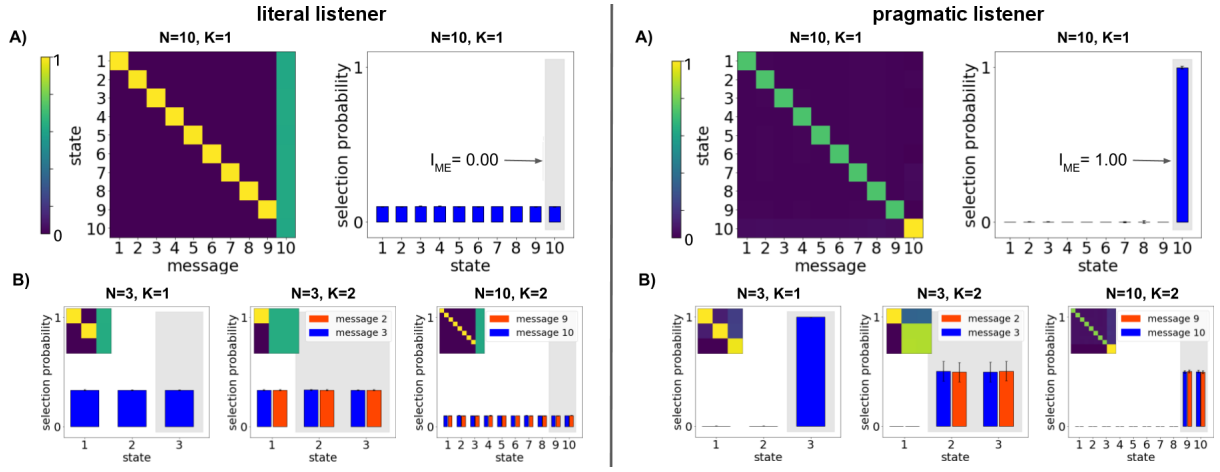

Figure 2: Average lexica and state selection probabilities given new input messages for literal (left) and pragmatic (right) listeners. N is the total number of states and messages, K the number of messages that were excluded from training. A) shows the results for $N = 10$ and $K = 1$, the other combinations are shown in B). The average lexicon is standardized such that the maximum value is 1. The bar plots depict the average selection probabilities, with standard deviations, for new input messages. Novel states are marked by a gray background. For $K = 2$ the probabilities are presented for both new input messages separately.

single agent setting. In summary, in both types of games pragmatic agents can exploit the ME bias from training onset and the bias grows stronger with an increasing certainty about the correct state-message mappings.

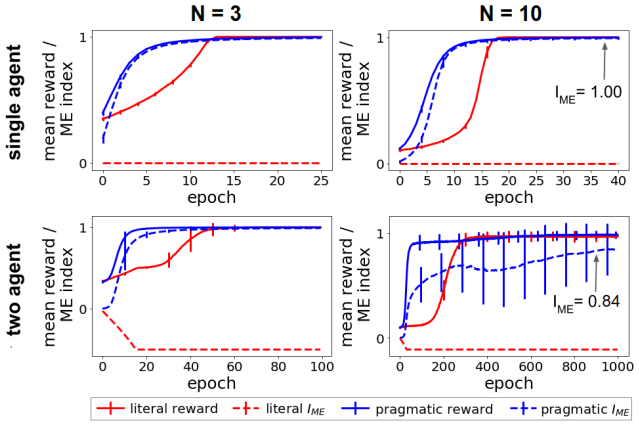

Figure 3: Average ME indices and rewards during training for literal and pragmatic agents. Shown are the simulations where one example was excluded from training in single and two agent setting (rows) with three or ten states (columns). While the rewards correspond to the performance on the training examples, the ME indices reflect the performance on the novel examples at test time.

Pragmatic Reasoning at Training Versus Test Time Here, we look at the results from the ablation test, where agents have pragmatic reasoning abilities only at training or test time. We report the mean ME indices of the four different training conditions as in Table 1 for a direct comparison. Starting with the single agent setting, the ME bias is very strong when using pragmatic inference at test time, regardless of the type of lexicon (mean ME indices between 0.99 and 1.00). Interestingly, with the lexicon of a pragmatic speaker, now also literal agents feature an ME bias (mean ME indices between 0.42 and 0.54). In short, in the single agent setting pragmatic reasoning at either training or test time leads to an ME bias. In the two agent setting the lexica of pragmatic speakers and listeners also allow their literal counterparts to display a small ME bias at test time (mean ME indices between 0.11 and 0.26). However, using the lexica of the literal agents never leads to an ME bias, regardless of the type of reasoning at test time. The literal listeners learn a lexicon where the rows of the missing states only have zero entries, which prevents the selection of novel states. Accordingly, in the two agent setting pragmatic reasoning during training always leads to an ME bias, whereas training literal agents makes an ME bias impossible.

Tracking the ME indices over the course of training (not shown here) further reveals that the ME bias always emerges at training onset, regardless of whether it is caused by lexical constraints or online inference. In general, using the bias strength to determine whether the effect of pragmatic reasoning at training or test time is more influential is not straightforward, as we found the quantitative results to depend on the optimality parameter α. We leave a detailed analysis of the role of α to future work and conclude with the main finding: pragmatic inference can lead to an early emerging ME bias in single agent word learning both by its effect on the lexicon during learning and by its effect on the online reasoning process.

Experiment 2: Convergence Time Scales

The pragmatic agents, in contrast to the literal agents, update all model weights with each training example they encounter. Here, we investigated whether this mechanism, aside from leading to an ME bias, also results in faster learning.

Methods

In a first step we compared the learning time scales of pragmatic and literal agents in the single and the two agent setting of a Lewis game. The update of additional state-message mappings by the pragmatic agents should be particularly effective for rare examples, where they can leverage the ME bias. Hence, in a second step we tested whether the pragmatic advantage is more pronounced if the frequency of input classes does not follow a uniform distribution but is rather inversely related to their frequency rank as suggested by Zipf's law, which corresponds to a more realistic setting (Zipf, 1949).

Uniform Input Distribution We trained 50 listeners (single agent setting) and 50 speaker-listener combinations (two agent setting) on $N = 10$ states and messages. The data set size was again 1000 and the agents' lexica were initialized randomly with $b_{i,j} \sim \mathcal{N}(0.1, 0.01)^2$. The Adam optimizer can update weights based on previous training steps even though their gradient in the current training step is zero. To remove this confounding effect we trained the agents with vanilla stochastic gradient descent (SGD). In the single agent setting the agents were trained for 100 epochs and in the two agent setting for 300 epochs.

Zipfian Input Distribution Let N be the number of states and i the frequency rank of state s_i, then the occurrence probability of s_i is given by $P(s_i) \propto 1/i$. The Zipfian input distribution for messages is analogous to the one for states. We ran the exact same simulations as for the uniform distribution except that training was extended to 500 epochs in the two agent setting.

Results

Figure 4 shows the rewards over time for the different game settings and input distributions. The pragmatic agents converge faster than the literal agents in all situations. The literal agents learn more slowly when the input data follows a Zipfian distribution than when it follows a uniform distribution while the pragmatic agents converge approximately equally fast independent of the input statistics. In short, the pragmatic agents learn faster than the literal agents and this advantage is stronger for data that follows a more natural distribution.

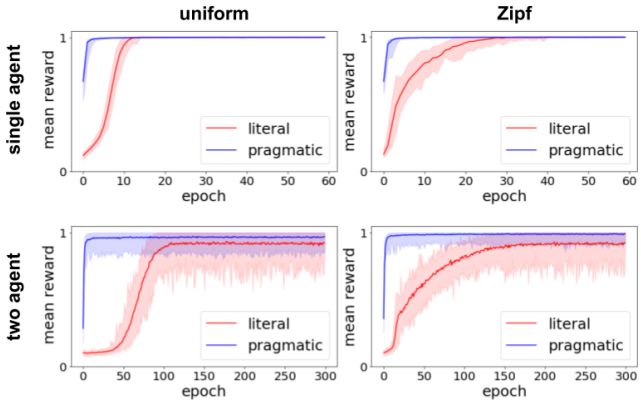

Figure 4: Average reward over time for literal and pragmatic agents trained in the single or two agent setting (rows) and with a uniform or Zipfian input distribution (columns). The shaded areas range from minimum to maximum reward.

Discussion

The main contribution of this work is the development of a framework in which explicit mental representations of semantic meaning evolve through reinforcement learning under pragmatic policies. Working with an explicit lexicon not only has the advantage that pragmatic agents can make very efficient updates but also makes the agents' behavior fully explainable as their knowledge is stored in a structured and accessible way. Yet, the matrix representation also imposes limitations: in some tasks it might be difficult to define states and messages to begin with. Moreover, for systems with a large number of states computational cost increases significantly. In contrast to other approaches applying RSA in gradient-based learning systems (e.g., Andreas & Klein, 2016; Monroe & Potts, 2015) we do not address the challenge of working with more flexible representations. Still, we think our framework is useful for language evolution and language-related AI research. For instance, future work could use the division of labor between semantics and pragmatics as an approach to transfer learning as the same representations can be learned and applied in different tasks.

Within our framework pragmatic agents develop an ME bias in word learning as well as language emergence. Importantly, this bias emerges at learning onset and can be exploited during the entire training process. At the same time pragmatic reasoning causes an acceleration of learning. We also establish that the positive effect on learning is stronger when word occurrence frequencies follow real-world statistics —a phenomenon that can be explained by mutual exclusivity. So far, we mainly evaluate the ME bias. In future work we would like to more closely investigate the role of ME in the actual learning process. So, while more functional details should be examined, a principled way of building artificial agents with an ME bias has been established.

Interestingly, both pragmatic reasoning during learning and during online inference can effect the ME bias. During learning the reasoning about alternatives causes lateral effects in the matrix updates, such that a *lexical* ME bias arises. Note, that such lateral effects used to be hard-coded in early models

[2] In the second experiment we used smaller initialization values, which speeds up the training (as most values converge to zero). Additional tests showed that swapping initializations in the two experiments affects the training duration but none of the qualitative results.

of language emergence (e.g., Steels, 1995). The early emergence of a lexical bias sheds new light on the debate whether the ME bias arises from lexical constraints or online pragmatic inference: we propose that both mechanisms can stem from the same principle of pragmatic reasoning.

Our work is an important step towards solving the challenge of integrating an ME bias into neural networks. Just like literal agents, neural networks have an anti-ME bias with negative effects on learning. Note that our system is implemented with modern deep learning algorithms and can easily be integrated with neural network modules. However, a problem arises if a neural network is used to classify input stimuli as certain types of states or messages which then serve as inputs to a pragmatic agent. If the network encounters stimuli that are not part of any training class it will fail to map them onto a new input type, preventing the agents from using the ME bias. A promising approach to this problem are methods for performing out-of-distribution recognition by extracting confidence estimates from neural networks (e.g., DeVries & Taylor, 2018; Liang, Li, & Srikant, 2018). Applied to our problem the network could then identify a novel stimulus class and map it onto a novel label, leaving the ME bias intact. In summary, we created gradient-based learning agents which develop an ME bias and with some additional work this idea may be integrated with neural networks.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - GRK 2340. We would like to thank the anonymous reviewers for their thoughtful and constructive feedback.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from `http://tensorflow.org/`

Andreas, J., & Klein, D. (2016). Reasoning about pragmatics with neural listeners and speakers. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1173–1182). Austin, Texas: Association for Computational Linguistics.

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language*, *15*(2), 317–335.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

DeVries, T., & Taylor, G. W. (2018). Learning confidence for out-of-distribution detection in neural networks. *arXiv preprint*, *arXiv:1802.04865*.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585.

Gandhi, K., & Lake, B. M. (2019). Mutual exclusivity as a challenge for neural networks. *arXiv preprint*, *arXiv:1906.10197*.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261–266.

Jacquet, B., Masson, O., Jamet, F., & Baratgin, J. (2019). On the lack of pragmatic processing in artificial conversational agents. In T. Ahram, W. Karwowski, & R. Taiar (Eds.), *Human Systems Engineering and Design* (pp. 394–399). Cham: Springer International Publishing.

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations.*

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253.

Lewis, D. (1969). *Convention*. Cambridge, MA: Harvard University Press.

Lewis, M., & Frank, M. C. (2013). Modeling disambiguation in word learning via multiple probabilistic constraints. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (CogSci)* (pp. 876–881).

Li, Y. (2017). Deep reinforcement learning: An overview. *arXiv preprint*, *arXiv:1701.07274*.

Liang, S., Li, Y., & Srikant, R. (2018). Enhancing the reliability of out-of-distribution image detection in neural networks. In *International Conference on Learning Representations.*

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275.

Monroe, W., & Potts, C. (2015). Learning in the Rational Speech Acts model. *arXiv preprint*, *arXiv:1510.06807*.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.

Scontras, G., Tessler, M. H., & Franke, M. (2018). *Probabilistic language understanding: An introduction to the Rational Speech Act framework*. Retrieved from `http://www.problang.org`

Smith, N. J., Goodman, N. D., & Frank, M. C. (2013). Learning and using language via recursive pragmatic reasoning about other agents. In *Advances in Neural Information Processing Systems 26* (pp. 3039–3047).

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, *2*(3), 319-332.

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

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.