Meaning in Communication Games

Andreas Blume

August 24, 2022

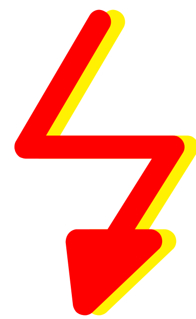
Milan - ESEM Talk

This paper addresses two related questions:

1. How can we model the **strategic use of a shared pre-existing language**?

2. How can we capture **constraints on and uncertainty about what is shared**?

Ludwig Wittgenstein: "Can I say 'bububu' and mean 'If it doesn't rain I shall go for a walk'? – It is only in a language that I can mean something by something." (*Philosophical Investigations*, p.18)



Game theorist (caricature): "Meaning is constituted in equilibrium (or via some other solution concept, if you prefer). No language is needed."

Semantic meaning: meaning in a pre-existing language, use (to accommodate Wittgenstein), literal meaning, conventional meaning, . . .

Equilibrium meaning: Message meaning constituted in equilibrium.

The paper aims to resolve the tension between **semantic meaning** and **equilibrium meaning** of messages.

A trivial example to illustrate **equilibrium meaning** and **semantic meaning**.

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	3,3	1,2

Coordination Game

- Two equally likely **payoff types**, t_1 and t_2.
- Three **receiver actions** a_1, a_2 and a_3.
- The sender sends a message m from the **message space** $M = \{m_1, m_2\}$.
- The receiver takes an action $a \in \{a_1, a_2, a_3\}$.
- Each cell in the payoff table indicates first the sender's and second the receiver's payoff for the corresponding type-action pair (t, a).
- Messages are "cheap talk" – they do not directly affect payoffs.

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	3,3	1,2

Coordination Game

Here are some of the equilibria of this game:

1. $((t_1 \mapsto m_1, t_2 \mapsto m_2); (m_1 \mapsto a_1, m_2 \mapsto a_2))$
2. $((t_1 \mapsto m_2, t_2 \mapsto m_1); (m_1 \mapsto a_2, m_2 \mapsto a_1))$
3. $((t_1 \mapsto m_1, t_2 \mapsto m_1); (m_1 \mapsto a_3, m_2 \mapsto a_3))$
4. $((t_1 \mapsto m_1 \text{ w/p } \alpha, t_2 \mapsto m_1 \text{ w/p } \alpha); (m_1 \mapsto a_3, m_2 \mapsto a_3))$
5. $((t_1 \mapsto m_1 \text{ w/p } \alpha, t_2 \mapsto m_1 \text{ w/p } \beta); (m_1 \mapsto a_3, m_2 \mapsto a_3))$ where $\alpha \approx \beta$

The **equilibrium meaning** of message m_1 is "$\{t_2\}$" in the second equilibrium, and "$\{t_1, t_2\}$" in the third and forth equilibria, and something close to "$\{t_1, t_2\}$" in the fifth equilibrium.

Equilibrium tells us next to nothing about how messages will be used.

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	3,3	1,2

Coordination Game

Add a **language** $\lambda : M \to A$, with $\lambda(m_i) = a_i$, $i = 1, 2$ (a very specific and somewhat impoverished language).

The language gives the semantic meaning of messages.

For now assume that the language is commonly known, i.e., it is a **common language**.

Observation: Simply adding the language does not affect any of the equilibria of the game – it can always be ignored.

The Puzzle

How does a pre-existing language combine with incentives, prior distribution, knowledge and belief, etc to determine equilibrium behavior, including use of the pre-existing language?

A simple idea: Iterate best replies from the pre-existing language until an equilibrium is reached:

1. The sender best responds to the language.
2. The receiver best responds to the sender's best response.
3. The sender best responds to the receiver's best responds to the sender's best response.
4. ....

Reminiscent of level-k reasoning anchored at the language (Crawford, *AER* 2003).

The simple idea has many pitfalls. The talk points out the most important ones and how to deal with them.

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	3,3	1,2

Coordination Game

Implementing the simple idea: In the example, iterating best replies from the common language $\lambda : M \to A$, with $\lambda(m_i) = a_i$, $i = 1, 2$ **trivially converges** to the equilibrium

$$((t_1 \mapsto m_1, t_2 \mapsto m_2); (m_1 \mapsto a_1, m_2 \mapsto a_2)).$$

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	3,3	1,2

Coordination Game

A problem with the simple idea: Iterating best replies from the common language $\lambda : M \to A$, with $\lambda(m_i) = a_i$, $i = 1, 2, 3$ **(where we have added a third message)** does not converge.

There are sequences of best replies in which the sender keeps varying the **mixing** probabilities for messages that induce identical receiver replies.

There generally are **multiple best replies**.

There are sequences of best replies that visit every pure separating strategy of the sender **infinitely often**.

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	3,3	1,2

Coordination Game

A fix for the problem with the simple idea:

- Iterate only over pure strategies.
- At every step pick a single best reply.
- At every step of the iteration, **drop unused messages**.

Then, in this example, for any common language $\lambda : M \to A$, with the property that for all $a \in A$ there is at least one message m with $\lambda(m) = a$,

1. the iterative procedure converges;
2. it converges to a separating equilibrium;
3. message use matches the messages' semantic meanings; and,
4. this works without any upper bound on the size of the message space.

Example: Dropping messages

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	**-1**,3	1,2

Dropping messages

Assume that sender and receiver have a common language $\lambda : M \to A$, with $\lambda(m_i) = a_i$, $i = 1, 2, 3$.
The sender's unique best reply against the receiver strategy $r_1 = \lambda$, defined by the language λ, is given by $s_1 = (t_1 \mapsto m_1, t_2 \mapsto m_3)$. The strategy s_1 does not use message m_2, which is therefore dropped. With message m_2 out of the picture, the receiver has a unique best reply $r_2 = (m_1 \mapsto a_1, m_3 \mapsto a_2)$ against s_1. The sender's unique best reply to r_2 is $\boldsymbol{s_2 = (t_1 \mapsto m_1, t_2 \mapsto m_1)}$. The unused message m_3 is (provisionally) dropped, the receiver's unique best reply in the game without messages m_2 and m_3 is the pooling action a_3, and the iterative procedure has converged.

Example: Dropping messages and message restoration

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	-1,3	1,2

Dropping messages

$\lambda : M \to A$, with $\lambda(m_i) = a_i$, $i = 1, 2, 3$.
The iterative procedure has the message space converge to $\{m_1\}$, has the sender use the strategy $\boldsymbol{s_2 = (t_1 \mapsto m_1, t_2 \mapsto m_1)}$, and has the receiver respond to message m_1 with action a_3.

At this point **messages** m_2 and m_3 are **restored** and a **language equilibrium** is defined as any equilibrium of the original game in which s_2 is the sender strategy and the receiver responds to message m_1 with action a_3.

Note: Message restoration is possible in every game with a common language.

Relation to prior work

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	-1,3	1,2

Dropping messages

$\lambda : M \to A$, with $\lambda(m_i) = a_i$, $i = 1, 2, 3$.

None of the earlier work predicts messages use in this game:

1. Farrell's **neologism proofness** criterion (*GEB*, 1993) rejects all equilibria – given any equilibrium (they all have the receiver take action a_3), type t_1 has a credible neologism. Strictly speaking, it doesn't even do that: his rich-language assumption is not satisfied.

2. There does not exist any **credible message profile** à la Rabin (*JET*, 1990).

Relation to prior work

	a_1	a_2	a_3
t_1	3,3	0,0	1,2
t_2	0,0	-1,3	1,2

Dropping messages

$\lambda : M \to A$, with $\lambda(m_i) = a_i$, $i = 1, 2, 3$.

3. Olszewski's (*JET*, 2006) **maximally rich language** criterion does not predict message use. A maximally rich language is entirely an equilibrium phenomenon – it leaves no role for semantic meanings.
4. Crawford's **level-k** construction (*AER*, 2003) is not aimed at singling out equilibria and needs to be amended (e.g., by having unused messages be dropped) to guarantee convergence.

Example: A role for prep-sets

	a_1	a_2	a_3
t_1	0,9	9,0	0,8
t_2	9,0	0,9	0,8

A role for Prep-sets

$\lambda : M \to A$, with $\lambda(m_i) = a_i$, $i = 1, 2, 3$.

The sender's unique best reply against the receiver's strategy $r_1 = \lambda$, defined by the language λ, is the strategy $s_1 = (t_1 \to m_2, t_2 \to m_1)$. Consider the *reduced game*, in which the unused message m_3 is provisionally dropped. The receiver's unique best reply against the sender's strategy s_1 in the reduced game is the strategy $r_2 = (m_1 \to a_2, m_2 \to a_1)$. **Iterating further generates a** sequence of best replies that are unique at every step and form a **cycle**.

Denote the set of pure strategies that support this cycle by $\mathsf{S}' \times \mathsf{R}'$.

The set of strategies $\mathsf{S}' \times \mathsf{R}'$ does not support an equilibrium of the reduced game, in either pure or mixed strategies.

To satisfy the desideratum of having the iterative procedure reach rest points that are equilibria, the procedure expands the set $\mathsf{S}' \times \mathsf{R}'$.

Voorneveld (GEB, 2004) defines a ***prep set*** as a set of pure strategy profiles that includes a best reply to every belief concentrated on that set.

This inspires the definition of an **$\mathsf{S}' \times \mathsf{R}'$-prep set** as a set of pure strategy profiles in the reduced game that includes $\mathsf{S}' \times \mathsf{R}'$ as well as a best reply to every belief concentrated on that set.

The procedure expands $\mathsf{S}' \times \mathsf{R}'$ to a *minimal* $\mathsf{S}' \times \mathsf{R}'$-*prep set*.

Minimality is with respect to set inclusion.

A *minimal* $\mathsf{S}' \times \mathsf{R}'$-*prep set* does not strictly contain another $\mathsf{S}' \times \mathsf{R}'$-prep set.

	a_1	a_2	a_3
t_1	0,9	9,0	0,8
t_2	9,0	0,9	0,8

A role for Prep-sets

$\lambda : M \to A$, with $\lambda(m_i) = a_i$, $i = 1, 2, 3$.

Recall that S' consists of $s_1 = (t_1 \to m_2, t_2 \to m_1)$ and $s_2 = (t_1 \to m_1, t_2 \to m_2)$.

Given a receiver belief that assigns equal probability to all sender strategies in S', the receiver's unique best reply in the reduced game is the strategy $\tilde{r} = (m_1 \to a_3, m_2 \to a_3)$.

Therefore, the strategy $\tilde{r}$ must be in any $\mathsf{S}' \times \mathsf{R}'$-prep set.

Indeed, once that strategy is included we have a minimal $\mathsf{S}' \times \mathsf{R}'$-prep set and that set includes an equilibrium of the reduced game. In any such equilibrium both messages m_1 and m_2 are used and the receiver responds to both messages with the action a_3. Finally, we can restore the unused message m_3 to the game.

Therefore, in every language equilibrium the sender mixes over the messages m_1 and m_2, and the receiver responds to all three messages with the pooling action a_3.

Example: Uncertainty about language

	a_1	a_2	a_3
t_1	10,10	9,0	0,9
t_2	9,0	10,10	0,9

Uncertainty about language

There is a language λ with $\lambda(m_i) = a_i$, $i = 1, 2$.

A **translation** $\theta : M \to M$ maps intended messages into interpreted messages.

Assume that there are **two translations** θ' and θ'', defined by $\theta'(m) = m$ and $\theta''(m) = m_2$ for both $m \in M$. Assume that there is **common prior μ over translations** with $\mu(\theta'') = p$, where p satisfies $\frac{1}{9} < p < 1$.

	a_1	a_2	a_3
t_1	10,10	9,0	0,9
t_2	9,0	10,10	0,9

Uncertainty about language

The sender's unique best reply against the receiver's strategy $r_1 = \lambda$, defined by the language λ, is the strategy $s_1 = (t_1 \to m_1, t_2 \to m_2)$.

As long as $p > \frac{1}{9}$, the receiver has a unique best reply $r_2 = (m_1 \to a_1, m_2 \to a_3)$ to the sender's strategy s_1.

Against r_2, the sender has a unique best reply $s_2 = (t_1 \to m_1, t_2 \to m_1)$.

At that point message m_2 is dropped from the iteration.

There is a unique λ-equilibrium strategy profile $(\sigma, \rho) = ((t_1 \to m_1, t_2 \to m_1), (m_1 \to a_3, m_2 \to a_3))$.

	a_1	a_2	a_3
t_1	10,10	9,0	0,9
t_2	9,0	10,10	0,9

Uncertainty about language

Rabin uses this example to illustrate how messages may only be **jointly credible**.

Stalnaker borrows the example to raise the possibility of **"ignorance or error about credibility"**.

Language equilibrium predicts message use and does not rely on credibility to do so.

	a_1	a_2	a_3
t_1	11,12	1,0	6,10
t_2	8,0	1,12	13,9

Uncertainty about language – another example

The message space is $M = \{m_1, m_3\}$.

There is a language λ with $\lambda(m_1) = a_1$ and $\lambda(m_3) = a_3$.

There are **two translations** θ' and θ'', defined by $\theta'(m) = m$ and $\theta''(m) = m_1$ for both $m \in M$. Assume that there is **common prior μ over translations** with $\mu(\theta'') = 1/2$.

In every language equilibrium both sender types send message m_1 exclusively.

There is another equilibrium. In that equilibrium both sender types send distinct messages. This equilibrium is the unique efficient equilibrium and the unique equilibrium that survives iterative deletion of dominated strategies (regardless of the order of deletion).

Four ideas that link a pre-existing language to its strategic use:

1. ***Anchoring:*** iterate best replies starting from the language;
2. ***Non-proliferation:*** provisionally drop messages, minimize message use (by using only minimal-message best replies), adjust best-replies only when necessary;
3. ***Expansion:*** expand limit sets reached via iteration to minimal prep sets;
4. ***Restoration***: restore provisionally dropped messages.

The iterative procedure always converges.

When there is no uncertainty about language, the limit **can always be extended to an equilibrium** in the entire game through restoration of dropped messages.

Proposition 1 *In a common-interest game with a rich shared language λ every λ-equilibrium profile (σ, ρ) achieves the maximal payoff and satisfies $\rho(m) = \lambda(m)$ for all messages $m \in M$ that are received with positive probability.*

With common interests, language equilibria are **efficient**, and **equilibrium meaning coincides with semantic meaning.**

Proposition 2 *For every game* $\Gamma(M)$ *with absence of a shared language, the set of* λ*-equilibrium strategies of the sender equals* $\{s \in \mathsf{S}(M) | s(t') = s(t''), \forall t', t'' \in T\}$.

With **maximal uncertainty about language**, there is **only pooling**.

Comment: This illustrates why the proposed iterative procedure should not be confused with a form of learning or evolution.

Absence of a shared language means that the translation is unknown but fixed. With a fixed translation, players could learn to communicate effectively in repeated interactions.

	a_1	a_2	a_3	a_4	a_5
t_1	5,2	1,6	-1,-1	-1,-1	4,3
t_2	1,5	5,2	-1,-1	-1,-1	4,3
t_3	-1,-1	-1,-1	5,2	1,6	4,3
t_4	-1,-1	-1,-1	1,5	5,2	4,3

Block-aligned preferences

$M = \{m_1, \ldots, m_5\}$, $\lambda(m_i) = a_i, i = 1, \ldots, 5.$

In every λ-equilibrium, types t_1 and t_2 mix over messages m_1 and m_2 and types t_3 and t_4 mix over messages m_3 and m_4. In all of these equilibria, the receiver responds to messages m_1 and m_2 with action a_2 and to messages m_3 and m_4 with action a_4. Note that the sender *ex ante* prefers pooling to any λ-equilibrium, that there is no credible message profile, and that λ-equilibria are not neologism proof: the set of types t_1 and t_3 has a credible neologism.

Proposition 3 *In games with block-aligned preferences and a rich and accessible language λ, if the sender learns the translation then every λ-equilibrium profile (σ, ρ) block conforms with the language λ.*

With partial incentive alignment, there is **coarse agreement between semantic meaning and equilibrium meaning**.

Finite CS games

Generic games for which there exist orderings of types and actions such that:

1. The functions u^i are unimodal in a for all $t \in T$ and $i = 1, 2$.
2. The sender's preference has an upward bias relative to the receiver: $a^R(t) < a^S(t), \forall t \in T$.
3. The receiver's ideal point is responsive: $a^R(t') \neq a^R(t)$ for all $t, t' \in T$ with $t \neq t'$
4. Each player i's payoff function u^i satisfies the single crossing condition:

$$t_2 > t_1 \text{ and } a_2 > a_1$$

implies

$$u^i(a_2, t_1) - u^i(a_1, t_1) > 0 \Rightarrow u^i(a_2, t_2) - u^i(a_1, t_2) > 0.$$

Proposition 4 *In any generic finite CS game with a rich and accessible language λ, if the sender learns the translation, then for every λ-equilibrium profile (σ, ρ) and all messages $m \in \Theta(M)$ that are received with positive probability,*

1. $\lambda(m) = a^S(t)$ *for some* $t \in T$, *and*
2. $\rho(m, h^R) < \lambda(m)$ *for all receiver signals* $h^R \in H^R$.

There is **language inflation:** Semantic meanings exceed equilibrum meanings in every language equilibrium.

The End