

The Emergence of Honest Signaling¹

David Catteeuw (dcatteeu@vub.ac.be)
AI lab, Vrije Universiteit Brussel, Brussels, Belgium

The emergence of honest (or reliable) signaling is a multi-disciplinary problem. Linguists and philosophers [5, 8] have long wondered how conventions, such as human language, can emerge without a pre-existing language. Biologists [6] noticed that the many signals in nature can only exist because they are honest. Otherwise they would be ignored and so, not worth the trouble sending. Economists [7] created a real breakthrough by recognizing that many interactions are characterized by asymmetric information (where one party knows more than the other and signals may, or may not, reveal what he knows). It explains, for example, why the free market does not work for health insurance: those willing to buy costly insurance are most likely those who expect to need it the most. I contributed to this research in three domains.

1 Common Interest

One reason why signals are honest is common interest: both the sender and the receiver of the signal benefit from conveying the correct information. Under common interest, the only question that remains is how a signal acquires its meaning. One explanation that may also explain the origins of language is that this happens by chance [8]. My findings [2] support this idea: (a) I introduced a simple, adaptive heuristic, win-stay/lose-inaction, and proved that it always reaches honest signaling in all Lewis signaling games [5] (the standard game-theoretic model to study the emergence of signaling under common interest). The expected number of iterations is only polynomial in the number of signals. No such algorithm was known before. (b) I showed that three well-known reinforcement algorithms (Q-learning, Roth-Erev learning, and Learning Automata) mimic win-stay/lose-inaction and are even robust to errors.

2 Costly Signals

Economists and biologists independently discovered that when interests conflict signals may be honest if they are costly. This is known as the ‘handicap principle’ and is almost exclusively studied by means of *equilibrium* analyses. I applied evolutionary and learning *dynamics* in finite populations and found [4, 3]: (a) In many cases where honest signaling is an equilibrium, it

does not emerge: equilibrium analyses wrongfully predict honest signaling. (b) Dynamics reveal (partially) honest signaling in some cases where it is not an equilibrium: equilibrium analyses fail to predict (partially) honest signaling.

3 Costly, Social Punishment

I studied [1] the effect of costly, social punishment—already known to favor the evolution of cooperation—on four behaviors deviating from honest signaling: (a) When punishment targets lying individuals, honest signaling emerges also for cost-free signals. So, punishment provides an alternative to the ‘handicap principle.’ (b) When punishment targets greedy individuals, honest signaling emerges also in cases with strong conflicts, similar to the punishment of defectors to promote cooperation. (c) Punishment of timid or worried individuals does not benefit the evolution of honest signaling.

References

- [1] D. Catteeuw, T. A. Han, and B. Manderick. Evolution of Honest Signaling by Social Punishment. In *Proceedings of the 2014 Genetic and Evolutionary Computation Conference*, pages 1–8, Vancouver, Canada, 2014. ACM Press.
- [2] D. Catteeuw and B. Manderick. The Limits and Robustness of Reinforcement Learning in Lewis Signaling Games. *Connection Science*, 26(2):1–19, 2014.
- [3] D. Catteeuw and B. Manderick. Honesty and deception in populations of selfish, adaptive individuals. *The Knowledge Engineering Review*, 31(2):1–15, 2016.
- [4] D. Catteeuw, B. Manderick, and T. A. Han. Evolutionary Stability of Honest Signaling in Finite Populations. In *Proceedings of the IEEE Congress on Evolutionary Computation*, pages 2864–2870, Cancun, Mexico, 2013. IEEE Computer Society.
- [5] D. K. Lewis. *Convention: A Philosophical Study*. Harvard University Press, Cambridge, MA, USA, 1969.
- [6] J. Maynard Smith and D. Harper. *Animal signals*. Oxford University Press, Oxford, UK, 2003.
- [7] J. G. Riley. Silver signals: Twenty-five years of screening and signaling. *Journal of Economic Literature*, 39(2):432–478, 2001.
- [8] B. Skyrms. *Signaling: Evolution, Learning and Information*. Oxford University Press, New York, NY, USA, 2010.

¹This abstract discusses my upcoming PhD thesis. I especially thank my supervisor, B. Manderick, and my co-authors, T. A. Han and J. De Beule, for their contributions.