Reinterpreting Signaling and Referential Games as Generative Models

Ryo Ueda
The University of Tokyo
Japan
ryoryoueda@is.s.u-tokyo.ac.jp

Abstract

Emergent Communication (EC) is a field that aims to unravel the evolution or dynamics of language by simulating its emergence. This paper reinterprets commonly used communication models in EC, such as signaling games and reference games, within the framework of generative models based on variational inference. Specifically, we formalize a game called a contextualized signaling game, which can be viewed as a type of Conditional Variational Autoencoder (CVAE). We then confirm that it generalizes generative versions of signaling and reference games.

1 Introduction

The purpose of this paper is to reinterpret signaling games and referential games; communication models frequently used in the field of Emergent Communication [EC, Lazaridou and Baroni, 2020, Peters et al., 2024, Boldt and Mortensen, 2024], as some form of generative models. In particular, we show that the games can be reinterpreted as a type of Conditional Variational Autoencoder [CVAE, Kingma et al., 2014, Sohn et al., 2015]. EC is a related field of evolutionary and computational linguistics that takes a constructive approach to providing insights into the emergence and dynamics of language. Although attempts to simulate the emergence of language have existed for a long time [Steels, 1999, Nowak and Krakauer, 1999, Briscoe, 2000, Kirby, 2002], recent advancements in representation learning and reinforcement learning have brought renewed attention. Simple communication models like Lewis' signaling game [Lewis, 1969] or its variant called referential game [Havrylov and Titov, 2017, Lazaridou et al., 2017] are often adopted for their simplicity, while various formulations are also possible depending on the focused aspect of communication dynamics [e.g., Foerster et al., 2016, Lowe et al., 2017, Jaques et al., 2019, Ebara et al., 2023, Lo et al., 2024].

In the signaling game, there are only two players (agents), a **sender** S and a **receiver** R. In each play, the sender S obtains an **observation** $\boldsymbol{x} \in \mathcal{X}$ randomly and converts it into a **message** $\boldsymbol{m} \in \mathcal{M}$. The receiver R then receives $\boldsymbol{m}$ and tries to guess the original observation $\boldsymbol{x}$, outputting a prediction $\hat{\boldsymbol{x}} \in \mathcal{X}$. The game is successful if $\boldsymbol{x} = \hat{\boldsymbol{x}}$. In the referential game, instead of choosing a prediction from the entire set $\mathcal{X}$, the receiver R tries to pick up the correct answer $\boldsymbol{x}$ from a candidate set $\{\boldsymbol{x}, \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(K-1)}\}$, which includes the incorrect candidates (**distractors**) $\boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(K-1)} \in \mathcal{X} \backslash \{\boldsymbol{x}\}$. The sender S and receiver R are typically represented as probabilistic models based on neural networks, optimized to make the game more likely to succeed. The communication protocol that emerges between the two agents can be considered "language" in that it serves as a symbolic system for transmitting information, which is often referred to as **emergent language**. However, within such simple games, it has often been pointed out that emergent languages lack certain properties of human languages [Kottur et al., 2017, Chaabouni et al., 2019, Ueda et al., 2023]. Previous work has attempted to mitigate this issue by modifying the framework, e.g., modeling humans' cognitive constraints [Ueda and Washio, 2021, Ri et al., 2023, Kato et al., 2024] or incorporating an evo-linguistic scenario [Graesser et al., 2019, Ren et al., 2020, Dagan et al., 2021].

Language Gamification Workshop at NeurIPS 2024 (LanGame @ NeurIPS 2024).

Recently, Ueda and Taniguchi [2024] presented a slightly different direction. They proposed to reinterpret the signaling game as a form of a generative model, specifically (beta-)VAE [Kingma and Welling, 2014, Higgins et al., 2017].[1] In this paper, we further extend this reinterpretation by considering not only signaling games but also reference games as types of generative models, by showing that these games can be uniformly reinterpreted as a form of Conditional VAE. Experimental justification for the "goodness" of this formulation is left for future work.

2 Background

2.1 Signaling Game and Referential Game

Conventional Objective of Signaling Game: Let $\mathcal{X}$ be an **observation space** and let $\mathcal{M}$ be a **message space**. The probability distribution of the observation is denoted as $P_X(X)$. A probabilistic model $S_{\boldsymbol{\phi}}(M|X)$, parametrized by $\boldsymbol{\phi}$, is referred to as a **sender**, while a probabilistic model $R^{\text{sig}}_{\boldsymbol{\theta}}(X|M)$, parametrized by $\boldsymbol{\theta}$, is referred to as a **receiver**. The intuitive procedure of the game is a unidirectional communication as described in Section 1, but typically, the following autoencoder-like objective function is adopted for optimization [Chaabouni et al., 2019, Rita et al., 2022]:

$$\mathcal{J}^{\text{sig}}(\boldsymbol{\phi}, \boldsymbol{\theta}) := \mathbb{E}_{P_X(\boldsymbol{x}), S_{\boldsymbol{\phi}}(\boldsymbol{m}|\boldsymbol{x})}[\log R^{\text{sig}}_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{m})]. \tag{1}$$

Conventional Objective of Referential Game: The observation space $\mathcal{X}$, message space $\mathcal{M}$, and sender $S_{\boldsymbol{\phi}}(M|X)$ are as above. A function $R^{\text{ref}}_{\boldsymbol{\theta}} : \mathcal{X} \times \mathcal{M} \to \mathbb{R}$ is referred to as a **receiver** agent in the referential game. Let K be an integer larger than 1. The objective function of the reference game is often defined as follows [Dessì et al., 2021, Chaabouni et al., 2022, Guo et al., 2022]:

$$\mathcal{J}^{\text{ref}}(\boldsymbol{\phi}, \boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{x}^{(1:K)} \overset{\text{w/o repl.}}{\sim} P_X(\cdot), S(\boldsymbol{m}|\boldsymbol{x}_1)} \left[\log \frac{\exp(R^{\text{ref}}_{\theta}(\boldsymbol{x}_1, \boldsymbol{m}))}{\sum_{i=1}^{K} \exp(R^{\text{ref}}_{\theta}(\boldsymbol{x}_i, \boldsymbol{m}))} \right], \tag{2}$$

where $\boldsymbol{x}^{(1:K)} \overset{\text{w/o repl.}}{\sim} P_X(\cdot)$ represents sampling K elements without replacement. Without loss of generality, $\boldsymbol{x}_1$ is regarded as the correct answer, and the others are distractors. It can be seen as InfoNCE [van den Oord et al., 2018], regarding $\boldsymbol{x}_1$ as a positive example and $\boldsymbol{x}_{2:K}$ as negative ones.

2.2 Reinterpretation of Signaling Game as (beta-)VAE

In contrast to the conventional formulations above, Ueda and Taniguchi [2024] argue that the objective function of the signaling game should be (re-)defined as **ELBO**:

$$\mathcal{J}^{\text{sig-elbo}}(\boldsymbol{\phi}, \boldsymbol{\theta}; \beta) := \mathbb{E}_{P_X(\boldsymbol{x})}[\mathbb{E}_{S_{\boldsymbol{\phi}}(\boldsymbol{m}|\boldsymbol{x})}[\log R^{\text{sig}}_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{m})] - \beta \,\mathrm{KL}(S_{\boldsymbol{\phi}}(M|\boldsymbol{x}) \mid\mid P^{\text{prior}}_{\boldsymbol{\theta}}(M))], \tag{3}$$

which can be transformed as:

$$= \mathcal{J}^{\text{sig}}(\boldsymbol{\phi}, \boldsymbol{\theta}) + \beta \,\mathbb{E}_{P_X(\boldsymbol{x}), S_{\boldsymbol{\phi}}(\boldsymbol{m}|\boldsymbol{x})}[\log P^{\text{prior}}_{\boldsymbol{\theta}}(\boldsymbol{m})] + \beta \,\mathbb{E}_{P_X(X)}[\mathcal{H}(S_{\boldsymbol{\phi}}(M|\boldsymbol{x}))], \tag{4}$$

where $\beta \geq 0$ is a hyper-parameter or annealed during training. This formulation essentially adds a prior term $\log P^{\text{prior}}_{\boldsymbol{\theta}}(\boldsymbol{m})$ and an entropy maximizer $\mathcal{H}(S_{\boldsymbol{\phi}}(M|\boldsymbol{x}))$, weighted by β, to the conventional objective function $\mathcal{J}^{\text{sig}}(\boldsymbol{\phi}, \boldsymbol{\theta})$.

Rationale for Introducing Prior: The first reason for introducing the prior is that the conventional objective function $\mathcal{J}^{\text{sig}}(\boldsymbol{\phi}, \boldsymbol{\theta})$ already contains an implicit, uniform prior distribution $P^{\text{prior}}_{\text{unif}}(M)$. This follows from the idea that, by appropriately choosing a prior such that $\nabla_{\boldsymbol{\phi},\boldsymbol{\theta}} \,\mathbb{E}_{P_X(\boldsymbol{x}), S_{\boldsymbol{\phi}}(\boldsymbol{m}|\boldsymbol{x})}[\log P^{\text{prior}}_{\text{unif}}(\boldsymbol{m})] = \mathbf{0}$, adding it to $\mathcal{J}^{\text{sig}}(\boldsymbol{\phi}, \boldsymbol{\theta})$ would have no impact on gradient-based optimization. In fact, $P^{\text{prior}}_{\text{unif}}(\boldsymbol{m})$ is the uniform distribution over messages, i.e.,

[1] Similar trends have been presented in (variational) information bottleneck-based emergent communication [Zaslavsky et al., 2018, Chaabouni et al., 2021, Tucker et al., 2022] and Metropolis-Hastings (MH) naming games [Taniguchi et al., 2023, Inukai et al., 2023, Okumura et al., 2023, Hoang et al., 2023]. The variational information bottleneck is known to be a generalization of beta-VAE [Alemi et al., 2017, Achille and Soatto, 2018]. The MH naming game adopts MCMC-based inference instead of variational inference.

$P^{\text{prior}}_{\text{msg}}(\boldsymbol{m}) = 1/|\mathcal{M}|$.[2] The second reason is that the implicit (uniform) prior $P^{\text{prior}}_{\text{unif}}(M)$ might have a negative influence on the properties of emergent languages. For instance, it could be one reason for a negative result reported by Chaabouni et al. [2019], who demonstrated that emergent languages, obtained by optimizing the conventional objective $\mathcal{J}^{\text{sig}}(\boldsymbol{\phi}, \boldsymbol{\theta})$, do not follow Zipf's law of abbreviation [ZLA, Zipf, 1935, 1949, Kanwal et al., 2017]. Suppose, as a natural assumption, that the message space $\mathcal{M}$ is defined as the set of all sequences up to length T over a finite alphabet $\mathcal{A}$. As a simple combinatorial matter, the number of longer messages is much larger than that of shorter ones in $\mathcal{M}$. Consequently, the uniform distribution over the message space $P^{\text{prior}}_{\text{unif}}(M)$ ends up assigning disproportionately large mass to longer messages, causing the emergent language to also become (unintentionally) longer. The third reason is that it is natural to reintroduce the prior explicitly as some form of language model to overcome the artifacts caused by the implicit prior. Ueda and Taniguchi [2024] claimed that the prior should be re-interpreted as a "language model" since it defines the parametrized probability distribution over the message space. Specifically, they proposed to redefine the prior as an auto-regressive neural network model $P^{\text{prior}}_{\boldsymbol{\theta}}(M)$ parametrized by $\boldsymbol{\theta}$. This allows the signaling game to naturally incorporate the concept of a language model, overcoming the artifacts of the unnatural implicit prior $P^{\text{prior}}_{\text{unif}}(M)$. Moreover, the term $\log P^{\text{prior}}(\boldsymbol{m})$ that appears in the ELBO corresponds to the (negative) **surprisal** in the field of computational psycholinguistics [Hale, 2001, Levy, 2008, Smith and Levy, 2013, Kuribayashi et al., 2022]. Introducing a prior as a language model can serve as a psycholinguistic analogy.

Rationale for Introducing Entropy Maximizer: The main reason for this is that since some ad hoc auxiliary function, such as an entropy regularizer (which is similar to the entropy maximizer), is often added to the conventional objective function, it would be more natural if such a term appears explicitly in the objective function from the beginning. From the perspective of the sender agent, a signaling game is a (non-stationary) Markov decision process, and the optimization method, considering only the sender agent, is equivalent to the policy gradient method. In policy gradient methods, regularizers are often introduced to prevent the policy entropy $\mathcal{H}(S_{\boldsymbol{\phi}}(\boldsymbol{m}|\boldsymbol{x}))$ from becoming too low. This encourages exploration by the agent, achieving a balance in the exploration-exploitation trade-off. In conventional signaling games, the entropy regularizer [Williams and Peng, 1991, Mnih et al., 2016] has often been used. On the other hand, the ELBO contains the entropy maximizer. Although the two are not exactly the same, they align in their motivation to increase the entropy of the policy and encourage exploration [Levine, 2018].

3 Contextualized Signaling Game as Generalization of Signaling and Referential Games

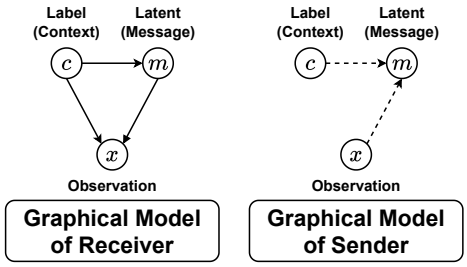

Figure 1: Bayesian Networks of Sender and Receiver in a Contextualized Signaling Game.

The goal of this paper is to reinterpret not only signaling games but also referential games within the framework of variational inference based on some generative models. It is not entirely straightforward: In signaling games, the receiver agent $R^{\text{sig}}(X|M)$ is a conditional probability model of $\boldsymbol{x}$ given $\boldsymbol{m}$, corresponding to the concept of a decoder in a (beta-)VAE. However, the receiver agent in reference games, $R^{\text{ref}} : \mathcal{X} \times \mathcal{M} \to \mathbb{R}$, is a real-valued function, and its objective function is expressed by InfoNCE. Thus, the receiver agent in referential games is more akin to a contrastive learning model rather than a VAE decoder. One might naively consider adding a (negative) KL term $-\beta\,\text{KL}(S_{\boldsymbol{\phi}}(M|\boldsymbol{x})||P^{\text{prior}}_{\boldsymbol{\theta}}(M))$ to $\mathcal{J}^{\text{ref}}(\boldsymbol{\phi}, \boldsymbol{\theta})$ as a new objective function. However, it is necessary to verify that this formulation is "generative" in some meaningful way. In fact, this problem can be resolved by extending (beta-)VAE to Conditional (beta-)VAE (CVAE). In what follows, we formulate a **contextualized signaling game (CSG)**, which can be regarded as a sort of CVAE. We then confirm that it indeed includes signaling games and referential games as special cases.

[2]Here, the message space $\mathcal{M}$ is assumed to be finite.

Let $\mathcal{C}$ be a **context space**. A probability distribution over the context is denoted by $P_C(C)$ and a conditional distribution of X given C is denoted by $P_{X|C}(X|C)$. The sender $S_{\boldsymbol{\phi}}(M|X)$, receiver $R^{\text{sig}}_{\boldsymbol{\theta}}(X|M)$, and prior $P^{\text{prior}}_{\boldsymbol{\theta}}(M)$ defined earlier will now be referred to as **context-agnostic**. In contrast, we consider **context-aware** counterparts: $S_{\boldsymbol{\phi}}(M|X,C)$, $R_{\boldsymbol{\theta}}(X|M,C)$, and $P^{\text{prior}}_{\boldsymbol{\theta}}(M|C)$, which are additionally conditioned on a context C. Figure 1 illustrates the graphical models of context-aware sender and receiver. Here, CSG is formulated as a game with the following procedure:

1. Sample a context: $\boldsymbol{c} \sim P_C(\cdot)$.
2. Sample an observation: $\boldsymbol{x} \sim P_{X|C}(\cdot|\boldsymbol{c})$.
3. The context-aware sender agent samples a message: $\boldsymbol{m} \sim S_{\boldsymbol{\phi}}(\cdot|\boldsymbol{x},\boldsymbol{c})$.
4. The context-aware receiver agent $R^{\text{sig}}_{\boldsymbol{\theta}}(X|\boldsymbol{m},\boldsymbol{c})$ predicts $\boldsymbol{x}$ from $\boldsymbol{m}$ and $\boldsymbol{c}$.

Based on this procedure, the objective function of CSG $\mathcal{J}^{\text{csg}}(\phi,\theta;\beta)$ is defined as follows:

$$\begin{aligned}&\mathcal{J}^{\text{csg}}(\boldsymbol{\phi},\boldsymbol{\theta};\beta)\\ &:= \mathbb{E}_{P_C(\boldsymbol{c}),P_{X|C}(\boldsymbol{x}|\boldsymbol{c})}[\mathbb{E}_{S_{\boldsymbol{\phi}}(\boldsymbol{m}|\boldsymbol{x},\boldsymbol{c})}[\log R^{\text{sig}}_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{c})] - \beta\,\mathrm{KL}(S_{\boldsymbol{\phi}}(M|\boldsymbol{x},\boldsymbol{c}) \mid\mid P^{\text{prior}}_{\boldsymbol{\theta}}(M|\boldsymbol{c}))].\end{aligned} \tag{5}$$

It can be seen as CVAE [Kingma et al., 2014, Sohn et al., 2015] where $\boldsymbol{c}$ serves as a class label.

Signaling Game as a Special Case: The signaling game can be regarded as a special case of CSG. As is evident from the definition, the context-agnostic sender, receiver, and prior are special cases of the context-aware sender, receiver, and prior. Replacing the context-aware models in Eq (5) with context-agnostic ones results in the objective function $\mathcal{J}^{\text{sig-elbo}}(\boldsymbol{\phi},\boldsymbol{\theta};\beta)$.

Referential Game as a Special Case: The referential game can also be regarded as a special case of CSG. Assume that the sender and prior are context-agnostic, while the receiver is context-aware.[3] Let the context space $\mathcal{C}$ and the probability $P_{X|C}(X|C)$ be defined as follows:

$$\mathcal{C} := \{(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_K) \in \mathcal{X}^K \mid \boldsymbol{x}_i \neq \boldsymbol{x}_j \text{ for } i \neq j\}, \quad P_{X|C}(\boldsymbol{x}|\boldsymbol{c}) := \frac{1}{K}\sum\nolimits_{i=1}^{K} \mathbb{1}_{\boldsymbol{x}=\boldsymbol{c}_i}. \tag{6}$$

Define the context-aware receiver agent as:

$$R^{\text{sig}}_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{c}) := \frac{\exp(R^{\text{ref}}_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{m}))\sum_{i=1}^{K}\mathbb{1}_{\boldsymbol{x}=\boldsymbol{c}_i}}{\sum_{\boldsymbol{x}'\in\mathcal{X}}\exp(R^{\text{ref}}_{\boldsymbol{\theta}}(\boldsymbol{x}',\boldsymbol{m}))\sum_{i=1}^{K}\mathbb{1}_{\boldsymbol{x}'=\boldsymbol{c}_i}}, \tag{7}$$

which can be transformed into:

$$= \begin{cases} \dfrac{\exp(R^{\text{ref}}_{\boldsymbol{\theta}}(\boldsymbol{c}_j,\boldsymbol{m}))}{\sum_{i=1}^{K}\exp(R^{\text{ref}}_{\boldsymbol{\theta}}(\boldsymbol{c}_i,\boldsymbol{m}))} & (\boldsymbol{x}=\boldsymbol{c}_j \text{ for some } j),\\ 0 & (\text{otherwise}). \end{cases} \tag{8}$$

$R^{\text{sig}}_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{c})$ can be identified with InfoNCE [van den Oord et al., 2018], as the "otherwise" case cannot occur from the definition of $P_{X|C}$. Also note that $R^{\text{sig}}_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{c})$ is a probability distribution, since $R^{\text{sig}}_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{c}) \geq 0$ and $\sum_{\boldsymbol{x}\in\mathcal{X}} R^{\text{sig}}_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{m},\boldsymbol{c}) = 1$. Thus, CSG includes the generative formulation of the reference game as a special case.

4 Discussion and Conclusion

In this paper, we formalized the contextualized signaling game (CSG) as a generalization of signaling and referential games. The referential game that uses context-aware receiver agents might more closely represent reality than the signaling game, in that speech acts occur within some context [Wittgenstein, 1953]. Note that the context $\boldsymbol{c} \in \mathcal{C}$ does not necessarily have to be a K-tuple consisting of a correct answer and distractors as described in Eq (6), nor does $P_{X|C}(X|C)$ necessarily have to be a distribution that samples an observation uniformly from a candidate set since we introduced $\boldsymbol{c} \in \mathcal{C}$ quite abstractly. It might be an interesting direction to explore different ways of defining the context C and observation X in the study on context-dependent communication. Future work is needed to conduct experiments to quantify the goodness of our formulation.

[3] If the sender and receiver are context-aware while the prior is context-agnostic, the game roughly corresponds to the ones discussed in e.g., Lazaridou et al. [2017], Bouchacourt and Baroni [2018].

Acknowledgements

This research was supported by the JSPS KAKENHI Grant Number JP23KJ0768.

References

Alessandro Achille and Stefano Soatto. Information dropout: Learning optimal representations through noisy computation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(12):2897–2905, 2018. doi: 10.1109/TPAMI.2017.2784440. URL https://doi.org/10.1109/TPAMI.2017.2784440.

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL https://openreview.net/forum?id=HyxQzBceg.

Brendon Boldt and David R. Mortensen. A review of the applications of deep learning-based emergent communication. *Trans. Mach. Learn. Res.*, 2024, 2024. URL https://openreview.net/forum?id=jesKcQxQ7j.

Diane Bouchacourt and Marco Baroni. How agents see things: On visual representations in an emergent language game. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 981–985, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1119. URL https://aclanthology.org/D18-1119.

Ted Briscoe. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. 76:245–296, 2000. doi: 10.2307/417657.

Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. Anti-efficient encoding in emergent communication. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 6290–6300, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/31ca0ca71184bbdb3de7b20a51e88e90-Abstract.html.

Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. Communicating artificial neural networks develop efficient color-naming systems. *Proceedings of the National Academy of Sciences*, 118(12):e2016569118, 2021. doi: 10.1073/pnas.2016569118. URL https://www.pnas.org/doi/abs/10.1073/pnas.2016569118.

Rahma Chaabouni, Florian Strub, Florent Altché, Eugene Tarassov, Corentin Tallec, Elnaz Davoodi, Kory Wallace Mathewson, Olivier Tieleman, Angeliki Lazaridou, and Bilal Piot. Emergent communication at scale. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=AUGBfDIV9rL.

Gautier Dagan, Dieuwke Hupkes, and Elia Bruni. Co-evolution of language and agents in referential games. In Paola Merlo, Jörg Tiedemann, and Reut Tsarfaty, editors, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 2993–3004. Association for Computational Linguistics, 2021. doi: 10.18653/V1/2021.EACL-MAIN.260. URL https://doi.org/10.18653/v1/2021.eacl-main.260.

Roberto Dessì, Eugene Kharitonov, and Marco Baroni. Interpretable agent communication from scratch (with a generic visual processor emerging on the side). In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 26937–26949, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/e250c59336b505ed411d455abaa30b4d-Abstract.html.

Hiroto Ebara, Tomoaki Nakamura, Akira Taniguchi, and Tadahiro Taniguchi. Multi-agent reinforcement learning with emergent communication using discrete and indifferentiable message. In *15th International Congress on Advanced Applied Informatics Winter, IIAI-AAI-Winter 2023, Bali, Indonesia, December 11-13, 2023*, pages 366–371. IEEE, 2023. doi: 10.1109/IIAI-AAI-WINTER61682.2023.00073. URL https://doi.org/10.1109/IIAI-AAI-Winter61682.2023.00073.

Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, and Shimon Whiteson. Learning to communicate to solve riddles with deep distributed recurrent q-networks. *CoRR*, abs/1602.02672, 2016. URL http://arxiv.org/abs/1602.02672.

Laura Graesser, Kyunghyun Cho, and Douwe Kiela. Emergent linguistic phenomena in multi-agent communication games. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 3698–3708. Association for Computational Linguistics, 2019. doi: 10.18653/V1/D19-1384. URL https://doi.org/10.18653/v1/D19-1384.

Shangmin Guo, Yi Ren, Kory Wallace Mathewson, Simon Kirby, Stefano V. Albrecht, and Kenny Smith. Expressivity of emergent languages is a trade-off between contextual complexity and unpredictability. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=WxuE_JWxjkW.

John Hale. A probabilistic earley parser as a psycholinguistic model. In *Language Technologies 2001: The Second Meeting of the North American Chapter of the Association for Computational Linguistics, NAACL 2001, Pittsburgh, PA, USA, June 2-7, 2001*. The Association for Computational Linguistics, 2001. URL https://aclanthology.org/N01-1021/.

Serhii Havrylov and Ivan Titov. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 2149–2159, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/70222949cc0db89ab32c9969754d4758-Abstract.html.

Irina Higgins, Loïc Matthey, Arka Pal, Christopher P. Burgess, Xavier Glorot, Matthew M. Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL https://openreview.net/forum?id=Sy2fzU9gl.

Nguyen Le Hoang, Tadahiro Taniguchi, Yoshinobu Hagiwara, and Akira Taniguchi. Emergent communication of multimodal deep generative models based on metropolis-hastings naming game. *Frontiers Robotics AI*, 10, 2023. doi: 10.3389/FROBT.2023.1290604. URL https://doi.org/10.3389/frobt.2023.1290604.

Jun Inukai, Tadahiro Taniguchi, Akira Taniguchi, and Yoshinobu Hagiwara. Recursive metropolis-hastings naming game: symbol emergence in a multi-agent system based on probabilistic generative models. *Frontiers Artif. Intell.*, 6, 2023. doi: 10.3389/FRAI.2023.1229127. URL https://doi.org/10.3389/frai.2023.1229127.

Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Çaglar Gülçehre, Pedro A. Ortega, DJ Strouse, Joel Z. Leibo, and Nando de Freitas. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 3040–3049. PMLR, 2019. URL http://proceedings.mlr.press/v97/jaques19a.html.

Jasmeen Kanwal, Kenny Smith, Jennifer Culbertson, and Simon Kirby. Zipf' s law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165:45–52, 2017. ISSN 0010-0277. doi: https://doi.org/10.1016/j.cognition.2017.05.001. URL https://www.sciencedirect.com/science/article/pii/S0010027717301166.

Daichi Kato, Ryo Ueda, Jason Naradowsky, and Yusuke Miyao. Emergent communication with stack-based agents. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 46(0), 2024.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014. URL http://arxiv.org/abs/1312.6114.

Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/d523773c6b194f37b938d340d5d02232-Paper.pdf.

Simon Kirby. Learning, bottlenecks and the evolution of recursive syntax. *Linguistic evolution through language acquisition: Formal and computational models*, pages 173–204, August 2002.

Satwik Kottur, José M. F. Moura, Stefan Lee, and Dhruv Batra. Natural language does not emerge 'naturally' in multi-agent dialog. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 2962–2967. Association for Computational Linguistics, 2017. doi: 10.18653/V1/D17-1321. URL https://doi.org/10.18653/v1/d17-1321.

Tatsuki Kuribayashi, Yohei Oseki, Ana Brassard, and Kentaro Inui. Context limitations make neural language models more human-like. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 10421–10436. Association for Computational Linguistics, 2022. doi: 10.18653/V1/2022.EMNLP-MAIN.712. URL https://doi.org/10.18653/v1/2022.emnlp-main.712.

Angeliki Lazaridou and Marco Baroni. Emergent multi-agent communication in the deep learning era. *CoRR*, abs/2006.02419, 2020. URL https://arxiv.org/abs/2006.02419.

Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. Multi-agent cooperation and the emergence of (natural) language. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. URL https://openreview.net/forum?id=Hk8N3Sclg.

Sergey Levine. Reinforcement learning and control as probabilistic inference: Tutorial and review. *CoRR*, abs/1805.00909, 2018. URL http://arxiv.org/abs/1805.00909.

Roger Levy. Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177, 2008. ISSN 0010-0277. doi: https://doi.org/10.1016/j.cognition.2007.05.006. URL https://www.sciencedirect.com/science/article/pii/S0010027707001436.

David K. Lewis. *Convention: A Philosophical Study*. Wiley-Blackwell, 1969.

Yat Long Lo, Biswa Sengupta, Jakob Nicolaus Foerster, and Michael Noukhovitch. Learning multi-agent communication with contrastive learning. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=vZZ4hhniJU.

Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In Isabelle Guyon, Ulrike von

Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 6379–6390, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/68a9750337a418a86fe06c1991a1d64c-Abstract.html.

Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 1928–1937. JMLR.org, 2016. URL http://proceedings.mlr.press/v48/mniha16.html.

Martin A Nowak and David C Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033, 1999.

Ryota Okumura, Tadahiro Taniguchi, Yoshinobu Hagiwara, and Akira Taniguchi. Metropolis-hastings algorithm in joint-attention naming game: experimental semiotics study. *Frontiers Artif. Intell.*, 6, 2023. doi: 10.3389/FRAI.2023.1235231. URL https://doi.org/10.3389/frai.2023.1235231.

Jannik Peters, Constantin Waubert de Puiseau, Hasan Tercan, Arya Gopikrishnan, Gustavo Adolpho Lucas De Carvalho, Christian Bitter, and Tobias Meisen. A survey on emergent language, 2024. URL https://arxiv.org/abs/2409.02645.

Yi Ren, Shangmin Guo, Matthieu Labeau, Shay B. Cohen, and Simon Kirby. Compositional languages emerge in a neural iterated learning model. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=HkePNpVKPB.

Ryokan Ri, Ryo Ueda, and Jason Naradowsky. Emergent communication with attention. In Micah B. Goldwater, Florencia K. Anggoro, Brett K. Hayes, and Desmond C. Ong, editors, *Proceedings of the 45th Annual Meeting of the Cognitive Science Society, CogSci 2023, Sydney, NSW, Australia, July 26-29, 2023*. cognitivesciencesociety.org, 2023. URL https://escholarship.org/uc/item/7dg8r8zk.

Mathieu Rita, Corentin Tallec, Paul Michel, Jean-Bastien Grill, Olivier Pietquin, Emmanuel Dupoux, and Florian Strub. Emergent communication: Generalization and overfitting in lewis games. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/093b08a7ad6e6dd8d34b9cc86bb5f07c-Abstract-Conference.html.

Nathaniel J. Smith and Roger Levy. The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3):302–319, 2013. ISSN 0010-0277. doi: https://doi.org/10.1016/j.cognition.2013.02.013. URL https://www.sciencedirect.com/science/article/pii/S0010027713000413.

Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 3483–3491, 2015. URL https://proceedings.neurips.cc/paper/2015/hash/8d55a249e6baa5c06772297520da2051-Abstract.html.

Luc Steels. The talking heads experiment. 1999.

Tadahiro Taniguchi, Yuto Yoshida, Yuta Matsui, Nguyen Le Hoang, Akira Taniguchi, and Yoshinobu Hagiwara. Emergent communication through metropolis-hastings naming game with deep generative models. *Adv. Robotics*, 37(19):1266–1282, 2023. doi: 10.1080/01691864.2023.2260856. URL https://doi.org/10.1080/01691864.2023.2260856.

Mycal Tucker, Roger P. Levy, Julie Shah, and Noga Zaslavsky. Trading off utility, informativeness, and complexity in emergent communication. In *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=O5arhQvBdH.

Ryo Ueda and Tadahiro Taniguchi. Lewis's signaling game as beta-vae for natural word lengths and segments. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=HC0msxE3sf.

Ryo Ueda and Koki Washio. On the relationship between zipf's law of abbreviation and interfering noise in emergent languages. In *Proceedings of the ACL-IJCNLP 2021 Student Research Workshop, ACL 2021, Online, JUli 5-10, 2021*, pages 60–70. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.acl-srw.6. URL https://doi.org/10.18653/v1/2021.acl-srw.6.

Ryo Ueda, Taiga Ishii, and Yusuke Miyao. On the word boundaries of emergent languages based on harris's articulation scheme. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/pdf?id=b4t9_XASt6G.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *CoRR*, abs/1807.03748, 2018. URL http://arxiv.org/abs/1807.03748.

Ronald J. Williams and Jing Peng. Function optimization using connectionist reinforcement learning algorithms. *Connection Science*, 3(3):241–268, 1991. doi: 10.1080/09540099108946587. URL https://doi.org/10.1080/09540099108946587.

Ludwig Wittgenstein. *Philosophical Investigations*. Wiley-Blackwell, New York, NY, USA, 1953.

Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. Efficient compression in color naming and its evolution. *Proc. Natl. Acad. Sci. USA*, 115(31):7937–7942, 2018. doi: 10.1073/PNAS.1800521115. URL https://doi.org/10.1073/pnas.1800521115.

George K. Zipf. *The psycho-biology of language*. Houghton Mifflin, 1935.

George K. Zipf. *Human Behaviour and the Principle of Least Effort*. Addison-Wesley, 1949.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the claims. Limitations are briefly discussed in Section 4.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations are briefly discussed in Section 4.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We gave clear definitions of terms in equations. We believe that the explanations of equation transformations are sufficient.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification: This paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: This paper does not include experiments requiring code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: This paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: This paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: This paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We reviewed and followed the NeurIPS Code of Ethics. We preserve anonymity.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: This paper only includes the mathematical discussions about abstract communication models, such as signaling games, for which we believe that there is no significant societal impact.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: This paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing or research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.