Trading off Utility, Informativeness, and Complexity in Emergent Communication

Mycal Tucker
M.I.T.
mycal@mit.edu

Roger Levy
M.I.T.
rplevy@mit.edu

Julie Shah
M.I.T.
julie_a_shah@csail.mit.edu

Noga Zaslavsky
M.I.T.
nogazs@mit.edu

Abstract

Emergent communication (EC) research often focuses on optimizing task-specific utility as a driver for communication. However, there is increasing evidence that human languages are shaped by task-general communicative constraints and evolve under pressure to optimize the Information Bottleneck (IB) tradeoff between the informativeness and complexity of the lexicon. Here, we integrate these two approaches by trading off utility, informativeness, and complexity in EC. To this end, we propose Vector-Quantized Variational Information Bottleneck (VQ-VIB), a method for training neural agents to encode inputs into discrete signals embedded in a continuous space. We evaluate our approach in multi-agent reinforcement learning settings and in color reference games and show that: (1) VQ-VIB agents can continuously adapt to changing communicative needs and, in the color domain, align with human languages; (2) the emergent VQ-VIB embedding spaces are semantically meaningful and perceptually grounded; and (3) encouraging informativeness leads to faster convergence rates and improved utility, both in VQ-VIB and in prior neural architectures for symbolic EC, with VQ-VIB achieving higher utility for any given complexity. This work offers a new framework for EC that is grounded in information-theoretic principles that are believed to characterize human language evolution and that may facilitate human-agent interaction.

1 Introduction

Good communication is a critical component of successful teams of humans and artificial agents, but differing notions of "good" makes training agents to develop such communication challenging. One view of communication focuses on *utility* by framing languages as successful to the extent they enable high task performance. Emergent communication (EC) literature, wherein agents learn to communicate while optimizing the utility associated with a task-specific goal, such as reaching a target landmark, emphasizes this view [1–6]. However, training via utility alone may lead to undesirable properties of the EC system, such as violations of linguistic universals [7], poor human interpretability [8], slow convergence rates [9], or high sensitivity to environment design [10]. This implies that some constraints may be needed to guide agents toward human-like communication.

Another view of communication, often considered in the cognitive science literature, focuses on task-agnostic communicative objectives as major forces shaping human languages [11, 12]. Most relevant to our work is a recent line of work [13–17] that argues that languages evolve under pressure to efficiently compress meanings into communicative signals by optimizing the Information Bottleneck (IB) principle [18]. In this context, IB can be interpreted as a tradeoff between the

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

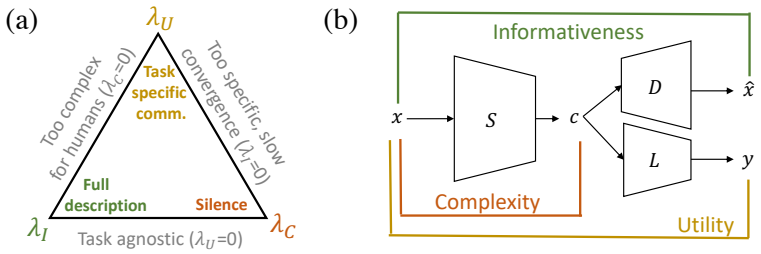


Figure 1: (a) In our theoretical framework, agents evolve to communicate with each other by optimizing a tradeoff between utility, informativeness, and complexity, weighted by λ_U, λ_I, and λ_C respectively. (b) Basic setup and agent architecture. A speaker, S, maps an input, x, to a signal, c. A listener, L, takes an action, y, and reconstructs x using a decoder D.

complexity and informativeness of the lexicon [13]. Intuitively, complexity corresponds to the number of bits allocated for communication, and informativeness corresponds to how well a listener can understand a speaker regardless of any task-specific goal (e.g., humans understand the term "blue" in "avoid the blue box" and "find a blue cup"). This information-theoretic framework for grounded semantic categories has been gaining broad empirical support across hundreds of languages and multiple semantic domains [13–16], and has been directly tied to language change over time [17]. However, the actual process that may drive a population of speakers — whether human or artificial — toward IB-efficient communication systems in a given environment remains largely unknown.

In this work, we integrate these two views of communication by training teams of interactive neural agents to optimize a tradeoff between maximizing utility, maximizing informativeness, and minimizing complexity, as depicted in Figure 1. We hypothesize that taking into account complexity and informativeness, which are cognitively-motivated task-general communicative facctors, in addition to task-specific utility, may help guide artificial agents toward communication systems that can be more naturally aligned with and adapted to human languages. To this end, building on prior work [19, 20], we propose Vector-Quantized Variational Information Bottleneck (VQ-VIB), a method for training neural agents to encode inputs into signals, while maintaining a discrete (symbolic) signal representation that is embedded in a continuous space, inspired by word embeddings [21–23]. We evaluate our approach in two multi-agent reinforcement learning (MARL) environments and in two variants of a color reference game. We show that: (1) by annealing a single tradeoff parameter, VQ-VIB agents can continuously adapt to changing communicative pressures and, in the color domain, align with human languages, without changing the agents' architecture or the characteristics of the training environment; (2) the emergent VQ-VIB embedding spaces capture semantically meaningful structures, grounded in the perceived environment; and (3) encouraging agents to be informative generally leads to faster convergence rates and improved utility, with VQ-VIB achieving higher utility for any given complexity compared to previous EC approaches that also support symbolic communication. Taken together, this work offers a new information-theoretic framework for EC that is grounded in general optimization principles that are believed to characterize human language evolution, and that has the potential of improving human-agent interaction.

2 Related work

While EC studies usually focuses on task utility, some studies also consider communicative constraints that are closely related to our measures of complexity and informativeness (defined in Section 3). For example, numerous works consider discretization, environmental noise, or penalty on large vocabulary sizes as proxies for reducing complexity [23–25]. In contrast, our notion of complexity is grounded in rate-distortion theory [26], and has been shown to capture the degree of semantic compression seen across languages [13]. Wang et al. [27] use the same IB complexity measure, but in different communicative settings that involve a centralized message processor. Other notions of linguistic complexity often focus on grammar, word forms, or compositionality [28–31]; these are beyond the scope of this work but offer promising directions for future work. Apart from complexity, some works consider auxiliary reconstruction losses in EC [9, 32]. In particular, Lin et al. [33] use an autoencoding loss which is closely related our notion of informativeness. However, they do not penalize complexity and note that their method may fall short in cases where "agents need to be more

selective of the information communicated." Lastly, there is a body of work on reinforcement learning with informational constraints similar to ours [e.g., 34–38], however not in the context of EC.

To evaluate our method, we consider particle-world environments that agents must navigate [3, 23], as well as reference games (related to Lewis signalling games [39]), both of which are popular types of EC environments. For reference games, we focus on the domain of color for several reasons: First, it has been a central test-case for cognitive theories of semantic categorization and language evolution [1, 40, 41]. Second, it enables comparison with the World Color Survey [WCS; 42], a dataset of color naming systems from 110 languages of non-industrialized societies. Third, it has been a focus of numerous studies in EC and machine learning more generally [1, 43–45, 10]. In particular, Chaabouni et al. [10] show that color signaling systems in artificial agents may reach the IB efficiency bound from [13] and capture some of the variation seen across human languages. However, to achieve this, they hand-craft varying characteristics of the training environment by selecting dissimilar distractor colors or by adding noise to the agents' communication. Rather than focusing on environment-based factors, we adopt a principled approach by directly optimizing agents with respect to a variational IB principle in a given environment.

3 Information-theoretic emergent communication

Here, we propose an information-theoretic EC framework that integrates task-general communicative objectives with task-specific utility maximization. We begin by reviewing the theoretical framework on which we build (Section 3.1.1) and the set of variational inference tools that we use in order to extend that framework to EC settings (Sections 3.1.2 and 3.1.3). We then define an integrative objective function for training EC agents (Section 3.2) and present Vector-Quantized Variational Information Bottleneck (VQ-VIB, Section 3.3), a method for training agents with respect to our objective function while maintaining structured symbolic representations that can be used for communication. The experimental setup and results for evaluating this new method are presented in Sections 4 and 5.

3.1 Preliminaries

3.1.1 The Information Bottleneck framework for grounded semantic systems

We build on and extend Zaslavsky et al. [13]'s information-theoretic framework for grounded semantic systems. In this framework, speakers and listeners jointly optimize the IB tradeoff [18] between the complexity and informativeness of their communication system. A speaker is characterized as a probabilistic encoder $S(c|x)$ that, given an input $x \sim p(x)$, mentally represents it as m_x and generates a communication signal, c. m_x is the speaker's belief state, which is grounded in the agents' representation of the world and captures any uncertainly the speaker may have about the input. More formally, m_x is defined by a probability distribution over a set of features $u \in \mathcal{U}$ that characterize the input space, i.e, $m_x(u) \equiv p(u|m_x)$. For example, x may be a target color and m_x a Gaussian distribution over the CIELAB perceptual color space (Figure 3c), capturing the speaker's perceptual uncertainly. Given c, the listener infers the speaker's belief state by performing Bayesian inference, i.e., it constructs an estimator $\hat{m}_c(u) = \sum_x S(x|c)m_x(u)$ using the posterior $S(x|c) \propto S(c|x)p(x)$.

Complexity is measured by the mutual information between the speaker's inputs and signals, $I_S(x, c)$, which, as in rate-distortion theory [26, 46], is roughly the number of bits that are needed for communication [18]. Informativeness corresponds to the similarity between the speaker's and listener's belief states. Because these states are defined by probability distributions, maximizing informativeness amounts to minimizing the expected Kullback-Leibler (KL) divergence between them, i.e., $\mathbb{E}_S[D_{\mathrm{KL}}[m_x \| \hat{m}_c]]$. Taken together, in this view, for any given tradeoff parameter $\beta \geq 0$ the speaker and listener should jointly minimize

$$I_S(x, c) + \beta \mathbb{E}_S[D_{\mathrm{KL}}[m_x \| \hat{m}_c]] \,, \tag{1}$$

which is equivalent to optimizing the IB principle [see 13, 47, for derivation]. For $\beta \leq 1$, the optimal systems are non-informative and minimally complex ($I_S(x, c) = 0$), yielding no communication. Gradually annealing β yields optimal systems with greater complexity, and as $\beta \to \infty$, the optimal systems become maximally informative but also maximally complex. Zaslavsky et al. [13] and colleagues have shown that human languages achieve near-optimal IB tradeoffs across multiple semantic domains (see Figure 6a for example), and that the continuous evolution of the optimal IB systems via annealing the tradeoff parameter β captures important aspects of language evolution [13, 17].

3.1.2 Variational Information Bottleneck

Variational Information Bottleneck [VIB, 19] is a variational approximation method for IB with continuous representations. In its typical form, a neural encoder $q_\theta(z|x)$ maps an input x to parameters of a d-dimensional Gaussian distribution, $\mu(x)$ and $\Sigma(x)$, from which a latent representation $z \in \mathbb{R}^d$ is sampled. Given z, VIB reconstructs a target feature u using $q_\phi(u|z)$, parameterized by a soft-max function. VIB optimizes a variational bound on the IB objective function, based on the fact that

$$I_{q_\theta}(x;z) \leq \mathbb{E}\left[D[q_\theta(z|x)\|r(z)]\right] \tag{2}$$

for any distribution $r(z)$, which is often taken to be $\mathcal{N}(0, I_d)$. VIB also bounds the IB informativeness term. However, because we consider here a different instantiation of IB, based on unsupervised rather than supervised learning, we use a different informativeness bound, as explained in Section 3.3.

3.1.3 Vector-Quantized Variational Autoencoder

While VIB learns continuous representations, Vector-Quantized Variational Autoencoder [VQ-VAE, 20] learns discrete, but potentially highly complex, representations in an unsupervised manner. More specifically, VQ-VAE defines a latent embedding space of K vectors $\zeta_i \in \mathbb{R}^d, i = 1, \ldots, K$. A deterministic encoder maps an input x to a discrete variable i by first passing x through a deterministic network to obtain a latent representation $z(x) \in \mathbb{R}^d$, and then discretizing it to the index of the nearest embedding vector, i.e., $i = \text{argmin}_j \|z(x) - \zeta_j\|$. The decoder receives the corresponding embedding vector, $\zeta_i(x)$, and passes it through another deterministic network to reconstruct the input. The set of trainable parameters Θ consists of the embedding vectors and the parameters of the encoder and decoder networks. Because gradients cannot be passed through the `argmin` operator, a straight-through estimator [48] is used. Similar to VAE [49], VQ-VAE's training loss (Eq. (3)) uses the evidence lower bound (ELBO), which is reduced in this case to the first term in Eq. (3). VQ-VAE adds to it two VQ terms, designed for learning the embedding space, yielding the overall training loss

$$l_{\text{VQ-VAE}} = \log p(x|\zeta_i(x);\Theta) + \|\texttt{sg}[z(x)] - \zeta_i(x)\|^2 + \alpha\|z(x) - \texttt{sg}[\zeta_i(x)]\|^2 \,, \tag{3}$$

where `sg` stands for the `stopgradient` operator, and following [20], $\alpha = 0.25$.

3.2 Integrating utility, informativeness, and complexity

The theoretical framework and variational tools reviewed in section 3.1 focus on general informational constraints and autoencoding reconstruction losses. However, EC settings are often based on multi-agent reinforcement learning (MARL), where agents are trained to maximize an expected reward, or utility, for a given task. In a basic cooperative setting, there are two agents: a speaker and a listener. The speaker receives an input, $x \in \mathcal{X}$, which is sampled from a distribution $p(x)$ and contains information about the task's goal which is hidden from the listener. The speaker's actions are restricted to communicating a vector $c \in \mathcal{C}$ to the listener. The listener receives c and a partial observation $o \in \mathcal{O}$ about the world state, and takes an action $y \in \mathcal{Y}$. We denote the speaker's policy by $S(c|x)$ and the listener's policy by $L(y|c, o)$. In each episode the agents receive a utility $U(x, y)$ based on their actions and aim to learn policies that maximize the expected utility $\mathbb{E}_{S,L}[U(x, y)]$.

Note that this setting is similar to that of Section 3.1.1, with the key difference being that here the listener takes an action in the world instead of reconstructing the speaker's belief state. To integrate the two settings, we add to the basic EC setting a decoder, D, as shown in Figure 1b. If the speaker has no uncertainty about the input, then the decoder maps c to a reconstruction $\hat{x}_c$. Otherwise, the decoder reconstructs a belief state, $\hat{m}_c$, as explained in Section 3.1.1. A similar approach can also be applied to more complex EC settings, which may involve sequential actions with discounted rewards, additional agents, or agents that can simultaneously communicate and act. In Sections 4 and 5, we consider navigation tasks with sequential actions, and reference games which correspond to the basic setting. We leave the exploration of other multi-agent settings to future work.

Our proposed objective function for training EC agents in these settings combines utility maximization with the IB objective from Eq. (1), yielding

$$\mathcal{L}_\lambda[S, L, D] = \lambda_U \mathbb{E}_{S,L}[U(x, y)] - \lambda_I \mathbb{E}_{S,D}[D_{\text{KL}}[m_x\|\hat{m}_c]] - \lambda_C I_S(x, c) \,, \tag{4}$$

such that agents should jointly maximize $\mathcal{L}_\lambda[S, L, D]$. The non-negative tradeoff parameters, or Lagrange multiplies, $\lambda = (\lambda_U, \lambda_I, \lambda_C)$, control the relative weights of utility, informativeness

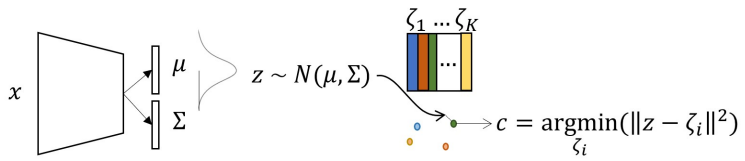


Figure 2: The VQ-VIB encoder architecture. The speaker encodes an input, x, as parameters to a Gaussian distribution, from which a vector is sampled and then discretized by mapping to the nearest encoding, to produce a communication vector, c.

(negatively related to the second term), and complexity. Intuitively, these three weights form a simplex (Figure 1a) that spans the range of communication systems that may emerge in our framework, from uninformative communication (minimizing complexity alone), to task-specific communication that may not generalize (maximizing utility alone), to complete descriptions (maximizing informativeness alone) that may be unnecessarily complex for many tasks and too complex for alignment with humans.

3.3 Vector-Quantized Variational Information Bottleneck

As in IB, directly maximizing Eq. (4) is intractable in large domains. To address this we take a variational approximation approach, where we assume parameterized agents and maximize a variational lower bound on Eq. (4). Additionally, we consider an encoder (speaker) architecture that learns symbolic representations that are embedded in a continuous space, inspired by word embeddings [22, and see also [23] for a related prototype-based EC approach]. We refer to this new method as Vector-Quantized Variational Information Bottleneck (VQ-VIB). Before defining our method precisely, we note that VQ-VIB builds on VIB and VQ-VAE, but at the same time differs from them in important ways. VQ-VIB learns a stochastic encoder, in contrast to VQ-VAE, and symbolic representations, in contrast to VIB. Additionally, while VQ-VIB's training objective borrows from VIB and VQ-VAE, it considers a different variational IB bound than VIB, uses that bound instead of the ELBO in VQ-VAE, and adds to that a symmetry-breaking term.[1] These difference are particularly crucial for human-like communication, as human languages employ symbolic representations (e.g., words), human speakers are often stochastic, and as discussed in Section 3.1.1, there is substantial empirical evidence that human languages are shaped by the IB principle.

3.3.1 VQ-VIB architecture

The VQ-VIB encoder (illustrated in Figure 2) is composed of two components: a codebook of K embedding vectors, $\mathcal{C} = \{\zeta_1, \dots, \zeta_K\} \subset \mathbb{R}^d$, similar to VQ-VAE, and a VIB-like stochastic encoder $q_\theta(z|x)$. Given an input x, the encoder $q_\theta(z|x)$ is used to sample a latent representation $z \in \mathbb{R}^d$, followed by a quantization layer that discretizes z into the nearest embedding vector. This encoding process generates a vector $c = \text{argmin}_{\zeta \in \mathcal{C}} \|z - \zeta\|^2$, which in our settings will be used for communication. We instantiate $q_\theta(z|x)$ by a feedforward network that maps x to parameters of a Gaussian distribution over the latent space. Overall, the VQ-VIB encoder, or speaker, is given by

$$S_\theta(c|x) = \mathbb{P}(c = \underset{\zeta \in \mathcal{C}}{\text{argmin}} \|z - \zeta\|^2 \big| x, \mathcal{C}, \theta) \,. \tag{5}$$

Finally, as shown in Figure 1b, we pair the VQ-VIB encoder with a simple decoder network $D_\psi(\hat{x}|c)$ that reconstructs the speaker's input given the communication vector (see Section 3.3.2 for justification), and with a task-specific policy network $L_\phi(y|c, o)$ from which the listener selects its actions. These two networks can also be paired with other speaker (encoder) architectures, as we do in subsequent sections in order to evaluate the VQ-VIB method.

3.3.2 Learning objective

To train VQ-VIB agents, we first define a variational bound on Eq. (4). We bound the complexity term (third term in Eq. (4)) by noting that $I_{S_\theta}(x; c) \leq I_{q_\theta}(x; z) \leq \mathbb{E}[D_{\text{KL}}[q_\theta(z|x) \| r(z)]]$, where the first inequality folows from the data processing inequality, and the second one follows from Eq. (2).

[1] VIB and VQ-VAE also do not optimize the utility term in Eq. (4), but that can be easily added to them.

(a) MARL: 9 points (b) MARL: Uniform (c) Reference game: 330 WCS colors

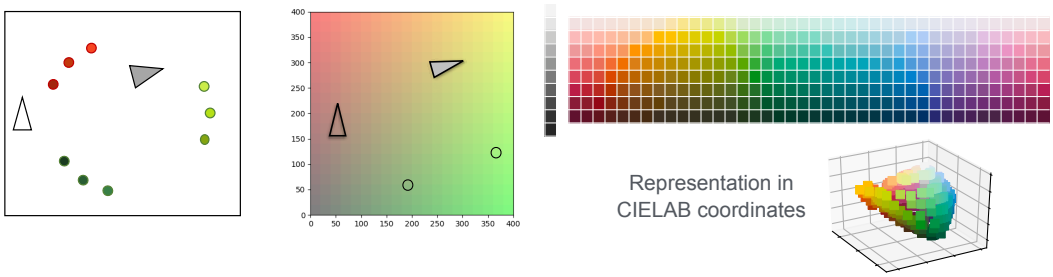


Figure 3: Test environments. In the MARL settings (a-b) a speaker (clear triangle) observes a target location drawn either from a fixed set of 9 points (a) or uniformly in a continuous space (b), communicates it to a listener (gray triangle), which then needs to navigate to that location (adapted from [23]). In the color reference game (c), a speaker observes a target color from the World Color Survey [WCS, 42] palette, represented in the 3D CIELAB perceptual color space, and communicates it to the listener, who then needs to identify the target color.

To bound the informativeness loss (second term in Eq. (4)), we first assume that given $x \in \mathbb{R}^n$, the speaker's uncertainty is captured by Gaussian noise and therefore its belief states are given by $m_x = \mathcal{N}(x, \Sigma)$. Under this assumption, it holds that $\mathbb{E}[D_{\mathrm{KL}}[m_x \| \hat{m}_c]] \leq \frac{1}{2}\,\mathbb{E}[\|x - \hat{x}_c\|^2] + const$ (see Appendix A for proof), allowing us to approximate the ideal decoder by a simplified decoder $D_\psi : \mathcal{C} \to \mathbb{R}^n$ that outputs $\hat{x}_c$. This yields the following lower bound on Eq. (4):

$$\tilde{\mathcal{L}}_\lambda[\theta, \phi, \psi, \mathcal{C}] = \lambda_U \mathbb{E}_{S_\theta, L_\phi}[U(x, y)] - \lambda_I \mathbb{E}_{S_\theta, D_\psi}[\|x - \hat{x}_c\|^2] - \lambda_C \mathbb{E}_{q_\theta}[D_{\mathrm{KL}}[q_\theta(z|x) \| r(z)]] \,. \quad (6)$$

To train the embedding vectors, we include the two VQ terms from Eq. (3), and based on [50], we also add a small regularization term, $\varepsilon(\theta, \mathcal{C})$, designed to break the symmetry between equally good solutions (see Appendix B). Taken together, we obtain the following objective function for VQ-VIB:

$$\mathcal{L}_{\text{VQ-VIB-}\lambda}[\theta, \phi, \psi, \mathcal{C}] = \tilde{\mathcal{L}}_\lambda[\theta, \phi, \psi, \mathcal{C}] - \mathbb{E}_{S_\theta}\left[\|\mathtt{sg}[z] - c\|^2 + \alpha \|z - \mathtt{sg}[c]\|^2\right] - \varepsilon(\theta, \mathcal{C}) \,, \quad (7)$$

which should be jointly maximized by the agents for any given λ. Following [19], we fix $r(z) = \mathcal{N}(0, I_d)$, and following [20], we fix $\alpha = 0.25$. To optimize this objective function, we use the VQ-VAE straight-through estimator to bypass the non-differentiablity of the `argmin` operator, and the reparameterization trick [49] for backpropogating through z. While the latter is internal to the speaker, the straight-through estimator is used to pass gradients from the decoder D_ψ and the listener's policy L_ϕ into the speaker S_θ. Here, we allow gradients to pass from the decoder, but take a more careful approach to passing gradients from the listener by comparing the straight-through estimator approach to methods that do not pass gradients, as explained in the next section.

4 Experimental setup

To evaluate the emergent VQ-VIB communication systems and how they are shaped by λ, we consider three test environments (Figure 3) and compare VQ-VIB with strong baselines, as described below. Full details of the architectures, hyperparameters, and environments are provided in Appendix D.[2]

Trials and baselines We consider three types of baseline speaker architectures that are derived from prior work: (1) a continuous speaker [e.g., 51], which unlike humans, communicates using continuous vectors without maintaining symbolic representations; (2) a one-hot speaker [e.g., 3], which communicates with discrete but unstructured vectors; and (3) a prototype-based speaker [Proto, 23], which maps one-hot vectors to an embedding space, and in that sense is similar to VQ-VIB, although its internal mechanism is different. We also considered a variant of the VQ-VIB speaker ('VQ-VIB After') which is more similar to Proto, but because it does not perform as well as VQ-VIB we leave its discussion to Appendix C. For a fair comparison, we train all speakers with the same decoder and listener architectures, allow them to be stochastic similar to the VQ-VIB speaker, and

[2]Code is available at `https://github.com/mycal-tucker/vqvib_neurips2022`

adjusted their training objective according to Eq. (6) to allow them to consider the same informational constraints as VQ-VIB. In all domains, agents are trained from scratch using 5 random seeds.

Particle worlds We train agents in the two particle world environments shown in Figure 3, which are adapted from Tucker et al. [23]. Both environments consist of two-dimensional worlds in which a speaker observes the exact target location that a listener needs to reach but cannot directly observe. In the 9 points environment, there are 9 landmarks at fixed locations, one of which is chosen at random as the target for each episode. In the Uniform environment, two landmarks are spawned in randomly drawn locations in the continuous space and one of them is specified as the target. Both agents receive a reward equal to the negative distance from the listener to the target and are trained using Multi-Agent Deep Deterministic Policy Gradient [3], which is a standard policy-gradient method without backpropagating from one agent into another. The listener cannot observe the target location, and the speaker can only communicate at the first timestep by broadcasting a vector that is observable at all future timesteps; thus, an optimal team policy consists of the speaker communicating about the target at the first timestep, and the listener navigating to the target based on the message.

Color reference games We consider two variants of a color reference game, a basic game and a common ground game (see Figure D.1). In the basic game, a speaker observes a target color and outputs a communication vector. A listener observes the speaker's communication vector and two colors — the target and a distractor, without knowing which is the target — and tries to guess the target. In the common ground game, the speaker also observes the distractor, in addition to the target, which in principle allows it to communicate in a more pragmatic way by taking into account how the distractor may influence the listener's choices [44, 52]. Apart from that, the two variants are the same. Following [41, 13], colors are represented in the 3D CIELAB color space (Figure 3c), in which small Euclidean distances between colors are correlated with human perceptual dissimilarities. To capture perceptual noise, the speaker's observations are corrupted by $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma^2 = 64$, as motivated by human perception [53]. The target and distractor are sampled independently from the 330 color chips of the World Color Survey [WCS, 42] shown in Figure 3c, according to a human-based prior from [13].[3] Agents are trained with respect to the binary crossentropy loss computed on the listener's output. We consider two training methods: (i) using straight-through gradient estimators to backpropogate from the listener into the speaker, which is common in EC and supports efficient training, although it is not biologically plausible; and (ii) using the REINFORCE algorithm [54], which does not pass gradients between agents but is known to suffer from slow convergence rates.

5 Results

5.1 Particle worlds

We begin with exploring the influence of λ on the emergent agent behavior in the two particle world environments. Without loss of generality, we take $\lambda_U = 1$ unless stated otherwise, and then evaluate the influence of λ_C and λ_I separately. Starting with λ_C, we simulate the evolution of the emergent VQ-VIB communication systems by fixing $\lambda_I = 1$ and varying λ_C using deterministic annealing [55] (see Appendix D for detail), much like the evolution of the IB representations in [50]. Figure 4 shows lower- and higher-complexity solutions for the two environments (see Movie 1 for the continuous evolution of the systems). The top panels (Figure 4a and c) show how the speaker partitions the environment into communication signals, together with the listener's trajectories in response to the speaker's communication. The lower panels (Figure 4b and d) show the emergent VQ-VIB embedding space $\mathcal{C}$. In both environments, lower complexity solutions form a coarser representation of the space, with fewer communication vectors (signals) used by the speaker and, respectively, fewer final positions reached by the listener. We wish to emphasize that the number of signals available to the speaker is fixed throughout these experiments ($K = 9$ in the 9 points environment; $K = 100$ in the uniform environment), as well as the agents' architectures and all environment parameters. The structural changes in the *effective* number of signals used by the speaker result from annealing λ_C, which dynamically controls the communicative bandwidth that the agents coordinate during self-play.

Remarkably, the emergent structure of the VQ-VIB embedding spaces (Figure 4b and d) captures semantically meaningful information that is grounded in the environment. That is, the embedding vectors roughly encode distances between "named" regions. We note that Tucker et al. [23]'s

[3] The WCS data are available at `www1.icsi.berkeley.edu/wcs` and the human-based color prior is available at `https://github.com/nogazs/ib-color-naming`.

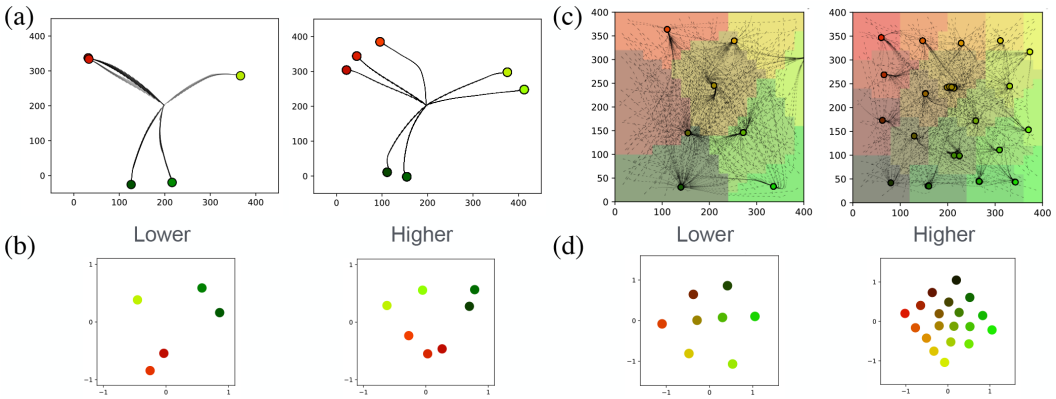


Figure 4: VQ-VIB solutions for the 9 points (a-b) and uniform (c-d) environments, with lower (left panels) and higher (right panels) complexity (see also Movie 1). Each communication vector is associated with the mean color of the positions it can refer to, based on the speaker's encoder and the colors in Figure 3. **(a)** Listener's trajectories (black curves) and final positions (dots) in response to a speaker's communication vector, indicated by the color of the final dot. **(b)** The emergent VQ-VIB communication vectors (trained with $d = 2$). **(c)** Same as (a) but for the uniform environment. Background colors show the modal communication vector for each target position in the continuous space. **(d)** 2D PCA of the VQ-VIB communication vectors (trained with $d = 3$).

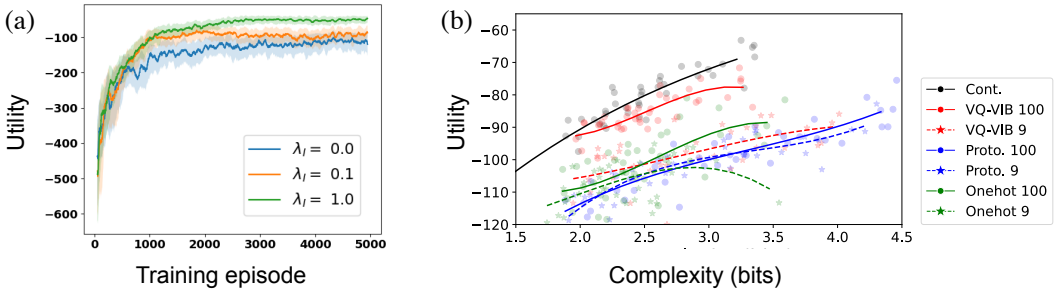


Figure 5: Uniform environment results. **(a)** Expected team utility during training of VQ-VIB agents with varying informativeness pressures. **(b)** Utility vs.complexity for the continuous architecture and the symbolic EC architectures with $K = 9$ (dashed curves and stars) and $K = 100$ (solid curves and dots) signals, excluding a few points below -120 for visualization only. Each curve corresponds to the cubic best fit to the data (stars/dots) generated across 5 random seeds for each configuration.

prototype-based approach is also capable of learning such grounded representations, although without informational constraints, thus resulting in maximally complex and task-specific communication. Therefore, as explained in Section 4, we compare VQ-VIB with an adjusted version of Tucker et al. [23]'s approach (Proto), in addition to a one-hot and a continuous architecture. While the latter is not human-like, we consider it here as an upper bound on the achievable performance in these environments. The results for the uniform environment are shown in Figure 5b, and similar trends also hold in the 9 points environment (Appendix E). It can be seen that VQ-VIB achieves higher utility for any given complexity and codebook size K compared to the two baselines that support discrete (symbolic) communication, i.e., Proto and one-hot. Additionally, increasing K from 9 to 100 substantially improves VQ-VIB, bringing it close to the performance of the continuous architecture, while the same increase in K yields much smaller improvements for Proto and one-hot.

Next, we evaluate the influence of the informativeness weight, λ_I. While informativeness and utility may seem similar, and they are indeed closely related, they are also importantly different: maximizing utility encourages agents to coordinate on a task-specific communication system, whereas maximizing informativeness encourages agents to develop a shared understanding regardless of any task-specific needs. Therefore, intuitively, informativeness may help agents coordinate faster and converge to more effective communication systems. To test this intuition, we eliminate the complexity constraint by setting $\lambda_C = 0$ and vary λ_I. Figure 5a shows that, indeed, increasing λ_I leads to higher utility and

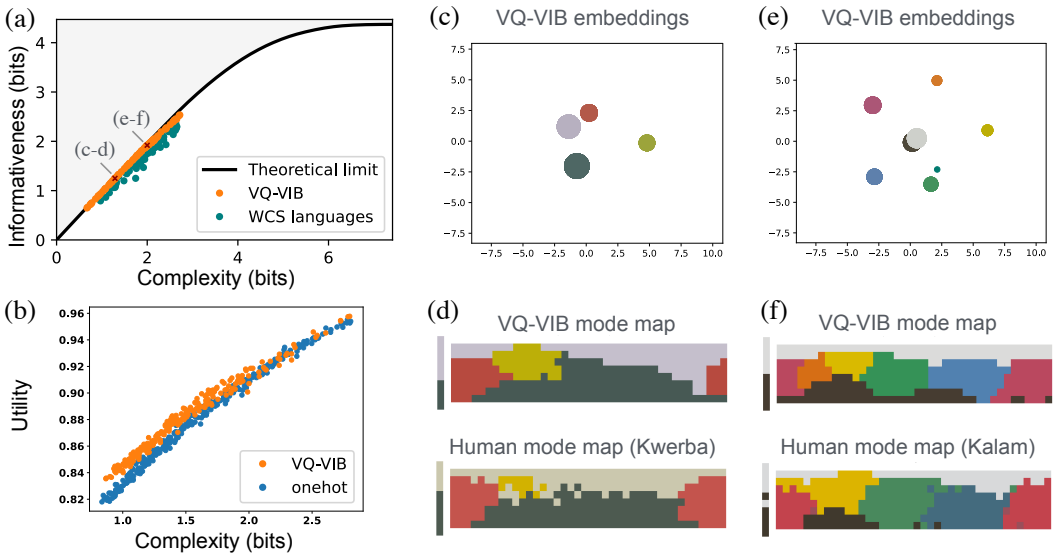


Figure 6: Color reference game results (basic setting, see also Movie 2). **(a)** VQ-VIB agents (orange dots) reach the theoretical bound from [13] (black curve) and capture the variation observed across the WCS languages (teal dots). **(b)** VQ-VIB achieves greater utility, for the same complexity, than one-hot. **(c-f)** The emergent VQ-VIB color naming systems and closest (min gNID) human systems in the WCS data, for agents with lower (c-d) and higher (e-f) complexity, as shown in (a). Each communication vector is associated with the mean color it can refer to. **(c, e)** 2D PCA of the VQ-VIB communication vectors, where the size of each point reflects its relative category size. **(d, f)** Mode maps, where each color chip from Figure 3c is colored according to its modal communication vector.

faster convergence rates. The results shown here are only for VQ-VIB in the uniform environment, but similar trends appear for all architectures in both environments (Appendix E).

5.2 Color reference games

Our results for the two types of color reference games (basic and common ground), and for the two types of training methods (REINFORCE and straight-through estimator) are similar, with REINFORCE being slightly worse (consistent with [10]). We therefore focus here on the basic game with straight-through estimator and provide the full set of results in Appendix F. In addition, because we wish to compare our agents' communication systems with actual human color naming systems, we focus only on VQ-VIB and one-hot. The continuous architecture cannot generate discrete color categories, and based on the previous section, Proto performs similarly to one-hot.

As before, we fix $\lambda_I = 1$ and train agents while annealing λ_C. We set $K = 330$ (the number of color chips) for all agents, which allows full expressivity, although that would require maximal complexity and is unlike attested human languages which typically use only a few color terms [40]. We therefore focus on the range of λ_C that produces the variation in complexity observed across languages. Following Zaslavsky et al. [13], we use the color naming data from the World Color Survey [WCS, 42] to estimate an encoder $S_l(c|x)$ for each WCS language l. Figure 6a replicates Zaslavsky et al. [13]'s finding that languages are near-optimal with respect to the IB bound, and further shows that VQ-VIB agents can also achieve that bound via self-play. Visualizing the emergent VQ-VIB communication systems (Figure 6c-f) shows, as before, the structural changes in the effective codebook size as complexity increases (see Movie 2 for the continuous evolution of these systems), as well as the similarity between the VQ-VIB speaker color-to-signal mapping and the nearest WCS language (Figure 6d and f). This recapitulates the theoretical evolutionary process derived by [13], while additionally taking into account local agent interactions and a learned embedding space.

This embedding space is particularly noteworthy, as it is grounded in human color perception and captures semantically meaningful information about the perceptual color space. At low complexity (Figure 6c), the embeddings reflect the irregularity of the CIELAB space due to the strong perceptual

Table 1: Quantitative evaluation of the color communication systems of annealed agents (VQ-VIB, One-hot), utility-only VQ-VIB agents, and the IB-optimal systems from [13]. Human-agent gNID is the average dissimilarity ($\pm$STD) between human color naming systems and nearest agent system. Agent efficiency loss is the average deviation from optimality ($\pm$STD) of the agents' systems, measured with respect to the theoretical bound from [13]. For both measures, lower values are better.

	VQ-VIB	One-hot	VQ-VIB (utility-only)	IB-optimal
Human-agent gNID	**0.16** (0.01)	**0.16** (0.00)	0.42 (0.02)	0.18 (0.10)
Efficiency loss	**0.03** (0.00)	0.05 (0.00)	0.11 (0.01)	N/A, 0 by definition

salience of yellow (Figure 3c). At higher complexity (Figure 6e), the embeddings form a hue circle, with achromatic colors (black and white) in the middle, as in CIELAB. Intriguingly, Shepard and Cooper [56] found that human similarity judgments for color words also recapitulate the hue circle, suggesting that VQ-VIB may potentially provide a possible mechanistic explanation for the emergence of grounded semantic representations both in humans and in artificial agents.

Finally, we measure the agents' deviation from the IB bound (efficiency loss, as in [13]) and their structural similarity to the nearest WCS languages (using gNID [13], a dissimilarity measure for encoders) for both VQ-VIB and one-hot agents. Table 1 shows that both VQ-VIB and one-hot agents learn human-like systems across the complexity spectrum, confirming the patterns in Figure 6. In addition, both are slightly better (lower human-agent gNID) compared to the IB-optimal systems from [13] that do not take into account utility. It is not surprising that both a stochastic one-hot encoder and a VQ-VIB encoder can perform well within the IB framework, because the traditional IB formulation with discrete representations does not assume any typology on the representation space. However, when assessed with respect to utility (Figure 6b), VQ-VIB outperforms the one-hot architecture, as in the particle world domains, suggesting that an embedding space may be particularly useful for downstream decision making. Training only to maximize utility ($\lambda_I = \lambda_C = 0$), however, causes VQ-VIB agents to converge to a narrow range of communication that fails to capture the diversity of human languages (Table 1, VQ-VIB (utility-only); training utility-only one-hot agents led to similar findings; see Appendix F). This suggests that neither communicative efficiency alone, at least as captured by IB, nor utility alone, but rather their combination may be important for the emergence of human-like communication.

6 Conclusions

In this work, we have proposed an information-theoretic framework for emergent communication that integrates two views: task-specific utility maximization, and general communicative constraints derived from the Information Bottleneck (IB) principle [18, 13]. To train agents within this framework, we have introduce a new method, VQ-VIB, that builds on VIB [19] and VQ-VAE [20], but also differs from them in ways that are particularly important for modeling human languages (Section 3.3). We have shown, across three types of test environments, that VQ-VIB agents can continuously evolve and adapt to changing communicative pressures via annealing a single tradeoff parameter, while learning grounded embedding spaces for communication that capture semantically meaningful information about the perceived environment, reminiscent of word embeddings. VQ-VIB also outperforms prior methods that support discrete (symbolic) communication, and in the domain of color, it can be aligned with human languages and possibly also with human similarity judgments. Natural extensions of our work include more complex domains and challenging tasks, as well as human-agent interaction experiments, which may further demonstrate advantages of VQ-VIB. More broadly, we believe our framework could be useful for further developing and improving both emergent communication in artificial agents and models of human language evolution.

Acknowledgments and Disclosure of Funding

M.T. was supported by an Amazon Science Hub Fellowship. N.Z. was supported by a K. Lisa Yang Integrative Computational Neuroscience Postdoctoral Fellowship. We thank Rahma Chaabouni for useful discussions on their prior work, and our anonymous reviewers for their helpful feedback.

References

[1] Luc Steels and Tony Belpaeme. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, 28(4):469–489, 2005.

[2] Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, and Shimon Whiteson. Learning to communicate with deep multi-agent reinforcement learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 2145–2153, Red Hook, NY, USA, 2016. Curran Associates Inc.

[3] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 6382–6393, Red Hook, NY, USA, 2017. Curran Associates Inc.

[4] Igor Mordatch and Pieter Abbeel. Emergence of grounded compositional language in multi-agent populations. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18. AAAI Press, 2018.

[5] 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.

[6] Angeliki Lazaridou and Marco Baroni. Emergent multi-agent communication in the deep learning era. *arXiv preprint arXiv:2006.02419*, 2020.

[7] Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. Anti-efficient encoding in emergent communication. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *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.

[8] Satwik Kottur, José Moura, Stefan Lee, and Dhruv Batra. Natural language does not emerge 'naturally' in multi-agent dialog. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2962–2967, Copenhagen, Denmark, September 2017. Association for Computational Linguistics. doi: 10.18653/v1/D17-1321.

[9] Tom Eccles, Yoram Bachrach, Guy Lever, Angeliki Lazaridou, and Thore Graepel. Biases for emergent communication in multi-agent reinforcement learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[10] 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.

[11] Edward Gibson, Richard Futrell, Steven Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Philip Levy. How efficiency shapes human language. *Trends in Cognitive Sciences*, 23:389–407, 2019.

[12] Charles Kemp, Yang Xu, and Terry Regier. Semantic typology and efficient communication. *Annual Review of Linguistics*, 4:109–128, 01 2018. doi: 10.1146/annurev-linguistics-011817-045406.

[13] Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31): 7937–7942, 2018. doi: 10.1073/pnas.1800521115.

[14] Noga Zaslavsky, Terry Regier, Naftali Tishby, and Charles Kemp. Semantic categories of artifacts and animals reflect efficient coding. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*, 2019.

[15] Noga Zaslavsky, Mora Maldonado, and Jennifer Culbertson. Let's talk (efficiently) about us: Person systems achieve near-optimal compression. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*, 2021.

[16] Francis Mollica, Geoff Bacon, Noga Zaslavsky, Yang Xu, Terry Regier, and Charles Kemp. The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, 118(49):e2025993118, 2021. doi: 10.1073/pnas.2025993118.

[17] Noga Zaslavsky, Karee Garvin, Charles Kemp, Naftali Tishby, and Terry Regier. The evolution of color naming reflects pressure for efficiency: Evidence from the recent past. *Journal of Language Evolution*, 04 2022. ISSN 2058-458X. doi: 10.1093/jole/lzac001. lzac001.

[18] Naftali Tishby, Fernando C. Pereira, and William Bialek. The Information Bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, 1999.

[19] 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.

[20] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 6309–6318, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

[21] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

[22] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1532–1543, 2014. doi: 10.3115/v1/d14-1162.

[23] Mycal Tucker, Huao Li, Siddharth Agrawal, Dana Hughes, Katia P. Sycara, Michael Lewis, and Julie A. Shah. Emergent discrete communication in semantic spaces. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 10574–10586, 2021.

[24] Ł ukasz Kuciński, Tomasz Korbak, PawełKoł odziej, and Piotr Mił oś. Catalytic role of noise and necessity of inductive biases in the emergence of compositional communication. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 23075–23088. Curran Associates, Inc., 2021.

[25] Igor Mordatch and Pieter Abbeel. Emergence of grounded compositional language in multi-agent populations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), Apr. 2018.

[26] Claude E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec*, 4(142-163):1, 1959.

[27] Rundong Wang, Xu He, Runsheng Yu, Wei Qiu, Bo An, and Zinovi Rabinovich. Learning efficient multi-agent communication: An information bottleneck approach. In *ICML*, pages 9908–9918, 2020.

[28] Matti Miestamo. On the feasibility of complexity metrics. In *FinEst linguistics, proceedings of the annual Finnish and Estonian conference of linguistics, Tallinn*, pages 11–26, 2004.

[29] Mathieu Rita, Rahma Chaabouni, and Emmanuel Dupoux. "LazImpa": Lazy and impatient neural agents learn to communicate efficiently. In *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 335–343, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.conll-1.26.

[30] Michael Cogswell, Jiasen Lu, Stefan Lee, Devi Parikh, and Dhruv Batra. Emergence of compositional language with deep generational transmission. *CoRR*, abs/1904.09067, 2019.

[31] Yi Ren, Shangmin Guo, Matthieu Labeau, Shay B. Cohen, and Simon Kirby. Compositional languages emerge in a neural iterated learning model. In *International Conference on Learning Representations*, 2020.

[32] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro Ortega, DJ Strouse, Joel Z Leibo, and Nando De Freitas. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In *International Conference on Machine Learning*, pages 3040–3049. PMLR, 2019.

[33] Toru Lin, Jacob Huh, Christopher Stauffer, Ser Nam Lim, and Phillip Isola. Learning to ground multi-agent communication with autoencoders. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 15230–15242. Curran Associates, Inc., 2021.

[34] Naftali Tishby and Daniel Polani. Information theory of decisions and actions. In Vassilis Cutsuridis, Amir Hussain, and John G. Taylor, editors, *Perception-Action Cycle: Models, Architectures, and Hardware*, pages 601–636. Springer New York, New York, NY, 2011.

[35] Susanne Still and Doina Precup. An information-theoretic approach to curiosity-driven reinforcement learning. *Theory in Biosciences*, 131(3):139–148, 2012.

[36] Pedro A. Ortega and Daniel A. Braun. Thermodynamics as a theory of decision-making with information-processing costs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 469(2153):20120683, 2013. doi: 10.1098/rspa.2012.0683.

[37] Roy Fox, Ari Pakman, and Naftali Tishby. Taming the noise in reinforcement learning via soft updates. In *32nd Conference on Uncertainty in Artificial Intelligence (UAI)*, 2016.

[38] Anirudh Goyal, Riashat Islam, DJ Strouse, Zafarali Ahmed, Hugo Larochelle, Matthew Botvinick, Sergey Levine, and Yoshua Bengio. Transfer and exploration via the information bottleneck. In *International Conference on Learning Representations*, 2019.

[39] David Lewis. *Convention: A philosophical study*. John Wiley & Sons, 2008.

[40] Brent Berlin and Paul Kay. *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley and Los Angeles, 1969.

[41] Terry Regier, Paul Kay, and Naveen Khetarpal. Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, 104(4):1436–1441, 2007. doi: 10.1073/pnas.0610341104.

[42] Paul Kay, Brent Berlin, Luisa Maffi, William R. Merrifield, and Richard Cook. *The World Color Survey*. Stanford: Center for the Study of Language and Information, 2009.

[43] Mike Dowman. Explaining color term typology with an evolutionary model. *Cognitive Science*, 31(1):99–132, 2007.

[44] Will Monroe, Robert X.D. Hawkins, Noah D. Goodman, and Christopher Potts. Colors in context: A pragmatic neural model for grounded language understanding. *Transactions of the Association for Computational Linguistics*, 5:325–338, 2017.

[45] Mikael Kågebäck, Emil Carlsson, Devdatt Dubhashi, and Asad Sayeed. A reinforcement-learning approach to efficient communication. *PLOS ONE*, 15(7):1–26, 07 2020.

[46] Toby Berger. *Rate distortion theory; a mathematical basis for data compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[47] Peter Harremoës and Naftali Tishby. The Information Bottleneck revisited or how to choose a good distortion measure. In *IEEE International Symposium on Information Theory*, pages 566–571, June 2007.

[48] Yoshua Bengio, Nicholas Léonard, and Aaron C. Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *CoRR*, abs/1308.3432, 2013.

[49] 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.

[50] Noga Zaslavsky. *Information-Theoretic Principles in the Evolution of Semantic Systems*. Ph.D. Thesis, The Hebrew University of Jerusalem, 2020.

[51] Sainbayar Sukhbaatar, Arthur Szlam, and Rob Fergus. Learning multiagent communication with backpropagation. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 2252–2260, Red Hook, NY, USA, 2016. Curran Associates Inc.

[52] Jennifer Hu, Roger Levy, and Noga Zaslavsky. Scalable pragmatic communication via self-supervision. In *ICML Workshop on Self-Supervised Learning for Reasoning and Perception*, 2021.

[53] WS Mokrzycki and M Tatol. Colour difference ΔE - a survey. *Machine Graphic and Vision*, 8, 2012.

[54] Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach. Learn.*, 8(3–4):229–256, may 1992. ISSN 0885-6125. doi: 10.1007/BF00992696.

[55] Kenneth Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proc. IEEE*, 86(11):2210–2239, 1998. doi: 10.1109/5.726788.

[56] Roger N. Shepard and Lynn A. Cooper. Representation of colors in the blind, color-blind, and normally sighted. *Psychological Science*, 3(2):97–104, 1992. ISSN 09567976, 14679280.

[57] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

[58] Chris J. Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See abstract, introduction, and findings
   (b) Did you describe the limitations of your work? [Yes] See Section 6, in which we note that we tested in limited domains and Appendix B, in which we note that an approximation we made may be improved upon in future work.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Section 6 in which we discuss how readers may misconstrue quantitative measures of languages as assigning value to one over the other.

(d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] See footnote for link to code
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Appendix D
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] See error bars in plots and Table 1, in Section 5. All scatter plots directly included the raw results from all 5 random trials.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Training times and computational resources are listed in Appendix D.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We use the World Color Survey (WCS) data, as well as the publicly-available code by Zaslavsky et al. [13]. Both sources are cited.
   (b) Did you mention the license of the assets? [Yes] We note that both are publicly-available and link to them in Footnote 1.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] Code is linked to in a footnote.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A] We only used the existing WCS dataset.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A] We only used the existing WCS dataset.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Variational bound on the informativeness loss

Here we prove the variational bound on the informativeness loss term (second term in Eq. (4)) that is used in Section 3.3.2 to derive the VQ-VIB learning objective. Recall that the speaker's belief states, m_x, are defined as probability distributions over a feature space $\mathcal{U}$ representing the domain $\mathcal{X}$. A decoder maps a communication signal $c \in \mathcal{C}$ to a reconstruction $\hat{m}_c$ of the speaker's belief. As shown in [13, SI Section 1.2], for any given speaker $S(c|x)$, the optimal decoder that minimizes $\mathbb{E}_S[D_{\mathrm{KL}}[m_x \| \hat{m}_c]]$ over all possible reconstructions $\hat{m}_c$ is Bayesian; that is, its reconstruction is given by $\hat{m}_c(u) = \sum_x S(x|c) m_x(u)$, where $S(x|c)$ is the posterior distribution with respect to $S(c|x)$ and $p(x)$. Recall that $\hat{m}_c(u)$ is exactly the listener's decoder in the IB framework (see Section 3.1.1). Therefore, any other decoder would lend an upper bound on the informativeness loss term. Here, we derive a particularly useful bound under the assumptions that $\mathcal{X} \subseteq \mathbb{R}^n$,[4] $\mathcal{U} = \mathbb{R}^n$, and that the speaker's uncertainty can be modeled with Gaussian noise, $\mathcal{N}(0, \Sigma)$. For simplicity, we assume that the noise covariance, Σ, does not depend on the input, however this assumption is not required for

[4] While $\mathcal{X}$ may be a continuous space, in our settings it is either a finite dataset or it is approximated by a finite sample during training. Therefore, we treat it here as a discrete set.

our general claim and could be easily relaxed. In this case, the speaker's belief states are given by $m_x \equiv \mathcal{N}(x, \Sigma)$. Let $D_\psi : \mathcal{C} \to \mathbb{R}^d$ be a function parameterized by ψ, then it holds that

$$\min_{\hat{m}'_c} \mathbb{E}_S[D_{\mathrm{KL}}[m_x \| \hat{m}'_c]] = \mathbb{E}_S[D_{\mathrm{KL}}[\mathcal{N}(x, \Sigma) \| \hat{m}_c]] \tag{8}$$

$$\leq \min_{\hat{\mu}(c), \hat{\Sigma}(c)} \mathbb{E}_S[D_{\mathrm{KL}}[\mathcal{N}(x, \Sigma) \| \mathcal{N}(\hat{\mu}(c), \hat{\Sigma}(c))]] \tag{9}$$

$$\leq \min_{\psi, \hat{\Sigma}(c)} \mathbb{E}_S[D_{\mathrm{KL}}[\mathcal{N}(x, \Sigma) \| \mathcal{N}(D_\psi(c), \hat{\Sigma}(c))]] \,. \tag{10}$$

While Eq. (8) follows from [13] and the inequality in Eq. (9) is straightforward, it has an interesting cognitive interpretation. Notice that under our assumptions, $\hat{m}_c$ is a Gaussian mixture, whereas the speaker's beliefs are simply Gaussian. Eq. (9) is equivalent to assuming that the listener's reconstructions are restricted to Gaussian distributions, similar to the speaker. Therefore, while this might not give a tight bound on the optimal IB solution, it may actually be more cognitively plausible. In addition, the optimal $\hat{\mu}(c)$ for Eq. (9) are given by the posterior centroids, i.e,

$$\hat{\mu}^*(c) = \mathbb{E}_S[x|c] = \sum_{x \in \mathcal{X}} S(x|c)x \,, \tag{11}$$

which means that the optimal D^*_ψ for Eq. (10) will approximate these centroids. Eq. (10) has a convenient form as minimizing the KL divergence between two Gaussians,

$$\min_{\psi, \hat{\Sigma}(c)} \frac{1}{2} \mathbb{E}_S \left[\left(\mathrm{tr}(\hat{\Sigma}(c)\Sigma^{-1}) + (x - D_\psi(c))^\top \hat{\Sigma}(c)^{-1}(x - D_\psi(c)) - n + \ln \frac{|\hat{\Sigma}(c)|}{|\Sigma|} \right) \right] \,, \tag{12}$$

which in principle, could be used in our framework with a proper parameterization of $\hat{\Sigma}(c)$. However, for simplicity, we take $\hat{\Sigma}(c) = I_d$, which, in combination with Eqs. (8)-(10), yields the following bound:

$$\mathbb{E}_S[D_{\mathrm{KL}}[m_x | \hat{m}_c]] \leq \min_\psi \mathbb{E}_S[D_{\mathrm{KL}}[\mathcal{N}(x, \Sigma) \| \mathcal{N}(D_\psi(c), I_d]] \tag{13}$$

$$\leq \frac{1}{2} \mathbb{E}_S \left[\|x - D_\psi(c)\|^2 \right] + a \,, \tag{14}$$

where $a = \mathrm{tr}(\Sigma^{-1}) - n - \ln |\Sigma|$ is a constant that does not affect the optimization.

B Symmetry breaking regularization

As shown in [50], for every complexity level there is a set of IB-optimal systems with varying *effective* lexicon sizes, $k = |\{c \in \mathcal{C} : \sum_x p(x) S_\theta(c|x) > 0\}|$. All the systems with the same k form an equivalence class and the canonical system within each class is the one with minimal k. These canonical systems are the natural one to prefer, because they can attain the optimum for a given complexity with a minimal codebook. We therefore aim to bias our agents toward these systems. One way of achieving that is by regularizing the entropy of the speaker's communication vectors. However, this is problematic in our settings given the non-differentiablity of the `argmin` operation used in the VQ-VIB discretization layer. Instead, we use the entropy of an approximated categorical distribution based on the euclidean distance to each quantized vector, i.e., $p_{\text{cat}}(\zeta_i|z) \propto e^{-(z-\zeta_i)^2}$ Because this entropy term is only used to break the symmetry between equally optimal systems, we use a small regularizing weight (0.05 in our experiments). We denote the weighted regularization term by $\varepsilon(\theta, \mathcal{C})$, as it depends both on the encoder $q_\theta(z|x)$ and on the embedding space $\mathcal{C}$.

C VQ-VIB After

We proposed an additional architecture, VQ-VIB After, that is closely related to our main VQ-VIB method. In VQ-VIB After, an encoder to generates a discrete representation ζ, whereupon a communication vector c is sampled from a Gaussian in the continuous space centered around ζ. This "sample-after" method is depicted in Figure 8, and corresponds to an intuitive notion of choosing a word (a discrete embedding) deterministically but then being uncertain about the form (by sampling in the continuous space around it).

This architecture fits into the standard Information Bottleneck framework, as complexity can be penalized via standard variational methods. However, it is distinct from discrete communication methods because communication vectors are ultimately sampled from a continuous space. Thus, this work does not fit naturally into the main focus of this work, but we believe it could have applications in other settings, such as speech processing (wherein discrete words are pronounced in a continuous auditory space).

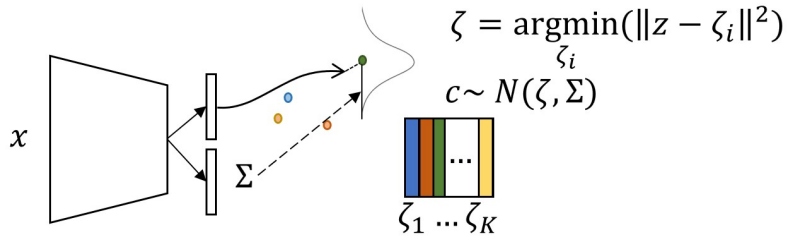


Figure C.1: In the VQ-VIB After architecture, the speaker deterministically maps x to the nearest quantized encoding, ζ, and then c is sampled from a Gaussian centered at ζ.

D Implementation Details

Here, we include implementation details that we omitted from the main paper that can be used to reproduce our experiments.

D.1 Multi-Agent Reinforcement Learning Experiments

We trained agents in the 9 points and Uniform environments with Multi-Agent Deep Deterministic Policy Gradient (MADDPG), a standard method for multi-agent reinforcement learning [3].

Agent architectures were identical in both environments. Regardless of the speaker architecture used, all teams used common decoder and listener agents (except for variations due to different communication dimensions, explained later). The decoder used 2 fully-connected ReLU layers with hidden dimension 64; the listener used 3 fully-connected ReLU layers with hidden dimension 64.

Because of the common architecture for the decoder and listener agents across teams, variation arose due to differences among the speaker agents. A speaker agent consisted of 2, 64-neuron fully-connected layers with ReLU activations, followed by a speaker "head." We dub the output of these first two layers the "encoder representation." Speaker heads implemented the specific communication method (e.g., continuous, VQ-VIB, etc.) by taking the 64-dimensional encoder representation and going through one of 5 possible architectures to produce a vector in the communication space, $\mathbb{R}^d$. (As discussed later, d varied by environment and architecture slightly.)

- **Continuous:** The continuous communication head consisted of two fully connected layers branching off the encoder representation to generate a μ and `logvar` (the log of the variance) from which the communication vector was sampled using the reparametrization trick. Here, as in other speaker architectures, we assumed a diagonal covariance matrix. Complexity was measured as the KL divergence between the parameters μ and `logvar` and a unit Gaussian.
- **one-hot:** The one-hot communication head consisted of one fully connected layer from the encoder representation to $\mathbb{R}^d$. One-hot vectors were generated using gumbel softmax sampling, with temperature 1. Complexity was measured as the KL divergence between the softmax distribution from which the one-hot vector was sampled and a uniform categorical distribution. Thus, complexity corresponds to the difference between the conditional and unconditional entropies of messages, unlike non-IB frameworks that often penalize unconditional entropy. We used greater communication dimensionality for one-hot agents and therefore rescaled the first layer of the decoder and listener agents to accept such communication.

- **VQ-VIB:** The VQ-VIB communication head first passed the encoder representation through two fully connected layers, branching off the encoding, to generate a μ and `logvar`, as in the continuous communication case. A communication vector was sampled from the distribution, whereupon it was discretized to the nearest quantized vector, ζ, in the communication space (as done in a standard VQ-VAE, without the sampling). Complexity was measured as the KL divergence between the parameters μ and `logvar` and a unit Gaussian.
- **VQ-VIB After:** The VQ-VIB After communication head first passed the encoder representation through two fully connected layers to generate an encoding, z, and `logvar`. The encoding, z, was discretized to the nearest quantized vector, ζ. Lastly, a communication vector, c, was sampled from a normal distribution centered at ζ with variance specified via `logvar`. Complexity was measured as the KL divergence between the parameters ζ and `logvar` and a unit Gaussian.
- **Proto:** The Proto head mapped the encoder representation through a single fully-connected layer to a categorical distribution over k neurons, for k representing the number of prototypes. A k-dimensional one-hot vector was sampled using gumbel softmax; a prototype, p was selected by multiplying the prototype matrix by the one-hot vector. A different fully-connected layer mapped from the encoder representation to `logvar`. Lastly, a communication vector, c, was sampled from a Gaussian centered at p with variance via `logvar`. Complexity was measured as the KL divergence between the parameters p and `logvar` and a unit Gaussian. This closely resembles the VQ-VIB After architecture, but with a different discretization mechanism.

All agents were trained with an Adam optimizer with learning rate 0.01 and batch size 1024.

D.1.1 9 points Environment

For all agent architectures except for one-hot, the communication dimension was 2; one-hot used communication dimension 9 so that the number of discrete representations allowed by the architecture was not the limiting factor for complexity. Similarly, the VQ-VIB and Proto agents used 9 discrete embeddings.

In our results with annealed complexity weight (presented in Figure E.3), agents were trained for 10,000 episodes, with each episode lasting 100 timesteps. The listener agent was always spawned at the origin, equidistant from each landmark. The informativeness loss weight, λ_I, was constant at 1.0. For the continuous and VQ-VIB communication methods, the complexity loss weight, λ_C increased by $8 * 10^{-6}$ per episode. The one-hot and Proto methods suffered from communication collapse at that rate, so results λ_C increased half as quickly: at $4 * 10^{-6}$ per episode.

Training a team in the 9 points environment for 10,000 episodes took approximately 15 minutes on a desktop computer with 16 Intel i9 cores for all architectures, except for one-hot, which took roughly 20 minutes. This difference in time is attributable to the larger communication dimension needed for one-hot.

D.1.2 Uniform Environment

In the Uniform environment, we used identical parameters as in the 9 points environment, except for the changes noted below.

For all agent architectures except for one-hot, the communication dimension was 3; for one-hot, we experimented with communication dimensions of 9 and 100. Similarly, we tested VQ-VIB and Proto agents when using 9 or 100 discrete embeddings. The listener agent was spawned at a starting location sampled uniformly at random from the map (bounded between -200 and 200 for both x and y). Otherwise, we used the same number of training episodes, episode duration, and complexity annealing rates as in 9 points.

Training a team in the Uniform environment for 10,000 episodes took approximately 15 minutes, except for one-hot, which took roughly 30 minutes for 100-dimensional communication.

D.2 Color Reference Game Experiments

Here, we include the implementation details used in our color reference game experiments.

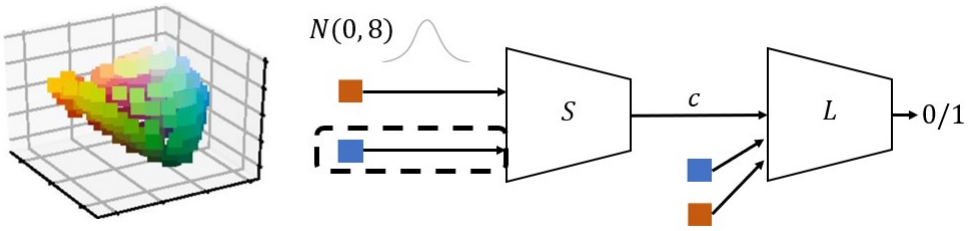


Figure D.1: Target and distractor colors were drawn uniformly without replacement from the set of 330 colors in CIELAB space (left). A speaker, S, observed the target and (in the Common Ground setting) distractor, corrupted by Gaussian noise. The listener, L, had to identify which of the shuffled colors it saw was the target color, given the speaker's communication, c.

Agents were trained in the overall architecture depicted in Figure D.1. The speaker agent was parametrized with 3 fully connected layers with ReLU activations and hidden dimension 64, followed by a communication head. The communication heads used the same architectures as specified in the particle world experiments: gumbel softmax was used for generating one-hot vectors, and a straight-through estimator was used for discretization in VQ-VIB architectures [57, 58]. In all experiments, the communication dimension was set to 32 for all agents except one-hot, for which the communication dimension was set to 330 (to match the expressive power of other discrete agents, that each had 330 discrete tokens in the 32 dimensional space). The decoder and listener agents used the same feed-forward architecture as in the particle world experiments.

Without annealing the complexity loss, 50 training epochs sufficed for agents to converge to a stable solution. Thus, in our experiments with $\lambda_C = 0$, we only trained for 50 epochs. For later experiments, wherein we annealed complexity, we first trained agents for 50 epochs before beginning to anneal the complexity term for the next 40 epochs (for a total of 90 epochs). We note that, for VQ-VIB, we set the complexity loss to 0 during this "burnin" period, but for one-hot communication, it appeared necessary to set the loss to 0.05 to prevent numerical instability during the first epoch after the burnin period. During complexity loss annealing, we increased λ_C by 0.01 during each epoch for VQ-VIB, and by 0.1 for one-hot. These different rates were necessary to induce similar ranges in complexity of learned communication.

In the Basic color reference game, the speaker observed a only the 3-dimensional, corrupted target color. In the Common Ground setting, the speaker observed a 6-dimensional vector, generated by concatenating noisy versions of the target and distractor colors. In both settings, the listener observed the communication output by the speaker, as well as the target and distractor colors (ordered randomly, without added noise). Results were analyzed using code partially-based on publicly-available code released by Zaslavsky et al. [13].

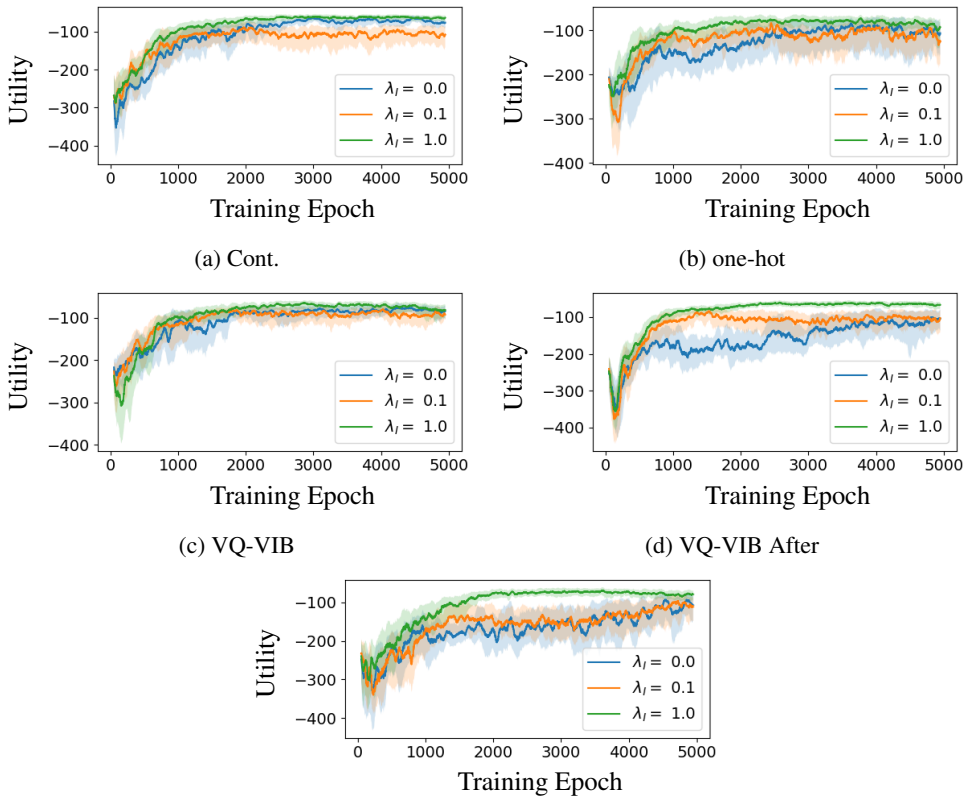
We trained agents on a desktop computer with 16 Intel i9 CPUs and 1 NVIDIA GeForce RTX 2080. Training for 90 epochs in both the Basic and Common Ground settings took approximately 3 minutes for all agent architectures. Agents were trained with an Adam optimizer with default parameters, with batch size 1024.

E Full Particle World Results

For brevity, we only included partial results from the particle-world experiments in the main paper. Here, we include the full results.

First, we experimented with varying λ_I for different environments and neural architectures, without annealing complexity. Training curves for both environments, using varying λ_I for each neural architecture, are included in Figures E.1 and E.2. For each configuration (a particular speaker architecture and λ_I), we trained 5 teams from scratch for 5,000 episodes and plotted the median and quartile rewards, smoothed over a 50 episode window.

A consistent trend emerged: higher λ_I led to faster convergence to a higher reward. Indeed, this corroborates findings by Lin et al. [33], who used an autoencoder architecture for communication in multiagent settings, but here we showed good convergence properties when mixing utility and



(a) Cont.

(b) one-hot

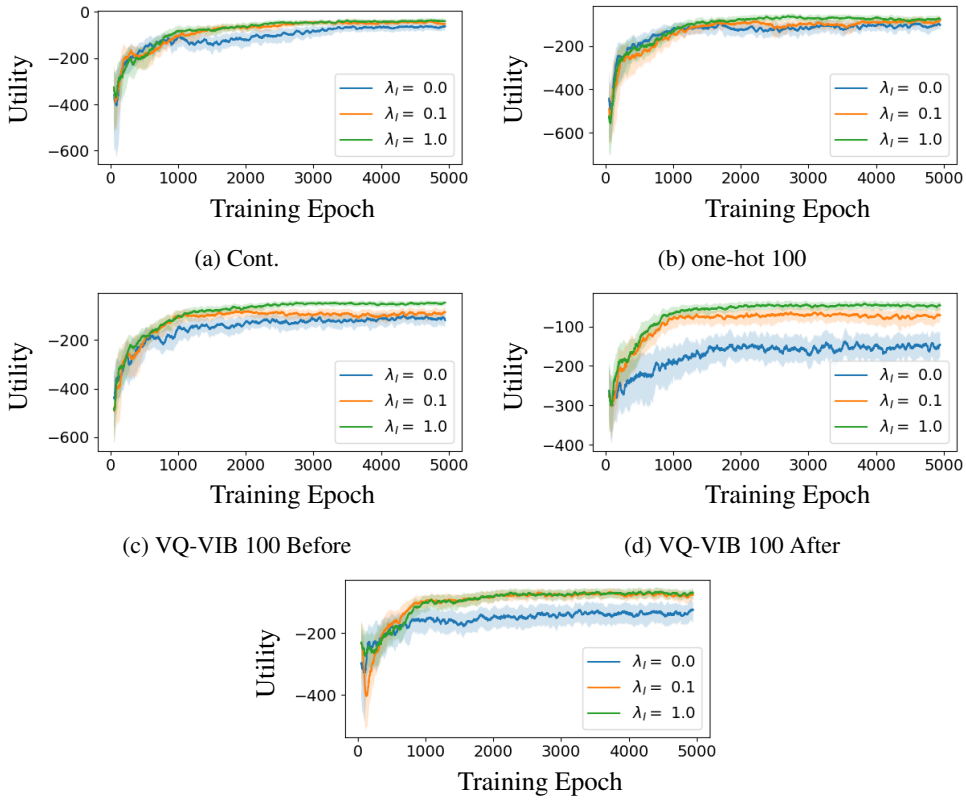
(c) VQ-VIB

(d) VQ-VIB After

(e) Proto

Figure E.1: In the 9 points environment, using a higher weight for informativeness (λ_I) resulted in faster convergence to higher rewards for each communication architecture. Medians and quartiles plotted over 5 trials for each configuration.

informativeness terms (whereas they reported worse results when adding utility losses). Overall, these findings motivated training agents to convergence with high λ_I before beginning to penalize complexity with λ_C.



(a) Cont.

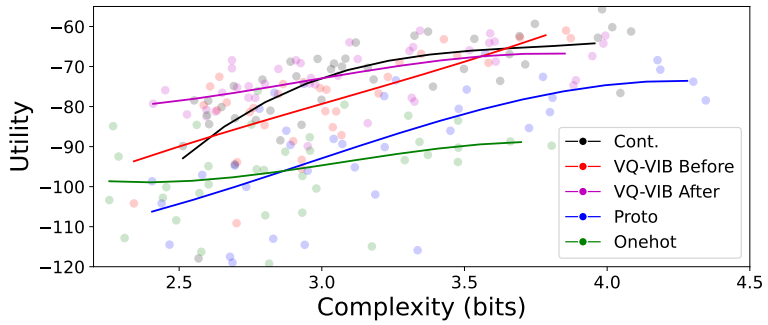
(b) one-hot 100

(c) VQ-VIB 100 Before

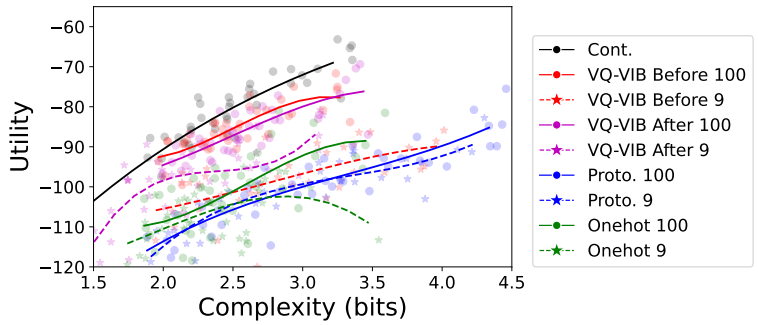
(d) VQ-VIB 100 After

(e) Proto 100

Figure E.2: In the Uniform environment, we observed similar trends as in 9 points: greater λ_I improved convergence rates and rewards. Medians and quartiles plotted over 5 runs.



(a) 9 points: Reward vs. Complexity



(b) Uniform: Reward vs. Complexity

Figure E.3: In the 9-points and Uniform environments, agent reward (utility) decreased with complexity. Each point represents the mean complexity and reward for agent checkpoints at 1,000 episode increments during training, over 5 random seeds. Curves are the best cubic fits to the data. VQ-VIB architectures typically outperformed other discrete methods by achieving higher reward for the same complexity, and over-parametrizing with additional embeddings (100 vs. 9) also helped.

Next, we include the results of annealing complexity in both the 9 points and Uniform environments, as depicted in Figure E.3. In both environments, as we increased λ_C, complexity and utility decreased. Comparing both VQ-VIB architectures to other discrete communication methods showed that VQ-VIB agents tended to obtain a higher mean reward, for a given complexity, than other methods. In the Uniform environment, we also compared different variants of VQ-VIB, Proto, and one-hot agents based on the maximum number of discrete messages allowed by the neural architectures. We tested agents parametrized with either 100 or 9 unique messages; we hypothesized that initializing agents with many discrete messages and then gradually decreasing complexity would result in better performance than training with few messages from the start. Indeed, we observed that was the case. As seen in Figure E.3 b, VQ-VIB agents parametrized with 100 embeddings consistently achieved higher reward for the same complexity than VQ-VIB agents parametrized with only 9 embeddings. This trend held even when the complexity was lower than $\log_2(9)$ bits.

Ultimately, results from both particle world environments corroborate the findings from the color reference games: encouraging informativeness improved convergence, VQ-VIB achieved greater utility than one-hot, and decreasing complexity led to coarser discretizations (main paper, Figure 4). The varied domains and training mechanisms (MARL vs. backpropagation for reference games) leading to consistent results suggest broad generalizability of our framework and VQ-VIB.

F Color Reference Game Full Results

We omitted some of the color reference game results from the main paper due to space constraints. Here, we include our complete results.

F.1 Basic Color Reference Game

First, we include results from the Basic setting, in which we trained agent with different values of λ_I without varying λ_C. These results are included in Figure F.1. Setting $\lambda_I = 0$ induced similar conditions to those used by Chaabouni et al. [10]; any positive λ_I favored more informative (and therefore more complex) communication. Given the monotonically increasing nature of the IB bound, an optimal agent trained with any positive λ_I would maximize utility and informativeness by maximizing complexity, up to a maximum of $\log_2(330)$ bits. Instead, we found that agents appeared to converge to lower complexity values, although increasing λ_I tended to somewhat increase complexity and informativeness. For example, one-hot communication increased in complexity from roughly 2.3 bits for $\lambda_I = 0.0$ to 3.0 bits for $\lambda_I = 1.0$. This alone is a novel result compared to Chaabouni et al. [10], who trained agents with no informativeness loss and observed lower-complexity communication. Curiously, VQ-VIB agents automatically converged to informativeness around 3.5 bits regardless of the λ_I value, although increasing λ_I did improve training stability. This difference between one-hot and VQ-VIB indicates that the differing architectures themselves have different inductive biases.

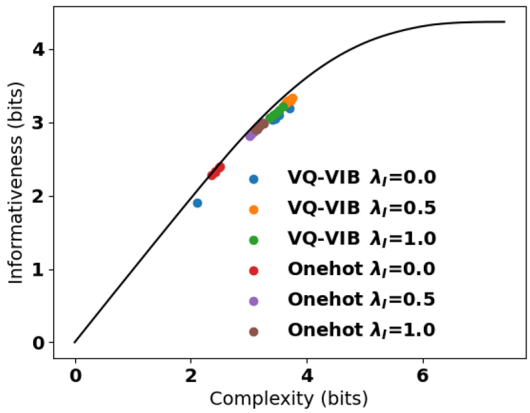
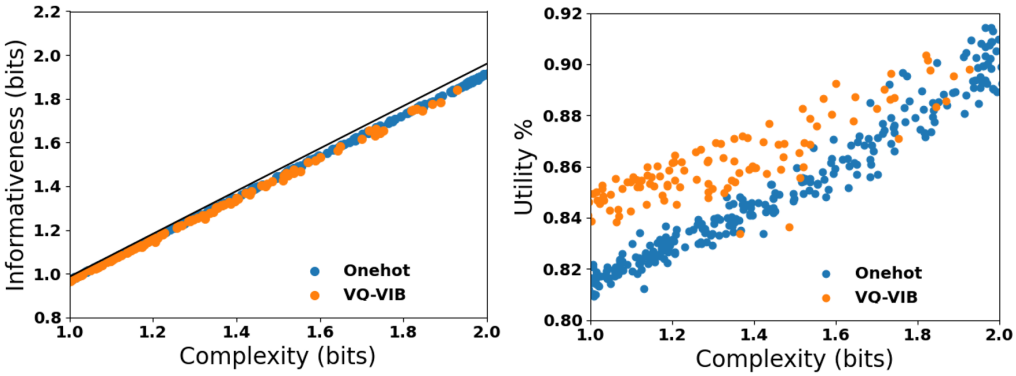


Figure F.1: Training agents in the Basic color reference game for various λ_I but no complexity loss. Without positive λ_I, one-hot learns low-informativeness communication, but increasing λ_I generated more complex agents, especially for one-hot.

The results in Figure F.1 were generated for varying λ_I and low (but fixed) λ_C. We wished to train agents with no penalty for complexity, but in practice we found that one-hot (but not VQ-VIB) agents often failed to converge to any meaningful communication if $\lambda_C = 0$ throughout training. Therefore, for both VQ-VIB and one-hot agents, we initially set $\lambda_C = 0.05$ for the first 30 epochs before setting $\lambda_C = 0$ for the remaining of training. We encountered similar convergence issues when training one-hot agents to maximize utility for Table 1 (to recreate the "VQ-VIB (max U)" column). One-hot agents did not converge for a constant $\lambda_C = 0$ but did if trained for 30 epochs with $\lambda_C = 0.05$ before being reset to 0. We leave exploration of these initialization weights as an area for future work.

We conducted further experiments to compare agents with the neural agents presented by Chaabouni et al. [10], wherein one-hot agents were parametrized with a communication dimension of 1024. We therefore tested VQ-VIB and one-hot agents with 1024 tokens, instead of the 330 used for the results in our main paper. This hyperparameter change largely did not alter results, which is unsurprising given that agents with 330 tokens did not use their full vocabulary. As shown in Figure F.2, both VQ-VIB and one-hot achieved nearly optimal informativeness for a given complexity, and VQ-VIB achieved higher utility than one-hot (results zoomed in to complexity values between 1 and 2 bits for clearest plotting of these trends).

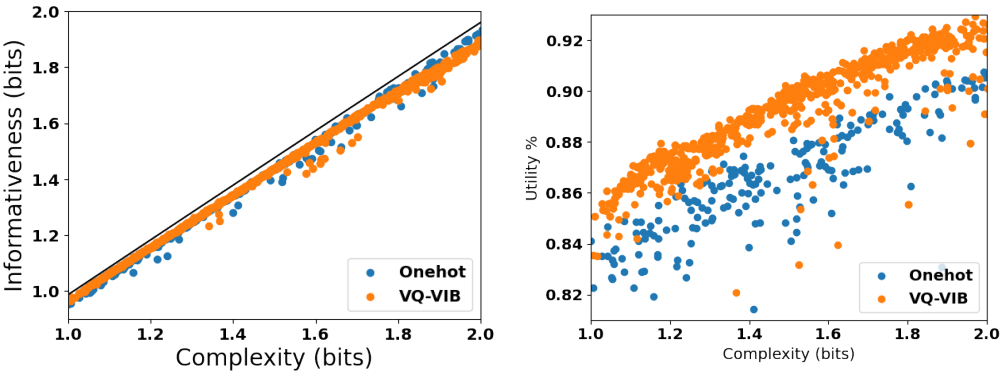
We performed additional experiments to investigate the effect of training with REINFORCE, instead of the supervised learning framework used in our main experiments (e.g., backpropagating through



(a) (b)

Figure F.2: Communicative informativeness (a) and utility (b) as a function of complexity for color agents trained with 1024 tokens in the Basic game. These results largely reproduced results from our main paper with 330 tokens, wherein VQ-VIB and one-hot were similarly efficient, but VQ-VIB achieved higher utility than one-hot.

gumbel softmax) [54]. Chaabouni et al. [10] had found that training with REINFORCE typically resulted in lower-complexity communication than when training with messages generated via a gumbel softmax layer, and that training with REINFORCE was less stable. Results from our REINFORCE experiments, using 330 tokens for one-hot and VQ-VIB, are included in Figure F.3. We trained agents with $\lambda_U = 1.0$, $\lambda_I = 50.0$, and λ_C was annealed from 0 to 0.1, using code adapted from Rita et al. [29].



(a) (b)

Figure F.3: Informativeness (a) and utility (b) versus complexity for agents trained with 330 tokens, using REINFORCE, in the Basic reference game. These results closely tracked prior experiments: both one-hot and VQ-VIB were efficient, and VQ-VIB achieved greater utility, for the same complexity, than one-hot.

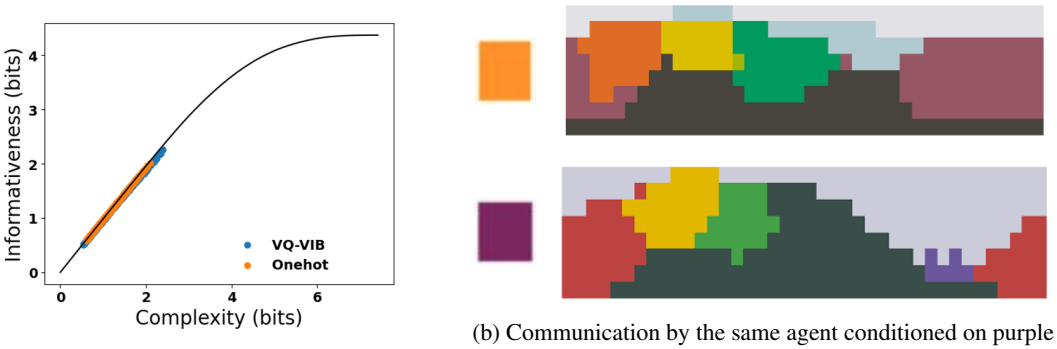
Our results from these experiments align with all prior findings: VQ-VIB and one-hot achieved high communicative efficiency, and VQ-VIB produced greater utility than one-hot did. Beyond those corroborating results, our experiments aligned with Chaabouni et al. [10]'s findings and further illustrated the importance of informativeness losses and the VQ-VIB architecture. With zero informativeness loss (i.e., $\lambda_I = 0$) one-hot agents often failed to converge to any useful communication (as indicated in prior art), but by increasing λ_I, agents consistently converged to useful communication. Indeed, with $\lambda_I = 50$ and $\lambda_C = 0$, one-hot agents learned to communication with complexity of up to 2.4 bits, greater than any complexity value found by Chaabouni et al. [10] using REINFORCE over 180 seeds (see Fig. 5 in that paper). Furthermore, we observed that VQ-VIB agents typically converged to higher informativeness (and higher complexity) communication than one-hot when complexity was not penalized, indicating that further investigation of inductive biases

is a promising direction for future work. Overall, this experiment largely replicates our findings from the main paper.

Thus, findings from our REINFORCE experiments indicate that the main conclusions of our paper hold true across several training mechanisms, and furthermore that our framework (including in particular the informativeness loss) can address issues already noted in prior work.

F.2 Common Ground Color Reference Game

Here, we include the full results from our Common Ground color reference game experiments. Crucially, these experiments tested agent behavior when utility and informativeness were not identical: utility could be maximized by the speaker only communicating about the target color, whereas informativeness was maximized by communicating about both target and distractor. Given the lack of human-labeled data for such a setting, we were unable to compare agent communication with human languages. However, by analyzing communication when varying the target for a fixed distractor, or by varying the distractor for a fixed target, we established that agents conditioned communication both on the target and distractor colors.



(a) Informativeness and Complexity for a gray distractor

(b) Communication by the same agent conditioned on purple or orange distractors differed, showing context-dependent communication.

Figure F.4: In Common Ground experiments, both VQ-VIB and one-hot communication demonstrated high informativeness for a fixed distractor (left), and varied communication according to distractor (right).

We trained agents in the Common Ground setting by randomly sampling target and distractor colors, as in the Basic setting, but here the speaker observed noisy versions of the distractor as well as the target. Based on insights that training to increase informativeness produced more complex communication, we trained agents for 30 episodes with $\lambda_I = 1.0$. For the next 70 episodes, we annealed the complexity weight, λ_C, using the same rates as in the Basic setting; this produced communication at different levels of complexity. As in the Basic setting, we found that VQ-VIB achieved higher utility, for the same complexity, than one-hot (Figure F.5).

We further analyzed communication in two ways: using a fixed distractor with varying targets and using a fixed target with varying distractors. First, for a fixed, gray distractor (#46 in the WCS data), we measured the standard metrics of informativeness and complexity. That is, we set the distractor color to gray, and we recorded, for each of the other 329 colors as the target, the speaker's communication. We then performed the same complexity and informativeness analysis as in the Basic setting, and found that both VQ-VIB and one-hot communication remained near-optimally efficient, as shown in Figure F.4 a.

Second, we measured whether agents conditioned their communication on the distractor color. Unlike the prior evaluation, in which we varied the target color for a fixed distractor, here, we evaluated communication for different distractors. Plots of a VQ-VIB agent's naming scheme, using either an orange (color #212) or purple (color #325) distractor are shown in Figure F.4 b.

As expected, the agent displayed different resolutions in regions of color depending upon the distractor. Specifically, for an orange distractor, the speaker had finer resolution around shades of orange and yellow; conversely, the speaker had finer resolution around shades of purples for a purple distractor.

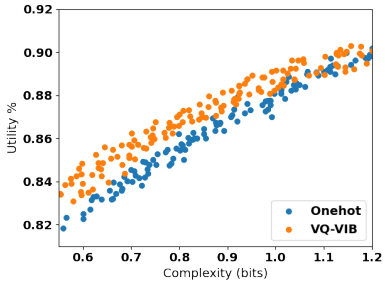


Figure F.5: In the Common Ground color reference game (as well as the Basic reference game, presented in the main paper), VQ-VIB achieved greater utility, for the same complexity, than one-hot communication.

This aligns with expectations that, when trying to differentiate between colors, one must use more complex or accurate words to distinguish similar colors.

Jointly, our findings indicate that many of the themes identified in the Basic setting – nearly-optimal communicative efficiency, the importance of annealing λ_C to generate varied complexity, and the advantages of VQ-VIB over one-hot – held in the Common Ground setting as well.