Regularized Conventions: Equilibrium Computation as a Model of Pragmatic Reasoning

Athul Paul Jacob
apjacob@mit.edu

Gabriele Farina
gfarina@mit.edu

Jacob Andreas
jda@mit.edu

Meaning is fluid and context-sensitive: speakers can use the word *blue* to pick out a color that in other contexts would be described as *purple*, or identify a friend as *the one with glasses* in a room in which everyone is wearing glasses (Figure 1). Such context-dependent meanings can arise as **conventions** among language users communicating repeatedly to solve a shared task (Clark and Wilkes-Gibbs, 1986). But they can also arise *without any interaction at all*, among language users who share only common knowledge of words' default meanings (Grice, 1975).

What makes this kind of context-dependent, pragmatic language use possible? Most existing computational models of pragmatics are implemented as **recursive reasoning** procedures, in which listeners interpret utterances by reasoning about the intentions of less-sophisticated speakers (e.g. Degen, 2023). We present an alternative model of pragmatic understanding based on **equilibrium search**. In this model (which we call Regularized Conventions, or ReCo), speakers and listeners solve communicative tasks like those in Figure 1 by searching for utterance–meaning mappings that are both close to a game-theoretically optimal communicative convention (a **signaling equilibrium**), and close to a shared initial semantics (which functions as a **regularizer**). In Figure 1, for example, ReCo assigns high probability to the use of *blue* to signal the intended color, and low (but nonzero) probability to the use of *purple* instead. This strategy is close to one of many optimal conventions (in which every utterance arbitrarily, but uniquely, picks out one color), and close to color terms' standard interpretation (in which the target color is improbably, but not impossibly, described as *blue*).

Formally, ReCo begins by modeling communication as a signaling game (Lewis, 1971), which features two players: the SPEAKER and the LISTENER. In this game, a **target meaning** (representing a communicative need) is first sampled from a space of possible meanings $m \in M$ with probability $p(m)$. To communicate this meaning, the SPEAKER produces an **utterance** $u \in U$ according to a policy $\pi_S(u \mid m)$. Finally, the LISTENER produces an **interpretation** according to a policy $\pi_L(m' \mid u)$. During communication, both players receive a **reward** if the interpretation matches the target meaning: $r(m, u, m') = \mathbf{1}[m' = m]$. Then, the *expected utility* of each player given policies (π_S, π_L) is defined as the expected reward when the meanings m are sampled from a prior distribution $p(m)$, and agents sample from their policies:

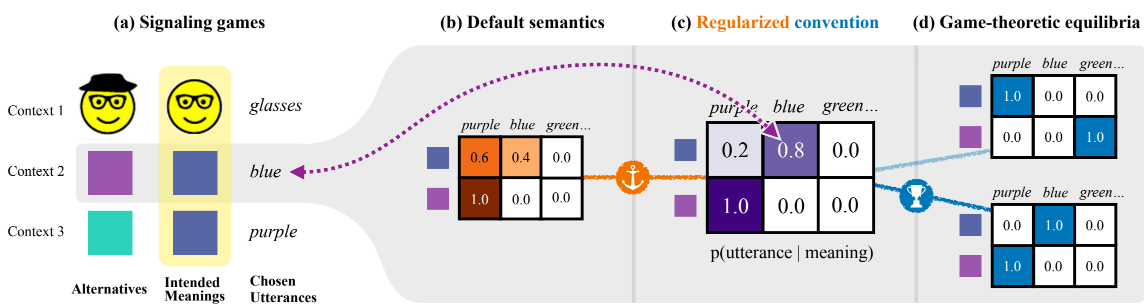

Figure 1: The ReCo model. To communicate (or resolve) an intended meaning from a set of possibilities **(a)**, language users search for distributions over utterances and interpretations that are close to some "default semantics" **(b)** and close to a (game-theoretically) optimal signaling convention **(d)**. The resulting "regularized conventions" **(c)** predict human judgments on a variety of pragmatic implicature tasks.

Proceedings of the Society for Computation in Linguistics (SCiL) 2024, pages 350-351.
Irvine, California, June 27-29, 2024

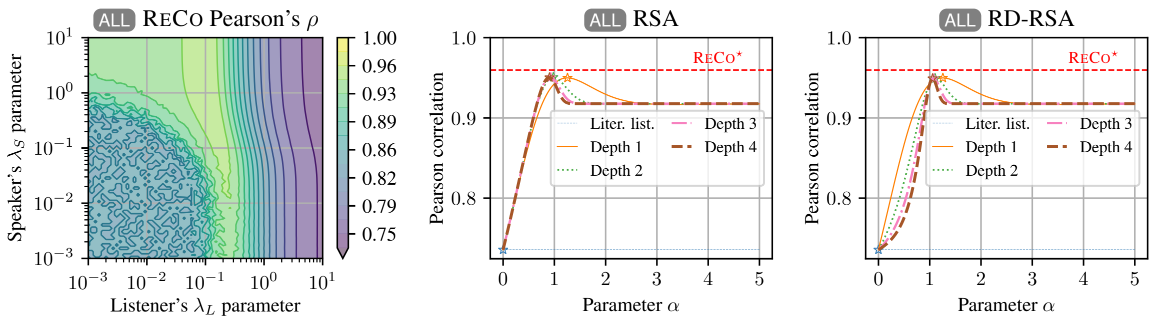

Figure 2: Pearson's correlation ρ on the full dataset of graded human judgments from (Frank, 2016). (Left) Correlation for RECO as a function of λ_L and λ_S represented as a contour plot. (Middle) Correlation between RSA at different levels of α and recursive depth (Right) Correlation between RD-RSA at different levels of α and recursive depth. (Middle, Right) RECO with the best setting of λ_L and λ_S is indicated with a red dashed line. Stars indicate the best α value at different depths.

	Literal LISTENER	BR SPEAKER	RSA	RD-RSA	RECO
ALL	73.57%	90.04%	95.07%	94.98%	**95.96%**
SIMPLE	70.10%	88.16%	**96.02%**	**96.02%**	**96.02%**
COMPLEX	83.86%	97.83%	94.74%	94.35%	**98.18%**
TWINS	97.61%	93.43%	97.61%	**98.98%**	97.61%
ODDMAN	**94.97%**	**94.97%**	**94.97%**	**94.97%**	**94.97%**

Table 1: Correlation across different methods with graded human judgements in four reference games Frank (2016) (with the best hyperparameter settings). RECO performs better than the alternatives in ALL.

$$u(\pi_\mathsf{S}, \pi_\mathsf{L}) \coloneqq \mathop{\mathbb{E}}_{\substack{m\sim p \\ u\sim\pi_\mathsf{S}(\cdot|m) \\ m'\sim\pi_\mathsf{L}(\cdot|u)}} r(m, u, m') \qquad (1)$$

RECO then posits that language users optimize modified utilities:

$$\tilde{u}_\mathsf{S}(\pi_\mathsf{S}, \pi_\mathsf{L}) \coloneqq u(\pi_\mathsf{S}, \pi_\mathsf{L}) - \lambda_\mathsf{S} \cdot \mathrm{D}_{\mathrm{KL}}(\pi_\mathsf{S} \,\|\, \tau_\mathsf{S}),$$
$$\tilde{u}_\mathsf{L}(\pi_\mathsf{S}, \pi_\mathsf{L}) \coloneqq u(\pi_\mathsf{S}, \pi_\mathsf{L}) - \lambda_\mathsf{L} \cdot \mathrm{D}_{\mathrm{KL}}(\pi_\mathsf{L} \,\|\, \tau_\mathsf{L}).$$

Here τ_S and τ_L represent the SPEAKER's and LISTENER's prior knowledge of language (independent of any specific communicative goal or context). We refer to these policies as the **default semantics** in the language used for communication. They play a similar role to the literal semantics used by RSA and other iterated response models. But here, we need not assume that they correspond specifically to literal semantics—instead, they model agents' prior expectations about how utterances are likely to be produced and interpreted in general by pragmatic language users.

We apply RECO to a family of four reference tasks introduced by Frank (2016). Frank gathered graded human judgments about the probability that particular utterances might carry particular meanings. We evaluate its predictions by measuring their *correlation* between human judgments. Comparisons between RECO and a variety of existing game-theoretic and iterated-response models of pragmatics are shown in Table 1. RECO improves upon the best predictions of RSA-family methods, both overall and on 3/4 tasks individually, and is robust across a range of speaker hyperparameters.

In summary, RECO provides an expressive, learnable model of pragmatic communication. Looking ahead, it might be used to study several related problems in context-dependent, multi-party communication—it could be applied to *iterated* conventions (Hawkins et al., 2017) established over multiple rounds of communication; or combined with tools for solving *extensive-form* games to model communicative strategies that play out over multiple turns of dialog.

References

Herbert H Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Judith Degen. 2023. The rational speech act framework. *Annual Review of Linguistics*, 9:519–540.

Michael C Frank. 2016. Rational speech act models of pragmatic reasoning in reference games.

Herbert P Grice. 1975. Logic and conversation. In *Speech acts*, pages 41–58. Brill.

Robert XD Hawkins, Mike Frank, and Noah D Goodman. 2017. Convention-formation in iterated reference games. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

David K Lewis. 1971. Convention: A philosophical study. *Philosophy and Rhetoric*, 4(2).