Equilibrium Selection in Cheap Talk Games: De Groot Ruiz, Offerman and Ondeerstal (2015)

Alec Tolentino (Tu er li)

Summary

Introduction

Cheap Talk Games

- Talk between two players (sender/Receiver)
- Communication doesn’t directly affect payoffs
- One player has information, other reacts
- Communication is (1) costless to transmit and receive (2) non-binding (3) unverifiable by a third party

Average Credibility Deviation Criterion (ACDC)

- ACDC is used to predict behavior in a wide range of cheap talk applications
- Allows for continuous interpretations, not just binary
- Makes predictions for cheap talk games where other models fail
- Assumes stability of an equilibrium is a decreasing function of its Average Credible Deviation (ACD)
- Lower ACD = better performing equilibrium
- Human behavior is not binary

Motivation

Motivation

Failure of other models to accurately predict a new cheap talk game for two reasons

1. Selection

2. Stability

Table 1

Game A.

	a_1	a_2	a_3	a_4	a_5
t_1 $(\frac{1-\delta}{2})$	1, 4	0, 0	0, 0	0, 0	2, $4-\delta$
t_2 $(\frac{1-\delta}{2})$	0, 0	0, $2+\delta$	3, 0	4, 2	2, 1
t_3 (δ)	0, 0	0, 0	$2+\varepsilon$, 3	2, 2	1, 1

Notes: The left column shows the Sender's type and between brackets the probability that it is drawn. The top row shows the Receiver's actions. The remaining cells provide the Sender's payoff in the first entry and the Receiver's payoffs in the second entry. $0<\delta<\frac{1}{3}$ and $0\leq\varepsilon<1$.

Two equilibrium outcomes

1. Pooling - all senders induce a_5
2. Partial separating equilibrium - t_1 induces a1 while t_2 and t_3 induce a_4

What do credible neologisms (Farrell, 1993) do in this game?

Neologisms (Farrell, 1993)

Out-of-equilibrium messages which are assumed to have a literal meaning in pre-existing natural language

Neologism is credible if and only if :

i.) All types t in N prefer $\tilde{a}$ *to their equilibrium action α(t)*

ii) All types t not in N prefer their equilibrium action α(t) to $\tilde{a}$*, and*

iii) The best reply of the Receiver after restricting the support of his prior to N is to play $\tilde{a}$

We will denote neologisms by $[\tilde{a}, N]$

Table 1
Game A.

	a_1	a_2	a_3	a_4	a_5
t_1 ($\frac{1-\delta}{2}$)	1, 4	0, 0	0, 0	0, 0	2, $4-\delta$
t_2 ($\frac{1-\delta}{2}$)	0, 0	0, $2+\delta$	3, 0	4, 2	2, 1
t_3 (δ)	0, 0	0, 0	$2+\varepsilon$, 3	2, 2	1, 1

Notes: The left column shows the Sender's type and between brackets the probability that it is drawn. The top row shows the Receiver's actions. The remaining cells provide the Sender's payoff in the first entry and the Receiver's payoffs in the second entry. $0<\delta<\frac{1}{3}$ and $0\leq\varepsilon<1$.

Neologism in Game A

- If ε = 0 : The pooling equilibrium admits the credible neologism ⟨*a*4,{*t*2,*t*3}⟩ (partially separating equilibrium is stable)
- If ε > 0 : The partially separating equilibrium also admits a credible neologism, so no stable equilibrium

Definition and Properties

$$\text{For each } t \in T,\ \mu(t) \in \arg\max_{m \in M} U^S(\alpha(m), t)$$

$$\text{For each } m \in M,\ \alpha(m) \in \arg\max_{a \in A} \int_T U^R(a, t)\beta(t|m)dt \qquad (1)$$

where $\beta(t|m)$ denotes the Receiver's posterior beliefs, which is derived from μ and β^0 using Bayes' rule wherever possible.

- *UR* : *A*×*T* →R be the utility function of the Receiver
- *US* : *A*×*T* →R that of the Sender
- Strategy function for Sender μ : *T* → *M*
- Strategy function for Receiver α : *M* → *A*
- Let {μ,α} be a strategy profile (Σ)
- These conditions form a "pure strategy perfect Bayesian equilibrium" -

Deviating Profile

Associate a deviating profile with each equilibrium σ ={μ,α,β}

$$CD(t,\sigma) \equiv \frac{U^S\left(t, \alpha^{\gamma(\sigma)}(\mu^{\gamma(\sigma)}(t))\right) - U^S(t, \alpha(\mu(t)))}{\bar{U}^S(t) - \underline{U}^S(t)}$$

CD has desirable properties

Invariant to affine transformations of payoffs:

Increasing in the difference between deviating and equilibrium payoff

ACD of an equilibrium σ : $ACD(\sigma) = E_t\,[CD(t, \sigma)]$

Definition 1. An equilibrium $\sigma*$ is an ACDC equilibrium if $ACD(\sigma*) \leq ACD(\sigma)$ for all $\sigma \in \Sigma*$

Proposition 1. *If the number of equilibrium outcomes is finite, the cheap talk game has an ACDC equilibrium.*

Proposition 2. *Let s be an equilibrium outcome and $ACD(s)$ the ACD of equilibria inducing s. Suppose the equilibrium outcome set S can be represented by a finite union of compact metric spaces $S = S$, such that $ACD(s)$ is continuous in s on all subsets S . Then, an ACDC equilibrium exists.*

Proof. $ACD(s)$ achieves a minimum on each compact subset Si and thus on S. Hence, min $*$ $ACD(\sigma)$ is nonempty and an ACDC equilibrium exists.

Table 1
Game A.

	a_1	a_2	a_3	a_4	a_5
t_1 ($\frac{1-\delta}{2}$)	1, 4	0, 0	0, 0	0, 0	2, $4-\delta$
t_2 ($\frac{1-\delta}{2}$)	0, 0	0, $2+\delta$	3, 0	4, 2	2, 1
t_3 (δ)	0, 0	0, 0	$2+\varepsilon$, 3	2, 2	1, 1

Notes: The left column shows the Sender's type and between brackets the probability that it is drawn. The top row shows the Receiver's actions. The remaining cells provide the Sender's payoff in the first entry and the Receiver's payoffs in the second entry. $0<\delta<\frac{1}{3}$ and $0\leq\varepsilon<1$.

$$ACD(\sigma)=\sum_{t\in T} f(t)\frac{U^S\left(t,\alpha^{\gamma(\sigma)}(\mu^{\gamma(\sigma)}(t))\right)-U^S\left(t,\alpha(\mu(t))\right)}{\bar{U}^S(t)-\underline{U}^S(t)},$$

Crawford- Sobel Game (CS Game) (1982)

Uses a leading uniform-quadratic case for a cheap talk game

Game is uniformly distributed on [0,1]

$U^R(a,t)=-(a-t)^2$; $U^S(a,t)= -(a-(t+b))^2$

GS has only a perfect Bayesian partition equilibrium

For each credible neologism $[\tilde{a}, N]$, the set of deviating types N turns out to be an interval between some $\underline{\tau}$ and $\bar{\tau}$

Characterize neologisms by $[\underline{\tau}, \bar{\tau}]$

Receivers best response is $\tilde{a} = (\underline{\tau} + \bar{\tau}) / 2$

Three types of neologism

When t = 0

b < ½ → credible neologism on the right- end of the type space

n - 1 credible neologism "in the middle"

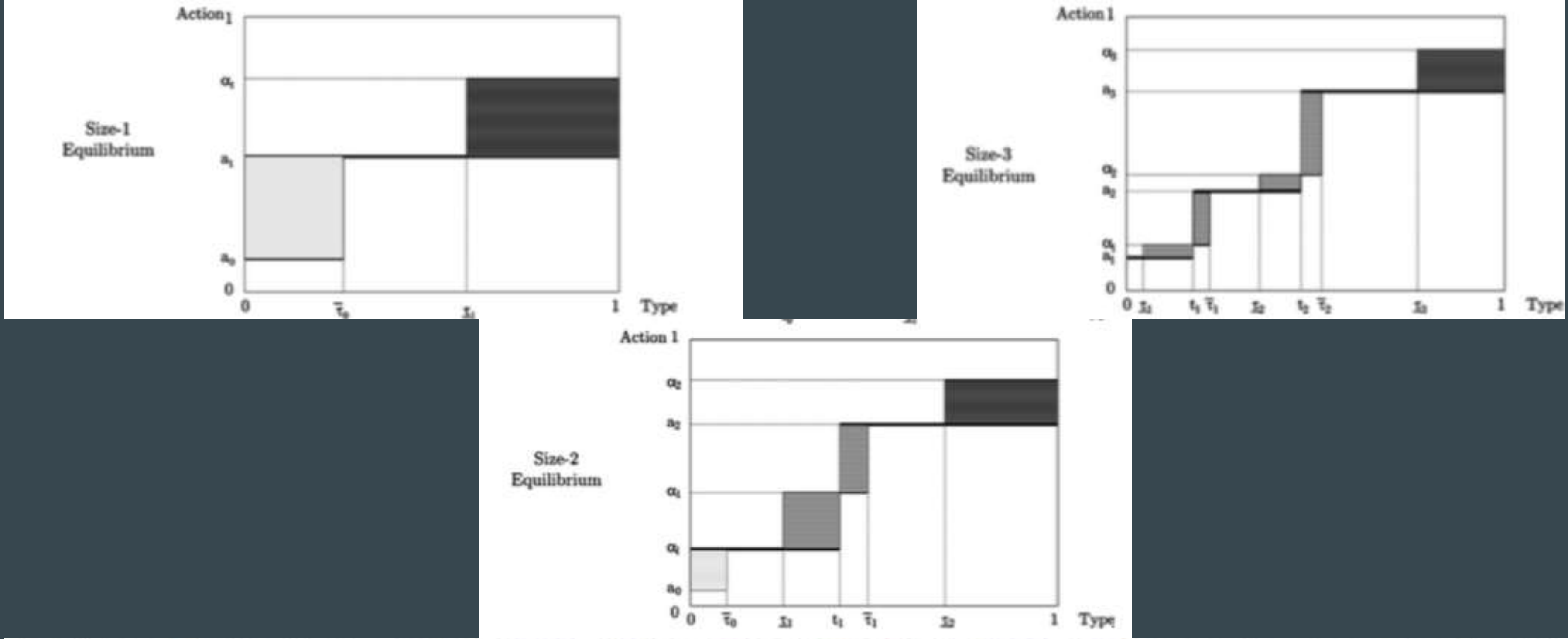

Fig. 1. The size-1, size-2 and (maximum) size-3 equilibria with the credible neologisms they admit for $b = \frac{1}{18}$. The area of the neologisms give an impression of their contribution to the ACD, although their height contributes quadratically to the ACD.

Proposition 3. *For all $b \in \{1/10000, 2/10000, \ldots, 1/4\}$ it holds that the ACD of the size-n equilibrium in the CS game is decreasing in n.*

Crawford-Sobel Game in Lab

ACDC supports : as bias parameter b decreases, the maximum size equilibrium becomes more stable

Table 2
ACDC in Baseline Treatments (Cai and Wang, 2006).

b[a]	Pooling equilibrium[b]			Most informative equilibrium			
	Credible neologisms	Error[c]	ACD	Equilibrium[d]	Credible neologisms	Error	ACD
0.5	⟨1, {1}⟩, ⟨3, {3}⟩, ⟨7, {7}⟩, ⟨9, {9}⟩	.916	.220	{1}, {3}, {5}, {7}, {9}		−.084	0
1.2	⟨1, {1}⟩, ⟨7, {5, 7, 9}⟩, ⟨8, {7, 9}⟩	.896	.181	{1, 3}, {5, 7, 9}	⟨3, {3}⟩, ⟨8, {7, 9}⟩	.146	.074
2	⟨7, {5, 7, 9}⟩	.734	.137	{1}, {3, 5, 7, 9}	⟨7, {5, 7, 9}⟩, ⟨8, {7, 9}⟩	.234	.099
4	⟨6, {3, 5, 7, 9}⟩	.391	.101	{1, 3, 5, 7, 9}	⟨6, {3, 5, 7, 9}⟩	.391	.101

- ACDC selects most informative equilibrium
- Most informative equilibrium has a lower ACD and becomes more stable as b decreases

Other Experimental Results

Blume et al. (2001) → Partial Common Interest (PCI)

4 Discrete cheap talk games

Games 1 and 3 → PCI and neologism proofness (and ACDC) are very much aligned

Game 2 → neologism predicts complete separation/ PCI predicts partial separation

Game 4 → no equilibrium is neologism proof/ PCI selects a unique equilibrium (2 equilibrium outcomes)

Table 3
Reproduction of Games 2 and 4 of Blume et al. (2001).

	a_1	a_2	a_3	a_4	a_5
t_1	800, 800	100, 100	0, 0	500, 500	0, 400
t_2	x, 100	y, 800	0, 0	500, 500	0, 400
t_3	0, 0	0, 0	500, 800	0, 0	0, 400

Notes: All the three types $\{t_1, t_2, t_3\}$ of the Sender are equally likely and the Receiver can implement one of the actions $\{a_1, \ldots, a_5\}$. Entry i, j represents $U^S(t_i, a_j), U^R(t_i, a_j)$. Games 2 and 4 are identical, except that $x = 100$, $y = 300$ in game 2, whereas $x = 300$, $y = 100$ in game 4.

Types *t*1 and *t*2 induce *a*4 while type *t*3 induces *a*3

Not full separation because t2 mimics t1

No neologism proofness

ACDC predicts partially separating equilibrium (ACD = ⅛) will be more observed than pooling equilibrium (ACD = ⅞), even if it's not even stable

Blume finds that 37% of outcomes are consistent with the partially separating

Conclusion

ACDC uses refinements that capture the behaviorally relevant aspects of equilibrium stability in cheap talk games

Can describe actual behavior in a large range of cheap talk games

Uses frequency and size of credible deviations

Measures stability, determines most plausible equilibrium

ACDC works in the general case

Behavior with lower ACD → more consistent with ACDC equilibrium

Limitation : ACDC doesn't predict how experimental subjects behave 'out of

QUESTIONS?