The evolution of weak bidirectional OT

Workshop
Games and Decisions in Pragmatics
October 24-25, 2003
ZAS Berlin

Gerhard Jäger
University of Potsdam
jaeger@ling.uni-potsdam.de
www.ling.uni-potsdam.de/~jaeger/

1. Overview

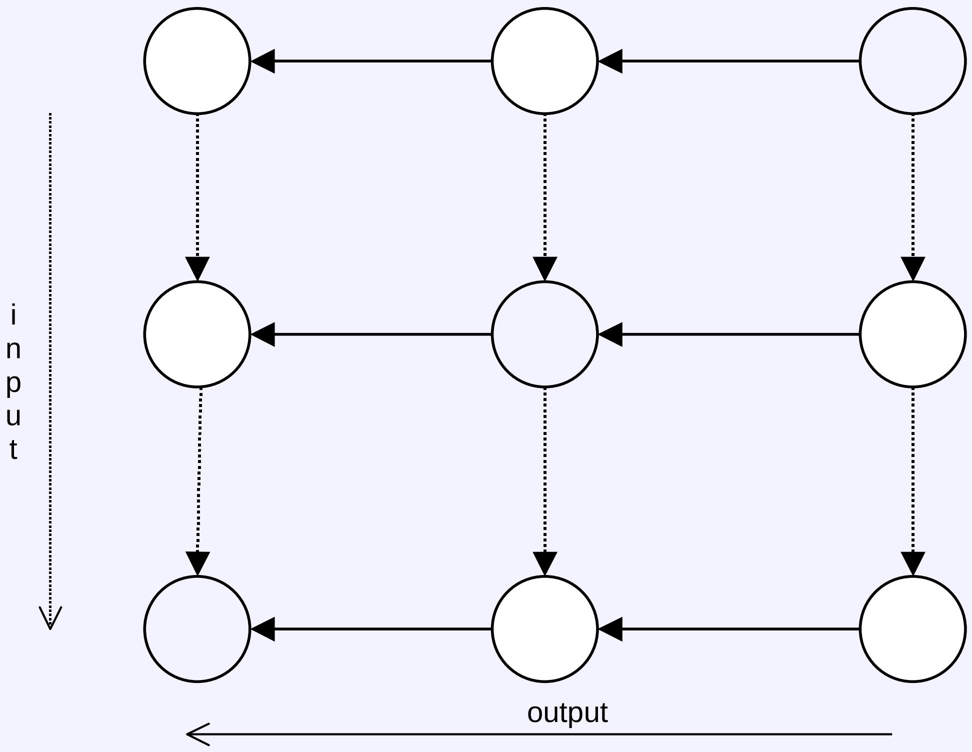
- Weak bidirectional OT: Synchrony and diachrony
- Game theoretic formalization
- Evolutionary Game Theory
- weak bidirectionality and evolutionary stability
- stochastic stability

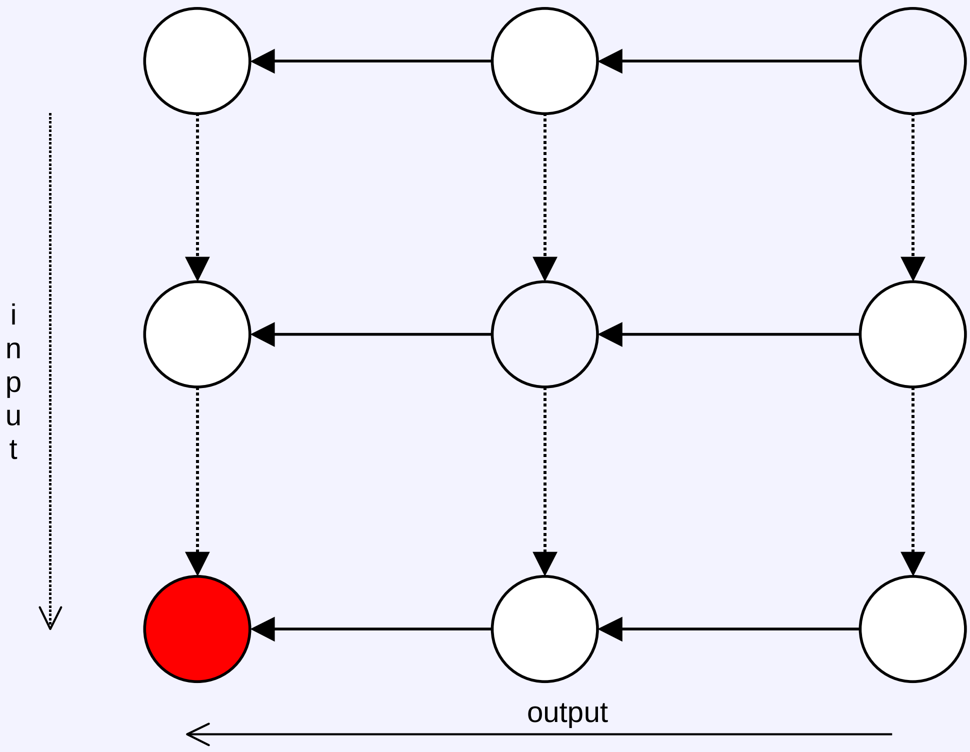
2. Weak Bidirectionality

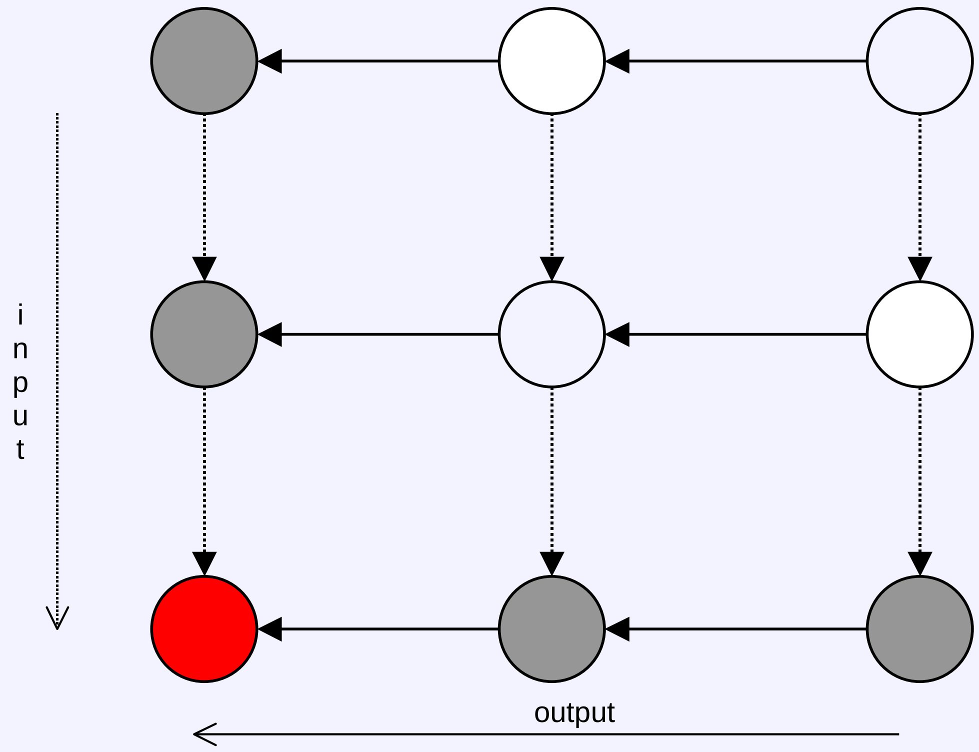
Definition 1 (Weak bidirectional optimality) *Let* $\mathcal{O} = \langle \textbf{GEN}, \textbf{CON} \rangle$ *be an OT-system. Then* $\langle i, o \rangle$ *is bidirectionally optimal iff*

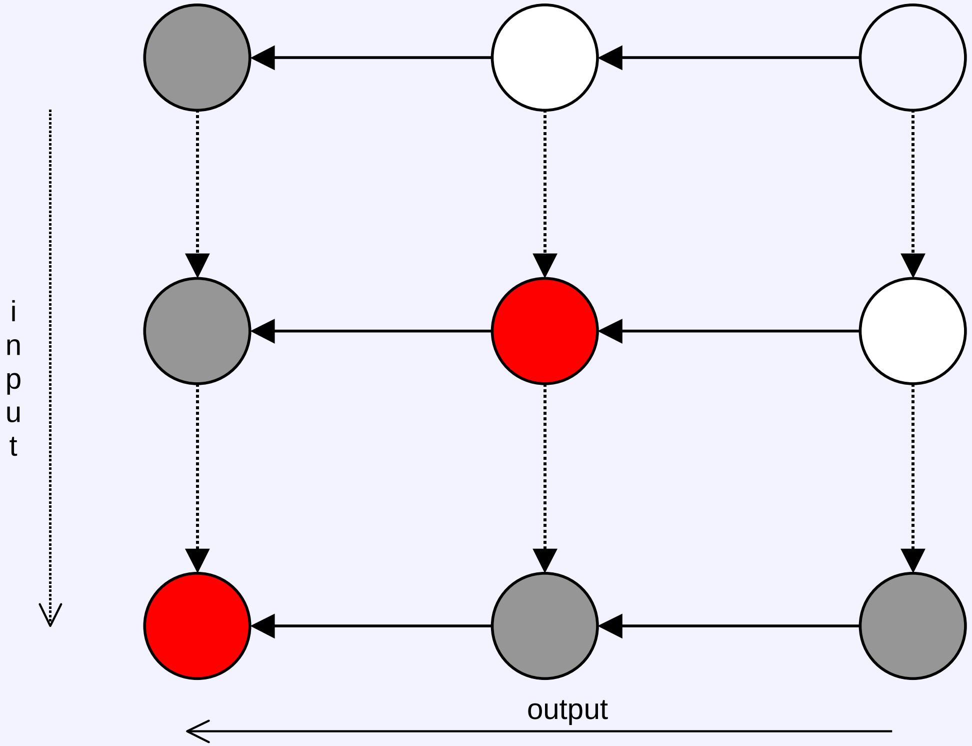
1. $\langle i, o \rangle \in$ **GEN**,
2. *there is no bidirectionally optimal* $\langle i', o \rangle \in$ **GEN** *such that* $\langle i', o \rangle \prec_{\mathcal{O}} \langle i, o \rangle$, *and*
3. *there is no bidirectionally optimal* $\langle i, o' \rangle \in$ **GEN** *such that* $\langle i, o' \rangle \prec_{\mathcal{O}} \langle i, o \rangle$.

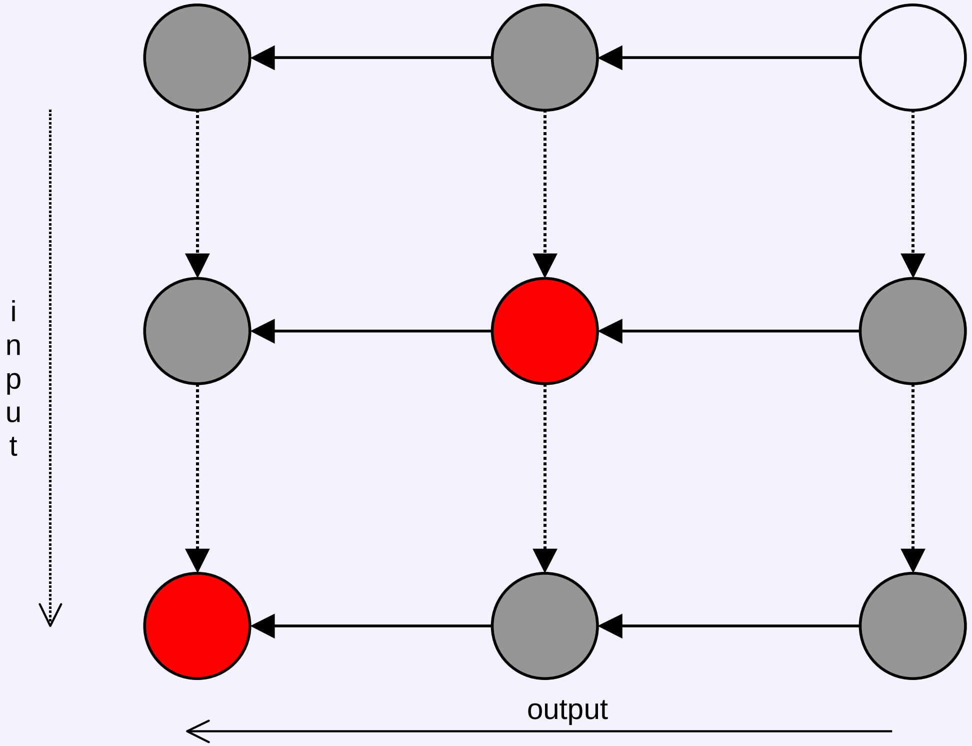
- predicts iconicity:
  - simple forms go with simple meanings
  - complex forms go with complex meanings

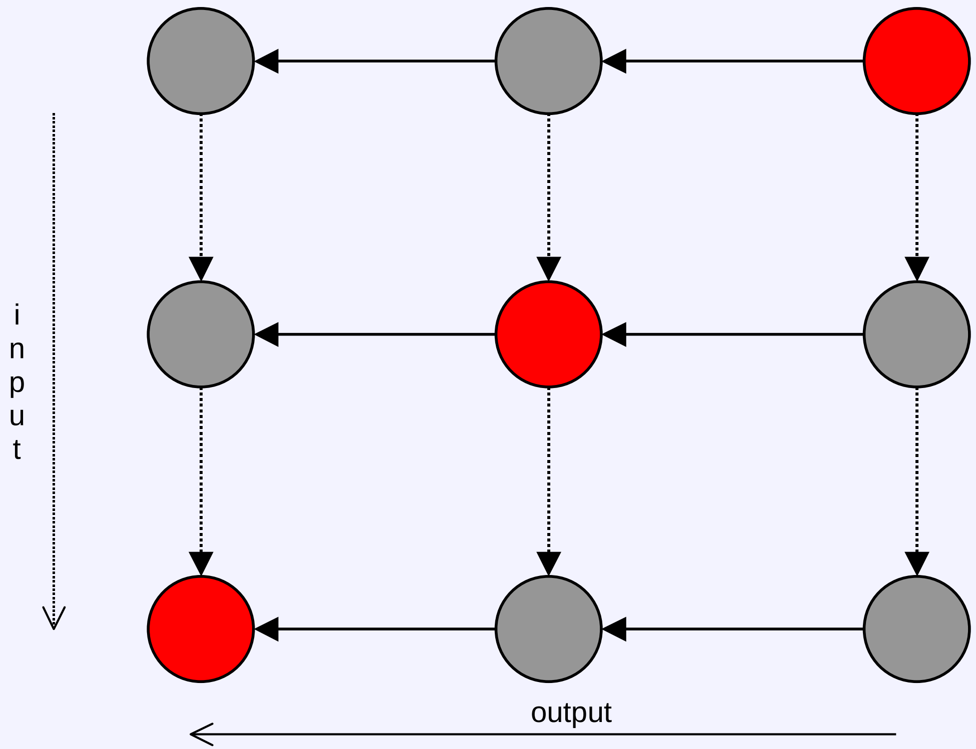

input
output


input
output


input
output


input
output


input
output



input
output

- not a synchronic rule:
  - *woman eats banana* $\prec$ *banana eats woman*
  - accusative case $\prec$ dative case
  - for feminine NPs in German, nominative = accusative
  - Still, both (1a) and (b) are translated as (2a), and (2b) is ungrammatical

(1) a. the banana that the woman eats
    b. the banana that eats the woman

(2) a. die Banane die die Frau isst
    THE BANANA WHICH[NOM/ACC] THE WOMAN[NOM/ACC] EATS
    b. *die Banane der die Frau isst
    THE BANANA WHICH[DAT] THE WOMAN[NOM/ACC] EATS

- But it does work in many cases!
- possible explanation (Benz, Blutner, Mattausch, van Rooy, ...):
  - Weak bidirectionality is not a synchronic rule but expresses a diachronic tendency
  - weakly bidirectional pairs are *evolutionary stable*
- possible formalization by means of *Evolutionary Game Theory*
- van Rooy: for 2-form-2-meaning games weak bidirectionality is in fact the only attractor

3. Game theoretic formalization

- (finite) sets M (meanings) and F (forms)
- relation **GEN**$\subseteq M \times F$
- two players (speaker and hearer)
- speaker strategy: function $S \subseteq \mathbf{GEN}$ from M to F
- hearer strategy: function $H \subseteq \mathbf{GEN}^{-1}$ from F to M

- speaker has to decide *what to say* and *how to say it*
- only latter decision is linguistically relevant
- idealization:
  - in each game, nature presents the speaker with a meaning m
  - speaker only has to decide how to express m
  - nature chooses meanings according to probability distribution p over M

Utilities

- hearer tries to decode intention of speaker from observed form
- speaker tries to communicate meaning with little effort
- measure of communicative success:

$$\delta_m(S,H) = \begin{cases} 1 & \text{iff } \; H(S(m)) = m \\ 0 & \text{else} \end{cases}$$

- hearer's only interest is to get the interpretation right:

$$u_h(m,S,H) \;=\; \delta_m(S,H)$$

- complexity of forms measured by means of function

$$\textit{cost} : F \mapsto (0, \infty)$$

- speaker has conflicting interest:
  - communicative success
  - little effort
- captured by speaker utility function

$$u_s(m, S, H) \ = \ \delta_m(S, H) - k \times \textit{cost}(S(m))$$

- k: positive coefficient that captures the preferences of the speaker
- present talk: k is always infinitesimally small

Average utilities

- averaging over many utterance situations:

$$
\begin{aligned}
u_s(S,H) &= \sum_m p_m \times (\delta_m(S,H) - k \times \textsf{\textit{cost}}(S(m))) \\
u_h(S,H) &= \sum_m p_m \times \delta_m(S,H)
\end{aligned}
$$

Communication as an asymmetric partnership game

- Note that strategy sets of speaker and hearer are disjoint!
- Communication is thus an asymmteric game
- speaker utility matrix and hearer utility matrix only differ by $-k \times$ *cost*$(S(m))$
- depends only on speaker strategy; hearer has no influence on it
- replacing u_h by u_s does not change the decision situation for hearer
- communication can be seen as *partnership game*
- revised utility function

$$u_s(S,H) = u_h(S,H) = \sum_m p_m \times (\delta_m(S,H) - k \times \textit{cost}(S(m)))$$

4. Evolutionary Game Theory

- two populations of players (in asymmetric two-person game)
- each individual is programmed for a strategy
- strategies with a high average utility increase their share of the population over time
- *Evolutionary Stable Strategy pair* (ESS):
  - stationary
  - immune against small invasions of mutant strategies

Evolutionary stability in asymmetric games

Definition 2 (Strict Nash Equilibrium) *A pair of strategies* (s, h) *is a* Strict Nash Equilibrium *iff*

$$\forall s'(s' \neq s \rightarrow u_s(s,h) > u_s(s',h))$$

and

$$\forall h'(h' \neq h \rightarrow u_h(s,h) > u_h(s,h'))$$

Theorem 1 (Reinhard Selten) (s, h) *is evolutionary stable if and only if it is a Strict Nash Equilibrium.*

- Remark: in asymmetric games only pure strategies can form Strict Nash Equilibria, so we can safely disregard mixed strategies

Bijections are evolutionary stable

- Suppose $|F| = |M|$.
- Then $\langle s, h \rangle$ is a Strict Nash Equilibrium iff
  - s and h are 1-1 maps, and
  - $s = h^{-1}$

Sketch of proof:

- $\Rightarrow$
  - suppose $\langle s, h\rangle$ is a SNE
  - then every $f \in F$ must be contained in range of s — otherwise every $h' \sim_m h$ would have the same utility as h
  - thus s is 1-1
  - thus no hearer strategy can be bettern than s^{-1}
- $\Leftarrow$
  - suppose s and h are 1-1 maps, and $s = h^{-1}$
  - every unilateral deviation would decrease average communicative success

5. Comparison

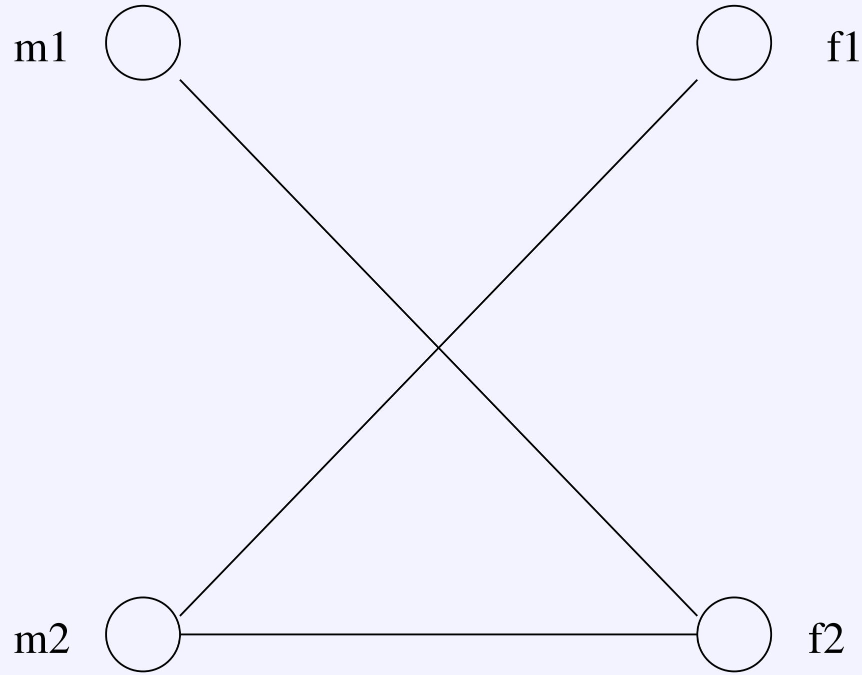
- weak bidirectionality also tends to favor bijective maps
- but how to relate GT-utilities and OT?
  - OT ordering of forms corresponds to GT costs
  - OT ordering of meanings corresponds to amount of information (in the sense of information theory)

$$\langle m_1, f_1 \rangle \; < \; \langle m_2, f_2 \rangle$$

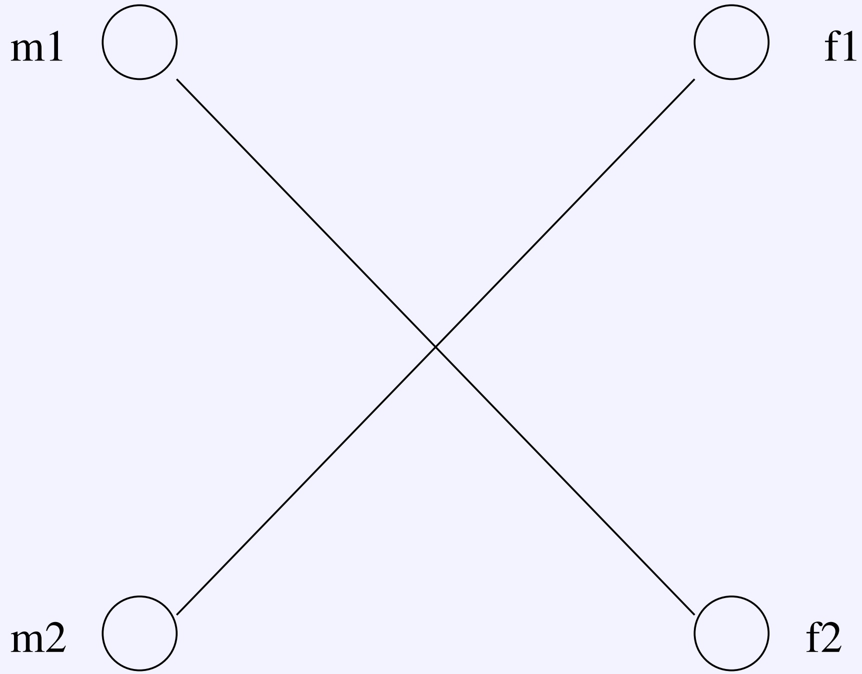
iff

$$-\log(p_{m_1}) \times \textit{cost}(f_1) \; < \; -\log(p_{m_2}) \times \textit{cost}(f_2)$$

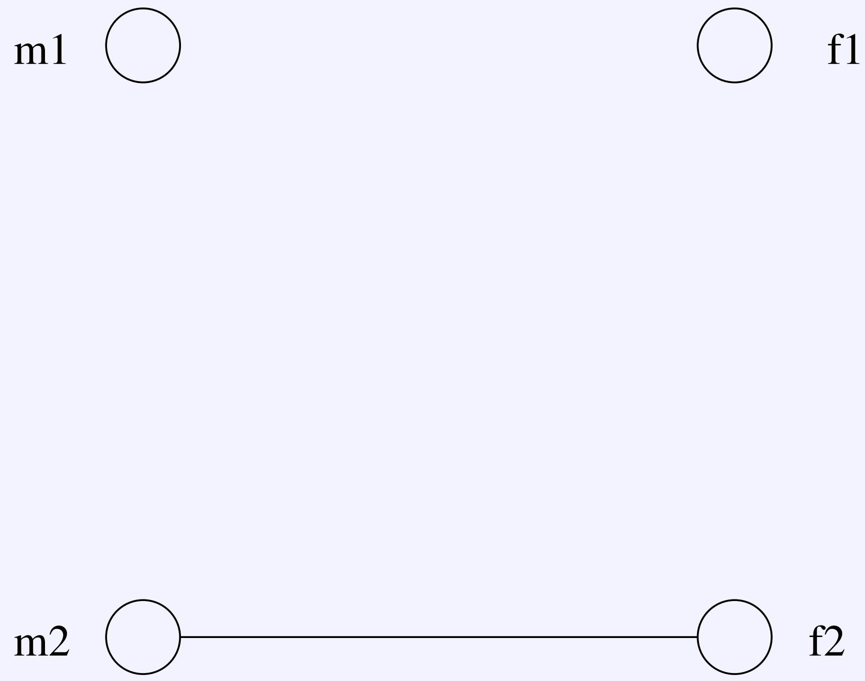
- suppose $p(m2) > p(m1)$, and $\textit{cost}(f2) < \textit{cost}(f1)$
- $\mathbf{GEN} = M \times F - \{m1, f1\}$
- graphically:



- There is only one 1-1 map contained in **GEN**, hence this is the only ESS
- prediction of EGT:



- weak bidirectionality predicts an incomplete map



6. Stochastic stability

(developed by Kandory, Mailath and Rob 1993 and Young 1993 in economics)

- EGT usually predicts several ESS
- "evolutionary stable" means "there is an invasion barrier"
- invasion barriers of multiple ESS are usually of varying height
- in finite populations, every invasion barrier is occasionally taken
- "jumping over" low barriers is more likely than jumping over high barriers
- hence system is most likely in the state with the highest invasion barrier
- this likelihood goes to 1 as the probability of a single mutation goes to 0

A state is *stochastically stable* if its probability converges to a positive value if the mutation probability goes to 0.

- In a 2×2 game, the risk-dominant Strict Nash Equilibrium is the only stochastically stable state (KMR 1993)
- partnership games: risk dominance $=$ Pareto efficiency
- no general recipes for games with more than two strategies per player
- *Conjecture:* in partnership games, Pareto-efficiency and stochastic stability coincide

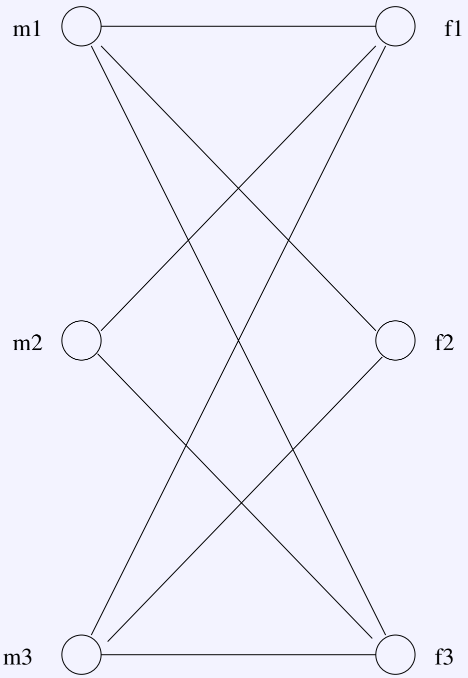
Stochastic stability and weak bidirectionality

- van Rooy 2002: in simple 2-form-2-meaning game, stochastic stability and weak bidirectionality coincide
- Does this generalize?
  - above example proves the opposite — if there is only one ESS, it is stochastically stable
  - but what if weak bidirectionality is a bijection?

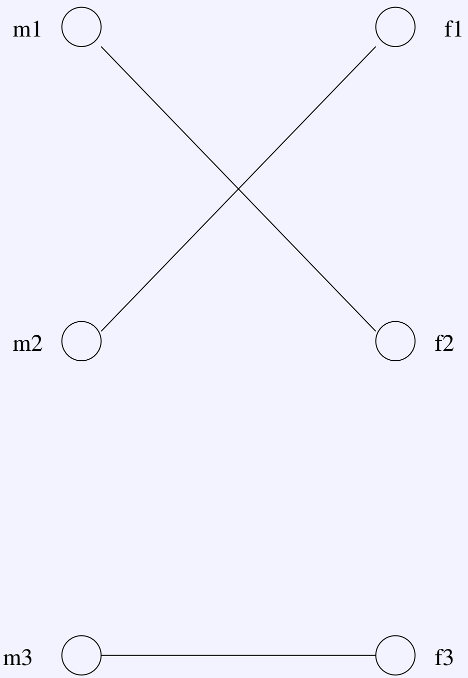
Even then weak bidirectionality and stochastic stability need not coincide:

- $M = \{m1, m2, m3\}$
- $F = \{f1, f2, f3\}$
- $\mathbf{GEN} = M \times F - \{\langle m2, f2 \rangle\}$
- $p_{m1} = 0.1, p_{m2} = 0.4, p_{m3} = 0.5$
- $\textit{cost}(f1) = 20, \textit{cost}(f2) = 11, \textit{cost}(f3) = 10$

- Generator:

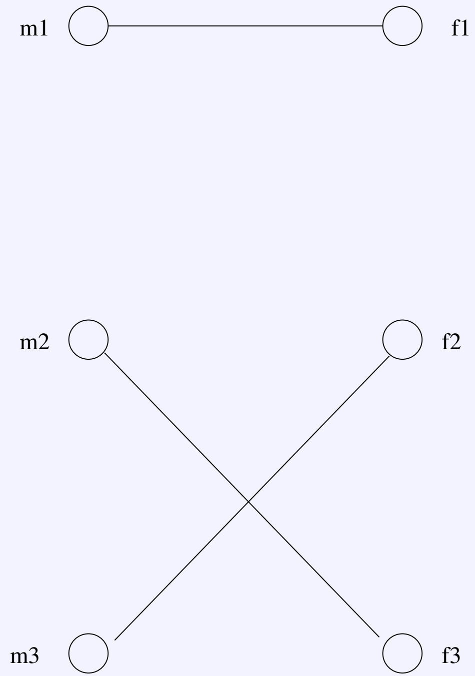


- weakly bidirectional map:



$$u(S, H) = 1 - k \times 14.1$$

- Pareto-efficient (and thus stochastically stable) state



$$u(S, H) = 1 - k \times 11.5$$

- examples all involved deficient **GEN**
- suppose $|F| = |M|$ and $\mathbf{GEN} = M \times F$
- suppose furthermore that there are no ties:

$$\forall m_1, m_2 : p(m_1) = p(m_2) \rightarrow m_1 = m_2$$
$$\forall f_1, f_2 : \textit{cost}(f1) = \textit{cost}(f2) \rightarrow f_1 = f_2$$

- then the isomorphic map (most frequent meaning goes with least costly form etc) is both Pareto-efficient and weakly bidirectionally optimal

7. Conclusion

- initial hypothesis: weak bidirectionality is a diachronic attractor
- formalized in terms of EGT
- first result: in EGT all 1-1 maps between forms and meanings are evolutionary stable
- refinement: stochastic evolution
- conjecture: exactly the Pareto-efficient 1-1-maps are the stochastically stable states
- weak bidirectionality and stochastic stability are guaranteed to coincide only under rather restrictive side conditions
- future work:
  - proof of the conjecture on stochastic stability and Pareto-efficiency
  - refined GT formalization of communication — beyond simple partnership games

Contents