Learning to refer informatively by amortizing pragmatic reasoning

Julia White[1], Jesse Mu[2], Noah D. Goodman[2,3]
Departments of [1]Electrical Engineering, [2]Computer Science, and [3]Psychology
Stanford University
{jiwhite,muj,ngoodman}@stanford.edu

Abstract

A hallmark of human language is the ability to effectively and efficiently convey contextually relevant information. One theory for how humans reason about language is presented in the Rational Speech Acts (RSA) framework, which captures pragmatic phenomena via a process of recursive social reasoning (Goodman & Frank, 2016). However, RSA represents ideal reasoning in an unconstrained setting. We explore the idea that speakers might learn to *amortize* the cost of RSA computation over time by directly optimizing for successful communication with an internal listener model. In simulations with grounded neural speakers and listeners across two communication game datasets representing synthetic and human-generated data, we find that our amortized model is able to quickly generate language that is effective and concise across a range of contexts, without the need for explicit pragmatic reasoning.

Keywords: reference, pragmatics, rational speech acts, emergent communication

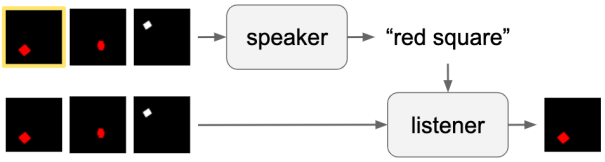

Figure 1: Reference game: Three images including one target (indicated by a gold box) are given to a speaker and listener. The speaker produces an utterance for the listener who then must choose the target image. In RSA, the speaker explicitly reasons about the listener's interpretation of alternative utterances; in our amortized model, the speaker is trained to generate informative utterances without such reasoning.

Introduction

When we refer to an object or situation using natural language, we easily generate an utterance that achieves a useful amount of information in the given context. Counterfactual reasoning about alternative utterances has been central to explanations of these pragmatic aspects of language (Grice, 1975; Searle, 1969). Yet these theories entail computations that appear slow and burdensome as cognitive processes, so how do we produce pragmatic utterances so easily?

One popular computational account of pragmatic reasoning is the Rational Speech Acts (RSA) model (Goodman & Frank, 2016). In RSA, speakers and listeners reason about the meaning of utterances by considering the other utterances that could have been produced, thus arriving at pragmatic interpretations that enrich literal semantics. In addition to referring expression generation (Frank & Goodman, 2012; Degen, Hawkins, Graf, Kreiss, & Goodman, 2020), RSA has seen success in modeling a wide variety of phenomena, including scalar implicature, metaphor, hyperbole, and politeness (for a review, see Goodman & Frank, 2016).

However, the successes of RSA have been in describing pragmatic language *competence*. In other words, within Marr's (1976) levels of analysis, RSA is a *computational* account of human language, describing only *what* is to be ideally computed, and not *how* humans produce language with limited resources. Indeed, the full computation specified by RSA involves reasoning over all possible utterances in context, which is intractable in all but the most trivial settings. Moreover, such computation is wasteful in that it is *memoryless*, and does not leverage past experience (Gershman & Goodman, 2014).

The impracticality of explicit pragmatic reasoning has had implications for both theory and modeling. Linguists have argued that pragmatic phenomena such as scalar implicature (Levinson, 2000) and metaphor (Lakoff & Johnson, 2008) are *conventionalized*, becoming the default interpretation of utterances unless specifically cancelled by the speaker. Others have argued that during pragmatic reasoning, listeners may only consider a subset of relevant interpretations (Sperber & Wilson, 1986; Fox & Katzir, 2011), or trade off between "slow" and "fast" processes (Degen & Tanenhaus, 2019). In natural language processing, RSA-based models for pragmatic referring expression generation use approximate inference methods, either sampling a subset of possible utterances from a learned model (Andreas & Klein, 2016; Monroe, Hawkins, Goodman, & Potts, 2017) or reasoning incrementally (Cohn-Gordon, Goodman, & Potts, 2019). Others use heuristics inspired from the psycholinguistics literature, with greedy and probabilistic search techniques to reduce the space of possible utterances (Krahmer & Van Deemter, 2012; van Gompel, van Deemter, Gatt, Snoeren, & Krahmer, 2019).

Given limited time and resources, it is indeed unlikely that we do exhaustive Bayesian reasoning every time we speak. An alternative is that experience with this slower reasoning process has led to amortized "habits of speech" that we use most of the time (Gershman & Goodman, 2014). In

this paper, we model this hypothesis through grounded language models that are trained with listeners in Lewis (1969)-style communication games (Fig. 1), measuring the degree to which they learn to speak pragmatically. Our models are able to acquire desirable pragmatic behavior while being far more efficient than naive RSA-based approaches. This suggests that agents can internalize pragmatic language production processes, providing a plausible *algorithmic* account of human language production in resource-constrained settings and a promising avenue for the development of more efficient pragmatic language models.

How to Speak Informatively

We explore communication within Lewis-style signalling games (1969) which allow us to analyze language use in grounded contexts. Formally, a reference game $(\mathbf{I}, t)$ consists of a set of N images $\mathbf{I} = (I_1, \ldots, I_n)$, with one target image I_t known only to the speaker. The speaker's job is to produce an utterance u which, when given to the listener, causes the listener to correctly identify t (Fig. 1). Crucially, the ideal language for image I_t changes depending on context. For example, in Fig. 1, it is not sufficient to simply say the color ("red") or shape ("square") of the target image; both must be provided to be unambiguous.

We consider a variety of speakers in this setting, each capturing different cognitive hypotheses about utterance production. The baseline **naive speaker** is a neural captioning model trained to generate observed descriptions of an image, completely ignoring context. To incorporate pragmatics, the naive speaker can be provided context objects as input, which we call the **contextual speaker**. In contrast, **RSA-based speakers** explicitly consider alternative utterances: the gold-standard **Full RSA** model does the complete reasoning process specified by RSA theory; the **sample-rerank** model approximates RSA by using a naive speaker to generate candidate utterances, reducing the space of alternatives. Our **amortized** speaker takes a different approach, training a context-aware language model with the RSA communication objective, without any pragmatic reasoning. The model is trained directly via backpropagation through a (white-box) internal listener model. In contrast, we also consider a **REINFORCE** speaker which learns from a sparser communicative reward signal given by a (black-box) external listener.

Naive speaker. The **naive speaker** S_0 is a standard image captioning (IC) model trained to generate referring expressions that completely ignore context. Given a reference game $(\mathbf{I}, t)$, S_0 embeds $\mathbf{I}_t$ with a convolutional neural network f_S, and uses this embedding to initialize a recurrent neural network decoder (RNN-IC) which provides a probability distribution over utterances u:

$$S_0(u \mid \mathbf{I}, t) = p_{\text{RNN-IC}}(u \mid f_S(I_t)). \tag{1}$$

RNN-IC is trained with a standard language modeling objective with teacher forcing: given an image I and ground-truth utterance u, the loss of a predicted caption $\hat{u}$ is

$$\mathcal{L}_{S_0}(\hat{u}, u) = \sum_t p_{\text{RNN-IC}}(\hat{u}_t = u_t \mid u_{<t}; f_S(I_t)). \tag{2}$$

Contextual speaker. The simplest way of incorporating context-sensitivity into S_0 is to condition it on the entire set of images $\mathbf{I}$ as opposed to just I_t. Given the same vision model f_S, let $\mathbf{c}_t$ be the embedding of the target image: $\mathbf{c}_t = f_S(I_t)$. Let $\mathbf{c}_{t'_1}$ and $\mathbf{c}_{t'_2}$ be embeddings of the distractor images for $t'_i \neq t$: i.e. $\mathbf{c}_{t'_1} = f_S(I_{t'_1})$. Then our **contextual speaker** S'_0 is trained identically to S_0, except it is conditioned on the concatenation of all three embeddings:

$$S'_0(u \mid \mathbf{I}, t) = p_{\text{RNN-IC}}(u \mid \mathbf{c}_t; \mathbf{c}_{-t'_1}; \mathbf{c}_{-t'_2}). \tag{3}$$

RSA speakers

RSA speakers generate utterances with the goals of accurate and efficient communication. Accuracy is based on how a pretrained internal listener model L_0 will interpret the utterance. Given a reference game $\mathbf{I}$ and a speaker utterance u, L_0 represents the probability of target t as proportional to the dot product between embeddings of I_t and u:

$$L_0(t \mid \mathbf{I}, u) \propto \exp\left(f_L(I_t)^\intercal g(u)\right), \tag{4}$$

where f_L is a CNN encoder (with the same architecture as f_S) and g is an RNN encoder. These encoders will be trained from observed target-utterance pairs.

The speaker S_{RSA} then chooses an utterance u which maximizes the probability of the listener identifying t, while also balancing the cost of the utterance $C(u)$:

$$S_{\text{RSA}}(u \mid \mathbf{I}, t) \propto \exp\left(\log L_0(t \mid \mathbf{I}, u) - C(u)\right). \tag{5}$$

When we can enumerate all possible utterances $\mathbf{U}$, we can compute Eq. 5 exactly, picking $\arg\max_{u \in \mathbf{U}} S_{\text{RSA}}(u \mid \mathbf{I}, t)$; we call this model **Full RSA** ($S_{\text{RSA-Full}}$) and treat it as an upper bound on the pragmatic quality of a speaker model. However, in many cases, there are an unbounded number of utterances to consider. Past work (Andreas & Klein, 2016; Monroe et al., 2017) uses the naive speaker S_0 as a *proposal* distribution from which a subset of M utterances $\mathbf{U}' = (u'_1, \ldots, u'_M)$ is sampled; exact inference is then performed with this subset. Since this involves *sampling* candidate utterances from S_0 then *reranking* them based on communicative utility, we refer to these approximations as **sample-rerank** models ($S_{\text{RSA-SRR}}$). For both models, our cost function $C(u)$ is simply a penalty linear in the length of the utterance: $C(u) = \lambda \|u\|$. We used a constant $\lambda = 1$ for $S_{\text{RSA-SRR}}$, $\lambda = 0.0001$ for $S_{\text{RSA-Full}}$, and $\lambda = 0.01$ for $S_{\text{RSA-Am}}$, and for $S_{\text{RSA-SRR}}$, set $M = 5$.[1]

[1] Increasing M resulted in little performance benefit.

Learning to produce informative utterances

Both versions of RSA speakers described so far generate informative utterances by explicitly considering alternatives, which can be slow and expensive. We next describe a model that *amortizes* this cost, learning to produce the utterances that RSA would prefer, without needing to explicitly consider alternatives at neither train nor test time.

Our **amortized speaker** $S_{\text{RSA-Am}}$ has the same architecture as the naive contextual model S_0', generating an utterance after encoding the target and distractors:

$$S_{\text{RSA-Am}}(u \mid \mathbf{I}, t) = p_{\text{RNN}}(u \mid \mathbf{c}_t; \mathbf{c}_{-t_1'}; \mathbf{c}_{-t_2'}) \tag{6}$$

However, while S_0' is trained as an IC model to match observed utterances, $S_{\text{RSA-Am}}$ is trained to directly optimize the RSA objective. Specifically, define the loss of a sampled caption $\hat{u}$ as the listener surprisal plus the cost of the utterance:

$$\mathcal{L}_{S_{\text{RSA-Am}}}(\hat{u} \mid \mathbf{I}, t) = -\log L_0(t \mid \mathbf{I}, \hat{u}) + C(\hat{u}). \tag{7}$$

Equation 7 looks similar to the RSA objective, Equation 5, but crucially omits the normalization term. This means we do not have to consider alternative utterances when training the model: for a fixed, pre-trained internal listener model L_0, each optimization step consists of sampling an utterance from $S_{\text{RSA-Am}}$, evaluating the listener model, then updating the parameters to minimize $\mathcal{L}_{\text{RSA-Am}}$. Note that unlike S_0', utterances are evaluated solely via communicative utility, and not compared to observed language.

A technical problem with training this model is the need to estimate the gradient of $\mathcal{L}_{\text{RSA-Am}}$ given discrete sequences sampled by our speaker. We use the Gumbel-Softmax trick (Jang, Gu, & Poole, 2017), which gives differentiable "samples" from a categorical distribution by adding Gumbel noise and applying a softmax with a temperature τ. To ensure that L_0 receives discrete inputs, we use the (biased) straight-through estimator: the utterance is discretized in the forward pass, but treated as the original continuous sample in the backwards pass. A constant $\tau = 1$ allowed our models to train successfully. Our differentiable cost function $C(u)$ is implemented as a penalty on not predicting the end of sentence token, which increases by λ after each sampled token.

$S_{\text{RSA-Am}}$ has an introspectable internal listener model that can be used for explicit pragmatic reasoning. Importantly, this means that the speaker model receives subtle word-by-word supervision during training. An alternative is an agent that has no internal listener model, but still attempts to maximize communication success by trial and error. We thus consider a reinforcement learning model, the **REINFORCE speaker**, S_{RL}, which is architecturally identical to $S_{\text{RSA-Am}}$, but trained with a black-box reward function rather than an explicit internal listener. A discrete utterance $\hat{u}$ is sampled from S_{RL} as before. We simulate a referent choice from an external listener by selecting $\hat{t} = \arg\max_t L_0(t \mid \mathbf{I}, \hat{u})$. S_{RL} receives a reward of +1 if the listener identifies the correct target and 0 otherwise[2]. The weights of S_{RL} are then updated via the REINFORCE gradient estimator (Williams, 1992). From a reinforcement learning perspective, S_{RL} and $S_{\text{RSA-Am}}$ have the same objective, but the former is model-free while the latter is model-based, with RSA as the model of communication.

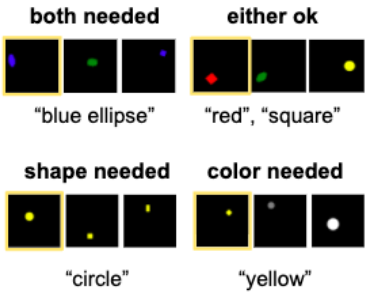

Figure 2: Example reference games with ideal utterances.

Experiments

Datasets

We run experiments on two reference game datasets representing synthetic and human-generated data. **ShapeWorld** (Kuhnle & Copestake, 2017) is an artificial dataset where each game consists a set of three images, each containing a single random shape with random color. Each game varies in the informativity of the utterance required for successful communication: either only the shape or color is sufficient, or both are needed (Fig. 2). This allows us to evaluate how our speakers modulate their utterances depending on the context. Our primary ShapeWorld dataset has 76000 games containing artificially generated utterances with a total vocabulary of 15 words. Additionally, to test systematic generalization to novel contexts, we constructed similar datasets where either a color (red), shape (square), or set of color-shape combinations (red circle, blue square, yellow ellipse, white circle, gray square) were held-out during training, but presented as targets in every test game.

Colors (Monroe et al., 2017) is a reference game dataset with real human language, where human speakers saw three patches of color and were asked to produce utterances that uniquely identify a target color. Here, contexts varied in their difficulty: either distractors were perceptually distinct from the target image (far) or were similar in color space (close) (Fig. 3). The language used in the ∼46000 games in this dataset is considerably more complex, with around 1400 unique vocabulary tokens; thus, reasoning over all possible utterances (as required by Full RSA) is infeasible.

[2]This is equivalent to stochastically sampling a reward from the RSA listener with a very low softmax temperature. A higher temperature led to more variance during training and overall worse performance.

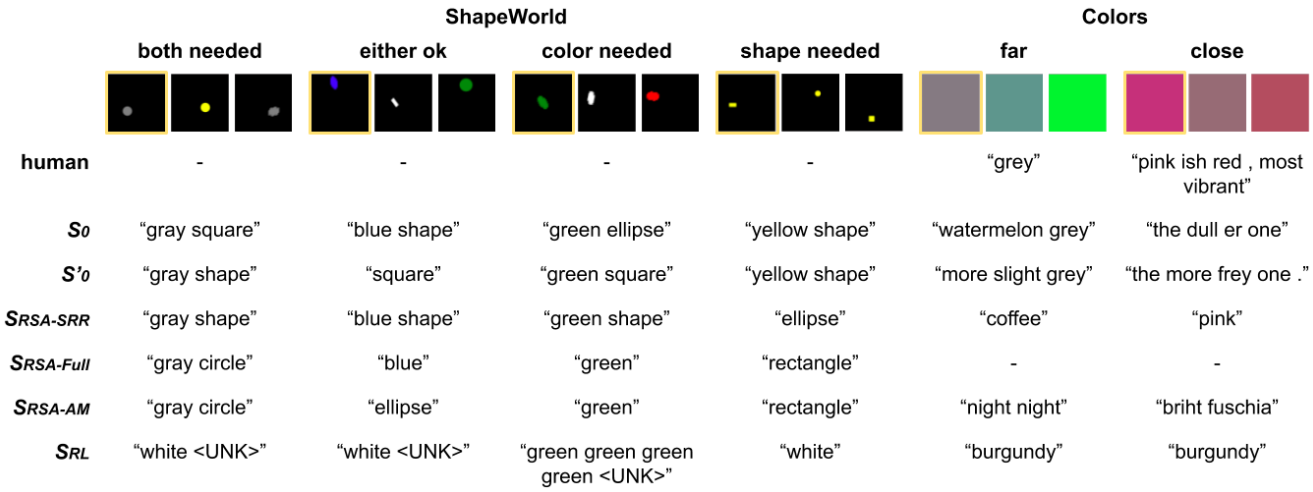

Figure 3: Output examples for ShapeWorld and Colors.

Training and Model Details

We split each dataset into a training, validation, and test split of 60:15:1 on ShapeWorld and 42:3:1 on Colors. The training datasets for each task were further subdivided into thirds. 1/3 was used to train our speakers (and speaker-internal listener L_0 if used). The second 1/3 of the training data was used to train a *validation listener*: training was stopped after a speaker obtained maximum accuracy with this listener on the validation set. Finally, the last 1/3 was used to train *evaluation listeners*. The ultimate communicative accuracy of speaker models was evaluated with these evaluation listeners on the test set. These divisions ensured that speakers did not overfit to a single listener, and instead were evaluated for behavior that generalized across different listeners.

Models were trained for a maximum of 100 epochs with the Adam optimizer (Kingma & Ba, 2014) with a batch size of 32 and a learning rate of 0.001 for speakers and 0.01 for listeners. After training, the models with the highest validation accuracy were kept. RNNs (both encoders and decoders) are 100-d unidirectional Gated Recurrent Unit (GRU) RNNs (Cho et al., 2014) with 50-d word embeddings learned from scratch, except for the contextual speakers S'_0, which have hidden size 300-d (since they consume 3 embeddings). The vision modules f_S and f_L are CNNs with 4 blocks, each block containing a 64-filter 3x3 convolutional layer, batch normalization, ReLU activation, and 2x2 max pooling. Given shapes and colors represented as $64 \times 64 \times 3$ input images, this produces 1024-d representations. In speakers, these are projected via a linear layer down to the GRU hidden size to initialize the GRU; for listeners, to compare image and text embeddings, we project the text representation into the 1024-d image space via a linear layer and use dot product to produce the target probability. Our code is available at `https://github.com/juliaiwhite/amortized-rsa`.

Results

We evaluate our speakers' utterances (see Fig. 3 for examples) in several ways. First, language should be **effective**: it should serve its communicative goal of helping the listener correctly identify the target image. We measure communication success with our evaluation listeners on unseen reference games. Second, language should be **concise**, saying as much as needed, but no more. We measure this by examining how the length and content of the utterance changes depending on the difficulty of the contexts as specified in either dataset. Finally, language should be **conventional**: to cause minimal confusion to a listener, it should look like conventional, grammatical English. Because the ShapeWorld dataset uses completely artificial language, we measure conventionality on the Colors dataset only.[3] As an imperfect proxy to conventionality, we measure the per-word probability of the utterances generated from our speakers, as reported by an unconditional language model trained on utterances in the training data.[4]

Efficacy. All speaker models, except for the REINFORCE model S_{RL}, perform well above chance on both test datasets (Fig. 4) though there are significant differences between them. The naive baseline S_0 attains an accuracy of 73% for ShapeWorld and 67% for Colors. While the additional context given to S'_0 results in some improvement, more substantial benefits come when models use pragmatic objectives. The sample-rerank model $S_{\text{RSA-SRR}}$ obtains much higher accuracy, only slightly behind $S_{\text{RSA-Full}}$ in ShapeWorld and even outperforming the human utterances in Colors.[5] Most notably,

[3] RSA and Amortized models used coherent one word utterances, e.g. “blue”, which due to the artificial nature of the dataset, did not exist in the training data (which had only “blue shape”). Thus, they had extremely low probability, making this evaluation nonsensical.

[4] Per-word probability avoids confounds with utterance length.

[5] Human utterance accuracy is imperfect because utterances are evaluated with respect to a neural listener, not a human listener. Our listener model is an imperfect approximation for a human listener,

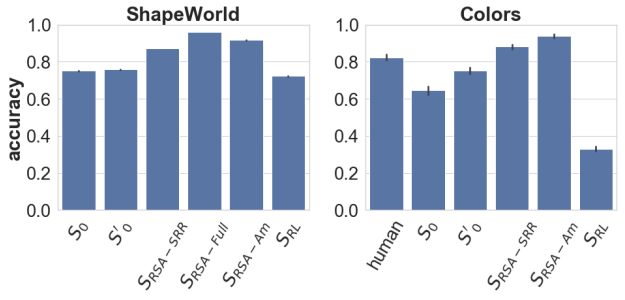

Figure 4: **Efficacy.** Accuracy for ShapeWorld (for 9 evaluation listeners) and Colors (for 1 evaluation listener). Error bars correspond to the 95% confidence intervals across the test games for all figures.

our amortized model $S_{\text{RSA-Am}}$ also performs very well, performing on par with $S_{\text{RSA-SRR}}$ in ShapeWorld and achieving the highest accuracy (94%) out of all the models tested for Colors. Meanwhile, the REINFORCE model S_{RL} struggles; it performs on par with S_0 for ShapeWorld and is unable to learn in the Colors setting.

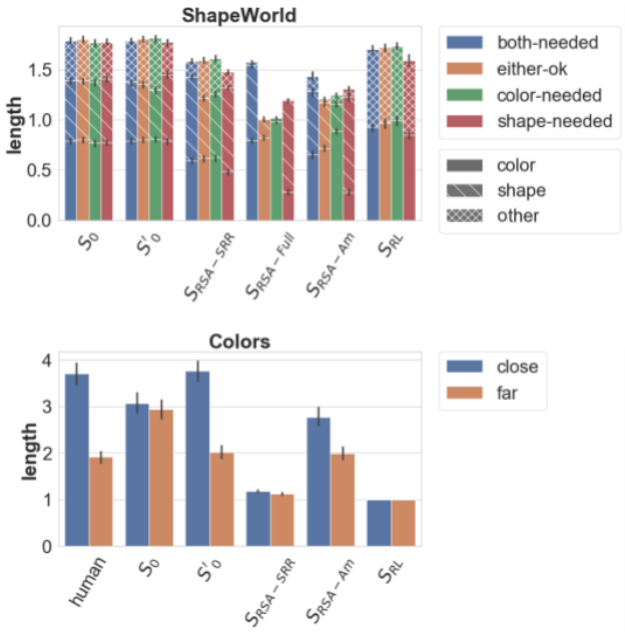

Figure 5: **Concision.** Average utterance lengths produced under different contexts for ShapeWorld and Colors. For Shapeworld, utterance length is split into the average number of colors, shapes, and other tokens (which include the unknown token, "<UNK>", and generic term "shape") .

Concision. We quantify concision by measuring the speaker's average utterance lengths for the different contexts of our reference game (Fig. 5). Analyzing utterance length within context, we see that RSA $S_{\text{RSA-Full}}$ and the amortized model $S_{\text{RSA-Am}}$ exhibit appropriately longer utterance lengths when both the shape and color are needed to identify the target image as opposed to either component separately. Looking more specifically at how the number of color and shape words per utterance differ based on the context, we see that in contexts where the shape is required, these models are less likely to say a color and vice versa.[6] This behavior is not reflected at all in the naive speaker S_0, and is less pronounced in the contextual (S'_0) and sample-rerank ($S_{\text{RSA-SRR}}$) models. In Colors, we see that humans tailor utterance length to game difficulty: for the difficult "close" contexts where colors were the same shade, utterances tended to be longer than the easier "far" contexts. Here, only the contextual speaker S'_0 and amortized model $S_{\text{RSA-Am}}$ demonstrate a similar significant difference in utterance length.

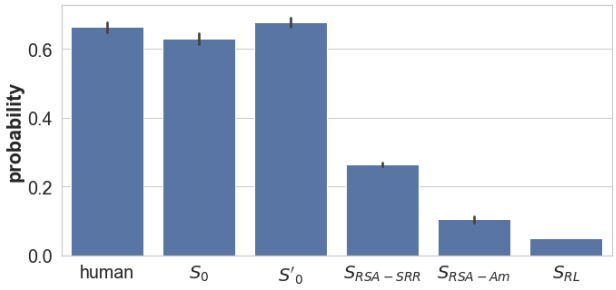

Figure 6: **Conventionality.** Average per-word probability of an utterance for Colors.

Conventionality. Finally, we quantify the extent to which generated utterances reflect conventional language, with an unconditional language model trained on human utterances from the Colors dataset. We explore the probability per word of utterances generated by our speaker models (Fig. 6). Unsurprisingly, the naive S_0 and contextual speakers S'_0, which are trained with language-modeling objectives, have high probability. The sample re-rank model $S_{\text{RSA-SRR}}$ sees a major decrease in probability, as it re-ranks sampled utterances with respect to an internal listener, which does not necessarily favor the most probable utterance.

The amortized model $S_{\text{RSA-Am}}$ is lower still: the communicative training objective results in *language drift* (see Fig. 3 for examples) whereby the model is able to sacrifice realism for communicative efficacy; this has been reported in similarly trained models (Lazaridou, Potapenko, & Tieleman, 2020). It is unclear whether this reflects creative and acceptable, but unconventional, use of language, or malformed language that would be hard for humans to understand. Regardless, the language is not overfit to the amortized speaker's

but is still a reasonable measure of efficacy given that we measure accuracy across multiple separately-trained listeners.

[6] Both RSA and amortized models also capture the well-attested tendency of humans to produce colors more often than shapes, all else being equal. This in turn follows from properties of the neural listener: a CNN is more accurate at detecting colors than shapes.

internal listener, given its ability to generalize to listeners trained on separate data.

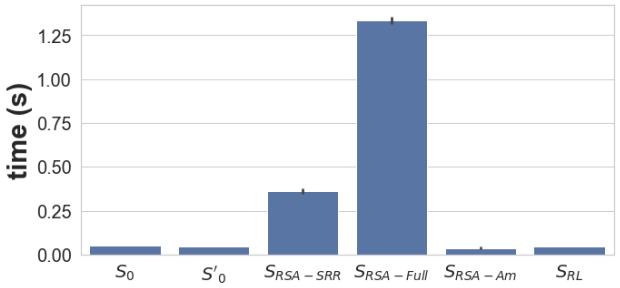

Figure 7: **Inference Speed.** Average time to generate an utterance for ShapeWorld.

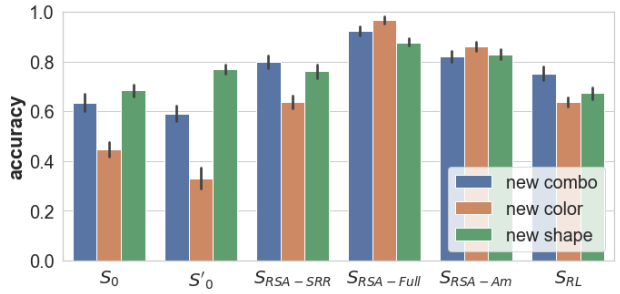

Figure 8: **Generalization.** Accuracy for ShapeWorld when generalizing to a held-out color, shape, or color-shape combination.

Inference Speed. Fig. 7 depicts average utterance generation time across 100 games for ShapeWorld. S_0, S'_0, $S_{\text{RSA-Am}}$, and S_{RL} are similarly fast, since they require only a single sample from an RNN. In contrast, $S_{\text{RSA-Full}}$ and $S_{\text{RSA-SRR}}$ must evaluate a listener across many utterances, resulting in much higher inference times. While these results are merely suggestive—our RSA models can likely be optimized—they point to the large gap in compute requirements between online pragmatic reasoning and amortized utterance production.

Generalization. Fig. 8 displays communication accuracy on the dataset variants with held-out colors, shapes, or combinations. The amortized model comes closest to the performance of Full RSA, followed by the sample-rerank and REINFORCE models. The contextual captioner's poor performance here likely indicates that, where it captures pragmatic behavior, it does so primarily by memorization. This indicates that communication-based training is needed to produce pragmatic language in novel contexts.

Conclusion and Discussion

We explored several models for speech production in reference games and evaluated them with respect to pragmatic efficiency, concision, conventionality, and processing time. The Full RSA model represents the gold standard at a competence level, capturing nuances of pragmatic behavior and generalizing well. But it requires large, often intractable, processing costs. In contrast, our *amortized* RSA model achieves performance close to Full RSA, while being far more efficient.

The contextual speaker, S'_0, performs poorly with respect to efficacy and concision, though well with respect to convention. Because the contextual model matches our amortized model in architecture but uses a language modeling objective, we conclude that some aspects of pragmatic communication are unattainable by simply observing surface-level linguistic cues. Despite recent work showing that language models may implicitly learn pragmatic phenomena from sufficient data (Schuster, Chen, & Degen, 2020), our simulations suggest that communication-based training is required for successful pragmatic communication, especially when generalizing to situations that have not been seen in the training distribution. The reinforcement learning approach provides a complementary contrast: it shared the communicative objective with the RSA models, but was forced to learn from the weaker signal of communicative success or failure. The poor performance of this model suggests that an internal model of the listener greatly helps communication-based training.

While common pragmatic behaviors may be amortized, our model must still learn from experience; highly novel situations may still require the full reasoning processes afforded by frameworks like RSA. Indeed, empirical evidence reports significant variability in processing time across instances of a pragmatic phenomenon (Degen & Tanenhaus, 2015). To better model this variability, one interesting extension of the work presented here is a model which reverts from amortized computation back to more sophisticated reasoning procedures in novel situations. Such a model would operationalize the trade-off between slow explicit reasoning processes and fast amortized computation that motivates recent theories of language processing (Degen & Tanenhaus, 2019) and amortized probabilistic inference (Gershman & Goodman, 2014).

An intriguing connection is to models of emergent communication (Lazaridou, Peysakhovich, & Baroni, 2017; Mordatch & Abbeel, 2018; Lazaridou et al., 2020). In order to ground our amortized speaker in actual language, the *listener* is pre-trained on real language and fixed. In most emergent communication models, either speaker and listener are co-trained from scratch, or in an attempt to ground the communication protocol, the *speaker* is pre-trained and/or given auxiliary language grounding tasks (Lazaridou et al., 2017).

In this paper we described an algorithmic model that forms "habits of pragmatic speech", internalizing the explicit pragmatic reasoning of Rational Speech Acts models into a fast but flexible speaker. This represents one solution to the puzzle of pragmatic speech: How do we produce pragmatically useful utterances so easily and quickly?

Acknowledgments

We thank anonymous reviewers and Christopher Potts for insightful comments. This research was supported in part by DARPA under agreement FA8650-19-C-7923, an NSF Graduate Research Fellowship for JM, and the Office of Naval Research grant ONR MURI N00014-16-1-2007.

References

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

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734).

Cohn-Gordon, R., Goodman, N., & Potts, C. (2019). An incremental iterated response model of pragmatics. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019* (pp. 81–90).

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A bayesian approach to "overinformative" referring expressions. *Psychological Review*.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science*, *39*(4), 667–710.

Degen, J., & Tanenhaus, M. K. (2019). Constraint-based pragmatic processing. In *The Oxford handbook of experimental semantics and pragmatics.*

Fox, D., & Katzir, R. (2011). On the characterization of alternatives. *Natural Language Semantics*, *19*(1), 87–107.

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

Gershman, S., & Goodman, N. (2014). Amortized inference in probabilistic reasoning. In *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci)* (Vol. 36).

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818-829.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3: Speech acts.* New York: Academic Press.

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

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR).*

Krahmer, E., & Van Deemter, K. (2012). Computational generation of referring expressions: A survey. *Computational Linguistics*, *38*(1), 173–218.

Kuhnle, A., & Copestake, A. (2017). ShapeWorld - a new test methodology for multimodal language understanding. *arXiv preprint*. doi: arXiv:1704.04517

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago Press.

Lazaridou, A., Peysakhovich, A., & Baroni, M. (2017). Multi-agent cooperation and the emergence of (natural) language. In *International Conference on Learning Representations (ICLR).*

Lazaridou, A., Potapenko, A., & Tieleman, O. (2020). Multi-agent communication meets natural language: Synergies between functional and structural language learning. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL).*

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

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

Marr, D., & Poggio, T. (1976). *From understanding computation to understanding neural circuitry*. Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

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 (TACL)*, *5*, 325–338.

Mordatch, I., & Abbeel, P. (2018). Emergence of grounded compositional language in multi-agent populations. In *Thirty-Second AAAI Conference on Artificial Intelligence.*

Schuster, S., Chen, Y., & Degen, J. (2020). Harnessing the richness of the linguistic signal in predicting pragmatic inferences. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL).*

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge University Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Cambridge, MA: Harvard University Press.

van Gompel, R. P., van Deemter, K., Gatt, A., Snoeren, R., & Krahmer, E. J. (2019). Conceptualization in reference production: Probabilistic modeling and experimental testing. *Psychological Review*, *126*(3), 345.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, *8*(3-4), 229–256.