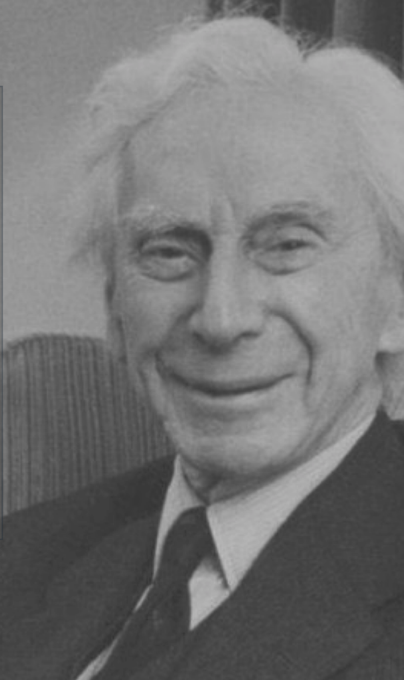
Models of Language Evolution

Evolutionary game theory & the evolution of meaning

Michael Franke

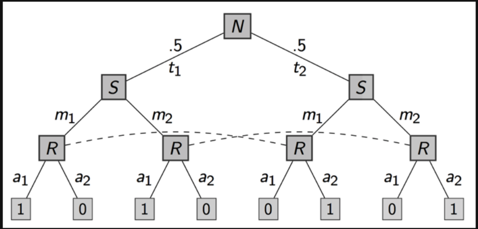
We can hardly suppose a parliament of hitherto speechless elders meeting together and agreeing to call a cow a cow and a wolf a wolf. The association of words with their meanings must have grown up ***by some natural process****, though at present the nature of the process is unknown.*

Bertrand Russell (1921) *The Analysis of Mind* p.190



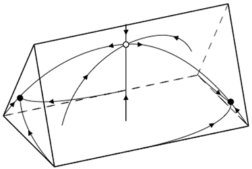
MEANING AS CONVENTION

equilibria of **signaling games**



David Lewis (1969) *Convention*

SIGNALING THEORY



evolutionary dynamics instead of equilibria

fitness-based selection **OR**
agent-level learning

meaning as **information content**

Brian Skyrms (2010) *Signals: Evolution, Learning, and Information*

Topics for today

1. evolutionary stability
2. meaning of signals
3. replicator dynamic

Population games

Evolutionary Stability

Meaning Evolution

(One-Population) Symmetric Game

A (one-population) symmetric game is a pair $\langle A, U \rangle$, where:

- A is a set of acts, and
- $U : A \times A \to \mathbb{R}$ is a utility function (matrix).

Example (Prisoner's dilemma)

$$U = \begin{array}{c} \\ a_c \\ a_d \end{array} \begin{array}{cc} a_c & a_d \\ \left(\begin{array}{cc} 2 & 0 \\ 3 & 1 \end{array}\right) \end{array}$$

Example (Hawk & Dove)

$$U = \begin{array}{c} \\ a_h \\ a_d \end{array} \begin{array}{cc} a_h & a_d \\ \left(\begin{array}{cc} 1 & 7 \\ 2 & 3 \end{array}\right) \end{array}$$

Symmetrizing asymmetric games

Example: signaling game

- big population of agents
- every agent might be sender or receiver
- an agent's strategy is a pair $\langle s, r \rangle$ of pure sender and receiver strategies
- utilities are defined as the average of sender and receiver role:

$$U(\langle s, r \rangle, \langle s', r' \rangle) = 1/2(U_S(s, r') + U_R(s', r)))$$

Example (Symmetrized 2-2-2 Lewis game)

		s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}
s_1	$\langle m_1, m_1, a_1, a_1 \rangle$	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
s_2	$\langle m_1, m_1, a_1, a_2 \rangle$	.5	.5	.5	.5	.75	.75	.75	.75	.25	.25	.25	.25	.5	.5	.5	.5
s_3	$\langle m_1, m_1, a_2, a_1 \rangle$	.5	.5	.5	.5	.25	.25	.25	.25	.75	.75	.75	.75	.5	.5	.5	.5
s_4	$\langle m_1, m_1, a_2, a_2 \rangle$	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
s_5	$\langle m_1, m_2, a_1, a_1 \rangle$	.5	.75	.25	.5	.5	.75	.25	.5	.5	.75	.25	.5	.5	.75	.25	.5
s_6	$\langle m_1, m_2, a_1, a_2 \rangle$	.5	.75	.25	.5	.75	1	.5	.75	.25	.5	0	.25	.5	.75	.25	.5
s_7	$\langle m_1, m_2, a_2, a_1 \rangle$	.5	.75	.25	.5	.25	.5	0	.25	.75	1	.5	.75	.5	.75	.25	.5
s_8	$\langle m_1, m_2, a_2, a_2 \rangle$	.5	.75	.25	.5	.5	.75	.25	.5	.5	.75	.25	.5	.5	.75	.25	.5
s_9	$\langle m_2, m_1, a_1, a_1 \rangle$	.5	.25	.75	.5	.5	.25	.75	.5	.5	.25	.75	.5	.5	.25	.75	.5
s_{10}	$\langle m_2, m_1, a_1, a_2 \rangle$	.5	.25	.75	.5	.75	.5	1	.75	.25	0	.5	.25	.5	.25	.75	.5
s_{11}	$\langle m_2, m_1, a_2, a_1 \rangle$	.5	.25	.75	.5	.25	0	.5	.25	.75	.5	1	.75	.5	.25	.75	.5
s_{12}	$\langle m_2, m_1, a_2, a_2 \rangle$	.5	.25	.75	.5	.5	.25	.75	.5	.5	.25	.75	.5	.5	.25	.75	.5
s_{13}	$\langle m_2, m_2, a_1, a_1 \rangle$	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
s_{14}	$\langle m_2, m_2, a_1, a_2 \rangle$	.5	.5	.5	.5	.75	.75	.75	.75	.25	.25	.25	.5	.5	.5	.5	.5
s_{15}	$\langle m_2, m_2, a_2, a_1 \rangle$	.5	.5	.5	.5	.25	.25	.25	.25	.75	.75	.75	.75	.5	.5	.5	.5
s_{16}	$\langle m_2, m_2, a_2, a_2 \rangle$	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5

Population games

Evolutionary Stability

Meaning Evolution

Mean-Field Population

- (nearly) infinite populations for each distinguishable role
- each population is entirely homogeneous
- agents play pure strategies
- each agent interacts purely at random with other agents
- strategy updates are rare

Evolutionary Stability (Intuition)

A strategy s is **evolutionarily stable** if a population that consists entirely/mostly of s-agents (the incumbents) cannot be invaded by any minority of mutants/invaders playing strategy t.

Evolutionary Stability (Derivation)

Intuition

s cannot be invaded by a minority of mutants t

$$\text{fitness of incumbent} > \text{fitness of mutant}$$
$$(1-\epsilon)\,U(s,s) + \epsilon\,U(s,t) > (1-\epsilon)\,U(t,s) + \epsilon\,U(t,t)$$

- if ϵ is infinitesimal, this holds when $U(s,s) > U(t,s)$
- but if $U(s,s) = U(t,s)$, then it also holds when $U(s,t) > U(t,t)$

Evolutionarily Stable Strategy (Definition)

A strategy s is **evolutionarily stable** iff for all t:

(i) $U(s,s) > U(t,s)$, or

(ii) $U(s,s) = U(t,s)$ and $U(s,t) > U(t,t)$.

Connection with NE

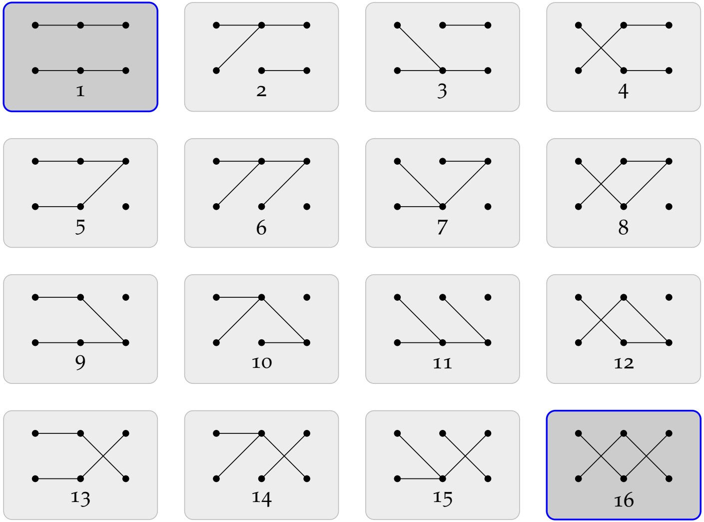
- strict-NEs $\subset$ ESSs $\subset$ NEs

Example (Symmetrized 2-2-2 Lewis game)

		s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}
s_1	$\langle m_1, m_1, a_1, a_1 \rangle$	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
s_2	$\langle m_1, m_1, a_1, a_2 \rangle$	.5	.5	.5	.5	.75	.75	.75	.75	.25	.25	.25	.25	.5	.5	.5	.5
s_3	$\langle m_1, m_1, a_2, a_1 \rangle$	.5	.5	.5	.5	.25	.25	.25	.25	.75	.75	.75	.75	.5	.5	.5	.5
s_4	$\langle m_1, m_1, a_2, a_2 \rangle$	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
s_5	$\langle m_1, m_2, a_1, a_1 \rangle$	.5	.75	.25	.5	.5	.75	.25	.5	.5	.75	.25	.5	.5	.75	.25	.5
s_6	$\langle m_1, m_2, a_1, a_2 \rangle$	.5	.75	.25	.5	.75	1	.5	.75	.25	.5	0	.25	.5	.75	.25	.5
s_7	$\langle m_1, m_2, a_2, a_1 \rangle$	.5	.75	.25	.5	.25	.5	0	.25	.75	1	.5	.75	.5	.75	.25	.5
s_8	$\langle m_1, m_2, a_2, a_2 \rangle$	.5	.75	.25	.5	.5	.75	.25	.5	.5	.75	.25	.5	.5	.75	.25	.5
s_9	$\langle m_2, m_1, a_1, a_1 \rangle$	.5	.25	.75	.5	.5	.25	.75	.5	.5	.25	.75	.5	.5	.25	.75	.5
s_{10}	$\langle m_2, m_1, a_1, a_2 \rangle$	.5	.25	.75	.5	.75	.5	1	.75	.25	0	.5	.25	.5	.25	.75	.5
s_{11}	$\langle m_2, m_1, a_2, a_1 \rangle$	.5	.25	.75	.5	.25	0	.5	.25	.75	.5	1	.75	.5	.25	.75	.5
s_{12}	$\langle m_2, m_1, a_2, a_2 \rangle$	.5	.25	.75	.5	.5	.25	.75	.5	.5	.25	.75	.5	.5	.25	.75	.5
s_{13}	$\langle m_2, m_2, a_1, a_1 \rangle$	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
s_{14}	$\langle m_2, m_2, a_1, a_2 \rangle$	.5	.5	.5	.5	.75	.75	.75	.75	.25	.25	.25	.5	.5	.5	.5	.5
s_{15}	$\langle m_2, m_2, a_2, a_1 \rangle$	.5	.5	.5	.5	.25	.25	.25	.25	.75	.75	.75	.75	.5	.5	.5	.5
s_{16}	$\langle m_2, m_2, a_2, a_2 \rangle$	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5

non-strict symmetric NE, ESS

All pairs of sender-receiver pure strategies for the 2-2-2 Lewis game



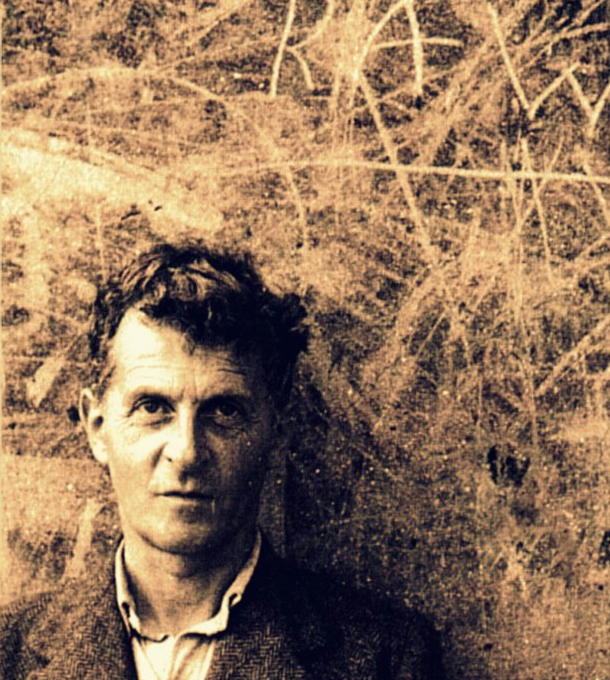
ESSs

Population games

Evolutionary Stability

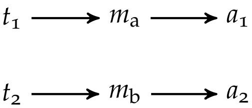
Meaning Evolution

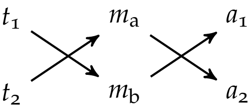
“Wenn sich alles so verhält, als hätte ein Zeichen Bedeutung, dann hat es auch Bedeutung.”



Meaning in Lewis games

Signaling systems of the 2-2-2 Lewis game





Fix an n-n-n Lewis game with SIGSYS $\langle s, r \rangle$ (i.e., ESS), and define:

indicative meaning

$$[\![m]\!]^T = \{t \in T \mid s(t) = m\}$$

imperative meaning

$$[\![m]\!]^A = \{a \in A \mid r(m) = a\}$$

(Lewis, 1969)

Natural vs. non-natural meaning

Natural meaning

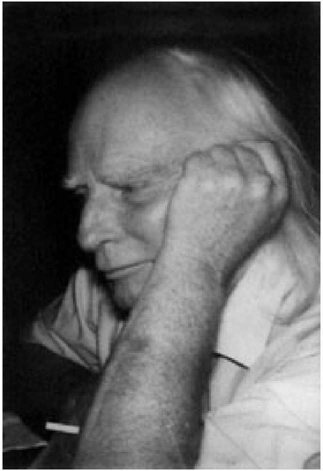
E.g.: smoke means fire

Non-natural meaning

E.g.: this gesture meant that the party is boring

Non-natural meaning: Grice's definition

"*A* meant$_{NN}$ something by x" is roughly equivalent to "*A* uttered x with the intention of inducing a belief [in his audience] by means of the recognition of this intention."



(Grice, 1957)

The Herod examples

(1) Herod presents Salome with the head of St. John the baptist.

(2) Herod says to Salome "He's dead."

(3) Herod leaves the head somewhere; Salome happens to see it.

(4) Herod leaves the head where he knows Salome will see it, correctly supposing she will not realize he left it for her to see.

(5) Herod leaves the head where Salome will see it, mistakenly supposing she will not realize he left it for her to see.

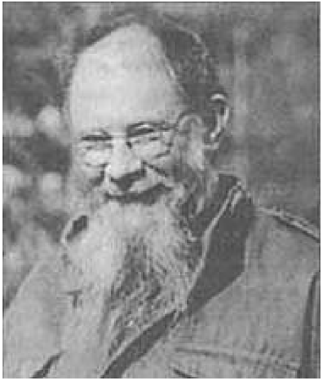
(Grice, 1957)

Meaning$_{NN}$ in signaling systems

Behavior in a SigSys is compatible with common belief in rationality.

We can then construct an infinite chain of rational intention recognition based on SigSys-behavior.

Meaning in SigSyss can be construed as meaning$_{NN}$ if the ascription of relevant mental states to agents is warranted.



(Lewis, 1969)

Practical reasoning justification in a SigSys

"Suppose I am the communicator and you are the audience (...) and having observed that t_1 holds, I give m_a in conformity to our convention. (...)

The intention with which I do m_a can be established by examining the practical reasoning that justifies me in doing it. I need not actually go through that reasoning to have an intention; actions done without deliberation are often done with definite intentions. (...)

My decision to do m_a, having observed t_1, is premised on my expectation that I can thereby produce a_1 and on my desire to produce a_1. So I do m_a with the intention to produce a_1.

I expect you to infer t_1 upon observing that I do m_a. I expect you to recognize my desire to produce a_1, conditionally upon t_1. I expect you to recognize my expectation that I can produce a_1 by doing m_a. So I expect you to recognize my intention to produce a_1, when you observe that I do m_a. (...)"

(Lewis, 1969, p.155)

Informational content of signals

Info-content of m about T

difference between:

info about T given m &

info about T without m

Merits

- applies to out-of-equilibrium behavior as well
- non-intentional, non-mentalistic
- applies to information flow between non-cognizing agents

(Skyrms, 2010, Chapter 3)

Informational content of signals

Info-content of m about T

difference between:

- info about T given m [$= P(t \mid m)$ posterior after signal reception]
- info about T without m [$= P(t)$ prior probability of states]

Kullback-Leibler divergence

Let $P, Q \in \Delta(X)$ be probability distributions over finite set X, then:

$$\mathrm{KL}(P \mid\mid Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$$

is the Kullback-Leibler divergence (measuring how many bits of information we would miss if we relied on Q rather than the (true) distribution P).

(Skyrms, 2010, Chapter 3)

Informational content of signals

Info-content of m about T

$$\mathrm{I}(m) = \sum_{t \in T} P(t \mid m) \log \frac{P(t \mid m)}{P(t)}$$

Informational content vector of m

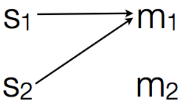
$$\mathrm{ICV}(m) = \left\langle \log \frac{P(t \mid m)}{P(t)} \mid t \in T \right\rangle$$

Propositional content of m

$$\mathrm{Prop}(m) = \left\{ t \in T \mid \log \frac{P(t \mid m)}{P(t)} > -\infty \right\}$$

(Skyrms, 2010, Chapter 3)

Examples



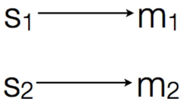
$$\left\langle \log\left[\frac{\Pr(s_1|m_1)}{\Pr(s_1)}\right], \log\left[\frac{\Pr(s_2|m_1)}{\Pr(s_2)}\right]\right\rangle = \left\langle \log\left[\frac{1/2}{1/2}\right], \log\left[\frac{1/2}{1/2}\right]\right\rangle$$

$$= \langle \log[1], \log[1]\rangle$$

$$= < 0, 0 >$$

Take the weighted sum to find the KL-divergence, and you'll see that m_1 carries 0 bits of information

---

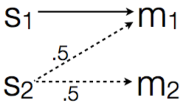


$$\left\langle \log\left[\frac{\Pr(s_1|m_1)}{\Pr(s_1)}\right], \log\left[\frac{\Pr(s_2|m_1)}{\Pr(s_2)}\right]\right\rangle = \left\langle \log\left[\frac{1}{1/2}\right], \log\left[\frac{0}{1/2}\right]\right\rangle$$

$$= \langle \log[2], \log[0]\rangle$$

$$= < 1, -\infty >$$

KL-divergence: m_1 carries 1 bits of information



$$\left\langle \log\left[\frac{\Pr(s_1|m_1)}{\Pr(s_1)}\right], \log\left[\frac{\Pr(s_2|m_1)}{\Pr(s_2)}\right]\right\rangle = \left\langle \log\left[\frac{3/4}{1/2}\right], \log\left[\frac{1/4}{1/2}\right]\right\rangle$$

$$= \langle \log[3/2], \log[1/2]\rangle$$

$$= < .58, -1 >$$

References

Grice, Paul Herbert (1957). "Meaning". In: *Philosophical Review* 66.3, pp. 213–223.

Kirby, Simon et al. (2014). "Iterated Learning and the Evolution of Language". In: *Current Opinion in Neurobiology* 28, pp. 108–114.

Lewis, David (1969). *Convention. A Philosophical Study*. Cambridge, MA: Harvard University Press.

Skyrms, Brian (2010). *Signals: Evolution, Learning, and Information*. Oxford: Oxford University Press.