# A Lewis Games Loss Decomposition : proofs

We provide all the proofs of the Lewis Games Loss Decomposition. We organize the proofs as follows:

- **Appendix A.1 - Reconstruction game**, we provide the proofs of the Decomposition for the reconstruction game.
  - **Appendix A.1.1 - log-likelihood reward**, we first prove the loss decomposition when the reward is the reconstruction log-likelihood (case of the main paper).
  - **Appendix A.1.2 - general reward**, we then extend the decomposition to a more general reward
- **Appendix A.2 - Extension to Lewis games**, we extend the Loss Decomposition to a more general class of Lewis games. We first describe the additional formalism (Appendix A.2.1), then we prove the decomposition when the reward is the listener's log-likelihood (Appendix A.2.2) and when the reward is more general (Appendix A.2.3). Eventually, we show how the classic discrimination game can be expressed under this formalism in Appendix A.2.4.
- **Appendix A.3 - Extension to agents optimizing different rewards**, we discuss how the decomposition is affected when the agents optimize different rewards.

## A.1 Proof of the Lewis Reconstruction Game Loss Decompositon

Let's first recall some notations that we will use throughout the proofs. We consider two agents: a speaker parameterized by $\theta$ and a listener parameterized by $\phi$. In the reconstruction game, the speaker observes objects denoted by $x$ and taken from a set $\mathcal{X}$. The random variable characterizing the object is denoted by $X$ and its distribution is denoted by $p$. Based on object $x$, the speaker then sends a message $m$ from a message space $\mathcal{M}$ according to its policy $\pi_\theta(.|x)$. The random variable $M_\theta$ characterizes the message that is sampled from the speaker's policy $\pi_\theta$. Eventually, the listener should reconstruct the original object $x$ based on the message $m$. The probability that the listener predicts the input $x$ given a message $m$ is denoted by $\rho_\phi(x|m)$.

For any probability distribution, we denote by $\mathrm{Supp}$ the support of the distribution.

In the reconstruction game, the two agents optimize the same loss:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[r_\phi(x,m)]$$

We will first prove the decomposition in the case where $r_\phi(x,m) = \log \rho_\phi(x|m)$ (reconstruction log-likelihood) for all $x$ and $m$ and then for a more general form of reward.

### A.1.1 Proof of the Decomposition when $r_\phi(x,m) = \log \rho_\phi(x|m)$

We first prove the decomposition in the case described in the main paper: $r_\phi(x,m) = \log \rho_\phi(x|m)$ for all $x$ and $m$.

**Optimal listener** For completeness, we recall the proof of Equation 2 of the expression of the listener that is optimal with respect to $\mathcal{L}_{\theta,\phi}$.

In the case $r_\phi(x,m) = \log \rho_\phi(x|m)$, the listener is optimizing a cross-entropy loss with respect to the joint variable $(X, M)$ where $X$ follows $p$ and $M$ follows speaker's policy $\pi_\theta$. The loss can be rewritten as:

$$\begin{aligned}\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho_\phi(x|m)] \\ \mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[\log \rho_\phi(x|m)]\end{aligned}$$

According to Gibbs inequality, the optimal distribution $\rho_{\phi^*}(\cdot|m)$ for all $m$ is $\rho_{\phi^*}(.|m) = \rho^{*(\theta)}(.|m)$ where $\rho^{*(\theta)}(.|m)$ is the speaker's posterior distribution with respect to the prior $p$ and the conditional distribution $\pi_\theta(\cdot|x)$:

$$\rho^{*(\theta)}(x|m) := \frac{p(x)\pi_\theta(m|x)}{\mathbb{E}_{x'\sim p}[p(x')\pi_\theta(m|x')]} \qquad \text{for all} \quad x \in \text{Supp}(p), m \in \text{Supp}(\pi_\theta(\cdot|x))$$

This concludes the proof of Equation 2.

**Loss Decomposition** The idea of the proof is to decompose the reward into the optimal reward (when the listener is optimal), denoted by $r^{*(\theta)}(x, m)$, and the residual that measures the optimality gap, denoted by $r_\phi(x, m) - r^\theta(x, m)$:

$$r_\phi(x, m) = r^{*(\theta)}(x, m) + (r_\phi(x, m) - r^{*(\theta)}(x, m)) \qquad \text{for all} \qquad x \in \text{Supp}(p), m \in \text{Supp}(\pi_\theta(.|x))$$

Due to the linearity of the expectation, it follows that:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r^{*(\theta)}(x, m)] - \mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r_\phi(x, m) - r^{*(\theta)}(x, m)]$$

In the case where the reward is taken as the listener's log-likelihood, we have:

$$\begin{aligned}
\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r^{*(\theta)}(x, m)] - \mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r_\phi(x, m) - r^{*(\theta)}(x, m)] \\
&= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho^{*(\theta)}(x|m)] - \mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}\left[\log \frac{\rho_\phi(x|m)}{\rho^{*(\theta)}(x|m)}\right] \\
&= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho^{*(\theta)}(x|m)] - \mathbb{E}_{m\sim\pi_\theta}\mathbb{E}_{x\sim\rho^{*(\theta)}(.|m)}\left[\log \frac{\rho_\phi(x|m)}{\rho^{*(\theta)}(x|m)}\right] \\
\mathcal{L}_{\theta,\phi} &= \underbrace{\mathcal{H}(X|M_\theta)}_{\mathcal{L}_{\text{info}}} + \underbrace{\mathbb{E}_{m\sim\pi_\theta} D_{KL}(\rho^{*(\theta)}(\cdot|m)||\rho_\phi(\cdot|m))}_{\mathcal{L}_{\text{adapt}}}
\end{aligned}$$

where $\mathcal{H}(X|M_\theta)$ is the conditional entropy of $X$ conditioned on $M_\theta$ and $D_{KL}(p||q)$ is the Kullback-Leiber divergence between two distributions $p$ and $q$.

This last computation concludes the proof of Equation 5.

**Remarks** The key ingredients of the loss decomposition are:

1. We isolate two sub losses: $\mathcal{L}_{\text{info}}$, independent from the listener ; $\mathcal{L}_{\text{adapt}}$ optimized both by the speaker and the listener.
2. $\mathcal{L}_{\text{info}}$ measures the degree of ambiguity in the communication protocol. If $\mathcal{L}_{\text{info}}$ is optimal, ie. $\mathcal{L}_{\text{info}} = 0$, messages are unambiguous: each message refers to a unique input. Otherwise, $\mathcal{L}_{\text{info}} > 0$ and ambiguities remain.
3. $\mathcal{L}_{\text{adapt}}$ measures the gap between the listener and its optimum (here the speaker's posterior distribution). When the listener is optimal, $\mathcal{L}_{\text{adapt}} = 0$ and the main loss is limited to its information part, otherwise $\mathcal{L}_{\text{adapt}} > 0$ and the speaker and listener should adapt to reduce the optimality gap.

### A.1.2 Decomposition with a General Reward

In order to generalize the loss decomposition to more general rewards, we adopt the following strategy:

- **Construction of the reward**: we first need to build a general expression of the communication reward. To do so, we describe the conditions that the cooperative reward should fulfill in the reconstruction game and then propose a general reward expression. For the sake of generality, we consider that the environment $\mathcal{X}$ and message space $\mathcal{M}$ may be continuous spaces and that all the probability distribution may not be discrete.

- **Examples of usual cases**: we show that our proposed general expression covers the rewards used in most emergent communication papers, e.g. log-likelihood and accuracy.
- **Loss decomposition in the general case**: we write the loss decomposition with this general form of reward, showing that the key properties of the loss decomposition still hold.

**Construction of the reward** The Lewis reconstruction game is a cooperative game: the more the listener is able to reconstruct the objects seen by the speaker, the better the task is solved both by the speaker and the listener. Therefore the reward of the Lewis reconstruction game should respect the following conditions:

- **C1**: For $x \in \text{Supp}(p)$ and $m \in \text{Supp}(\pi_\theta(.|x))$, the expected reward $r_\phi(x, m)$ is maximum when $\rho_\phi(.|m) = \mathbf{1}_x$, where ie. $\mathbf{1}_x$ denotes the indicator function on $\mathcal{X}$ taken on $x$: the listener predicts $x$ with probability 1 when it receives $m$.
- **C2**: For $x \in \text{Supp}(p)$ and $m \in \text{Supp}(\pi_\theta(.|x))$, the expected reward $r_\phi(x, m)$ is sub-optimal when $\rho_\phi(.|m) \neq \mathbf{1}_x$, ie., the listener has a non-negative probability to predict the wrong object $x' \neq x$.

Given these assumptions, we propose the following general reward expression:

$$r_\phi(x, m) = -D(\mathbf{1}_x||\rho_\phi(.|m)) + K \tag{11}$$

where $D$ is such that $D(p||q) = 0$ iff $p = q$,$D(p||q) > 0$ otherwise. $\mathbf{1}_x$ is the indicator function on $\mathcal{X}$ taken in $x$ and $K$ is a real number that fixes the highest value of the reward. Note that $D(p||q)$ is close to a divergence, but has less assumptions.

**Usual rewards as special instances of the general expression**

We show that Equation 11 recovers most rewards used in the emergent communication literature, specifically:

- **Reconstruction log-likelihood** [9, 10, 40, 68, 67, 11] This case is the one used in the main paper and which is standarly used in reconstruction settings. With the following parameters
  - $D(p||q) = D_{KL}(p||q)$,
  - $K = 0$·

  we have:

$$r_\phi(x, m) = -D_{KL}(\mathbf{1}_x||\rho_\phi(.|m)) = \log \rho_\phi(x|m)$$

- **Accuracy** [65, 53, 52, 48, 56, 28, 23] Accuracy is the most commonly used reward in the emergent communication literature. It corresponds to agents receiving reward 1 if the prediction sampled according to the listener's output probability $\rho_\phi(\cdot|m)$ matches the original object $x$. The pointwise accuracy depends on the specific sample drawn from the listener's distribution. We are interested in how good the listener's prediction is on average, and thus in the expected accuracy, which is expressed as in Equation 11 with:
  - $D(p||q) = 1 - \mathbb{E}_p[q]$
  - $K = 1$

  The expected accuracy is then defined as:

$$r_\phi(x, m) = D(\mathbf{1}_x||\rho_\phi(.|m)) = \rho_\phi(x|m)\cdot$$

**Loss Decomposition**

With this definition of the reward, the speaker and listener loss can be written:

$$\begin{aligned}\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[r_\phi(x|m)] \\ &= -K + \mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_x||\rho_\phi(.|m))]\end{aligned}$$

We first need to define the optimal listener. Note that the expectation can be re-formulated:

$$\mathcal{L}_{\theta,\phi} = -K + \mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(\cdot|m)}[D(\mathbf{1}_x||\rho_\phi(.|m))]$$

The optimal listener is the listener $\rho_\phi$ that minimizes $\mathbb{E}_{x\sim\rho^{*(\theta)}(\cdot|m)}[D(\mathbf{1}_x, \rho_\phi(.|m))]$ for all $m$. In the general case, there is no close-formed expression of the optimal listener. The optimal listener policy is dependent of the function $D$ and the posterior distribution $\rho^{*(\theta)}(\cdot|m)$. We next denote the optimal listener policy $\rho_{\phi^*}(\cdot|m)$ that is fully characterized by $\rho^{*(\theta)}(\cdot|m)$ and $D$ and is **independent** of $\phi$.

As in Appendix A.1.1, we denote $r^{*(\theta)}(x, m)$ the reward of the optimal listener. We can then apply the same reward decomposition as in Appendix A.1.1:

$$r_\phi(x, m) = r^{*(\theta)}(x, m) + (r_\phi(x, m) - r^{*(\theta)}(x, m)) \qquad \text{for all} \qquad x \in \text{Supp}(p), m \in \text{Supp}(\pi_\theta(.|x))$$

which is equal to:

$$r_\phi(x, m) = -D(\mathbf{1}_x||\rho_{\phi^*}(.|m)) - [D(\mathbf{1}_x||\rho_\phi(.|m)) - D(\mathbf{1}_x||\rho_{\phi^*}(.|m))] - K$$

where $\rho_{\phi^*}(.|m)$ is the optimal listener distribution that is independent of $\phi$.

We can then rewrite the loss by taking the expectation of this reward and isolate an information and co-adaptation component:

$$\begin{aligned}
\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r^{*(\theta)}(x, m)] - \mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r_\phi(x, m) - r^{*(\theta)}(x, m)] \\
&= \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m))]}_{\mathcal{L}_{\text{info}}} \\
&+ \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_\phi(.|m)) - D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m))]}_{\mathcal{L}_{\text{adapt}}} - K
\end{aligned}$$

To be an information/co-adaptation decomposition, this loss decomposition should fulfill the following conditions:

1. $\mathcal{L}_{\text{info}}$ should be independent from the listener's weight $\phi$; $\mathcal{L}_{\text{adapt}}$ should be optimized both by the speaker and the listener.
2. $\mathcal{L}_{\text{info}}$ should be optimal ($\mathcal{L}_{\text{info}} = 0$) when the communication protocol is unambiguous, ie. each message refers to a unique input, sub-optimal ($\mathcal{L}_{\text{info}} > 0$) otherwise.
3. $\mathcal{L}_{\text{adapt}}$ should be 0 when the listener matches its optimum value with respect to the current object-message joint distribution, otherwise $\mathcal{L}_{\text{info}} > 0$ and the speaker and listener should adapt to reduce the optimality gap.

Let's prove that all those conditions hold:

1. The optimal listener policy $\rho_{\phi^*}(\cdot|m)$ is independent of $\phi$. It turns out that $\mathcal{L}_{\text{info}}$ is independent from the listener. On the contrary, $\mathcal{L}_{\text{adapt}}$ is dependent both on $\theta$ and $\phi$ and therefore is optimized both by the speaker and listener.
2. Let first show that when $\mathcal{L}_{\text{info}} = 0$, the speaker language is unambiguous. The language is considered unambiguous iff each message refers to a unique input. Formally, let $x$ be in the support of $p$ and

   $$\mathcal{M}_x = \{m \in \mathcal{M} \quad | \quad \rho^{*(\theta)}(x|m) > 0\},$$

   be the set of messages referring to $x$.

This set is non empty because $\mathbb{E}_{m\sim\pi_\theta}[\rho^{*(\theta)}(x|m)] = p(x) > 0$. The emergent language is considered unambiguous iff for all $x$ and $x'$ in the support of $p$:

$$x \neq x' \Rightarrow \mathcal{M}_x \cap \mathcal{M}_{x'} = \emptyset,$$

This property is equivalent of having a speaker posterior distribution $\rho^{*(\theta)}(.|m)$ being a Dirac distribution for all $m$ (otherwise, there is at least one message that refers to more than one object).

Let's demonstrate that $\mathcal{L}_{\text{info}} = 0$ iff $\rho^{*(\theta)}(.|m)$ is a Dirac distribution for all $m$.

First, when the speaker's posterior distribution is not a Dirac distribution, we have: $\mathcal{L}_{\text{info}} > 0$. Let $m$ be a message in the support of $\pi_\theta$. If $\rho^{*(\theta)}(\cdot|m)$ is not a Dirac distribution, there exists $x$ such that $\rho^{*(\theta)}(x|m) > 0$ and $D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m)) > 0$. Indeed, if there exists $x'$ such that $D(\mathbf{1}_{x'}||\rho_{\phi^*}(\cdot|m)) = 0$, we have $\rho_{\phi^*}(\cdot|m) = \mathbf{1}_{x'}$ by definition of $D$ and thus: if $x \neq x' \quad \Rightarrow \quad D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m)) = D(\mathbf{1}_x||\mathbf{1}_{x'}) > 0$ by definition of $D$. It implies that when $\rho^{*(\theta)}(\cdot|m)$ is not a Dirac distribution : $\mathbb{E}_{x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m))] > 0$.

Reciprocally, if for all $m \in \text{Supp}(\pi_\theta)$, $\rho^{*(\theta)}(\cdot|m)$ is a Dirac distribution: $\rho^{*(\theta)}(\cdot|m) = \mathbf{1}_{x_m}$ (with $m$ referring to $x_m$ and all $x \in \text{Supp}(p)$ covered by the messages) and the corresponding optimal listener is also the Dirac distribution $\rho_{\phi^*}(\cdot|m) = \mathbf{1}_{x_m}$, we have:

$$\mathcal{L}_{\text{info}} = \mathbb{E}_{m\sim\pi_\theta}[D(\mathbf{1}_{x_m}||\rho_{\phi^*}(\cdot|m))] = \mathbb{E}_{m\sim\pi_\theta}[D(\mathbf{1}_{x_m}||\mathbf{1}_{x_m})] = 0$$

Therefore, $\mathcal{L}_{\text{info}}$ is equal to 0, ie. is minimum, if and only if the speaker has a posterior which is Dirac distribution, ie. the speaker develops an unambiguous language.

3. When the listener is optimal with respect to its loss, $\rho_\phi(.|m) = \rho_{\phi^*}(.|m)$ for all $m$ and as a direct consequence, $\mathcal{L}_{\text{adapt}} = 0$. When the listener is not optimal with respect to its loss, $\mathcal{L}_{\text{adapt}} > 0$ by definition of the optimal listener which is the listener that minimizes $\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(\cdot|m)}[D(\mathbf{1}_x||\rho_\phi(.|m))]$.

In conclusion, in the case of a general reward, we keep the main ingredients of the information/co-adaptation decomposition.

### A.2 General Proof of the Lewis Games Loss Decomposition

In the previous section, we provided a proof of the loss decomposition for the Lewis Reconstruction Game with a general cooperative reward. The goal of this Section is to extend this decomposition to a more general definition of Lewis Games:

- **Appendix A.2.1 - Formalism**: We first describe the additional formalism.
- **Appendix A.2.2 - Log-likelihood reward**: We prove the decomposition for the general Lewis Game when the reward is the listener's log-likelihood.
- **Appendix A.2.4 - General cooperative reward** We prove the decomposition for the general Lewis Game with a general cooperative reward.
- **Appendix A.2.4 - Discrimination game** : Eventually, we show how the widely studied discrimination game [12, 60, 21, 30, 65, 53, 52, 33, 56, 58] can be expressed under this formalism.

#### A.2.1 Formalism

In the general form, we consider inputs $x$ from a set $\mathcal{X}$ where $x$ is drawn from $p_X$. We consider a random feature $F$ of $X$ (in the reconstruction game $F = X$) that is distributed following $p_F(.|X)$. A draw of $F$ is denoted $f$ and the set of potential features $\mathcal{F}$. We here consider that the listener may have access to an auxiliary input $y$. We denote $Y$ the random variable of this auxiliary input and $p_Y(\cdot|X, F)$ its probability distribution. The task is here the communication of the feature $f$. To this end, the speaker still sends messages $m$ from the message space $\mathcal{M}$. The random variable $M_\theta$ characterizes the messages that are sampled from the speaker's policy $\pi_\theta(\cdot|X)$. Eventually, the probability that the listener predicts the correct feature $f$, given message $m$ and auxiliary features $y$ is denoted by $\rho_\phi(f|m, y)$.

### A.2.2 Proof with the log-likelihood reward: $r_\phi(f, m, y) = \log \rho_\phi(f|m, y)$

We first prove the decomposition in the case: $r_\phi(f, m, y) = \log \rho_\phi(f|m, y)$ for all $f$,$m$ and $y$, ie. the reward is the listener's log-likelihood of predicting the good feature. The agents' loss becomes

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(f|m, y)]\cdot$$

**Optimal listener** The optimal listener is the listener that optimally minimizes $L_{\theta,\phi}$ for a fixed speaker policy $\pi_\theta$. It is obtained by noting that:

$$\begin{aligned}\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(f|m, y)]\\ \mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{(m,y)\sim p_{M_\theta,Y}}\mathbb{E}_{f\sim\rho^{*(\theta)}(\cdot|m,y)}[\log \rho_\phi(f|m, y)]\cdot\end{aligned}$$

where $\rho^{*(\theta)}(f|m, y) = \frac{\mathbb{E}_{x\sim p_X}[\pi_\theta(m|x)p_F(f|x)p_Y(y|f,x)]}{\mathbb{E}_{x\sim p_X, f\sim p_F(.|x)}[\pi_\theta(m|x)p_Y(y|f,x)]}$ for all $f$, $m$ and $y$ and $p_{M_\theta,Y}(m, y) = \mathbb{E}_{x\sim p_X, f\sim p_F(.|x)}[\pi_\theta(m|x)p_Y(y|f, x)]$.

It follows from Gibbs inequality that the optimal listener is $\rho^{*(\theta)}(\cdot|m, y)$ for all $m$ and $y$.

We can apply the reward decomposition of Appendix A.1.1:

$$r_\phi(x, m) = r^{*(\theta)}(f, m, y) + (r_\phi(f, m, y) - r^{*(\theta)}(f, m, y)) \qquad \text{for all} \qquad f \in \mathcal{F}, m \in \mathcal{M}, y \in \mathcal{Y}\cdot$$

Plugging this decomposed reward in our loss, and applying the exact same steps as in Appendix A.1.1, we get

$$\mathcal{L}_{\theta,\phi} = \underbrace{\mathcal{H}(F|M_\theta, Y)}_{\mathcal{L}_{\text{info}}} + \underbrace{\mathbb{E}_{(m,y)\sim p_{M_\theta,Y}} D_{KL}(\rho^{*(\theta)}(\cdot|m, y)||\rho_\phi(\cdot|m, y))}_{\mathcal{L}_{\text{adapt}}} \tag{12}$$

which is the Loss Decomposition for a general game.

**Remarks** You note that the decomposition is close to the Loss Decomposition in the reconstruction case (Equation 5). Indeed, since the listener should predict a given feature $F$, the information task is to build an unambiguous message protocol with respect to this feature and the optimal listener becomes the posterior distribution of the speaker with respect to this feature. The co-adaptation loss is once again a Kullback-Leiber distribution between the listener and the speaker's posterior. $\mathcal{L}_{\text{info}}$ and $\mathcal{L}_{\text{adapt}}$ respects the conditions states in Appendix A.1.1.

### A.2.3 Proof with the general reward $r_\phi(f, m, y) = -D(\mathbf{1}_f||\rho_\phi(.|m, y)) + K$

To study the general case, we use the reward definition provided in Appendix A.1.2:

$$r_\phi(f, m, y) = -D(\mathbf{1}_f||\rho_\phi(.|m, y)) + K$$

where $D(p||q)$ is a function that is null when $p = q$, greater than 0 otherwise, $\mathbf{1}_f$ the indicator function on $\mathcal{F}$ taken in $f$ and $K$ is a real number that fixes the highest value of the reward.

Agents' loss becomes:

$$\begin{aligned}\mathcal{L}_{\theta,\phi} &= -\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[r_\phi(f, m, y)]\\ \mathcal{L}_{\theta,\phi} &= \mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_f||\rho_\phi(.|m, y))] - K\end{aligned}$$

Denoting $\rho_{\phi^*}(.|m, y)$ the listener that optimally minimises $\mathcal{L}_{\theta,\phi}$ and $r^\theta(f, m, y)$ the reward of the optimal listener, the loss can be decomposed:

$$\mathcal{L}_{\theta,\phi} = -\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[r^{*(\theta)}(f,m,y)]$$
$$- \mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[r_\phi(f,m,y) - r^{*(\theta)}(f,m,y)]$$
$$\mathcal{L}_{\theta,\phi} = \underbrace{\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_f||\rho_{\phi^*}(.|m,y))]}_{\mathcal{L}_{\text{info}}}$$
$$+ \underbrace{\mathbb{E}_{x\sim p_X, f\sim p_F(\cdot|x), y\sim p_Y(\cdot|x,f), m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_f||\rho_\phi(.|m,y)) - D(\mathbf{1}_f||\rho_{\phi^*}(.|m,y))]}_{\mathcal{L}_{\text{adapt}}} - K$$

For the same arguments as in Appendix A.1.2, $\mathcal{L}_{\text{info}}$ is only optimized by the speaker and is optimal when the speaker develops an unambiguous message protocol with respect to $F$ given $Y$, $\mathcal{L}_{\text{adapt}}$ is null when the listener is optimal, otherwise it is $> 0$, ie. sub-optimal. Therefore, we recover the key ingredients of the Loss Decomposition: when the listener is optimal, speaker's loss is limited to $\mathcal{L}_{\text{info}}$, when the listener is not optimal, the speaker has the additional task to help the listener matching its optimum.

### A.2.4 Case of the Discrimination Game

Recall that in a discrimination game, as in a reconstruction game, the speaker observes an input, $x$ and sends a message $m$ to the listener. The listener is then provided with both the message $m$, and a list of $N+1$ candidate inputs, containing input $x$, along with $N$ other inputs, or *distractors*. The goal of the listener is then to give the index of the candidate that corresponds to the actual input.

To formally define discrimination games as instances of the general Lewis game described above, we define $X_1, \dots, X_N$ to be i.i.d. samples from the inputs distribution $p$. These inputs will be used as the distractors. We additionally set $X_0 = X$. We then define a random permutation $\Sigma$, drawn uniformly from the set of $N+1$ element permutations, and independently from all other random variables. We then set our auxiliary input $Y = (X_{\Sigma(0)}, \dots, X_{\Sigma(N)})$, which provides the listener with a permuted list, containing both the correct input at a random position, as well as the distractors. Finally, we set the feature to be predicted as $F = \Sigma^{-1}(0)$. The task of the listener becomes to identify the index of the correct input among all distractors, and we recover a discrimination game.

## A.3 Speaker and Listener Optimizing Different Rewards

In this paper, we only discuss the case where the agents are fully cooperative, ie. they are optimizing exactly the same reward. When the agents are not aligned on the same objective, the system should be decoupled and an additional *alignement bias* is added to the loss of the speaker. For example, in the reconstruction game where the speaker is optimizing a general reward $r_\phi(x,m) = -D(\mathbf{1}_x, \rho_\phi(\cdot|m)) + K$ and the listener a cross-entropy loss, the system becomes:

$$\begin{cases} \mathcal{L}_\theta &= \mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[D(\mathbf{1}_x, \rho_\phi(\cdot|m))] - K \\ \mathcal{L}_\phi &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(\cdot|x)}[\log \rho_\phi(x|m)]. \end{cases} \tag{13}$$

where $\mathcal{L}_\theta$ is the speaker's loss and $\mathcal{L}_\phi$ the listener's loss.

By denoting $\rho_{\phi^*}(\cdot|m)$ the optimal listener for all $m$ with respect to $\mathcal{L}_\theta$ (which is fully determined by the speaker's posterior and $D$) and $\rho^{*(\theta)}(\cdot|m)$ the optimal listener for all $m$ with respect to $\mathcal{L}_\phi$ (in this case, the speaker posterior), the speaker loss now decomposes into:

$$\mathcal{L}_\theta = \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho^{*(\theta)}(\cdot|m))]}_{\mathcal{L}_{\text{info}}} + \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_\phi(.|m)) - D(\mathbf{1}_x||\rho_{\phi^*}(\cdot|m))]}_{\mathcal{L}_{\text{adapt}}}$$
$$+ \underbrace{\mathbb{E}_{m\sim\pi_\theta, x\sim\rho^{*(\theta)}(.|m)}[D(\mathbf{1}_x||\rho_{\phi^*}(.|m)) - D(\mathbf{1}_x||\rho^{*(\theta)}(\cdot|m))]}_{\text{alignment bias}} - K$$

Compared to the standard decomposition, there is an additional term, that we name the *alignment bias*, linked to the gap between the listener optimum of $\mathcal{L}_\theta$ and the listener optimum of $\mathcal{L}_\phi$. If those

optima are close, the amplitude of this term is negligible compared to $\mathcal{L}_{\text{info}}$ and $\mathcal{L}_{\text{adapt}}$. If those optima are very different (eg. competitive game), the information and co-adaptation terms could have a significantly smaller amplitude compared to the *alignment bias*. We leave to future work the theoretical study of this *alignment bias* which echoes some empirical studies [61].

# B Method: Additional Computations

In Section 3.2, we propose a protocol to balance the importance of the information and co-adaptation losses in the speaker's training loss. To do so, we use the probe listener's estimate of the speaker's posterior on the train set $\rho^{\text{train}}_{\omega^*}(x|m) = \log \rho^{\text{train}}_{\omega^*}(x|m)$ and build the following reward:

$$r_\phi(x, m; \alpha) = (1 - 2\alpha) \times \underbrace{\log \rho^{\text{train}}_{\omega^*}(x|m)}_{\text{probe listener reward}} + \alpha \times \underbrace{\log \rho_\phi(x|m)}_{\text{standard listener reward}}$$

where $\alpha$ is a weight in $[0; 0.5]$.

The loss equality defined in Section 3.2 is then recovered with the following computations:

$$\begin{aligned}
\mathcal{L}_\theta(\alpha) &= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[r_\phi(x, m; \alpha)] \\
&= -\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[(1 - 2\alpha) \times \log \rho^{\text{train}}_{\omega^*}(x|m) \ + \ \alpha \times \log \rho_\phi(x|m)] \\
&= -(1 - 2\alpha)\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho^{\text{train}}_{\omega^*}(x|m)] - \alpha\mathbb{E}_{x\sim p, m\sim\pi_\theta(.|x)}[\log \rho_\phi(x|m)] \\
&= (1 - 2\alpha)\hat{\mathcal{L}}^{\text{train}}_{\text{info}} + \alpha(\hat{\mathcal{L}}^{\text{train}}_{\text{info}} + \hat{\mathcal{L}}^{\text{train}}_{\text{adapt}}) \\
\mathcal{L}_\theta(\alpha) &= (1 - \alpha)\hat{\mathcal{L}}^{\text{train}}_{\text{info}} + \alpha\hat{\mathcal{L}}^{\text{train}}_{\text{adapt}}
\end{aligned}$$

**Remark** In the paper, we only consider the case $\alpha \in [0; 0.5]$ and do not explore larger values of $\alpha$. Indeed, controlling the co-adaptation rate $\alpha$ is made by re-weighting $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ (estimated with a probe listener). However, two issues occur when $\alpha > 0.5$:

- First, the goal of computing $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ is to indirectly balance the weight of the training information loss $\mathcal{L}^{train}_{\text{info}}$. By taking the loss of the probe listener close to optimality, we get an upper bound estimate $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ of the training information loss $\mathcal{L}^{\text{train}}_{\text{info}}$. Therefore, it theoretically ensures that we minimize $\mathcal{L}^{\text{train}}_{\text{info}}$ when optimizing $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$. However, when $\alpha > 0.5$, the weight of $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ is negative. In this case, since $\hat{\mathcal{L}}^{\text{train}}_{\text{info}}$ is an upper bound of $\mathcal{L}^{\text{train}}_{\text{info}}$, we do not have the guarantee that the speaker minimizes $-\mathcal{L}^{\text{train}}_{\text{info}}$ anymore.
- Second, we empirically experimented $\alpha > 0.5$ even if theoretical conditions are not reached. In practice, if the system converged for values of $\alpha$ closed to $0.5$, the system quickly became unstable for larger values of $\alpha$. Our main hypothesis is that the speaker cannot start structuring its messages when the weight of $\hat{\mathcal{L}}^{\text{train}}_{\text{adapt}}$ is too strong. Indeed, agents start with random weights. It implies that, at the beginning of the training, if the weight of $\hat{\mathcal{L}}^{\text{train}}_{\text{adapt}}$ is too strong, it pressures the speaker to have an almost uniform posterior, ie. to develop a fully ambiguous language. In short, if $\alpha$ is too large, the speaker has too little pressure on developing meaningful messages and therefore succeeding in the communication task.

# C Regularization

We here provide:

- the parameters used for the listener's regularization (Appendix C.1)
- the results obtained when regularizing the speaker (Appendix C.2)

## C.1 Parameters of the Listener's Regularization

Regularization parameters have been tuned in order to get the best average generalization scores while having a convergence success rate greater or equal to $75\%$. When regularizing with the layer

normalization (noted *No LN.* in Table 1), we remove the layer normalization applied of the listener's LSTM cell. Dropout rate is set to $0.2$ and weight decay penalty is set to $0.01$ both when layer normalization is kept (noted *Weight decay* in Table 1) and when layer normalization is removed (noted *No LN. + WD* in Table 1).

### C.2 Comparison with Speaker's Regularization

**Parameters** For the sake of completeness, we also study the impact of regularizing the speaker. Here, we only report the results with the weight decay penalty. Indeed, removing the layer normalization makes the training slow and unstable while results with dropout are worse than those with weight decay. Weight decay penalty has been fine-tuned to $0.005$ to get the best average generalization performances while having $> 75\%$ successful experiments.

**Results** In Table 2, we compare the generalization and compositionality of emergent languages with and without regularization applied on the speaker. First, when we regularize the speaker without any regularization on the listener, we see that the gain of generalization and compositionality is negligible and inferior to the gain obtained when regularizing the listener. Moreover, we note that when we regularize both the speaker and the listener, scores of generalization and compositionality are similar to those obtained when only regularizing the listener. It suggests that regularizing the speaker has little impact on generalization and compositionality.

These results support the claim of Section 5.2: the listener is the main contributor of the co-adaptation overfitting in the reconstruction game.

| **No Speaker reg.** | Gen. ↑ | Compo. ↑ |
|---|---|---|
| Continuous | $0.58_{\pm 0.05}$ | $0.22_{\pm 0.02}$ |
| No LN. | $0.70_{\pm 0.03}$ | $0.24_{\pm 0.02}$ |
| Weight decay | $0.72_{\pm 0.03}$ | $0.25_{\pm 0.03}$ |
| No LN. + WD | $0.87_{\pm 0.07}$ | $0.30_{\pm 0.03}$ |

| **Speaker with WD** | Gen. ↑ | Compo. ↑ |
|---|---|---|
| Continuous | $0.62_{\pm 0.02}$ | $0.22_{\pm 0}.03$ |
| No LN. | $0.68_{\pm 0.07}$ | $0.23_{\pm 0.01}$ |
| Weight decay | $0.74_{\pm 0.05}$ | $0.26_{\pm 0.04}$ |
| No LN. + WD | $0.82_{\pm 0.07}$ | $0.32_{\pm 0.04}$ |

Table 2: Performance comparison: (left) without speaker regularization ; (right) with speaker regularization. Weight decay penalty on the speaker is set to $0.005$. Parameters of regularization methods for the listener are reported in Appendix C.1.

## D Image Discrimination Games

We here complete Section 5.3 by presenting the rules and experimental settings of the image discrimination game (Appendix D.1), reporting the results of compositionality (Appendix D.2) and completing generalization results of Table 1 with regularization experiments (Appendix D.3).

### D.1 Experimental Settings

For the implementation of the image discrimination game, we mostly follow the protocol proposed by [12].

#### D.1.1 Game Rules and notations

In the Lewis image discrimination game, the speaker observes an image. Then, the speaker sends a descriptive message to the listener. Based on this message, the listener should retrieve the correct image among a set of candidates.

Formally, the image observed by the speaker is denoted by $x$ and belongs to a set $\mathcal{X}$. The intermediate message sent by the speaker is denoted by $m$ and belongs to a set a potential messages $\mathcal{M}$. The speaker follows a policy $\pi_\theta$ which samples a message $m$ with probability $\pi_\theta(m|x)$ conditioned on image $x$. The listener encodes the message $m$ into a representation $t_\phi(m)$. The set of candidates received by the listener are denoted $\mathcal{C}$ and the listener encodes each candidates $x' \in \mathcal{C}$ by a representation $t_\phi(x')$. The probability of a candidate $x'$ to be the correct image is : $\rho_\phi(x'|m, \mathcal{C})$. It is obtained by comparing the message encoding $t_\phi(m)$ with the image encoding $t_\phi(x')$ of all candidates.

### D.1.2 Environment

**Datasets** We perform the discrimination game on ImageNet [19, 69] and CelebA [57]. We work with image pre-processed encodings $f(x)$ of size 2048 that have been open-sourced by [12]. In the two datasets, each image has been center-cropped and processed by a ResNet-50 encoder pretrained on ImageNet with the self-supervised method BYOL [29].

**Train/val/test splits** For building our custom training sets, we first considered the splits provided by [12]. From the respective $1400k$ and $200k$ labelled images of ImageNet and CelebA, they slitted the dataset in train, validation and test with the ratio $80/10/10$.

To test agents generalization capacities, we also build subsets of the training set provided by [12]: ImageNet $\frac{1}{20}$, ImageNet $\frac{1}{100}$, CelebA $\frac{1}{20}$ and CelebA $\frac{1}{100}$. For each of those sub-training sets, we randomly selected a small fraction of the training set, approximatively corresponding to $1/20$-th and $1/100$-th of the total training set. The corresponding number of samples are reported in Table 3.

| | Training samples | | | Training samples | |
|---|---|---|---|---|---|
| **CelebA** | $1/20$ | $1/100$ | **ImageNet** | $1/20$ | $1/100$ |
| | 8492 | 2123 | | 50732 | 12683 |

Table 3: Number of training samples for the four training subsets considered: ImageNet $\frac{1}{20}$, ImageNet $\frac{1}{100}$, CelebA $\frac{1}{20}$ and CelebA $\frac{1}{100}$

All our experiments on images are run with those $4$ small training sets. We keep the original validation and test sets from [12].

### D.1.3 Agent Models

**Speaker model** The speaker is a neural network that takes the pre-processed representation of an image $f(x)$ as input of size 2048 and returns a message $m = (m_i)_{1 \leq i \leq T}$ of length $T$.

The speaker follows a recurrent policy: given the image representation $f(x)$, it samples for all $t \in [1, T]$ a token $m_t$ with probability $\pi_\theta(m_t | m_{<t}, f(x))$. The image representation $f(x)$ is first projected by a linear layer to get an object embedding of size 256 that is used to initialize a LSTM of size 256 with layer normalization. At each time step, the LSTM's output is fed into a linear layer of size $|\mathcal{V}|$, followed by a softmax, to produce $\pi_\theta(m_t | m_{<t}, f(x))$.

In our experiments, the following parameters have been chosen: $T = 10, |V| = 10$ meaning that the message space is of size $10^{10}$ preventing any channel capacity bottleneck.

**Listener model** The listener is a neural network that takes the speaker's message $m$ and a set of image candidates $\mathcal{C}$ containing the target image $x$ and outputs the probability for each candidate $x' \in \mathcal{C}$ to be the target image $x$.

The listener is composed of two modules: one that encodes the message ; the other that encodes images. For a message $m = (m_1, ..., m_T)$, the listener passes each symbol $m_t$ through an embedding layer of dimension 256 followed by a LSTM of size 256 with layer normalization. The final recurrent state $h_T^1$ is then passed to a linear layer that produces the image encoding $t_\phi(m)$ of size 256. In parallel, each candidate $x'$ is first pre-processed by $f$ and then passed through a linear layer producing an image encoding $t_\phi(x')$ of size 256.

The message representation $t_\phi(m)$ is then compared to each candidate representation $t_\phi(x')$ with the following score function: $\text{score}(m, x', \phi) := t_\phi(m) t_\phi(x')^T$. Note that contrary to [12], we rather use a dot-product score function [53] instead of a cosine similarity because we empirically got better results and more stable trainings. The probability distribution over the candidates $\mathcal{C}$ of being the target image $x$ is then obtained by normalizing the scores with a softmax. This probability distribution is denoted by $\rho_\phi(\cdot | m, \mathcal{C})$ and the listener guess is $\hat{x} = \underset{x'}{\text{argmax}} \rho_\phi(x' | m, \mathcal{C})$.

In our experiments, the number of candidates is $|\mathcal{C}| = 1000$.

### D.1.4 Agents Training

We follow the same principle as in the reconstruction game: the listener is trained to best predict the target image among the set of candidates, while the speaker takes the opposite of the listener's loss as reward:

**Listener loss** The listener is trained to predict the target image among the set of candidates $\mathcal{C}$. When receiving a batch of inputs $x$, a set of candidates $\mathcal{C}$ is sampled for each input $x$. The sampling is uniform without replacement over $\mathcal{X} - \{x\}$ meaning that the target image $x$ cannot be duplicated into the candidates. The listener is then trained to optimized the average InfoNCE loss [62]:

$$\mathcal{L}_\phi = \sum_{x \in \text{batch}} -\log \rho_\phi(x|m, \mathcal{C})$$

**Speaker loss** When the speaker observes an image $x$, sends a message $m$ and the listener has to choose among a set of candidates $\mathcal{C}$, the speaker's reward is defined as:

$$r_{\phi,\mathcal{C}}(x, m) = \log \rho_\phi(x|m, \mathcal{C})$$

The speaker is trained to maximize its cumulative reward: $\mathbb{E}_{x,m,\mathcal{C}}[\log \rho_\phi(x|m, \mathcal{C})]$ which means that the speaker and the listener have the same loss.

**Optimization** The agents are optimized using `Adam` [42] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The speaker's learning rate is $5 \cdot 10^{-4}$ while the listener's learning rate is $1 \cdot 10^{-3}$. Agents are trained on batches of size of 2048. For the speaker, we use policy gradient [76], with a baseline computed as the average reward within the minibatch, and we add an entropy regularization of 0.01 to the speaker's loss [82].

## D.2 Topographic Similarity Results

We report results of topographic similarity for experiments of Section 5.3. To be complete, we add the scores when applying listener regularization (corresponding generalization performances are reported in Appendix D.3).

Scores of topographic similarity are reported in Table 4. Here, the distance used to compare images is the cosine distance between the vector representations of the ResNet-50 encoder pretrained on ImageNet. The distance used to compare messages remains the edit-distance. As mentioned in the main paper, we can see that there is not any compositionality trend when agents communicate about images. Moreover, when comparing with Table 6 that reports generalization performances, we see that gains of generalization do not correlate with gains of topographic similarity. It suggests that the topographic similarity does not capture agents' language structure in image based settings, as already observed in previous work [12, 1].

| | Topographic similarity ↑ | | | Topographic similarity ↑ | |
|---|---|---|---|---|---|
| **CelebA** | 1/20 | 1/100 | **ImageNet** | 1/20 | 1/100 |
| Continuous | $\mathbf{0.28_{\pm 0.03}}$ | $\mathbf{0.32_{\pm 0.03}}$ | Continuous | $0.17_{\pm 0.03}$ | $0.17_{\pm 0.03}$ |
| No LN. | $0.26_{\pm 0.04}$ | $0.29_{\pm 0.03}$ | No LN. | $0.18_{\pm 0.01}$ | $0.16_{\pm 0.02}$ |
| No LN. + WD | – | $0.30_{\pm 0.03}$ | No LN. + WD | $\mathbf{0.19_{\pm 0.03}}$ | $\mathbf{0.21_{\pm 0.03}}$ |
| Weight decay | $0.27_{\pm 0.03}$ | $0.28_{\pm 0.04}$ | Weight decay | $0.17_{\pm 0.02}$ | $0.15_{\pm 0.02}$ |
| Early stopping | $0.27_{\pm 0.04}$ | $0.30_{\pm 0.03}$ | Early stopping | $0.18_{\pm 0.04}$ | $0.20_{\pm 0.03}$ |

Table 4: Topographic similarity of emergent languages in the image discrimination game where images are compared with a cosine similarity. *No LN.* refers to the removal of the layernorm on the listener's LSTM cell ; *Weight decay* to the addition of weight decay on the listener with penalty equal to 0.01 ; *No LN. + WD* refers to the removal of layernorm and addition of weight decay on the listener. No result for *No LN. + WD* are reported with Celeba $\frac{1}{20}$ because experiments did not converge with the regularization parameters chosen.

In addition, we also test whether scores of topographic similarities are improved when using another distance to compare images. In Table 5, we use the attributes provided in CelebA to compare the

images. The distance between two images is computed as $1 -$ (propotion of common attributes). For the message comparison, we keep the edit-distance. Once again, no topographic similarity trends emerge, sustaining results already observed in [12].

| | Topographic similarity (with attributes) ↑ | |
|---|---|---|
| **CelebA** | $1/20$ | $1/100$ |
| Continuous | $0.13_{\pm 0.02}$ | $\mathbf{0.15_{\pm 0.03}}$ |
| No LN. | $\mathbf{0.14_{\pm 0.02}}$ | $0.14_{\pm 0.03}$ |
| No LN. + WD | – | $\mathbf{0.15_{\pm 0.04}}$ |
| Weight decay | $\mathbf{0.14_{\pm 0.02}}$ | $\mathbf{0.15_{\pm 0.01}}$ |
| Early stopping | $0.13_{\pm 0.02}$ | $\mathbf{0.15_{\pm 0.02}}$ |

Table 5: Topographic similarity of emergent languages in the image discrimination game where images are compared with CelebA attributes. *No LN.* refers to the removal of the layer normalization on the listener's LSTM cell ; *Weight decay* to the addition of weight decay on the listener with penalty equal to $0.01$ ; *No LN. + WD* refers to the removal of layer normalization and addition of weight decay on the listener. No result for *No LN. + WD* are reported with Celeba $\frac{1}{20}$ because experiments did not converge with the regularization parameters chosen.

### D.3 More Results with Listener Regularization

To complete the generalization scores of Table 1 in the main paper, we report in Table 6 the generalization scores in the image discrimination game for various regularization methods applied on the listener. We observe the same trends as in the reconstruction game. Indeed, listener regularization consistently improves the performances. It means, that a large gain of performance can be obtained in those games by regularizing the listener. The *Early stopping listener* remains a top line in image based experiments.

| | Generalization ↑ | | | Generalization ↑ | |
|---|---|---|---|---|---|
| **CelebA** | $1/20$ | $1/100$ | **ImageNet** | $1/20$ | $1/100$ |
| Continuous | $0.67_{\pm 0.02}$ | $0.39_{\pm 0.07}$ | Continuous | $0.77_{\pm 0.01}$ | $0.51_{\pm 0.03}$ |
| No LN. | $0.67_{\pm 0.03}$ | $0.44_{\pm 0.02}$ | No LN. | $0.77_{\pm 0.01}$ | $0.53_{\pm 0.03}$ |
| No LN. + WD | – | $0.50_{\pm 0.07}$ | No LN. + WD | $0.75_{\pm 0.01}$ | $0.59_{\pm 0.04}$ |
| Weight decay | $0.77_{\pm 0.04}$ | $0.60_{\pm 0.06}$ | Weight decay | $0.79_{\pm 0.03}$ | $0.62_{\pm 0.02}$ |
| Early stopping | $\mathbf{0.80_{\pm 0.03}}$ | $\mathbf{0.69_{\pm 0.04}}$ | Early stopping | $\mathbf{0.81_{\pm 0.01}}$ | $\mathbf{0.64_{\pm 0.01}}$ |

Table 6: Comparison of generalization performances between the *Continuous listener*, *Early stopping listener* and listeners with regularization on the image discrimination game. *No LN.* refers to the removal of the layer normalization on the listener's LSTM cell ; *Weight decay* to the addition of weight decay on the listener with penalty equal to $0.01$ ; *No LN. + WD* refers to the removal of layer normalization and addition of weight decay on the listener. No result for *No LN. + WD* are reported with Celeba $\frac{1}{20}$ because experiments did not converge with the regularization parameters chosen.