COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 46 (2021) e13069

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13069



# Mutual Exclusivity in Pragmatic Agents

Xenia Ohmer,[a] Michael Franke,[a,b] Peter König[a,c]

[a]*Institute of Cognitive Science, University of Osnabrück*
[b]*Department of Linguistics, University of Tübingen*
[c]*Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg- Eppendorf*

Received 8 October 2020; received in revised form 11 October 2021; accepted 30 October 2021

**Abstract**

One of the great challenges in word learning is that words are typically uttered in a context with many potential referents. Children's tendency to associate novel words with novel referents, which is taken to reflect a mutual exclusivity (ME) bias, forms a useful disambiguation mechanism. We study semantic learning in pragmatic agents—combining the Rational Speech Act model with gradient-based learning—and explore the conditions under which such agents show an ME bias. This approach provides a framework for investigating a pragmatic account of the ME bias in humans but also for building artificial agents that display an ME bias. A series of analyses demonstrates striking parallels between our model and human word learning regarding several aspects relevant to the ME bias phenomenon: online inference, long-term learning, and developmental effects. By testing different implementations, we find that two components, pragmatic online inference and incremental collection of evidence for one-to-one correspondences between words and referents, play an important role in modeling the developmental trajectory of the ME bias. Finally, we outline an extension of our model to a deep neural network architecture that can process more naturalistic visual and linguistic input. Until now, in contrast to children, deep neural networks have needed indirect access to (supposed to be novel) test inputs during training to display an ME bias. Our model is the first one to do so without using this manipulation.

*Keywords:* Deep learning; Mutual exclusivity; Pragmatics; Rational Speech Act model; Reinforcement learning

## 1. Introduction

Word learning is central to language acquisition. The core problem of word learning is that novel words are typically encountered in situations that offer a multitude of potential refer-

---

Correspondence should be sent to Xenia Ohmer, Institut für Kognitionswissenschaft, Universität Osnabrück, Wachsbleiche 27, 49090 Osnabrück, Germany. Email: xenia.ohmer@uni-osnabrueck.de

ents. To learn the meaning of a new word, children must understand whether the word refers to an object in the scene, and if so whether it refers to the object as a whole or rather to a specific feature or part of it. Next to social, linguistic, and attentional information, inductive biases help children disambiguate the meanings of novel words (e.g., Bloom, 2000; Hollich, Hirsh-Pasek, & Golinkoff, 2000; Markman, 1991). The mutual exclusivity (ME) bias accords with the property of language that word-meaning mappings tend to be bijective (Clark, 1987). In the now classical ME paradigm, Markman and Wachtel (1988) showed that 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. In additional experiments, they demonstrated that this inference mechanism not only helps children to learn labels for whole objects but also for object parts and features. Accordingly, the ME bias supports the identification of referents in ambiguous context and applies to a wide variety of situations, making it a key word learning mechanism.

With some due simplification, the ME-bias behavior observed in the classical ME paradigm is this. An agent is familiar with and able to recognize words $\{w_1, \ldots, w_n\}$ and objects $\{o_1, \ldots, o_n\}$. The agent has learned to associate, for simplicity, $w_i$ with $o_i$ for all $1 \leq i \leq n$ in a one-to-one mapping. The agent has never encountered the word $w_{n+1}$ before, but recognizes it as different from any word in $\{w_1, \ldots, w_n\}$. Similarly, mutatis mutandis, for novel object $o_{n+1}$. The agent now perceives a speaker use $w_{n+1}$ in a referential context $C$, where $C \subseteq \{o_1, \ldots, o_n, o_{n+1}\}$ with $o_{n+1} \in C$. The agent shows an ME bias in the classical ME paradigm if they show an *ME response* by associating the novel object $o_{n+1}$ with the novel word $w_{n+1}$. Notice that the ME bias, thus delineated, presupposes that objects and words are perceived as familiar or novel, and that the speaker's goal of using word $w_{n+1}$ is to refer to exactly one element from the referential context $C$.

There has been a long-standing discussion about the mechanism underlying the ME bias (for an overview, see Lewis, Cristiano, Lake, Kwan, & Frank, 2020). Two proposals dominate the literature. Under the first proposal, the ME bias is a manifestation of an innate or early emerging constraint. One important version of this proposal is the *lexical constraint account*. It posits that children are biased to consider only lexica with one-to-one mappings between words and objects (Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003). If only one-to-one mappings between words and objects are feasible, then from the assumed one-to-one associations of $w_i$ to $o_i$ for all $1 \leq i \leq n$, the only plausible mapping of novel word $w_{n+1}$ is to novel meaning $o_{n+1}$. A second prominent approach uses a pragmatic explanation. Pragmatics studies how humans reason about each other's intentions and take into account contextual factors when producing and interpreting utterances (Clark, 1996). Social-pragmatic reasoning abilities, such as understanding others' intentions and theory-of-mind reasoning, play an important role in language acquisition (e.g., Bloom, 2000; Bohn & Frank, 2019; Clark & Amaral, 2010; Tomasello, 2001). Under the traditional *pragmatic inference account*, the ME bias is based on the assumption that the speaker follows cooperative principles of communication. According to Clark (1988), the relevant principles are the *Principle of Conventionality* (speakers use the same words to refer to the same objects) and the *Principle of Contrast* (every two types of objects contrast in meaning). Following a pragmatic explanation along these lines, the ME response, associating novel word $w_{n+1}$ to novel meaning $o_{n+1}$, is supported by the pragmatic argument that *if* the speaker would have wanted to refer to a known object $o_i \neq o_{n+1}$, they would have used word $w_i$ (by contrast and conventionality).

The classical ME paradigm constitutes a context-dependent inference task. The ME bias phenomenon, however, is embedded in the long-term learning process of language acquisition. In general, it has been argued that long-term word learning and online inference in situations of referential ambiguity operate on different time scales and are not straightforwardly dependent on each other (Frank, Goodman, & Tenenbaum, 2009; Gulordava, Brochhagen, & Boleda, 2020; McMurray, Horst, & Samuelson, 2012). For example, given the classical ME paradigm, children are able to identify the correct referent, arguably via online inference, but show poor retention of this novel word-meaning mapping when tested five minutes later (Horst & Samuelson, 2008). While accurate referent selection can be achieved by excluding competitors in a given context, retention requires encoding the association between the novel word and object (Axelsson, Churchley, & Horst, 2012). In addition, several studies provide insights into the developmental trajectory of the ME bias. When children of different age groups are tested within the same ME experiment, the ME bias consistently increases with age (Bion, Borovsky, & Fernald, 2013; Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016; Grassmann, Schulze, & Tomasello, 2015; Halberda, 2003; Lewis et al., 2020). This effect seems to be driven by two factors that increase with age: vocabulary size and linguistic exposure, that is, familiarity with "familiar" objects and labels (Grassmann et al., 2015; Lewis et al., 2020). To provide a full account of the ME bias phenomenon, it is important to consider these relations between online inference and long-term learning as well.

To address the ME bias puzzle, together with modulating developmental effects such as vocabulary size and exposure, we develop a computational model that comprises both pragmatic reasoning and long-term associative learning. In line with probabilistic pragmatic word learning models, we rely on the Rational Speech Act (RSA) framework as a computational mechanism for the ME bias. In the RSA framework, speakers and listeners recursively reason about each other's intention to enrich the literal meanings of utterances, using Bayesian inference. It has successfully modeled various pragmatic phenomena (e.g., Scontras, Tessler, & Franke, 2018), and probabilistic pragmatic word learning models either rely on the framework itself (Smith, Goodman, & Frank, 2013) or similar formalizations (Frank et al., 2009; Lewis & Frank, 2013). In all these models, agents take into account the speaker's perspective when inferring the most likely lexicon from a history of observed word-meaning pairs. Lacking situation-time dynamics, there is no differentiation between long-term learning and online inference. Our model, in contrast, embeds (pragmatic) online inference into a long-term learning process, where lexical associations are formed incrementally in a gradient-based learning process. As a result, the model can account for processes at different time scales and can be compared to behavioral data from the classical, inferential ME paradigm as well as developmental studies. This comparison allows us to evaluate which of the mechanisms hypothesized to play a role in the ME bias, and more generally word learning, agree or disagree with psychological reality.

We introduce different pragmatic agent models with explicit lexical representations and use them for an in-depth investigation of the ME bias phenomenon from a pragmatic perspective. One implementation we evaluate is the same as in our earlier work (Ohmer, König, &

Franke, 2020), where agents have a fixed lexicon size, corresponding to the number of words and objects in the data set. In addition, we test a novel implementation with a dynamically growing lexicon. There is an important conceptual difference between the two implementations. The pragmatic inference account describes how pragmatic reasoning during online inference leads to an ME bias. This *inferential ME bias* occurs in both implementations. However, pragmatic reasoning has different long-term learning effects on the two lexicon types. In the fixed lexicon, but not the dynamic lexicon, an additional *lexical ME bias*, as proposed by the lexical constraint account, arises. (Section 2.3 provides a detailed explanation of how the two bias components relate to the two lexicon types.) Using these two models, we study the ME bias, its developmental trajectory, and its role in long-term learning; and simultaneously evaluate the influence of inferential and lexical bias components.

The ME bias is also of interest to the machine learning community. Recently, Gandhi and Lake (2020) showed that neural networks lack an ME bias and even have the reverse tendency of selecting familiar labels for novel objects. They further demonstrated that this anti-ME bias slows down learning in various types of networks. As a result, they proposed the *ME bias challenge* for neural networks, which is not only a technical challenge applying to general classification or translation models but concerns any artificial agent design based on standard neural network architectures. Compared to traditional probabilistic pragmatic models, the gradient-based learning mechanism used by our approach is compatible with neural network training, and we explore it as a potential solution to the challenge.

Recently, deep neural word learning models have been introduced, which try to address the ME bias challenge (Gulordava et al., 2020; Vong & Lake, 2020). These models operate with two networks: a visual module processing raw images and a language module processing words. Whether a word maps to an object is determined by the similarity of their embeddings. In the classical ME paradigm, the child's ability to recognize familiar words and objects, as well as to recognize the novel word and object as such, is given. While the two networks can learn to recognize familiar words and objects, due to the anti-ME bias, they are not guaranteed to recognize the novel word, $w_{n+1}$, as different from any familiar word or the novel object, $o_{n+1}$, as different from any familiar object. So far, deep neural word learning models use *negative sampling* to solve this problem. In the classical ME paradigm, the bias is tested with words and objects the child is entirely unfamiliar with. In contrast, these models rely on presenting the test items as negative examples, so mismatching combinations of words and objects with explicit negative feedback during training. Negative examples make the network learn that the novel word (object) is not associated with any of the familiar objects (words). However, negative sampling is not an empirically justified assumption about human word learning because it is far from obvious that negative samples are available at a sufficient rate. Neither does it solve the ME bias challenge of building neural networks that map novel inputs onto novel, hitherto unseen outputs. Consequently, it is important to find mechanisms for generating ME-like behavior in neural network architectures that do not require negative sampling.

The remainder of the paper is structured as follows. Section 2 introduces the computational pragmatic models with explicit lexical representations and investigates their behavior in a classical ME paradigm, developmental effects, as well as the relation between online inference and long-term learning. Section 3 explores a proof-of-concept extension of our approach to a deep neural network architecture. We follow existing joint-embedding space architectures and test whether pragmatic reasoning can also achieve ME in deep neural networks (without negative sampling). Section 4 critically assesses the approach taken here and the results obtained from it before Section 5 concludes.

## 2. Mutual exclusivity in pragmatic agents with explicit lexical representations

We set out to explain not only the ME bias behavior as it is observed in the classical ME paradigm but also interactions between ME bias and long-term learning, in particular, the effects of vocabulary size and linguistic exposure on bias strength, and whether the ME bias supports learning. In the following experiments, we explore these questions from a computational pragmatic perspective. We develop two different computational pragmatic models, one where the lexicon size is fixed and one where it grows dynamically, both of which embed pragmatic inference into an associative long-term learning process. As this is a novel approach, we initially analyze the mechanisms by which these implementations lead to an ME bias on a theoretical level. We find that they make different assumptions about how pragmatic reasoning causes an ME effect. The computational experiments therefore not only test whether our pragmatic model can account for empirical findings but also which of these assumptions are more plausible.

In our first analysis, we test whether both implementations lead to an ME bias in a long-term learning context. We then proceed to investigate the development of the ME bias over time. As vocabulary size and linguistic exposure have been identified as the main drivers behind the increasing bias strength in early development, we conduct two separate analyses to examine the influence of these two variables on our model. In an additional analysis, we test whether, even though making the correct inference is not sufficient for long-term learning, it could still serve as a supporting factor. In a final analysis, we use differences in prediction between the two implementations to identify the pressures of pragmatic reasoning on learning and inference that can best explain empirical findings. To account for the difference between the two processes of long-term learning and online inference, they are monitored separately throughout the analyses.

### 2.1. Pragmatic agent model

The agents in our model feature two main components: (a) explicit lexical representations and (b) rules of pragmatic behavior telling them how to use these representations to produce and interpret messages. We consider agents with a fixed lexicon size (Ohmer et al., 2020) and also explore an implementation with a dynamically growing lexicon.

### 2.1.1. Lexical representations

The lexicon $B_A$ of agent $A$ is a matrix providing a mapping between words and objects. Each matrix entry $B(o_i, w_j) \in \mathbb{R}^+$ is an unnormalized value of how appropriate (in a semantic sense) word $j$ is for object $i$. The matrix entries are the only trainable parameters of the model.

*Fixed lexicon* Working with a fixed lexicon size is in line with other word learning models (e.g., Lewis & Frank, 2013; McMurray et al., 2012; Regier, 2005). The agents' lexicon is a matrix of size $N \times N$, where $N$ is the total number of word–object pairs in the training data. Because agents are pragmatic, their reasoning process during learning encompasses all words and objects in the lexicon, even those that have not yet appeared in the learning process, which is similar to the negative sampling strategy employed in neural word learning models. In this implementation, however, the use of negative examples is not modeled explicitly but arises naturally from pragmatic reasoning. While the idea that agents reason about unfamiliar words and objects is questionable when trying to draw a connection to human word learning, one can defend this approach by arguing that agents are aware that there are unknown states and messages in the world. A fixed lexicon size implements the idea that agents extend their reasoning to future novel inputs for which they reserve lexicon space.

*Dynamic lexicon* Alternatively, the dynamic lexicon only encompasses familiar items and is extended for novel inputs. This type of implementation has also been used before (e.g., Kachergis, Yu, & Shiffrin, 2012). The agents start out with a minimal lexicon of size $2 \times 2$. Every time the agents encounter a novel object or word, the lexicon is extended by one row or column, respectively. The newly created lexicon entries are initialized with the mean of the old entries. As the mean of the lexicon entries changes throughout the word learning process, learning has an indirect effect on associations between novel words and objects. But, unlike with a fixed lexicon, they are not updated in the training process itself. Using the average of the old lexicon as initialization for the new slots is a natural choice because the lexicon entries are not upper-bounded. A constant initialization value would have different effects depending on the hyperparameter setting (which influences the range of values the lexicon entries take on) and the stage of the training process as lexicon entries tend to keep changing over time.

### 2.1.2. Rules of pragmatic behavior

For modeling pragmatic behavior, the vanilla RSA model is used. In the RSA model, conditional probabilities describe how a speaker maps a state, $s$, onto an utterance, $u$, and how a listener maps an utterance onto a state, while they take into account each other's perspective.

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

$$P_{PS}(u \mid s) \propto \exp(\alpha \times [\log P_{LL}(s \mid u) - C(u)]) \tag{2a}$$

$$P_{PL}(s \mid u) \propto P_{PS}(u \mid s) \times P(s) \tag{3a}$$

At the basis of the recursive reasoning process is a literal listener (LL: 1a) who maps an utterance onto any state for which it is true, at the same time considering the prior probability of that state. In (1a), $[\![u]\!](s)$ is the denotation function returning the truth value of utterance $u$ for state $s$. A pragmatic speaker (PS: 2a) chooses her utterance such that the probability of being correctly understood by a literal listener is maximized while production cost, $C(u)$, stays low. The parameter $\alpha \in \mathbb{R}^+$ regulates the speaker's optimality. For $\alpha = 0$, the speaker's choices are random, and for $\alpha \to \infty$, she will always select the utterance that yields the maximal probability of being correctly understood by the literal listener. The pragmatic listener (PL: 3a), in turn, interprets an utterance as if coming from a pragmatic speaker, also considering the prior probability of states. We model our agents as pragmatic listeners.

We adapt the vanilla RSA model in several ways. The vanilla model assumes that agents have access to the lexicon, which is a truth table of utterances across states. In our case, the agent learns the (non-negative, real-valued) entries of its lexicon, which it also uses for its internal reasoning process. We assume a flat prior over states, zero costs for every utterance, and set $\alpha = 5$.[1] As our agents only reason about words and objects, we change the notation accordingly:

$$P_{LL}(o \mid w, B_{LL}) \propto B_{LL}(o, w) \tag{1b}$$

$$P_{PS}(w \mid o, B_{PS}) \propto P_{LL}(o \mid w, B_{PS})^{\alpha} \tag{2b}$$

$$P_{PL}(o \mid w, B_{PL}) \propto P_{PS}(w \mid o, B_{PL}) \tag{3b}$$

The agents are myopic with respect to the possibility of different lexica, that is, they only consider their own current lexicon and do not reason about which lexicon their interlocutor might likely have as in some pragmatic-inferential accounts (Frank & Goodman, 2014; Frank et al., 2009; Lewis & Frank, 2013). We do not use literal listener or pragmatic speaker models directly; they only appear as part of the pragmatic listener's inference process.

### 2.2. *Reinforcement learning*

We train our agents with reinforcement learning.[2] Following the pragmatic reasoning process in 3b, the agents map an input word onto a probability distribution over objects, $P_{PL}(o \mid w, B_{PL})$, which defines their policy. The agents' selection is sampled from this policy. If they select the right object, they obtain a positive reward, $R = 1$; otherwise, they obtain zero reward, $R = 0$. The loss function is defined as the negative expected reward,

$$\mathcal{L}(\theta) = -\mathbb{E}[R] \tag{4}$$

and the parameters to be optimized correspond to the agents' lexicon entries, $\theta = B_{PL}$. The parameters are updated using REINFORCE (Williams, 1992), which belongs to the family of policy gradient algorithms. In our case, gradients are calculated as

$$\nabla_{\theta}\mathcal{L} = -\mathbb{E}[R\, \nabla_{\theta} \ln P_{PL}(o \mid w, \theta)] \tag{5}$$

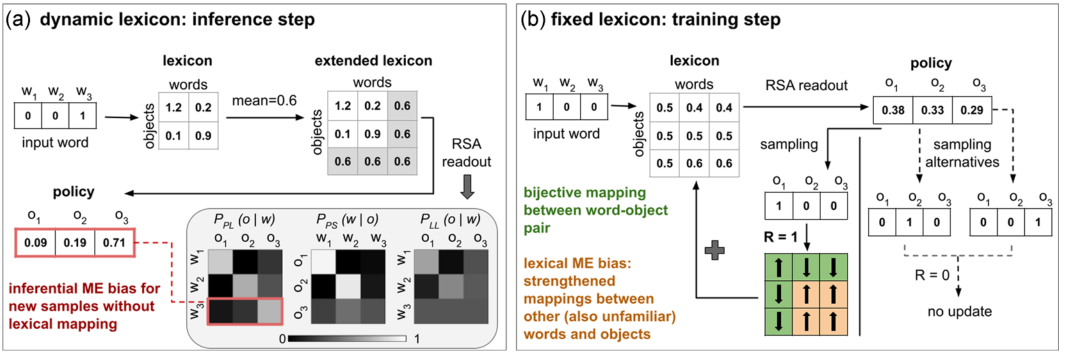


Fig. 1. (a) Example of the (pragmatic listener) agent's inference step with a dynamic lexicon. When the agent encounters a new word–object pair, its lexicon is extended by one row and one column. The "RSA readout" shows the recursive reasoning steps that lead to the agent's policy, $P_{PL}(o \mid w)$. The agent reasons about the policy of a pragmatic speaker, $P_{PS}(w \mid o)$, which, in turn, reasons about the policy of a literal listener, $P_{LL}(o \mid w)$. Note that the matrices here visualize policies and not lexica. Following this inference process, the selection probability for the unfamiliar object is highest when receiving an unfamiliar word. Pragmatic reasoning leads to an inferential ME bias. (b) Example of a training step with a fixed lexicon, adapted from Ohmer et al. (2020). If the listener selects the correct object, it obtains a positive reward, $R = 1$, else $R = 0$. The weight update following a correct selection reinforces one-to-one mappings between the current word-object pair (green) and strengthens associations between all other words and objects in the lexicon (orange). With a fixed lexicon, the update includes unfamiliar words and objects, leading to a lexical ME bias.

Gradient-based updates lead to incremental changes in the lexical associations between words and objects and simulate the long-term learning process.

### 2.3. Mutual exclusivity in pragmatic reasoning

The consideration of alternative words and meanings by the pragmatic listener, as in 3b, leads to an ME bias during online inference. This inferential ME bias applies equally to implementations with a fixed and a dynamic lexicon. Fig. 1a illustrates this effect, with the help of an example, for the dynamic lexicon implementation. Here, the agent encounters a new word and a new object in the context of familiar objects, which is why the lexicon is extended. The entries in the new column and row are all identical to the average over the existing matrix. Nevertheless, the rows corresponding to familiar objects have entries with above-average values because associations between familiar words and objects have already been formed. That is, a pragmatic speaker would probably have chosen one of the familiar words if referring to one of the familiar objects (see the speaker mapping $P_{PS}$ in the RSA readout shown in Fig 1a). Taking the reasoning process of the pragmatic speaker into account, the pragmatic listener can infer that the new word probably refers to the new object (see the listener mapping $P_{PL}$ in the RSA readout in Fig. 1a) although the last row in the lexicon has no information on the choice of utterance for a new object. Thus, even without a learned lexical association for the input word, the agent has a high probability of selecting the correct object. In other words, pragmatic reasoning causes an ME bias that is independent of lexical constraints.

In our model, pragmatic inference is embedded into a long-term learning process. During each inference, the pragmatic agent consults all word–object associations to infer a policy. Accordingly, the learning signal created by a single inference affects all lexicon entries, as shown for a fixed lexicon implementation in Fig. 1b. In particular, a one-to-one mapping between the current word–object pair is reinforced (green updates), and at the same time, associations between all other words and objects in the lexicon are strengthened (orange updates). As the lexicon has a different structure in the two implementations, the learning process has a different effect. In the dynamic lexicon, only associations between familiar words and objects are updated. In the fixed lexicon, in contrast, the learning process affects all associations, which induces assumptions about associations between unfamiliar words and objects. These assumptions implement a lexical ME bias: Associations between novel words and novel objects become increasingly more likely than associations between novel words and familiar objects or novel objects and familiar words. In conclusion, pragmatic reasoning causes an inferential ME bias regardless of lexicon type, and an additional lexical ME bias for the fixed but not the dynamic lexicon.

The learning process does not induce a lexical ME bias if agents have a dynamic lexicon. Still, it has an indirect effect on the associations between novel words and objects via the initialization mechanism. During learning, associations between words and their referents are strengthened. At the same time, associations between words and other objects are weakened. As true associations are sparse, a majority of the lexicon entries converge to zero. As a consequence, the mean of all lexicon entries—with which new slots are initialized—moves further away from the true association weights, such that under pragmatic principles, the novel word becomes an increasingly less likely choice for any of the familiar samples. Thus, the initialization mechanism in the dynamic lexicon gradually builds up evidence for one-to-one-mappings between words and objects.

In both implementations, the agent reasons pragmatically during learning and inference but the consequences on the lexicon are different. With a fixed lexicon, pragmatic reasoning during learning and inference divides the ME bias into lexical and inferential components. Other fixed lexicon implementations, combining literal listener and pragmatic listener components, are conceivable. A literal learner, performing pragmatic inference will have an exclusively inferential ME bias. A pragmatic learner, performing literal inference will have an exclusively lexical ME bias. And indeed, pragmatic inference can lead to an early emerging ME bias both by its effect on the lexicon during learning and by its effect on the online reasoning process (Ohmer et al., 2020). A fixed lexicon with literal learning and pragmatic inference provides an interesting alternative to the dynamic lexicon implementation, in that it implements an inferential ME bias component without the indirect effects of the initialization mechanism. First and foremost, we are interested in the predictions of fully pragmatic agents with a fixed or dynamic lexicon. In a separate analysis, however, we use different literal–pragmatic combinations for the fixed lexicon to disentangle the contribution of lexical and inferential pressures on the ME bias.

### 2.4. Methods

Agents and training were implemented with Tensorflow 2.0 (Abadi et al., 2015).

#### 2.4.1. Training

*Main setup* We train agents on a word learning task to simulate the long-term learning process. We use a simple learning scenario in which the frequencies of words follow Zipf's law (Zipf, 1949).[3] Objects and words are represented by one-hot vectors and have a predetermined one-to-one mapping such that object $i$ is associated with word $i$, for $1 \leq i \leq 100$. We define a *sample* as a word–object pair that belongs together. The training set contains all words and objects that the agent can encounter and grows monotonically with the number of epochs. At the beginning, agents are exposed to a single sample, and every $k$ epochs a new sample is added until a total number of $N = 100$ is reached. Given that word–object pairs are always added together, the dynamic lexicon is always extended by one row and column simultaneously. Every epoch consists of a fixed number of trials. At each trial, a word from the current training set is randomly selected and presented to the agent. The agent maps the input word to a policy over all objects in its current lexicon. The lexicon entries are updated using reinforcement learning, as described above. The exposure interval $k$ determines after how many epochs new samples are added to the training set, and thereby regulates how well the agent has learned familiar words when it encounters a novel word. If not mentioned otherwise we report results for $k = 15$, which provides near-optimal learning conditions in that agents have enough time to learn familiar words almost perfectly. We consider other exposure intervals ($k \in \{1, 3, 6, 9, 12, 15\}$) to investigate the relation between linguistic exposure and ME bias and how differences in ME bias relate to differences in learning success. To obtain robust results, we run 100 simulations for every value of $k > 1$ for the dynamic and the fixed lexicon, respectively. For $k = 1$, we run a total of 500 simulations each.

*Hyperparameters* Training hyperparameters such as batch size, learning rate, lexicon initialization values, and samples per epoch were selected based on model performance after running a small grid search (for details see Appendix A). In general, we found the qualitative results to be very robust across different hyperparameter values. The fixed lexicon entries are initialized as $b_{i,j} = 0.001$, $1 \leq i, j \leq 100$, and the dynamic lexicon entries as $b_{i,j} = 0.1$, $1 \leq i, j \leq 2$. In every epoch, 1000 word–object pairs are sampled randomly from the current training set. Agents are trained with vanilla stochastic gradient descent (SGD) on batches of 32 examples and with learning rate $\gamma = 0.1$, with one exception where we reduce the learning rate (see below).

#### 2.4.2. Evaluation

We evaluate long-term learning and online inference. Word learning progress can be monitored by the average reward achieved per epoch. As the reward for each trial is either zero or one, the average reward directly corresponds to the proportion of words that are mapped onto the correct object. When evaluating online inference, we are interested in the ME bias. We consider those time points in the learning process where the agent encounters a novel word for the first time. This happens every $k$ epochs as determined by the exposure interval. By tracking the agent's inference at all these time points, we can analyze the ME bias throughout development.

*Mutual exclusivity index* We use the ME index (Ohmer et al., 2020), $I_{ME}$, to quantify the ME bias formally. An ME bias exists if the probability of selecting a novel object upon receiving a novel word is greater than chance:

$$I_{ME} = \frac{p(\text{new object selected} \mid \text{new word}) - p(\text{new object})}{p(\text{familiar object})}$$

If the probability of selecting a novel object given a novel word is at chance level, the ME index is equal to zero; and if the entire conditional probability mass is on the new object(s), it is equal to one. We add samples incrementally and evaluate the ME index separately for each new word. With a fixed lexicon, there are various novel objects that the agent can select, and their number decreases as a function of epochs. With a dynamic lexicon, there is only one novel object. The exact formulas for both cases are provided in Appendix B.

*General versus specific referential contexts* We use two different settings to evaluate the ME bias. In the *general-context evaluation*, the referential context $C$ comprises all known words and objects, as well as any novel ones presented currently. Consequently, the agent's selection policy encompasses all objects in the lexicon. The ME bias strength, based on this policy, can be taken to quantify the agent's general assumption that a novel word must refer to a novel object rather than an old one. In contrast, the *specific-context* simulates the classical ME paradigm. The agent is presented with a novel word in the context of just one familiar object (distractor) and one novel object (target), and the agent's policy only encompasses these two objects. At the same time, we limit the pragmatic reasoning process to the novel word, and words that the agent considers plausible for the familiar object. These candidate words are determined by sampling 25 times (independently, with replacement) from the policy of a pragmatic speaker given the familiar object as input. The agent's reasoning process then includes all words that were sampled at least once.[4] Thus, not only the objects but also the words under consideration are contextualized in the specific context. The results for the specific-context evaluation are obtained by aggregating: for every new word–object pair, we create a specific context with every other object in the lexicon as distractor. Differences between the two evaluations provide insights into the role of contextualization in the ME bias phenomenon.

## 2.5. Analyses and results

### 2.5.1. Does pragmatic reasoning lead to ME in a long-term associative learning process?

Pragmatic reasoning can explain the ME bias and has been implemented successfully using the RSA model. We start by testing whether both our RSA-based agent models, the fixed lexicon implementation and the dynamic lexicon implementation, successfully realize this ME bias mechanism in a long-term word learning context. To measure whether the agents

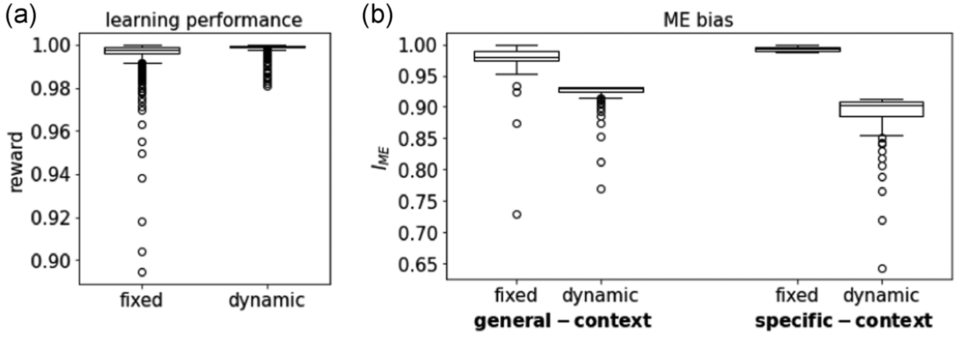


Fig. 2. Distributions of rewards (a) and ME indices (b) across training. The distributions are averaged across 100 simulations with maximal exposure ($k = 15$). Results are shown for the fixed lexicon implementation as well as the dynamic lexicon implementation, and the ME index is evaluated in general context and specific context. The outliers in both plots represent low values that arise at the beginning or at the end of training (see Section 2.5.2).

have an ME bias throughout learning, we calculate the ME index distributions over the course of training.

Fig. 2a shows that agents with both types of lexica achieve very high rewards throughout the training process, which means that they manage to consolidate all the words they encounter. Fig. 2b shows the ME indices for the general-context and the specific-context evaluation. Agents with a fixed lexicon have a stronger ME bias than agents with a dynamic lexicon in both types of evaluation. Still, the ME bias is very strong for either lexicon type, with all mean ME indices lying above 0.88. All distributions lie completely above zero, which means that the agents display an ME bias throughout the entire training process. The results extend our earlier finding that pragmatic agents with a fixed lexicon have an ME bias in a general-context evaluation (Ohmer et al., 2020), to agents with a dynamic lexicon and to a specific-context evaluation simulating the classical ME paradigm.

*2.5.2. How does vocabulary size influence the ME bias?*

We are interested in the predictions of the two implementations regarding the developmental trajectory of the ME bias. To test whether they are in line with the empirical observation that the ME bias increases as children grow older and have a larger vocabulary, we look at the agents' word learning performance and ME bias over time. Under the near-optimal learning conditions considered here, the agents map novel words onto the correct referents almost as soon as they have been added to the training set. We can therefore approximate the words that are part of the agent's vocabulary by the words in the training set. As a new sample is introduced every 15 epochs, the vocabulary size grows monotonically with the number of training epochs.

Fig. 3 shows average rewards and ME indices over the course of training. We begin by clarifying technical peculiarities that arise with a fixed lexicon size. Early in the training process, rewards drop more strongly for agents with a fixed lexicon (top left) than for agents

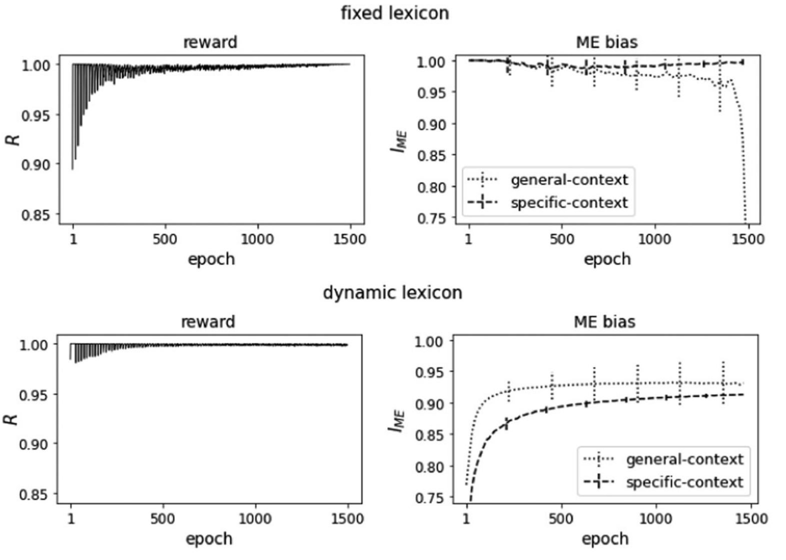


Fig. 3. Reward and ME bias strength over the course of training for both lexicon types. Shown are rewards (left) and ME indices for the general- and the specific-context evaluation (right). For the specific-context evaluation, we average the ME index across all referential games with the same target. Reported are mean values across 100 runs with exposure interval $k = 15$, and for the ME indices, we include standard deviations.

with a dynamic lexicon (bottom left) when new words are added to the training set. At training onset, agents with a fixed lexicon have far more selection options than agents with a dynamic lexicon. Over time, as more associations are established, the lexical ME bias reduces the number of potential referents and the meaning of novel words can more easily be inferred. Looking at the ME indices over time, the ME bias is relatively stable except for the initial and the final training phases. With a fixed lexicon (top right), the ME bias drops for the final samples in the general-context but not the specific-context evaluation. The sudden drop is due to a ceiling effect of the ME index calculation. For a maximal ME index, the selection probability for each new sample must be equal to $1/l$ where $l$ is the number of free slots. When the number of free slots reduces to one, an extreme increase in the selection probability of the remaining novel samples is required for the ME bias to be maximal. In the specific-context evaluation, in contrast, only the relative difference in selection probability between the new sample and a randomly selected old sample is important. Both effects, the increasing rewards as well as the drop in ME bias, are an effect of the fixed lexicon design and are not conceptually relevant.

Let us now return to the question of interest. What predictions do the two implementations make about the relationship between vocabulary size and ME bias? For the fixed lexicon, we find that the ME index is high throughout training in both evaluations (ignoring the ceiling effect) (Fig. 3, top right). In essence, the size of the lexicon does not influence the ME bias. In contrast, for the dynamic lexicon, the ME bias increases with the size of the lexicon following a curve with decreasing incline (Fig. 3, bottom right). While this effect only holds for the initial training phase in the general-context evaluation (dotted line), it holds throughout the entire training process in the specific-context evaluation (dashed line). An additional test reveals that if the policy from the general-context evaluation is used to contrast the selection probability of the novel object separately against the selection probability of each familiar object, the ME bias converges and does not increase continuously. Hence, for the ME bias to keep increasing, not only the target-distractor contrast itself is important, but also the contextualization of the pragmatic reasoning process, in that the agent only considers the objects present in the scene and only alternative words that would be a sensible choice for these objects. As associations between familiar words and objects are learned increasingly well, the agent considers fewer alternative words for the distractor, which facilitates the correct inference. In summary, agents with a fixed lexicon size have a constant ME bias, regardless of vocabulary size, while for agents with a dynamic lexicon, the ME bias increases with vocabulary size, in particular when the inference process is contextualized.

### *2.5.3. How does the amount of linguistic exposure influence the ME bias?*

The ME bias has also been shown to increase with the amount of exposure to familiar words, which can be measured by varying the child's familiarity with the distractor label in the classical ME paradigm. Children's linguistic exposure is based on a developmental history of word learning experiences, whereas direct manipulation in the lab is only possible over a short time span (e.g., Lewis et al., 2020). In our simulations, in contrast, we can directly regulate the amount of exposure by varying the exposure interval $k$ and can do so throughout the long-term learning process. The higher the exposure interval, the more time the agents have to reinforce the relation between familiar words and referents, and thus, a relation between word knowledge level and ME bias can be established.

Fig. 4 shows the results for different levels of exposure, $k \in \{3, 6, 9, 12, 15\}$, with larger intervals corresponding to darker shades. With respect to the agents' word learning performance (left column), we find that performance is lower when the agents have little exposure to the training samples (light-colored lines) as compared to high exposure to the training samples (dark-colored lines). Hence, the exposure interval can, in fact, be used to regulate the word-level knowledge for familiar words. Looking at the rewards over time, agents with a fixed lexicon improve to near-optimal performance regardless of exposure level. As explained for the exposure interval $k = 15$ above, the increase in reward stems from a continuous elimination of potential referents that arises when a lexical ME bias acts on a fixed lexicon. The performance of agents with a dynamic lexicon, however, decreases under low exposure, with a faster decrease for smaller intervals. With respect to the agents' ME bias (center and right column), there is a consistent pattern for both types of lexica as well as both evaluations: More exposure to familiar samples increases the ME bias for novel samples. Visually, this pattern is reflected in the monotonically darker shades of red toward higher ME indices.

We examine the role of linguistic exposure in more detail by comparing the agents' selection probabilities when tested for ME under maximal ($k = 15$) and minimal ($k = 1$) exposure levels, as shown in Fig. 5. In particular, we are interested in why the agents sometimes map novel words onto familiar objects. For the general context, the policy provides a full selec-

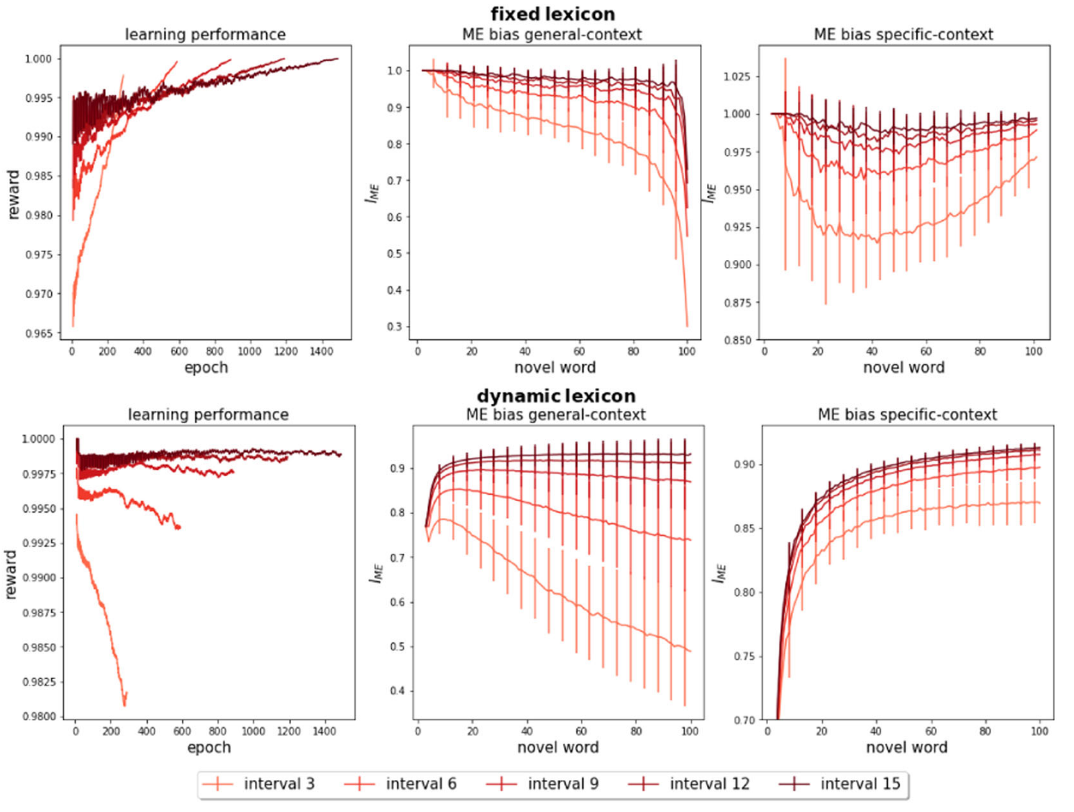


Fig. 4. Performance and ME bias strength over the course of training for different amounts of linguistic exposure: $k \in \{3, 6, 9, 12, 15\}$. Plotted are means and standard deviations (only for the ME bias) across 100 runs for each exposure level. The darkest red lines repeat the results in Fig. 3 apart from different scaling on the $x$-axis. Shown are performance (left), bias strength as given by the ME index in the general-context evaluation (center), and the specific-context evaluation (right). Rewards are smoothed by calculating the moving average across 19 epochs.

tion distribution across all objects. For the specific context, we calculate the policy for a referential game with each familiar object as distractor, respectively. Differences in the target selection probability indicate how much each object competes with the target object. Given a high amount of exposure, the samples in the training set are learned almost perfectly before novel samples are encountered. Therefore, selection probability mass is concentrated almost exclusively on the novel object(s). In the general-context evaluation, the high selection probability for novel objects is indicated by the probability mass lying on the upper right triangular matrix for the fixed lexicon, or on the diagonal for the dynamic lexicon (Fig. 5a, left column). In the specific-context evaluation, the target selection probability is consistently very high across all referential games (Fig. 5b, left column). Given a small amount of exposure, the samples in the training set have not been learned perfectly by the time a novel sample is introduced. Accordingly, in an ME bias evaluation, agents may select the novel object but

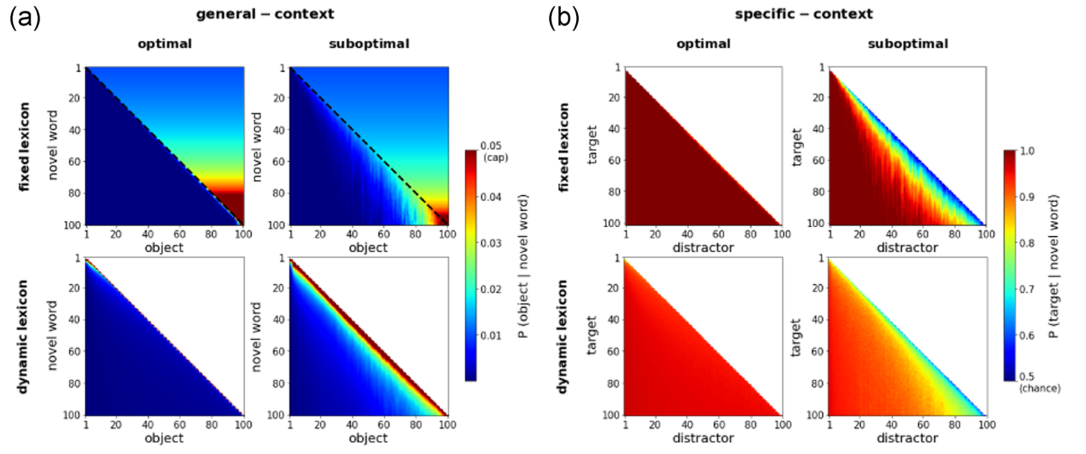


Fig. 5. Comparison of the ME bias in terms of agents' selection policies, averaged across 100 runs.

Note: Selection probabilities are evaluated for every novel word–object pair that is added to the training set. In both figure parts, results are shown for the fixed (top) and the dynamic (bottom) lexicon when learning is optimal ($k = 15$; left) or suboptimal ($k = 1$; right). (a) For the general context, we display the agents' selection probabilities given the novel word as input. We cap the probabilities at 0.05 to make differences below that value more visible. The $y$-axis indicates the introduction of a new sample to the training set. In the *optimal* learning condition, this happens every 15 training epochs and in the *suboptimal* learning condition every epoch. For every sample on the $y$-axis, the selection probabilities in the ME bias evaluation are plotted along the $x$-axis. (b) For the specific-context evaluation, we display the target selection probability for each referential game, with the novel object as target on the $y$-axis, and each (familiar) distractor object on the $x$-axis.

they may also select any of the objects they have not learned. In the general-context evaluation, some of the selection probability mass is allocated to other recently introduced objects (Fig. 5a, right column), and in the specific-context evaluation, agents perform much worse if the distractor is one of these more recent objects (Fig. 5b, right column). Overall, the ME bias increases with the amount of exposure to familiar samples and insufficient exposure makes the selection probabilities leak from novel to unconsolidated objects.

### *2.5.4. Does an ME bias during online inference support long-term learning?*

The ME bias supports the fast mapping of words to objects during online inference. We investigate whether this inferential advantage also supports long-term learning, by relating differences in ME bias strength to differences in learning success. For this purpose, we divide the training data into two categories, words for which the agent has a weak ME bias ($0 < I_{ME} < 0.5$) upon first encounter, and words for which the agent has a strong ME bias ($I_{ME} \geq 0.5$) upon first encounter.[5] We consider a scenario with little linguistic exposure ($k = 1$) to make sure that not all samples are learned immediately, such that there is enough variation in learning success. For the simulations with a fixed lexicon, we additionally reduce the learning rate to $\gamma = 0.0001$ to achieve that. We then measure learning success, in terms of whether a word–object association is learned in the long run, and learning duration, in terms of the

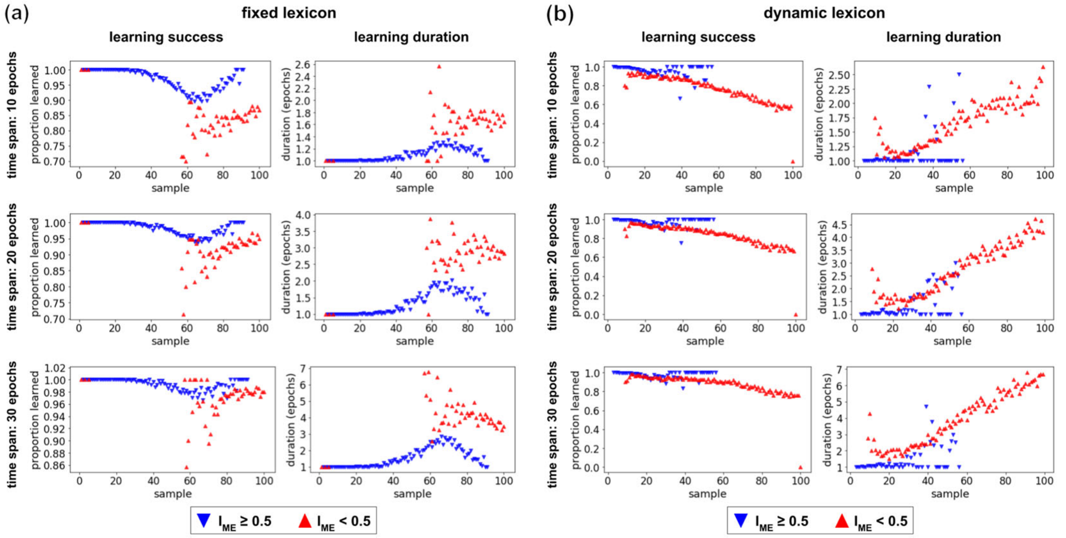


Fig. 6. Learning success and learning durations with respect to ME bias strength. Shown are averages across 500 simulations with an exposure interval of $k = 1$ for the fixed lexicon implementation (a) and the dynamic lexicon implementation (b), respectively. For the fixed lexicon, the learning rate was reduced to $\gamma = 0.0001$ to obtain a significant number of samples that are not learned immediately. The rows indicate different time spans in which the samples had to be learned after they were first encountered: 10, 20, or 30 epochs. Learning success is measured as the percentage of samples learned across simulations and learning duration as the average number of epochs until samples were learned. For a sample to be learned, it must be mapped correctly 99% of the time in all remaining epochs. To display meaningful statistics, we only calculate learning success and duration for a minimum of five samples (1% of the simulations) with either strong or weak ME bias.

number of epochs it takes until the association is stable. A word–object association counts as learned when the word is mapped onto the correct object in more than 99% of the cases in all remaining epochs.

Fig. 6 shows the results for different time spans in which the agent must learn the samples, with each row corresponding to one time span (10, 20, and 30 epochs). Results for the fixed lexicon are given in Fig. 6a and results for the dynamic lexicon in Fig. 6b. Blue triangles correspond to words with a strong bias and red triangles to words with a weak bias. When children infer a certain word-meaning mapping in context, they do not necessarily remember this association (Horst & Samuelson, 2008). We find a similar behavior in our model. Even words for which the agent has a strong ME bias ($I_{ME} > 0.5$, blue triangles) are not always learned, in the sense that they are not consistently mapped onto the correct referent in the long run. The difference between inference and learning success arises because changes to the lexicon, based on the associative learning process, are incremental and operate on much slower time scales than the inference process. If the agent does not map the novel word sufficiently often onto the correct referent, the association is not reinforced strongly enough, even though the mapping upon first exposure was correct. But the longer the available time span, the more objects are learned. Looking at the relation between ME bias and long-term learning success, we find that a strong ME bias increases the learning success rate and decreases the learning duration. These improvements can be observed for both implementations and for all time spans.

Experiments with adults show that word learning is less successful and slower when referential uncertainty is high (Smith, Smith, & Blythe, 2011). The current analysis allows us to establish a link between uncertainty and learning success via the ME bias. There is high referential ambiguity under suboptimal learning conditions: the agent does not know whether the novel word refers to the novel object or any of the old objects that it has not learned. This uncertainty, in turn, has a negative impact on the ME bias, and as a consequence on long-term learning.

### *2.5.5. How do lexical and inferential pressures influence the ME bias?*

As discussed in Section 2.3, pragmatic reasoning can cause an ME bias via lexical and inferential pressures. The inferential pressure exists regardless of lexicon type but the lexical pressure arises only in the fixed lexicon because associations of unknown words and objects are updated in the learning process. Still, the dynamic lexicon accumulates evidence for one-to-one mappings via its initialization mechanism. Our goal is to identify how differences in lexical and inferential pressures for ME influence the developmental trajectory of the ME bias. Considering the results above, the main difference between the two implementations is that the dynamic lexicon predicts an increase in ME bias strength across development, whereas the fixed lexicon does not. We perform an ablation study for the fixed lexicon implementation by removing the agent's pragmatic reasoning ability during either learning or inference. The pragmatic reasoning process is replaced by that of a literal listener in the RSA model, as in 1b. A literal learner performing pragmatic inference will only have an inferential ME bias, while a pragmatic learner performing literal inference will only have a lexical ME bias. This allows us to disentangle the role of lexical and inferential pressures toward ME from other factors.

Fig. 7 (top row) shows the word learning performance (left) and ME bias strength (right) across training for an agent that uses pragmatic reasoning during learning but not for inference. On average, word learning is successful throughout training as rewards remain constantly near-optimal. Initially, ME indices of both evaluations are also maximal, due to lexical constraints induced by pragmatic considerations during learning. However, further into the training, both ME indices start to decrease. The general-context ME index continues to decrease monotonically, whereas the specific-context ME index recovers. This pattern is very similar to the results for a fully pragmatic agent with a fixed lexicon under suboptimal learning conditions (see Fig. 4, fixed lexicon, $k = 3$). For the fully pragmatic agent, the decrease in general-context ME bias arises because one-to-one mappings between familiar words and objects are not strengthened sufficiently to fully exclude familiar objects as referents. For the pragmatic-literal agent, the decrease arises because familiar objects cannot be excluded as confidently in a literal reasoning process, which does not take into account that familiar objects are already "occupied" by a familiar word. The increase in specific-context ME bias is due to the fixed lexicon size, and not conceptually relevant. Even though the lexical constraint account does not commonly assume the involvement of pragmatic reasoning, the simulations

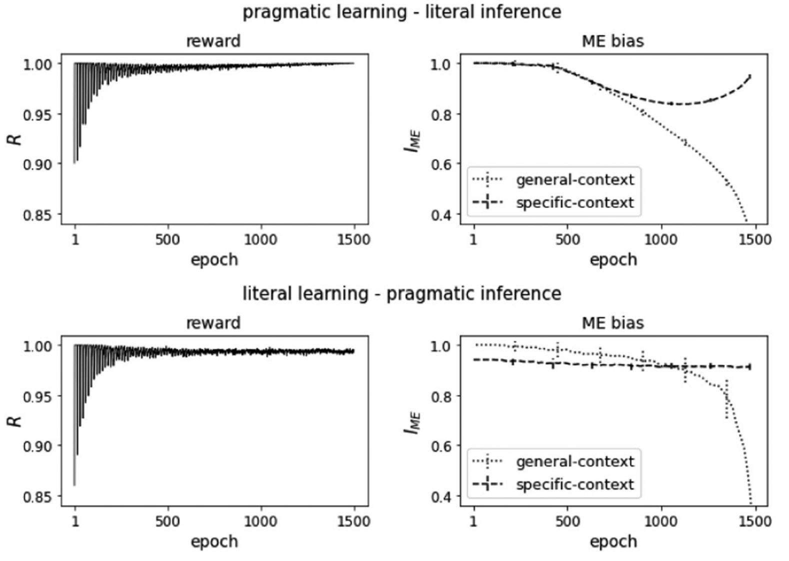


Fig. 7. Average rewards and ME indices across training for an agent with a fixed lexicon applying pragmatic reasoning during learning but not inference (top row) or applying pragmatic reasoning during inference but not learning (bottom row). Training was conducted with an exposure interval of $k = 15$ epochs. We show average results across 100 runs, including standard deviations for the ME indices.

underline that while lexical constraints can cause an ME effect, additional assumptions must be made to explain why it increases in strength.

Fig. 7 (bottom row) shows word learning performance (left) and ME bias strength (right) across training for an agent that uses pragmatic reasoning only in the online inference process but not during learning. High rewards indicate that learning is also successful for this combination. We compare the agent's ME bias to that of an agent with a dynamic lexicon (see Fig. 3, dynamic lexicon). The general-context ME bias is relatively constant throughout development for both implementations. Yet, while the dynamic lexicon predicts an increasing ME bias in the classical ME paradigm, the literal–pragmatic combination predicts a constant ME bias. Even though both agents largely rely on inferential pressures toward ME, agents with a dynamic lexicon collect evidence for one-to-one mappings via the initialization mechanism, whereas agents with a fixed lexicon do not. In conclusion, pragmatic online inference can cause an ME bias throughout development without a need for additional lexical pressures; but the developmental trajectory of the ME bias can only be accounted for if the pragmatic inference process includes the increasing evidence for one-to-one mappings in the lexicon.

## 3. Mutual exclusivity in pragmatic neural network agents

The models discussed in Section 2, from now on called *explicit lexicon models*, are limited by the simplicity of their input representations. They neither capture how human or artificial agents can learn new words from raw visual and linguistic input nor how the ME bias arises from such inputs. There are two different parts to the ME bias phenomenon. In a first step, the agent must recognize that visual and linguistic input represent novel *types* and not novel instances of familiar types. This process is closely related to the problem of out-of-distribution detection in machine learning models (e.g., DeVries & Taylor, 2018; Hendrycks & Gimpel, 2017; Liang, Li, & Srikant, 2018). In a second step, the agent must map the novel word to the novel object, which is the actual ME effect. The explicit lexicon models can capture the second step, but not the first one, which must rely on perceptual similarities and possibly common-sense knowledge.

The explicit lexicon models can in principle be combined with neural network modules as they rely on the same gradient-based learning mechanism. For example, objects displayed in images and words recorded as text could be processed by dedicated networks mapping them onto the corresponding row or column in the lexicon. This approach faces two immediate problems. First, end-to-end training is difficult as mapping onto specific slots of the lexicon requires using the non-differentiable argmax function. So, either training is not end-to-end or the argmax operation must be approximated, for example, using the *Gumbel-softmax trick* (Jang, Gu, & Poole, 2017). The second problem is specific to modeling the ME bias. If neural networks are trained to process the visual or linguistic input, they will fail to recognize novel inputs due to the diagnosed anti-ME bias. For example, an image classification network will map objects from novel categories with high confidence onto familiar categories. However, if the agent does not recognize an object as novel, it cannot use the ME bias.

To overcome these problems, we use continuous word and object representations that can exploit similarity relations in the input space. To perform pragmatic reasoning on these representations, their association strength (corresponding to the lexicon entries) is determined by calculating the similarities between these representations in a joint embedding space. With this architecture, end-to-end training is possible. We expect pragmatic reasoning to cause an ME bias also in this setup. As certain word and object representations become very similar over the course of training, pragmatic reasoning makes the use of novel words for these objects unlikely. We run an experiment to test this hypothesis and additionally examine the influence of negative sampling on our model.

### *3.1. Neural pragmatic agent model*

The model, as shown in Fig. 8, consists of three main components: a vision module (orange), a language module (blue), and a pragmatic reasoning module (green). The agent learns new word–object mappings by receiving a word and trying to select the correct referent from several objects given by the context. The vision and the language module map their respective inputs into a joint embedding space, where lexical association strength is determined by the similarity of the embeddings. In the pragmatic reasoning module, the RSA model is used to calculate a policy for the different objects under the given input word, taking into account alternative input words that could have been used.

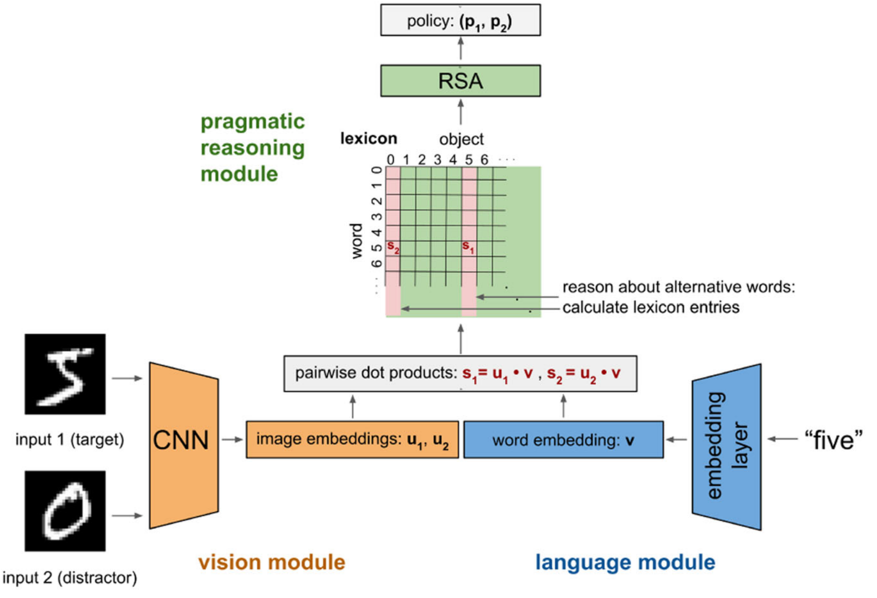


Fig. 8. Visualization of architecture and training setup for the neural network model.

### 3.1.1. Vision and language modules

*Vision module* The vision module maps raw pixel input onto an image embedding. We pretrain a convolutional neural network (CNN) on the input data using supervised learning.[6] The CNN consists of two convolutional layers followed by a dense layer and the final output layer (see Appendix C for details on model and training hyperparameters). We use the activations of the fully connected layer to extract the image features. These features are mapped into the joint embedding space by an additional fully connected layer with sigmoid activation function.

*Language module* The language module consists of an embedding layer mapping the integer symbol inputs onto continuous vectors. An additional fully connected layer with sigmoid activation function maps these word representations into the joint embedding space.

*Bounding the embedding space* Often representations are unbounded in a joint embedding space. When learning a lexicon, associations between learned words and objects can take on extreme values over time. In the dynamic lexicon implementation, we initialize novel lexicon entries with the lexicon's mean value. Here, in contrast, we cannot control how novel slots are initialized, that is, what values the embeddings of unknown words and objects take on. As a consequence, also the similarities (associations) between unknown words and objects are unconstrained. We find that bounding the embedding space by using a sigmoid output function, instead of a linear one, is in our case sufficient to provide a working initialization.

*3.1.2. Pragmatic module*

Again, our agent is implemented as a pragmatic listener in the RSA framework. As the pragmatic reasoning process involves the literal listener and the pragmatic speaker, we need to express these formulas based on our neural network architecture. We assume that the agent receives a single input word, $w$ in a context with multiple objects, $C = \{o_1, \dots, o_k\}$. Given word embedding $\mathbf{v} \in \mathbb{R}^m$ and image embeddings $\mathbf{u}_1, \dots, \mathbf{u}_k \in \mathbb{R}^m$, we can calculate the similarity values between the word embedding and each image embedding $s_1, \dots, s_k \in \mathbb{R}$ with $s_i = \exp(\mathbf{u}_i^T \mathbf{v})$. With a generic optimality parameter $\alpha$, this leads to the following reformulation of 1b–3b:

$$P_{LL}(o \mid w, C) \propto s \tag{1c}$$

$$P_{PS}(w \mid o, C) \propto P_{LL}(o \mid w, C)^{\alpha} \tag{2c}$$

$$P_{PL}(o \mid w, C) \propto P_{PS}(w \mid o, C) \tag{3c}$$

*3.2. Methods*

Again, agents and training were implemented with Tensorflow 2.0.

*3.2.1. Training*

*Main setup* We train our agents on a referential game (see Fig. 8). During each round of the game, the agent receives a word, the target object (referred to by the word), and a distractor object as input. The agent outputs a selection probability for the two objects, and the actual selection is sampled from this policy. If the agent selects the target object, it receives a positive reward, $R = 1$; otherwise, it receives zero reward, $R = 0$. Again, the agent is trained with REINFORCE using Eqs. 4 and 5, with the only difference that the parameters to be optimized, $\theta$, correspond to the neural network weights instead of the lexicon entries. We use the images of the MNIST data set (LeCun, Cortes, & Burges, 2010) as objects. These images contain 70,000 handwritten examples of the digits 0–9, with a train/test ratio of 60,000/10,000 and a size of 28 × 28 pixels. In our setup, the world consists of 10 different objects, corresponding to the different digits and 20 possible words, corresponding to 20 distinct symbols (0–19), both of which are uniformly distributed. Our training and test sets contain digits 0–8, and nine randomly selected, distinct words, which are assigned to these objects. Selection and assignment of words vary between, but not within runs. The test set is used to measure how well the network generalizes to novel examples of digits 0–8. The remaining object, digit 9, and the remaining words are reserved for evaluating the ME bias and form a separate data set. Reserving multiple words and a novel object for the evaluation simulates a world in which there are many potential names for an object. By holding out the images of digit 9, we have train/test sets of 54,051/8,991 images. To generate the referential games, each image is used as the target once and paired with a random distractor showing a different digit. At every training trial, the agent's input consists of one of the nine words in the training set, an image of the digit that word refers to (target), and an image of a different digit.

*Hyperparameters* The embedding layer of the language module as well as the two fully connected sigmoid layers mapping word and object representations into the joint embedding space each have dimensionality 32. All network parameters are initialized randomly. The network is trained with Adam optimizer, learning rate $\gamma = 0.0001$, and batch size 64. Training proceeds for 100 epochs. All parameters were selected by hand.

#### 3.2.2. *Evaluation*

To evaluate the ME bias, we compile referential games with a novel input word and number 9 as the target. We measure the ME bias as the correct selection probability in this test setup. Precisely, the correct selection probability is calculated by pairing each of the examples of digit 9 with a random distractor from the test set as well as a random novel input word, and averaging the results across these test games. Pairing the novel object with different potential novel names provides a more robust ME bias estimate as random differences in the embeddings of novel words are averaged out.

#### 3.2.3. *Experiments*

We train the agent on the referential games as described above and evaluate whether it has an ME bias. By default, the agent's pragmatic reasoning step encompasses the word and objects given by the context as well as all other words in its lexicon, so the remaining words in the training set (assigned to digits 0–8). As the number of words and objects is small, taking into account all alternative messages is not too costly; for more complex worlds, sampling may become necessary. For the explicit lexicon models, we know $\alpha = 5$ to be a suitable optimality parameter from earlier work, and the grid search further showed that results are not very sensitive with respect to optimality. Without such information for the neural network implementation, we test different optimality values, $\alpha \in \{5, 10, 15\}$. In addition, to evaluate whether the ME bias can be attributed to the pragmatic reasoning ability of the agent, we run the same experiment with a literal agent as given by 1c and compare the results. For the pragmatic agent with optimality $\alpha = 5$, we test modified versions of the negative sampling strategies employed by Gulordava et al. (2020). For negative sampling of words, the agent takes into account *all* possible words in its reasoning process, not just the ones in the training set. For negative sampling of objects, the images of number 9 already appear as distractors during training. We also test negative sampling of both words and objects. In total, we run 25 simulations for each variation.

### 3.3. *Results*

Fig. 9 shows the results for the pragmatic neural network agents. The top row shows the training rewards and the bottom row the strength of the ME bias, both over time. All runs converge to maximal accuracy on the training data and reach test accuracies (not in the figure) greater than 99.8%. Looking at the rewards (top left), it is not surprising that pragmatic

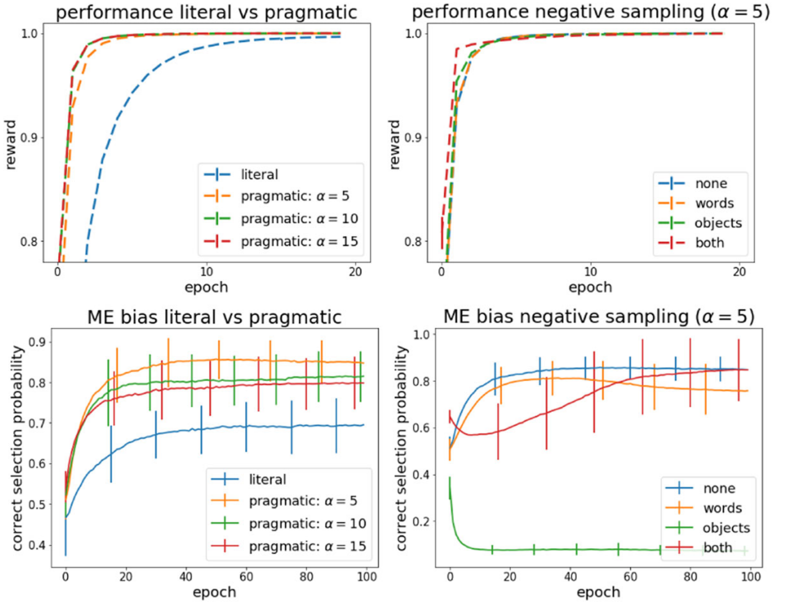


Fig. 9. Rewards (top) and ME bias (bottom) for the neural network architecture. All values are averaged across 25 runs, and standard deviations are displayed by error bars; no error bars are visible for the rewards due to very little variation. The left column compares pragmatic agents with different optimality parameters and a literal agent. The right column compares different negative sampling strategies for a pragmatic agent with optimality $\alpha = 5$: no negative sampling, negative sampling of words, negative sampling of objects, or negative sampling of words and objects (both).

agents learn faster than literal agents. For a literal listener, the learning update only affects the representations of the current training input. The pragmatic listener samples alternative words the speaker could have used. Accordingly, the learning update not only affects the representations of the currently present word and objects but also the representations of these alternative words. Pragmatic agents with different optimality parameters (top left) as well as different negative sampling strategies (top right) learn approximately equally fast, with a slight advantage for higher optimality parameters.

There is no clear relationship between performance and ME bias. While all agents achieve maximal rewards, the correct selection probabilities in the ME paradigm vary strongly. The different agents trained without negative sampling (left column) all display an ME bias. The strength of the bias varies with the optimality parameter, with higher optimality leading to a lower bias. With higher optimality parameters, small distances in the embedding space are amplified strongly by the exponentiation with a large $\alpha$. Then both word–target and word–distractor similarity may become zero, such that pragmatic reasoning cannot take effect. Interestingly, even the literal agents have a weak bias and make correct selections on average 69.5% of the trials at the end of training. The ME bias of the literal agent arises from the structure of the embedding space. Looking at the literal agent's lexicon (see Appendix D), it turns out that representations of the novel words lie closer to representations of the novel object than to those of familiar objects. Further research is needed to understand why this pattern arises. Still, the pragmatic agents have a consistently higher bias, with 79.8% ($\alpha = 15$), 81.5% ($\alpha = 10$), and 84.7% ($\alpha = 5$) at the end of training.

Looking at the different sampling strategies (bottom right), training without negative sampling surprisingly leads to the strongest bias, with negative sampling of words and negative sampling of both words and objects being close or equal. If negative sampling is used only for objects, the agent develops a strong negative ME bias. Appendix D shows visualizations of the learned lexica using different sampling strategies. Overall, it seems that pragmatic reasoning alone induces enough competition between novel and familiar objects. As discussed by Gulordava et al. (2020), negative sampling of objects introduces an anti-polysemy bias, while negative sampling of words introduces an anti-synonymy bias. Given that the ME bias is an anti-synonymy bias, this distinction explains why negative sampling of words or both words and objects leads to a much higher ME bias than negative sampling of objects. When negative sampling of objects is used, the ME bias drops far below chance. The representations for these objects move so far away in the embedding space that novel words eventually lie closer to the distractors than the target. In conclusion, our main result is that pragmatic reasoning alone is sufficient for the agent to develop an ME bias and negative sampling leads to no further increase in bias strength.

## 4. Discussion

We provide a new computational pragmatic model of the ME bias that combines insights from cognitive models of language use and modern machine learning techniques. We use agent models with explicit lexical representations to demonstrate that pragmatic reasoning not only leads to an ME bias in the classical ME paradigm but can also capture important aspects of the relation between ME bias and long-term word learning. In line with empirical findings, our model makes the following predictions: (a) the ME bias increases with the agent's exposure to familiar words and objects, (b) the ME bias increases with the agent's vocabulary size, and (c) correct inference does not guarantee long-term learning. The different implementations with fixed and dynamic lexica allow the modeler to choose between different assumptions on how pragmatic reasoning causes ME—only via an inferential bias or also via a lexical bias. Further analysis of these competing pressures reveals that an inferential ME bias alone is sufficient to predict an increase in ME bias strength across development if evidence for one-to-one mappings—collected throughout learning—is used in the inference process. We additionally show that a strong ME bias during online inference positively influences learning success and duration. Pragmatic reasoning may therefore constitute a useful ME bias mechanism for machine learning models. As a proof-of-concept, we demonstrate a transformation of our approach to a deep neural network architecture working with raw visual inputs and show that pragmatic reasoning also leads to an ME bias in such deep neural network agents. Together, our results open up new possibilities for research on the ME bias in word learning and deep neural networks.

### 4.1. Word learning

If pragmatic reasoning processes as formalized by the RSA framework play a role in infant, child, or adult word learning, they can be captured by our model, at least at the *computational theory* level of explanation (Marr, 1982). Over the past years, there has been growing evidence on the pervasive role of pragmatics in (early) language learning. It has been shown that preverbal infants already understand the communicative nature of language (Martin, Onishi, & Vouloumanos, 2012; Vouloumanos, Onishi, & Pogue, 2012). A recent review by Bohn and Frank (2019) maps out how young children use pragmatic inferences in word learning and how language understanding becomes increasingly more subtle as these inferences grow more complex over time. On the contrary, several studies have found that even at 5 years of age children often fail to perform certain types of pragmatic inferences (Huang & Snedeker, 2009). In the classical ME bias paradigm, alternative utterances can be derived from the context, $C$. The novel word, $w_{n+1}$, is contrasted with the label of the familiar object, $w_i \in \{w_1, \dots, w_n\}$. It turned out that in cases where pragmatic inference fails, children struggle with generating alternative utterances because they are less clear from the context, rather than computing the inference per se (Barner, Brooks, & Bale, 2011). Supporting this, several experiments by Frank and Goodman (2014) suggest that children and adults do indeed make RSA-like inferences to infer novel word meanings in context.

Next to pragmatic accounts of ME, constraint and bias accounts form a major strand of theories. They propose that infants have an innate or early emerging lexical bias toward one-to-one mappings between words and meanings. In principle, this bias can be specific to word learning or result from domain-general processes (Markman, 1992). In the dynamic lexicon implementation, pragmatic reasoning affects the agent's inference process. As such it can be seen as a computational model for a pragmatic inference account. In the fixed lexicon implementation, pragmatic reasoning not only affects the inference process but also induces a lexical ME bias. This lexical bias emerges at learning onset and explains why agents with a fixed lexicon but not agents with a dynamic lexicon have a strong ME bias already at the beginning of training. So, with a fixed lexicon implementation, lexical constraint accounts and pragmatic inference accounts can be accommodated by the same general principle of pragmatic reasoning—applied to learning and inference.

Our ablation analysis provides important insights into the role of lexical and inferential pressures toward ME. In line with the probabilistic pragmatic model by Lewis and Frank (2013), we find that both lexical and inferential pressures are sufficient but not necessary for ME. In addition, innate or early emerging lexical biases alone cannot account for the fact that children's ME bias increases with their vocabulary size. It follows that lexical constraint accounts must identify additional factors responsible for this development (e.g., Halberda, 2003). But inferential pressures must also use an evidence accumulation mechanism that reflects increasing certainty about the justification of a one-to-one assumption to model the increasing bias (cf., Lewis & Frank, 2013).

By using gradient-based learning in pragmatic agents, our model combines aspects of probabilistic pragmatic and associative word learning models (e.g., Kachergis et al., 2012; McMurray et al., 2012; Regier, 2005). Our agents use pragmatic reasoning to infer the meaning intended by the speaker among different alternatives but use gradient-based learning. While associative models typically hard code a competition mechanism to achieve ME (Yurovsky, Yu, & Smith, 2013), in our case, such competition arises naturally from the consideration of alternative meanings and utterances. Compared to probabilistic models, gradient-based learning allows our model to separate online inference and long-term learning and thereby to account for interactions between them. At the same time, our model loses the ability to learn from few examples, a disadvantage that is shared by many gradient-based word learning models (e.g., McMurray et al., 2012; Najnin & Banerjee, 2018; Vong & Lake, 2020). From a technical perspective, it opens up the integration of a pragmatic reasoning module with neural network components for processing visual or linguistic input. In sum, our approach might inspire new pragmatic word learning models as it differentiates between long-term learning and online inference and can operate on raw inputs when implemented with a deep neural network architecture.

Aside from dedicated models, general learning theory in the form of the Rescorla–Wagner (R-W) model has also been applied to word learning (e.g., Ramscar, Dye, & Klein, 2013; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). The R-W model can be considered a reinforcement learning model where learning is driven by reward prediction errors. Importantly, associations between referents (cues) and words (outcomes) are updated for both words that are present and words that are not present. However, it is unclear how a learner identifies relevant non-outcomes (Hollis, 2019). Both our pragmatic model and R-W models of word learning have the effect that children reason about the informativity of a word compared to other words (Ramscar et al., 2013), but the pragmatic reasoning process provides a natural explanation of how relevant alternative utterances can be identified in terms of what a listener thinks a speaker could have said.

This paper set out to explain ME bias behavior as selecting the novel referent $o_{n+1}$ given a novel word $w_{n+1}$. What is left unaddressed is how the set of relevant referents is to be construed in the first place by the learning agent. If the learning agent knows the word "dog" and also knows that the dog in front of them is called "Fido", a novel word for a novel object (e.g., a cat) could contrast with the known object at the level of kinds or at the level of individual names. Just like knowledge of how to construe relevant utterance alternatives is crucial for ontogenetically developing pragmatic reasoning abilities (see above), so, too, is it necessary to construe which meaning distinctions are relevant in the given context. Linguistic theory models relevance of meaning distinctions as (possibly implicit) questions under discussions (e.g., Roberts, 2012), essentially using partitions of objects into equivalence classes based on which distinctions matter to the conversation. Recent natural language applications similarly have started to integrate such partition-based approaches to modeling discourse relevance (e.g., Nie, Cohn-Gordon, & Potts, 2020). By extending the work presented here to include different levels of partitioning objects into relevance-guided equivalence classes to which novel words might refer, the present approach could be extended to go beyond considering one-to-one relationships between words and objects, thereby capturing a hierarchical organization of word meanings at different levels of granularity. Further challenges for extending the approach in this paper to the full flexibility of natural language lexical meanings include dealing with polysemy, ambiguity and context-dependence, vagueness, and, though arguably very infrequent (since no two expressions are absolutely equivalent in meaning and use), synonymy.

### 4.2. Deep neural networks and outlook

Apart from the pragmatic reasoning module, our deep neural network implementation is very similar to existing deep word learning models (Gulordava et al., 2020; Vong & Lake, 2020). Gulordava et al. (2020) even try a pragmatic inference-based approach at test time. Still, these models rely on negative sampling during training to induce competition and achieve an ME bias. As children can map entirely novel words to entirely novel objects in the ME paradigm, the use of negative sampling undermines the explanatory potential of these models. We demonstrate that using pragmatic reasoning at training and test time is sufficient to cause an ME bias such that negative sampling is not necessary.

In future work, it is important to establish if and how our approach can scale to more complex data sets and learning scenarios. Various works apply the RSA framework to deep learning problems. The resulting neural RSA models are used for different supervised learning tasks, such as generating and interpreting referential expressions (Andreas & Klein, 2016; Cohn-Gordon, Goodman, & Potts, 2018; Monroe & Potts, 2015; Monroe, Hawkins, Goodman, & Potts, 2017; Zarrieß & Schlangen, 2019), generating and following instructions (Fried, Andreas, & Klein, 2018; Fried et al., 2018), machine translation (Cohn-Gordon & Goodman, 2019), and text generation (Shen, Fried, Andreas, & Klein, 2019). Most of these models are different from ours in that they pretrain a literal listener or a literal speaker on a labeled data set, and then add pragmatic reasoning on top of this "base agent" at test time, whereas our agent applies pragmatic reasoning during training. None of these publications addresses the ME challenge. Notably, Zarrieß and Schlangen (2019) consider the problem from the speaker's perspective and use pragmatic reasoning to create better referring expressions for scenes including novel objects. Without negative sampling, regulating the embeddings of novel words and objects becomes crucial for pragmatic reasoning to induce an ME effect (Gulordava et al., 2020). Our results suggest that bounding the embedding space is one option to achieve this, but it may at the same time limit the model's flexibility. Even though pragmatic reasoning in neural networks is successful in complex domains, and can, in principle, induce an ME bias, the question of whether the proposed ME mechanism generalizes to such applications remains.

Several points should be considered when working with a more complex deep learning setup. First, given the limited number of words in the lexicon, our agent can iterate over all of them in its reasoning process. In a more realistic scenario with an ever-growing vocabulary size, this iteration is computationally too demanding. Many of the approaches mentioned above apply sampling techniques to limit the search space of speaker and listener. Yuan, Monroe, Bai, and Kushman (2018) show that using only the most promising word or object candidates in the pragmatic reasoning process even improves the agent's success. Second, a more challenging training setup can be investigated. How does the agent behave when facing multiple words or several distractor objects? Third, the behavior of the embedding space in relation to architecture and parameter choice must be better understood, such that more general solutions to regularizing the embeddings of novel words and objects can be developed. Our setup is well suited as proof of concept but any research trying to push these ideas to a full word learning model or to a deployable machine learning architecture must factor in these points.

In general, our work is in line with a trend toward building artificial agents with pragmatic reasoning abilities. This trend can be observed in language emergence research (e.g., Choi, Lazaridou, & de Freitas, 2018; Kang, Wang, & de Melo, 2020; Yuan et al., 2020), amongst others. Language emergence research often employs cooperative games that require a speaker and listener to develop a communication protocol. Our model can also be applied in a multiagent language emergence setting (Ohmer et al., 2020). Given that the pragmatic listener reasons about the speaker, a speaker agent is already part of the model. In addition, language emergence models typically also use reinforcement learning. Although the focus of this paper is on the ME bias, future work should apply our pragmatic agent models to other language learning and language emergence phenomena.

## 5. Conclusion

We have developed a model of learning in pragmatic agents, which can be parameterized by lexicon entries or neural network weights. We show that pragmatic inference combined with learning can account for the ME bias phenomenon and (at least qualitatively) its developmental trajectory, also under the influence of modulating factors. The neural network model demonstrates how pragmatic reasoning in semantic learning can implement an ME bias mechanism in deep word learning models. In future work, we would like to investigate the model behavior for many-to-one and one-to-many associations between words and objects, include a mechanism for determining relevant meaning distinctions, and find more general solutions to structuring the embedding space in the neural network model.

## Data availability

Materials and code are openly available at the Open Science Framework: https://osf.io/2wz9x/?view_only=154564daa91c4ce9a6398ea641ae598d

## Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—GRK 2340. We would like to thank the two anonymous reviewers and the editor for their extremely helpful comments and suggestions.

Open Access funding enabled and organized by Projekt DEAL.

## Conflicts of interest

The authors have no conflicts to disclose.

## Notes

1 Because we use the model in a learning context, $\alpha$ also influences the trade-off between exploration and exploitation. While $\alpha = 5$ provides a good balance, most qualitative results are robust across other values ($\alpha \in \{2.5, 10\}$). Additional explanations and analyses are provided in our Open Science Framework project.
2 Under a supervised learning regime, which provides a stronger feedback signal, the qualitative results stay the same. The main quantitative differences are that (a) training is faster, (b) performance is higher, and (c) performance increases more substantially for agents with a fixed lexicon where all possible word-referent mappings are updated at every training step.
3 The Zipfian input distribution is arguably a natural assumption about the relative frequency of meanings to be communicated. In previous work, we showed that pragmatic agents develop an ME bias regardless of whether words follow a uniform or a Zipfian distribution. However, we found the advantage of pragmatic reasoning and the resulting ME bias in terms of learning speed to be stronger for a Zipfian input distribution.
4 This sampling procedure implements a probabilistic version of applying a relative threshold criterion to the word probability under the speaker's policy. One can think of it as the speaker constructing a context model (fixing which words and objects are salient alternatives for pragmatic reasoning) by collecting a number of candidate words that easily come to mind.
5 The two categories cover all words since $I_{ME} > 0$ without exception.
6 Pretraining the CNN facilitates training the remaining model parameters. It can be done without loss of generality, given that we are interested in how the ME bias arises when associations between words and objects are learned, independent of how the visual features of these objects are extracted.

## 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).

Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, *3*(88), 1–8.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.

Bion, R., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39–53.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, *1*(1), 223–249.

Choi, E., Lazaridou, A., & de Freitas, N. (2018). Compositional obverter communication learning from raw visual input. In *International Conference on Learning Representations (ICLR)*.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. B. MacWhinney, *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Clark, E. V., & Amaral, P. M. (2010). Children build on pragmatic information in language acquisition. *Language and Linguistics Compass*, *4*(7), 445–457.

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

Cohn-Gordon, R., & Goodman, N. (2019). Lost in machine translation: A method to reduce meaning loss. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)* (pp. 437–441).

Cohn-Gordon, R., Goodman, N., & Potts, C. (2018). Pragmatically informative image captioning with character-level inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)* (pp. 439–443).

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. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96.

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.

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Development*, *17*(1), 1–17.

Fried, D., Andreas, J., & Klein, D. (2018). Unified pragmatic models for generating and following instructions. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)* (pp. 1951–1963).

Fried, D., Hu, R., Cirik, V., Rohrbach, A., Andreas, J., Morency, L.-P., … Darrell, T. (2018). Speaker-follower models for vision-and-language navigation. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Gandhi, K. & Lake, B. M. (2020). Mutual exclusivity as a challenge for neural networks. In Advances in Neural Information Processing Systems (NeurIPS) (pp. 14182-14192).

Grassmann, S., Schulze, C., & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology*, *6*, 1–8.

Gulordava, K., Brochhagen, T., & Boleda, G. (2020). Deep daxes: Mutual exclusivity arises through both learning biases and pragmatic strategies in neural networks. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society (CogSci)* (pp. 2089–2095).

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23–B34.

Hendrycks, D., & Gimpel, K. (2017). A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations (ICLR)*.

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, *65*(3), i–vi, 1–123.

Hollis, G. (2019). Learning about things that never happened: A critique and refinement of the Rescorla-Wagner update rule when many outcomes are possible. *Memory and Cognition*, *47*, 1415–1430.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157.

Huang, Y. T., & Snedeker, J. (2009). Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology*, *45*(6), 1723–1739.

Jang, E., Gu, S., & Poole, B. (2017). Categorical reparameterization with gumbel-softmax. In *International Conference on Learning Representations (ICLR)*.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word–referent mappings. *Psychonomic Bulletin & Review*, *19*(2), 317–324.

Kang, Y., Wang, T., & de Melo, G. (2020). Incorporating pragmatic reasoning communication into emergent language. In *Advances in Neural Information Processing Systems (NeurIPS)* (pp. 10348–10359).

LeCun, Y., Cortes, C., & Burges, C. (2010). MNIST handwritten digit database. *ATT Labs*. http://yann.lecun.com/exdb/mnist, *2*.

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, *198*, 104191.

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).

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 (ICLR)*.

Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. S. A. Gelman & J. P. Byrnes, *Perspectives on language and thought: Interrelations in development* (pp. 72–106). Cambridge: Cambridge University Press.

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. M. R. Gunnar & M. Maratsos, *Modularity and constraints in language and cognition: The Minnesota Symposia on Child Psychology* (Vol. 25, pp. 59–101). Lawrence Erlbaum Associates, Inc.

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.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Henry Holt and Co., Inc.

Martin, A., Onishi, K. H., & Vouloumanos, A. (2012). Understanding the abstract role of speech in communication at 12 months. *Cognition*, *123*(1), 50–60.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877.

Monroe, W., Hawkins, R. X. D., Goodman, N. D., & Potts, C. (2017). Colors in context: A pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics*, *5*, 325–338.

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

Najnin, S., & Banerjee, B. (2018). Pragmatically framed cross-situational noun learning using computational reinforcement models. *Frontiers in Psychology*, *9*(5), 1–18.

Nie, A., Cohn-Gordon, R., & Potts, C. (2020). Pragmatic issue-sensitive image captioning. In *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 1924–1938).

Ohmer, X., König, P., & Franke, M. (2020). Reinforcement of semantic representations in pragmatic agents leads to the emergence of a mutual exclusivity bias. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society (CogSci)* (pp. 1779–1785).

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science*, *24*(6), 1017–1023.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909–957.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*(6), 819–865.

Roberts, C. (2012). Information structure in discourse: Towards an integrated theory of pragmatics. *Semantics & Pragmatics*, *5*(6), 1–69.

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

Shen, S., Fried, D., Andreas, J., & Klein, D. (2019). Pragmatically informative text generation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)* (pp. 4060–4067).

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480–498.

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 (NeurIPS)* (pp. 3039–3047).

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. M. Bowerman & S. Levinson, *Language acquisition and conceptual development* (pp. 132–158). Cambridge: Cambridge University Press.

Vong, W. K., & Lake, B. M. (2020). Learning word-referent mappings and concepts from raw inputs. *arXiv preprint*, *arXiv:2003.05573*.

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, *109*(32), 12933–12937.

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

Yuan, A., Monroe, W., Bai, Y., & Kushman, N. (2018). Understanding the Rational Speech Act model. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society (CogSci)* (pp. 2759–2764).

Yuan, L., Fu, Z., Shen, J., Xu, L., Shen, J., & Zhu, S.-C. (2020). Emergence of pragmatics from referential game between theory of mind agents. *arXiv preprint*, *arXiv:2001.07752*.

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, *37*(5), 891–921.

Zarrieß, S., & Schlangen, D. (2019). Know what you don't know: Modeling a pragmatic speaker that refers to objects of unknown categories. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 654–659).

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

## Appendix A: Hyperparameter search—Explicit lexicon

To find good hyperparameters for the models with an explicit lexicon, we conducted a grid search for the fixed and dynamic lexicon simulations, respectively. We ran the search for an intermediate exposure interval of $k = 10$, and varied the following hyperparameters:

- data set size: $\{100, 1,000\}$
- batch size: $\{16, 32\}$
- learning rate: $\{0.001, 0.01, 0.1\}$
- lexicon initialization: $\{0.0001, 0.001, 0.01, 0.1\}$

So, for each lexicon, we tested $2 \times 2 \times 3 \times 4 = 48$ different combinations. Out of data set size, batch size, and learning rate, we used the hyperparameters that worked best across both lexicon types, to allow for a more direct comparison. The initial lexicon size is very different between the two lexicon types, which is why we used the lexicon initialization only for within implementation comparison. Choosing some parameters based on best performance across implementations is not problematic, since the final parameter combinations achieve (with negligible differences) the same performance as the best parameters for each implementation. The full list of results and details of the selection procedure can be found in our OSF project.

## Appendix B: ME index formulas

Words and objects are indexed $1 \leq i \leq N$, respectively, with words and objects of the same index belonging together. New word–object pairs are added to the training set in order of their indices. In the following formulas, the ME index is evaluated with respect to the word and object with index $j$, so $w_j$ and $o_j$. For the fixed lexicon, until the last word–object pair is added, there are always several novel objects the agent can select:

$$I_{ME}(B_{PL}, w_j) = \frac{\sum_{i=j}^{N} PL(o_i \mid w_j, B_{PL}) - \frac{N-(j-1)}{N}}{\frac{j-1}{N}}$$

In case of a dynamic lexicon, there is only one novel object the agent can select leading to the following simplification:

$$I_{ME}(B_{PL}, w_j) = \frac{PL(o_j \mid w_j, B_{PL}) - \frac{1}{M}}{\frac{M-1}{M}}$$

where $M <= N$ is the current lexicon size (after being extended for $w_j$ and $o_j$).

## Appendix C: Convolutional neural network for feature extraction

The vision module of the deep neural network implementation has the following architecture. First, there are two convolutional layers, each with 32 filters of size $3 \times 3$. Then a fully connected layer with 64 units follows and finally the output layer with 10 units. Hidden units use a relu activation function and output units a softmax activation function. Each convolutional layer is followed by a max-pooling layer with pooling size 2, and every layer apart from the output layer uses dropout with a probability of 0.3. Weights were initialized randomly, and training proceeded until an early stopping criterion with patience 3 was reached on the validation loss. We used a batch size of 32 and the Adam optimizer with a learning rate of 0.001. The model achieved the following training, validation, and test accuracies: 97.97%, 98.87%, 99.08%.

## Appendix D: Example lexica of the neural network agent

The matrices in Fig. D1 show the literal agent's lexicon after training, for three different random seeds. Learned associations are given by high lexicon entries, corresponding to strong similarities between word and object representations in the embedding space. It can be seen that the agent always learns the one-to-one correspondences between familiar objects (0–8) and familiar words (nine randomly selected words). The bottom row of the lexica shows how strongly the novel object, digit 9, is associated with each word. It turns out that associations with novel words are not necessarily stronger than with familiar words. However, novel words

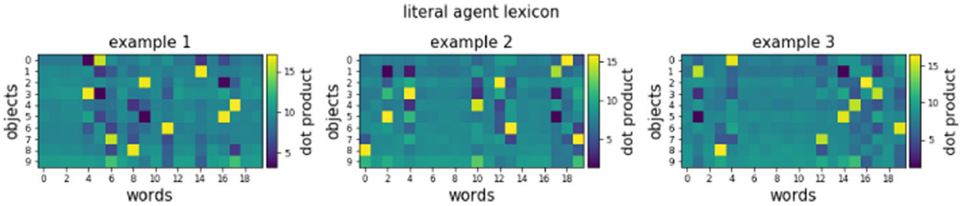


Fig. D1. Lexica of the literal agent from three randomly selected runs. Lexicon entry $ij$ is calculated as the average dot product between the embeddings of training examples showing object $i$ and the embedding of word $j$.

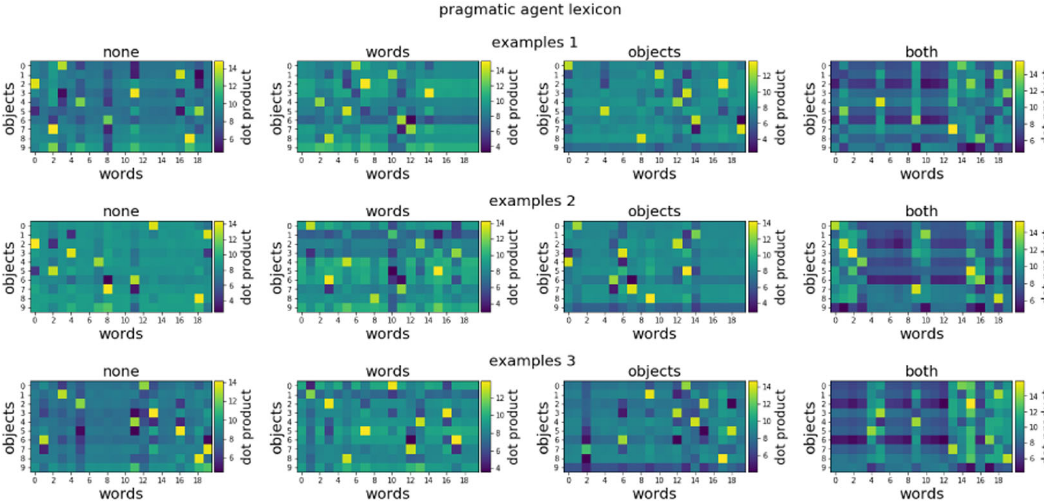


Fig. D2. Randomly selected lexica of the pragmatic agent ($\alpha = 5$) trained with different negative sampling strategies. Each row contains one example for each sampling strategy. Lexicon entry $ij$ is calculated as the average dot product between the embeddings of training examples showing object $i$ and the embedding of word $j$.

tend to lie closer to novel object 9 than to familiar objects, that is, lexicon entries for novel words (e.g., {0, 1, 2, 3, 7, 10, 12, 13, 15, 18, 19} in the first example) are relatively high in the bottom row. The structure of the joint embedding space illustrates why also the literal agent displays an ME bias in our simulations. After training, embeddings of novel words happen to be more similar to embeddings of the novel object than to embeddings of the familiar objects.

The matrices in Fig. D2 visualize the pragmatic agent's lexicon after training, for different negative sampling strategies, and for three different random seeds. Again, the learned one-to-one mappings between familiar words and familiar objects are clearly visible. The lexica without negative sampling (first column) and with negative sampling of words (second column) are similar to the lexica of the literal agent. With negative sampling of objects (third column), however, the agent learns that the novel object is not associated with any of the familiar words. Driving the embedding of the novel object away from the embeddings of familiar words simultaneously increases the distance to the embeddings of novel words, resulting in low values throughout the bottom row. This side effect can be mitigated by negative sampling of both words and objects (fourth column). Embeddings of novel words/objects move away from those of familiar objects/words. Within the bounded embeddings space, embeddings of novel words and novel objects stay close together in the process.