Noisy Population Dynamics Lead to Efficiently Compressed Semantic Systems

Nathaniel Imel
University of California, Irvine

Richard Futrell
University of California, Irvine

Michael Franke
University of Tübingen

Noga Zaslavksy
University of California, Irvine

Abstract

Converging cross-linguistic evidence suggests that that human vocabularies are shaped for efficient communication, but we know little about the agent-based dynamics that could explain their evolution. In this paper, we show that very general population dynamics of signaling games lead to the emergence of information-theoretically efficient meaning systems. In numerical simulations, we observe that noisy perception of meaning can result in evolved systems with higher efficiency.

1 Introduction

A prominent idea in semantic typology is that human vocabularies categorize the world efficiently, by optimally balancing the costs of mental representation with the need to communicate accurately [19, 20, 23]. The intuition behind this idea is that out of the logically possible languages, natural languages are near-optimal solutions to the problem of trading off complexity and accuracy. Empirically, there is robust support for various formulations of this view across many diverse semantic domains [19, 33, 48, 45, 46, 50, 38, 39, 25, 44, 51, 12, 13, 16, 8], indicating that shared properties of human meaning systems emerge naturally from pressures for efficient communication.

However, as an explanation about how languages evolve under external constraints, the communicative efficiency account is incomplete. We do not yet have a clear agent-based explanation of the mechanisms that drive efficiency for linguistic communities in evolutionary time. Zaslavsky et al. [48] argued that lexica evolve along efficient trajectories along the theoretical bound of efficiency, defined by the Information Bottleneck (IB) tradeoff [41] between the complexity and accuracy of the lexicon, and this idea has been supported empirically using both synchronic [48, 50, 51, 25] and diachronic data [52]. However, this account does not provide an agent-based mechanistic explanation. A few studies have explored agent-based simulations in the context of efficient communication, leveraging deep reinforcement learning [7, 3, 22, 42, 43] and Bayesian iterated learning paradigms [6, 21, 5, 4]. However, the precise link between the agent-based population dynamics and the efficiency of emergent lexica is still largely unknown. Here, we ask: can languages achieve efficient trajectories from general, independently motivated population dynamics?

Specifically, we draw on a population dynamics of behavioral imitation that is derived from the well-known replicator equation [40, 34], which has previously been applied to language evolution [37, 27, 29] and has its origins in the signaling games literature to explain the emergence of vagueness in language [9]. We show how this dynamical model may be related to the IB principle, following recent work quantifying the efficiency of emergent semantic systems using the IB framework [42, 43, 4]. In numerical simulations, we observe that speakers and listeners converge to lexicons that are near the IB theoretical bounds of compression, and that noisy perception of meanings leads to more efficient meaning systems.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

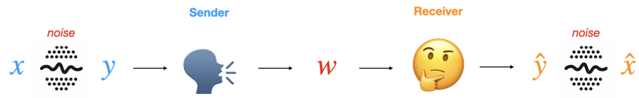

Figure 1: A round of play in the signaling game with confusability (noise) of states.

2 Evolutionary model

Our dynamics is an aggregate, population-level description of how behaviors optimize for communicative success in the presence of perceptual noise [9]. It describes the most likely evolutionary path of agents optimizing (via learning, imitation, or some other mechanism) strategies, but confuse similar perceptual states for one another [14]. The core of our model is a *sim-max game* which describes communication in terms of strategic interactions between a Sender and Receiver (depicted in Fig. 1) [24, 37, 17, 18, 30] as follows. First, nature randomly selects a state of the world $x \in \mathcal{X}$ to present to Sender S. Depending on Sender's level of perceptual certainty, Sender then observes this state or a similar one $y \in \mathcal{X}$, and chooses a signal $w \in \mathcal{W}$ to send to Receiver, who can only observe the signal. Upon observing this signal, Receiver chooses an interpretation $\hat{y} \in \mathcal{X}$, which may possibly be confused as state $\hat{x}$. Nature awards payoff to both players to the extent that the interpretation $\hat{y}$ is similar to state y. We model evolution in a version of the discrete-time replicator dynamics (see Appendix A.2). Formally, the behavioral updates to the average sender S and average receiver R are given by the following[1]:

$$S'(w \mid x) \;\propto\; \sum_{y \in \mathcal{X}} C(y \mid x) \cdot S(w \mid y) \cdot f_S(w, y) \tag{1}$$

$$R'(\hat{x} \mid w) \;\propto\; \sum_{\hat{y} \in \mathcal{X}} C(\hat{y} \mid \hat{x}) \cdot R(\hat{y} \mid w) \cdot f_R(w, \hat{y}), \tag{2}$$

where the probability of confusing state z for state z' is defined by their (symmetric) similarity:

$$C_\alpha(z' \mid z) \propto \text{similarity}_\alpha(z, z'), \tag{3}$$

and

$$\forall z, z' \in \mathcal{X} \quad \text{similarity}_\gamma(z, z') = \exp(-\gamma\, (z - z')^2). \tag{4}$$

Discriminative Need For our model of perceptual similarity, we follow [14, 33] in using an independently motivated model from mathematical psychology [36, 26]. In Eq. 4, γ is an inverse temperature parameter representing a tolerable level of pragmatic slack, playing a similar role to the concept of *discriminative need* deployed by [7]. The minimum discriminative need is assumed when $\gamma = 0$, which lets any pair of Sender and Receiver meanings receive identical payoff. Perfect discrimination between states is enforced when $\gamma \to \infty$, in which only perfect guesses by Receiver yield nonzero payoff.

Noise / Perceptual certainty Following [14], we assume that the probability of confusing states depends on their similarity (Eq. 3). To do this, we supply the similarity function (Eq. 4) with a perceptual certainty parameter α, and manipulate the level of 'noise' in the game dynamics. Note that fitness (via discriminative need) and perceptual noise use the same similarity function, but their inverse temperature parameters (γ and α, respectively) vary independently.

Fitness As an instantiation of the replicator dynamics, our model assumes that the frequency of a signaling behavior evolves according to its current frequency and its fitness, and that evolution is driven by the fact that successful behaviors become more frequent. The fitnesses f_R, f_S of Sender and Receiver's strategies are defined respectively in terms of how well the strategies maximize similarity between states $y, \hat{y} \in \mathcal{X}$:

$$f_S(w, y) = \sum_{\hat{y}} R(\hat{y} \mid w) \cdot \text{similarity}_\gamma(y, \hat{y}), \tag{5}$$

$$f_R(w, \hat{y}) = \sum_{y} Pr(y) \cdot S(w \mid y) \cdot \text{similarity}_\gamma(y, \hat{y}). \tag{6}$$

[1]We omit only the normalizing constants in Eqs. 1, 2, and 3, hence the proportionalities.

Each Sender S and Receiver R are defined by a set of deterministic strategies, so that population averages describe probabilistic strategies. Thus, in Eq. 1, $S(w \mid x)$ denotes the probability that a randomly sampled sender signals $w \in \mathcal{W}$ to communicate state $y \in \mathcal{X}$, and in Eq. 2 $R(\hat{y} \mid w)$ denotes the probability that a randomly sampled receiver chooses to interpret signal w to mean world state $\hat{y}$. We also need to specify a prior over meanings, $\Pr(\cdot)$, which reflects how often language users need to communicate about particular states of the world. To begin to explore the efficiency of this dynamical process, we consider a very simple setting in which the world states for Sender and Receiver is a set of n contiguous integers $\mathcal{X} = \{0, 1, \ldots, n\}$, to model a situation in which states of the environment stand in minimally interesting physical distance relations.

3 Measuring efficiency of emergent semantic systems

To explain the evolution of efficient meaning systems, we require a domain-agnostic method of quantifying the complexity/accuracy trade-off for semantic systems. Following [48, 47], we measure the trade-off in terms of the Information Bottleneck (IB) Framework. IB is a special case of Rate-Distortion Theory (RDT), the branch of information theory concerned with optimizing data compression under bounded resources [35, 11]. In information-theoretic terms, languages minimize both *rate* (the resources necessary to compress a thought into a word, quantified in bits) and *distortion* (the error a listener makes in reconstructing a speaker's intention). Here we briefly describe the IB objective, keeping the same notation as [48]. Consider a language equipped with words $\mathcal{W}$ to communicate about meanings $\mathcal{M}$, which are modeled as distributions $p(u \mid m)$ over world states $\mathcal{U}$. This language can be represented by an aggregate (stochastic) speaker $q(w \mid m)$, the *encoder*. The language's semantic system is *efficient* to the extent that q minimizes:

$$\mathcal{F}_{q(w|m)} = I(M; W) - \beta I(W; U) \text{ s.t. } \beta \geq 1. \tag{7}$$

The information rate $I(M; W)$ quantifies the complexity of the encoder, and is to be minimized. Maximizing accuracy, quantified by $I(W; U)$, amounts to minimizing the distortion between speaker and listener meanings, i.e. the KL-divergence $D_{\mathrm{KL}} \left[M \| \hat{M} \right]$. In our game setting, the confusability of states C (Eq. 3) corresponds to the perceptually uncertain meaning distributions M. The states of the signaling game ($\mathcal{X}$) are straightforwardly identified with world states in the IB framework ($\mathcal{U}$), and similarly words and signals ($\mathcal{W}$) play identical roles. A Sender $S(w \mid m)$ in the sim-max game corresponds to an encoder $q(w \mid m)$ in the IB framework; we therefore identify each emergent semantic system with aggregate Sender population behavior.

4 Simulation Results

We simulated evolution in the noisy sim-max game for a variety of parameter configurations and measured the efficiency of the resulting systems with respect to the IB theoretical bound, which we estimated using the IB-method algorithm [41]. Specifically, we simulate the population dynamics for up to 200 time steps, varying discriminative need, the degree of perceptual noise, and random initialization of agent populations, resulting in 343 distinct runs. We fix the signaling game and IB bound resulting from $|\mathcal{U}| = |\mathcal{W}| = 100$ world states and signals, a uniform prior over meanings, and setting the meaning distributions $p(u \mid m) = C_\alpha(u \mid m)$ for $\alpha = 1.0$ (Eq. 3) for all simulations. We report the results of evolution after 200 time steps; most runs converged early.

The outcomes of our simulations are displayed in Figs. 2, 3 (See Appendix B.1 for more details of our parameter sweep). We catalogue two main empirical findings. First, most emergent systems converge close to the IB theoretical limit, spaning a diverse set of near-optimal solutions. Second, the complexity of emergent systems is predicted by the degree of perceptual noise ($R^2 = 0.66$, $p \approx 0$). We observe that (i) with no noise in the dynamics, many higher complexity systems have low efficiency; (ii) with some noise, systems all achieve high efficiency and the complexity and accuracy is bounded from above; (iii) with a high degree of noise, systems evolve to be among the simplest possible, including systems with one word for all 100 meanings.

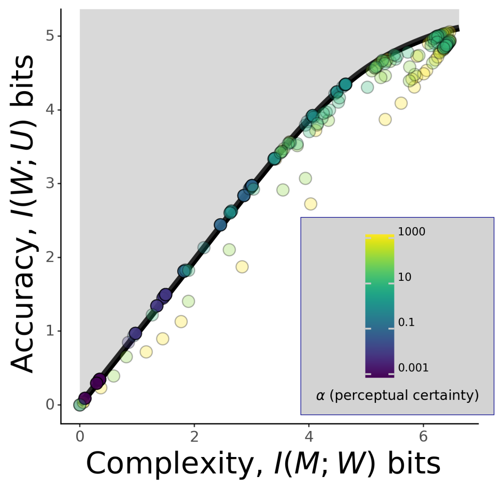

Figure 2: Emergent semantic systems from evolutionary dynamics of sim-max games with 100 states and signals. Black line: the information curve, i.e. the set of optimal solutions to the IB objective (Eq. 7). Circles: emergent systems. Color: perceptual certainty parameter controlling the probability of confusing similar states.

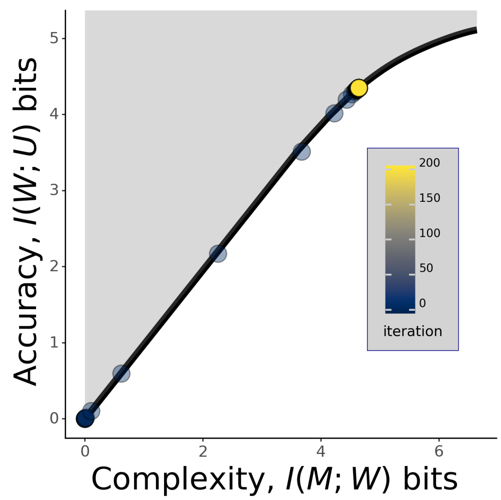

Figure 3: The simulated evolutionary trajectory of one semantic system in a sim-max game with $\gamma = 0.01$ and $\alpha = 1000$. Circles: the semantic category system at one time step. Color: the discrete time step in evolution, from $1 - 200$. Most simulations had similar trajectories.

5 Discussion

The emergent systems span a large range of near-optimal solutions. *Regularizing* noise in the perception and interpretation of meaning, however, appears to constrain this variation and increase efficiency. In actual linguistic communities, this noise may result from specific cognitive architectures and learning biases, or other sources of randomness in the replication of behavior. This prediction, that noisy perception of meaning can drive efficiency in the lexicon, echoes results in the literature, that transmission errors in (iterated) language learning can drive simplicity [4, 5]. Our predictions are also consistent with the perspective of the IB framework, which takes meanings to be probabilistic. Additionally, the evolutionary paths we observe (Fig. 3) are compatible with the proposal that annealing processes well-describe the trajectories of systems along the information plane [48, 49]. There are several limitations of the current study. First, focusing on how noise (α) can drive language efficiency assumes that language must co-evolve with perception. Future work may address this limitation by formulating a version of the dynamics that does not identify α directly with perception, or by manipulating only discriminative need (γ). Also, our dynamics yields systems at unnatural extremes of complexity, a limitation that some previous work does not suffer from [7, 4, 42, 43]. However, the diversity of emergent optimal systems is intrinsically interesting. It invites questions about possible connections between game theory and IB; e.g., are there relationships between Nash equilibria in sim-max games, and rate-distortion optima? Another limitation of our work is that we focused on a simple, 1D semantic space, but future work could easily explore real domains (as in [10]). Lastly, our evolutionary model assumes two disjoint, infinite populations, and changes that occur at abrupt, evenly spaced, countable time intervals. While unrealistic, our model follows much prior work in game theory in these idealizing assumptions, which sometimes lead to easier mathematical investigation into precise, general results.

6 Conclusion

Using evolutionary game theory, we have explored a simple, general process by which meaning systems can evolve in a population along the solutions to an information-theoretic efficiency objective. In simulations, we find that noise in the perception of meaning can drive optimal complexity/accuracy trade-offs. Our results suggest promising connections between two influential ideas in language evolution: efficiency in the Information Bottleneck and the replicator dynamics of signaling games.

7 Acknowledgments

We thank Fausto Carcassi, Roger Levy, Shane Steinert-Threlkeld, and Brian Skyrms for valuable feedback. This paper is an extension of previous work, [15], which appeared in the Proceedings of the 45th Annual Meeting of the Cognitive Science Society.

References

[1] Immanuel M. Bomze. Lotka-Volterra equation and replicator dynamics: A two-dimensional classification. *Biological Cybernetics*, 48(3):201–211, October 1983. ISSN 1432-0770. doi: 10.1007/BF00318088. URL https://doi.org/10.1007/BF00318088.

[2] Immanuel M. Bomze. Lotka-Volterra equation and replicator dynamics: new issues in classification. *Biological Cybernetics*, 72(5):447–453, April 1995. ISSN 1432-0770. doi: 10.1007/BF00201420. URL https://doi.org/10.1007/BF00201420.

[3] Emil Carlsson, Devdatt Dubhashi, and Fredrik D. Johansson. Learning Approximate and Exact Numeral Systems via Reinforcement Learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43(43), 2021. URL https://escholarship.org/uc/item/0kx279mj.

[4] Emil Carlsson, Devdatt Dubhashi, and Terry Regier. Iterated learning and communication jointly explain efficient color naming systems. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society*, 2023.

[5] Jon W. Carr, Kenny Smith, Jennifer Culbertson, and Simon Kirby. Simplicity and informativeness in semantic category systems. *Cognition*, 202:104289, September 2020. ISSN 0010-0277. doi: 10.1016/j.cognition.2020.104289. URL https://www.sciencedirect.com/science/article/pii/S0010027720301086.

[6] Alexandra Carstensen, Jing Xu, Cameron T. Smith, and Terry Regier. Language evolution in the lab tends toward informative communication. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36(36), 2014. URL https://escholarship.org/uc/item/8mq448m7.

[7] 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, March 2021. doi: 10.1073/pnas.2016569118. URL https://www.pnas.org/doi/10.1073/pnas.2016569118. Publisher: Proceedings of the National Academy of Sciences.

[8] Sihan Chen, Richard Futrell, and Kyle Mahowald. An information-theoretic approach to the typology of spatial demonstratives. *Cognition*, 240:105505, November 2023. ISSN 0010-0277. doi: 10.1016/j.cognition.2023.105505. URL https://www.sciencedirect.com/science/article/pii/S0010027723001397.

[9] José Pedro Correia. The bivalent trap: Vagueness, theories of meaning and identity. Master's thesis, Universiteit van Amsterdam., 2013.

[10] José Pedro Correia and Michael Franke. Towards an Ecology of Vagueness. In Richard Dietz, editor, *Vagueness and Rationality in Language Use and Cognition*, Language, Cognition, and Mind, pages 87–113. Springer International Publishing, Cham, 2019. ISBN 978-3-030-15931-3. doi: 10.1007/978-3-030-15931-3_6. URL https://doi.org/10.1007/978-3-030-15931-3_6.

[11] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006. ISBN 978-0-471-24195-9.

[12] Milica Denić, Shane Steinert-Threlkeld, and Jakub Szymanik. Complexity/Informativeness trade-off in the domain of indefinite pronouns. In *Proceedings of Semantics and Linguistic Theory (SALT 30)*, volume 30, pages 166–184, 2020. doi: 10.3765/salt.v30i0.4811.

[13] Milica Denić, Shane Steinert-Threlkeld, and Jakub Szymanik. Indefinite Pronouns Optimize the Simplicity/Informativeness Trade-Off. *Cognitive Science*, 46(5):e13142, 2022. doi: 10.1111/cogs.13142.

[14] Michael Franke and José Pedro Correia. Vagueness and Imprecise Imitation in Signalling Games. *The British Journal for the Philosophy of Science*, 69(4):1037–1067, December 2018. ISSN 0007-0882. doi: 10.1093/bjps/axx002. URL `http://www.journals.uchicago.edu/doi/full/10.1093/bjps/axx002`. Publisher: The University of Chicago Press.

[15] Imel. The evolution of efficient compression in signaling games. In *Proceedings of the 45th Annual Meeting of the Cognitive Science Society*, 2023. to appear.

[16] Nathaniel Imel and Shane Steinert-Threlkeld. Modal semantic universals optimize the simplicity/informativeness trade-off. In *Proceedings of semantics and linguistic theory (SALT 32)*, 2022.

[17] Gerhard Jäger. The evolution of convex categories. *Linguistics and Philosophy*, 30(5):551–564, October 2007. ISSN 1573-0549. doi: 10.1007/s10988-008-9024-3. URL `https://doi.org/10.1007/s10988-008-9024-3`.

[18] Gerhard Jäger, Lars P. Metzger, and Frank Riedel. Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals. *Games and Economic Behavior*, 73(2):517–537, November 2011. ISSN 0899-8256. doi: 10.1016/j.geb.2011.03.008. URL `https://www.sciencedirect.com/science/article/pii/S0899825611000595`.

[19] Charles Kemp and Terry Regier. Kinship categories across languages reflect general communicative principles. *Science (New York, N.Y.)*, 336(6084):1049–1054, 2012. doi: 10.1126/science.1218811.

[20] Charles Kemp, Yang Xu, and Terry Regier. Semantic typology and efficient communication. *Annual Review of Linguistics*, pages 1–23, 2018. doi: 10.1146/annurev-linguistics-011817-045406.

[21] Simon Kirby, Monica Tamariz, Hannah Cornish, and Kenny Smith. Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102, August 2015. ISSN 0010-0277. doi: 10.1016/j.cognition.2015.03.016. URL `https://www.sciencedirect.com/science/article/pii/S0010027715000815`.

[22] Mikael Kågebäck, Emil Carlsson, Devdatt Dubhashi, and Asad Sayeed. A reinforcement-learning approach to efficient communication. *PLOS ONE*, 15(7):e0234894, July 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0234894. URL `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0234894`. Publisher: Public Library of Science.

[23] Stephen C. Levinson. Kinship and Human Thought. *Science (New York, N.Y.)*, 336(6084):988–989, 2012. doi: 10.1126/science.1222691. URL `https://www.science.org/doi/abs/10.1126/science.1222691`.

[24] David Kellogg Lewis. *Convention: A philosophical study*. Cambridge, MA, USA: Wiley-Blackwell, 1969.

[25] 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, December 2021. doi: 10.1073/pnas.2025993118. URL `http://www.pnas.org/doi/full/10.1073/pnas.2025993118`. Publisher: Proceedings of the National Academy of Sciences.

[26] Robert M. Nosofsky. Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115:39–57, 1986. ISSN 1939-2222. doi: 10.1037/0096-3445.115.1.39. Place: US Publisher: American Psychological Association.

[27] Martin A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, 2006. ISBN 978-0-674-02338-3. doi: 10.2307/j.ctvjghw98. URL `https://www.jstor.org/stable/j.ctvjghw98`.

[28] Martin A. Nowak and David C. Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033, July 1999. doi: 10.1073/pnas.96.14.8028. URL `https://www.pnas.org/doi/full/10.1073/pnas.96.14.8028`. Publisher: Proceedings of the National Academy of Sciences.

[29] Martin A. Nowak, Natalia L. Komarova, and Partha Niyogi. Evolution of Universal Grammar. *Science*, 291(5501):114–118, January 2001. doi: 10.1126/science.291.5501.114. URL `https://www.science.org/doi/10.1126/science.291.5501.114`. Publisher: American Association for the Advancement of Science.

[30] Cailin O'Connor. The Evolution of Vagueness. *Erkenntnis*, 79(4):707–727, April 2014. ISSN 1572-8420. doi: 10.1007/s10670-013-9463-2. URL `https://doi.org/10.1007/s10670-013-9463-2`.

[31] Karen M.. Page and Marting A. Nowak. Unifying Evolutionary Dynamics. *Journal of Theoretical Biology*, 219(1):93–98, November 2002. ISSN 0022-5193. doi: 10.1006/jtbi.2002.3112. URL `https://www.sciencedirect.com/science/article/pii/S0022519302931127`.

[32] Christina Pawlowitsch. Why evolution does not always lead to an optimal signaling system. *Games and Economic Behavior*, 63(1):203–226, May 2008. ISSN 0899-8256. doi: 10.1016/j.geb.2007.08.009. URL `https://www.sciencedirect.com/science/article/pii/S0899825607001480`.

[33] Terry Regier, Charles Kemp, and Paul Kay. Word meanings across languages support efficient communication Informativeness and simplicity as competing principles. In Brian MacWhinney and William O'Grady, editors, *The Handbook of Language Emergence*, pages 237–263. Wiley, 2015.

[34] Peter Schuster and Karl Sigmund. Replicator dynamics. *Journal of Theoretical Biology*, 100 (3):533–538, February 1983. ISSN 0022-5193. doi: 10.1016/0022-5193(83)90445-9. URL `https://www.sciencedirect.com/science/article/pii/0022519383904459`.

[35] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948. ISSN 0005-8580. doi: 10.1002/j.1538-7305.1948.tb01338.x. Conference Name: The Bell System Technical Journal.

[36] Roger N. Shepard. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22(4):325–345, December 1957. ISSN 1860-0980. doi: 10.1007/BF02288967. URL `https://doi.org/10.1007/BF02288967`.

[37] Brian Skyrms. *Signals: Evolution, Learning, and Information*. Oxford University Press, Oxford, 2010. ISBN 978-0-19-958082-8. doi: 10.1093/acprof:oso/9780199580828.001.0001. URL `https://oxford.universitypressscholarship.com/10.1093/acprof:oso/9780199580828.001.0001/acprof-9780199580828`.

[38] Shane Steinert-Threlkeld. Quantifiers in natural language optimize the Simplicity/Informativeness trade-off. In Julian J Schl\"{o}der, Dean McHugh, and Floris Roelofsen, editors, *Proceedings of the 22nd Amsterdam Colloquium*, pages 513–522, 2020.

[39] Shane Steinert-Threlkeld. Quantifiers in Natural Language: Efficient Communication and Degrees of Semantic Universals. *Entropy. An International and Interdisciplinary Journal of Entropy and Information Studies*, 23(10):1335, 2021. doi: 10.3390/e23101335.

[40] Peter D. Taylor and Leo B. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1):145–156, July 1978. ISSN 0025-5564. doi: 10.1016/0025-5564(78)90077-9. URL `https://www.sciencedirect.com/science/article/pii/0025556478900779`.

[41] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottlneck method. *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, pages 368–377.

[42] Mycal Tucker, Roger P. Levy, Julie Shah, and Noga Zaslavsky. Trading off utility, informativeness, and complexity in emergent communication. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in neural information processing systems*, 2022. URL `https://openreview.net/forum?id=O5arhQvBdH`.

[43] Mycal Tucker, Roger P. Levy, Julie Shah, and Noga Zaslavsky. Generalization and Translatability in Emergent Communication via Informational Constraints. In *NeurIPS 2022 Workshop on Information-Theoretic Principles in Cognitive Systems*, 2022. URL `https://openreview.net/forum?id=yf8suFtNZ5v`.

[44] Wataru Uegaki. The informativeness / complexity trade-off in the domain of Boolean connectives. *Linguistic Inquiry*, 2021.

[45] Yang Xu and Terry Regier. Numeral systems across languages support efficient communication: From approximate numerosity to recursion. In *Cognitive science society (cogsci-2014)*, pages 1802 – 1807, 2014.

[46] Yang Xu, Terry Regier, and Barbara C. Malt. Historical Semantic Chaining and Efficient Communication: The Case of Container Names. *Cognitive Science*, 40(8):2081–2094, November 2016. ISSN 0364-0213. doi: 10.1111/cogs.12312.

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

[48] 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.

[49] Noga Zaslavsky, Karee Garvin, Charles Kemp, Naftali Tishby, and Terry Regier. Evolution and efficiency in color naming: The case of Nafaanra. In *CogSci*, page 68, 2019.

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

[51] 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.

[52] 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*, page lzac001, April 2022. ISSN 2058-458X. doi: 10.1093/jole/lzac001. URL `https://doi.org/10.1093/jole/lzac001`.

A Additional modeling details

A.1 Reproducing results

A script for reproducing all simulation results and figures can be found at `https://github.com/nathimel/ibsg`. This script was run on a 8 CPU core laptop in a matter of hours.

A.2 The replicator dynamics

The standard replicator dynamics describes change in mean behavior in a population of game players. Change in proportion x_i of strategy i is defined by the differential equation in Eq. 8 and one of its discretizations in Eq. 9:

$$\dot{x_i} = x_i[f_i(x) - \phi(x)], \tag{8}$$

$$x_i' = x_i \left[\frac{(f(x))_i}{\phi(x)} \right]. \tag{9}$$

The continuous time formulation is mathematically equivalent to a number of equations from evolutionary theory [31, 1, 2]. The discrete-time version is easier to work with in simulations. In both Eqs. 8 and 9, $x \in \mathbb{R}^n$ is a vector of the distribution of n possible types in the population, x_i is the proportion of type $i \in [n]$ in the population, $f_i(x)$ is the fitness of this type (relative to the population), and $\phi(x) = \sum_{j=1}^{n} x_j f_j(x)$ represents the average fitness of the population.

In contrast to [28, 32], we allow for changes at individual choice-points, rather than requiring changes at the level of entire contingency plans. In this way, our approach is more compatible with ideas of imitation and learning than with biological evolution among agents predisposed to inflexibly execute their native behavior. Therefore, in Eq. 9, x_i represents the frequency of a signaling behavior, corresponding to a particular strategy, e.g. for S to send w in state x or R to interpret w as $\hat{x}$. The analogous discrete-time updates used in this work (adding detail to Eqs. 1, 2) are:

$$S'(w \mid y) \; = \; \sum_{x \in \mathcal{X}} C(y \mid x) \cdot \frac{S(w \mid x) \cdot f_S(w, x)}{\sum_{w \in \mathcal{W}} S(w \mid x) \cdot f_S(w, x)} \tag{10}$$

$$R'(\hat{y} \mid w) \; = \; \sum_{\hat{x} \in \mathcal{X}} C(\hat{y} \mid \hat{x}) \cdot \frac{R(\hat{x} \mid w) \cdot f_R(w, \hat{x})}{\sum_{w \in \mathcal{W}} R(w \mid \hat{x}) \cdot f_R(w, \hat{x})} \tag{11}$$

For a derivation of the dynamics we deploy (Eqs. 10, 11) from Eq. 9, we refer readers to Section 5.2 of [9]. See [14] for discussion of how the current dynamics are related to, but distinct from, the famous replicator-mutator dynamics [31, 27].

A.2.1 Vectorized dynamics

For ease of replication, we include the simple form of the vectorized discrete-time update steps in Eqs. 1, 2. Let S, R represent the conditional distributions for Sender and Receiver, respectively, C represent the transition matrix describing the probability of confusing one state with another, F_S, F_R the fitness matrices for Sender and Receiver, respectively, and $pr \in \mathbb{R}^{|\mathcal{W}| \times |\mathcal{U}|}$ a matrix consisting of $|\mathcal{W}|$ row-wise concatenations of the stochastic vector corresponding to the communicative need distribution. The update steps for the Sender and Receiver populations can be expressed as:

$$S' = CS \odot \; (RF_S)^\top \tag{12}$$

$$R' = R \odot \; pr \odot (F_R S)^\top C \tag{13}$$

After each step, the rows of S and R need to be normalized so that they sum to 1.

A.3 Quantifying efficiency loss

We follow [48] in quantifying the inefficiency of an emergent semantic system in terms of its *efficiency loss*, ϵ, which is the difference between the value of the IB objective function for the emergent system's associated aggregate Sender $q = S(w|m)$ and that of the nearest optimal encoder $q^*(w|m)$ lying on the theoretical limit:

$$\epsilon = \Delta \mathcal{F}_q = \mathcal{F}_q - \mathcal{F}_{q^*} \tag{14}$$

and define the nearest optimal encoder to a given Sender as:

$$1 - \min_{q^* \in \mathcal{Q}^*} \mathrm{gNID}(q, q^*), \tag{15}$$

out of the set of all optimal encoders $\mathcal{Q}^*$. Here, gNID is a generalization of the Normalized Information Distance to unimodal, soft probabilistic partitions. We refer readers to [48], Supplementary Information, Section 3 for full definition and details of gNID.

B Simulation parameter sweep

In addition to the level of perceptual noise in the dynamics, we explored several other factors commonly assumed to be important in language evolution play a shaping role in the efficiency of emergent systems; namely, the initial distribution of Sender and Receiver strategies in the population and discriminative need (Eq. 4).

We randomly initialize the populations of Senders and Receivers for the beginning of each simulation using an energy-based initialization:

$$S = \text{softmax}(\delta \cdot N) \quad \text{with } N \in \mathbb{R}^{|\mathcal{W}| \times |\mathcal{M}|}, \tag{16}$$

$$R = \text{softmax}(\delta \cdot N) \quad \text{with } N \in \mathbb{R}^{|\mathcal{M}| \times |\mathcal{W}|}. \tag{17}$$

where N is a matrix of numbers sampled from the standard normal distribution $\mathcal{N}(0, 1)$ and δ is a parameter for controlling how uniformly the population initial strategies will be distributed; as $\delta \to -\infty$, the population will be initialized uniformly; as $\delta \to \infty$ the population will begin with a fully deterministic category system.

Recall from Section 2 that we vary discriminative need by varying the similarity parameter, γ:

$$\text{similarity}_\gamma(x, x') = \exp(-\gamma \cdot (x - x')^2). \tag{18}$$

We vary perceptual noise by varying the parameter α independently from γ, which in this case represents perceptual certainty instead of discriminative need. The rows of the similarity matrix are normalized to form a confusion matrix specifying the probability of confusing one world state for another:

$$C_\alpha(x'|x) = \frac{\text{similarity}_\alpha(x, x')}{\sum_y \text{similarity}_\alpha(x, y)}. \tag{19}$$

B.1 Results

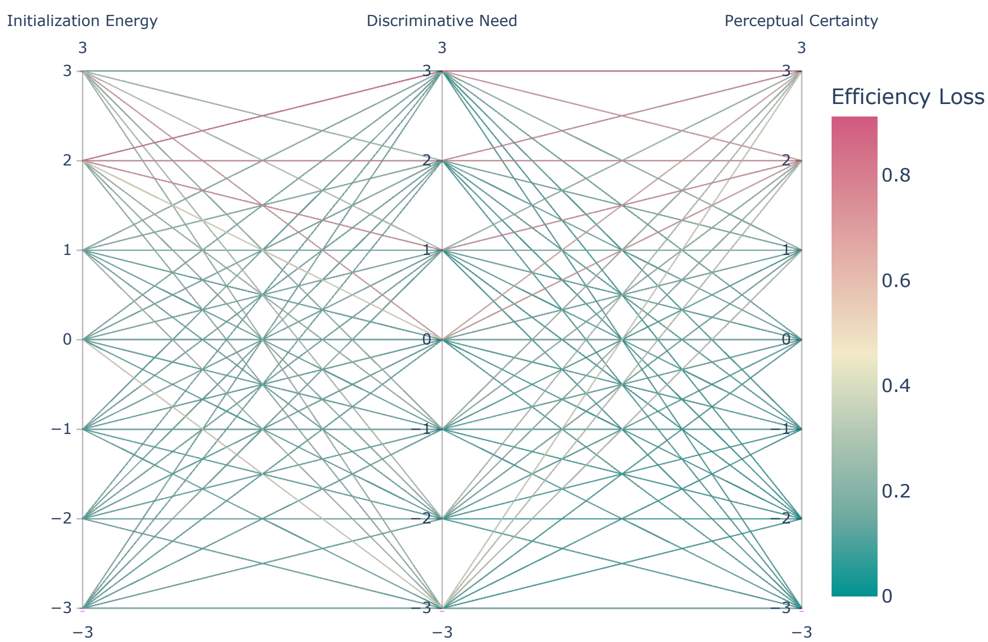

Figure 4: Results of parameter sweep over initial population distributions, discriminative need, and perceptual certainty on efficiency loss across 343 simulations. A line connecting each parameter coordinate represents a single simulation. Parameter values are labeled on a $\log_{10}$ scale. Color of lines represents the efficiency loss of the emergent semantic category system (Eq 14), with green corresponding to lowest (0.0) values (better) and red corresponding to highest (0.91) efficiency loss. The majority of simulations lead to low inefficiency (median $\epsilon = 0.07$). Meanwhile, runs that (i) start with very biased (non-uniform) initial population dispositions, (ii) involve very high discriminative need in the signaling game, and (iii) assume no perceptual noise in the dynamics represent simulations that lead to the most inefficient semantic systems.

parameter	β	R^2	p
initialization energy	0.01	0.046	5.50×10^{-5}
discriminative need	0.005	0.006	1.32×10^{-1}
perceptual certainty	0.02	0.175	5.09×10^{-16}

Table 1: OLS regression results of population initialization energy, discriminative need, and perceptual certainty on efficiency loss (parameters in log scale).

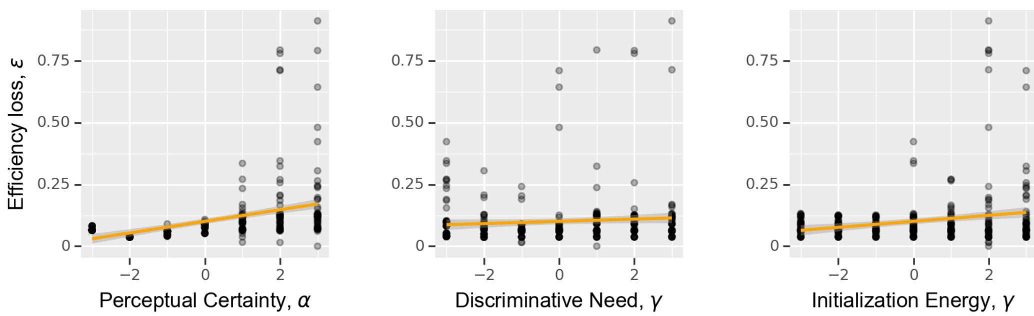

Figure 5: OLS regression fits of population initialization energy, discriminative need, and perceptual certainty on efficiency loss. x-axis in log scale.

Predicting efficiency For each of these factors, there are parameter values that lead to inefficiency as quantified by Eq. 14, with the results depicted in Fig. 4. In particular, higher values of discriminative need result in lower efficiency. This is somewhat intuitive and parallel to the trend with noise in the dynamics: when payoffs are awarded only for perfectly precise communication, success is rarer; furthermore, stimuli cannot be generalized based on similarity in order to lead to regular partitions that support efficient compression. Initial entropy of the Sender and Receiver population distributions also has a significant effect on efficiency, such that more uniform initial populations tend to result mean systems with higher efficiency. The results of performing linear regressions of each factor (individually) on efficiency loss are reported in Table 1 and Fig. 5.

B.1.1 Predicting complexity

When visualizing all systems on the complexity/accuracy plane, initialization and perceptual noise appear to appear to constrain emergent complexity in a systematic way (Figures 7, 9), while the effects of discriminative need are less obvious (Fig. 8). Linear regressions also revealed this relationships numerically, reported in Table 2 and Fig. 6.

parameter	β	R^2	p
initialization energy	-0.21	0.04	2.66×10^{-4}
discriminative need	0.07	0.00	2.23×10^{-1}
perceptual certainty	0.89	0.66	4.04×10^{-81}

Table 2: Individual OLS regressions of population initialization energy, discriminative need, and perceptual certainty on complexity (parameters in log scale).

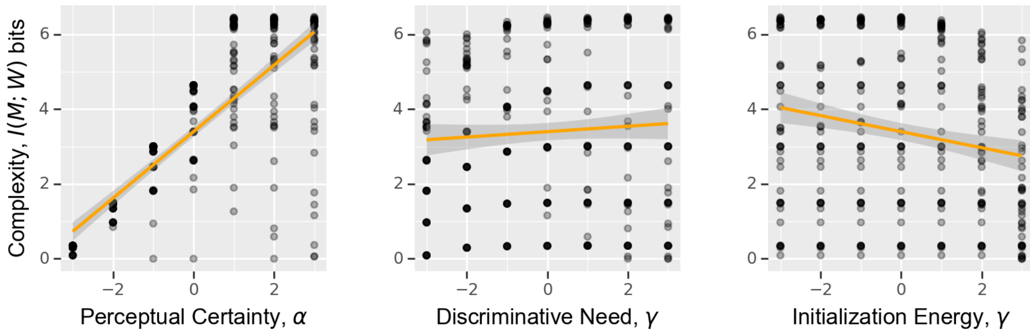

Figure 6: OLS regression fits of population initialization energy, discriminative need, and perceptual certainty on complexity. x-axis in log scale.

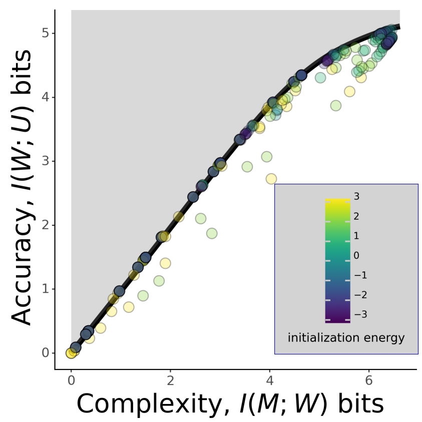

Figure 7: Emergent systems on the complexity/accuracy plane, colored according to initialization energy with blue being more random and yellow being deterministic. Initializations that are more random tend to produce simpler, and more efficient systems.

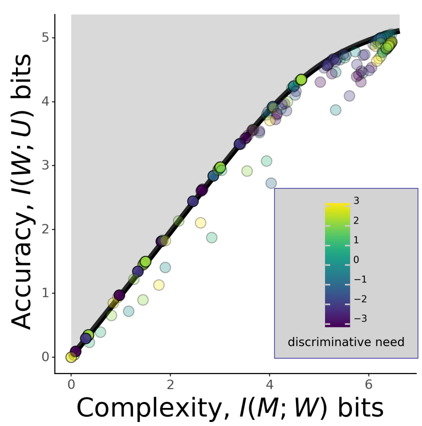

Figure 8: Emergent systems on the complexity/accuracy plane, colored according to discriminative need with blue corresponding to games with unlimited pragmatic slack, and yellow corresponding to games with all-or-nothing payoffs. Complexity does not appear to systematically change with discriminative need (at least when the other parameters are allowed to vary.)

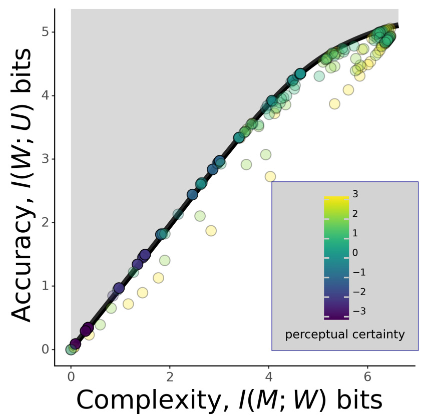

Figure 9: Emergent systems on the complexity/accuracy plane, colored according to perceptual certainty (note this is Fig. 2 repeated with different legend). Here, blue systems result from populations with very low certainty w.r.t. world state perception (i.e., a large amount of noise in the dynamics) and yellow systems result from populations with very high certainty (no noise). Perceptual noise tends to produce simpler, and more efficient systems.