

# LET'S TALK IT OVER: COORDINATION VIA PREPLAY COMMUNICATION WITH LEVEL-K THINKING

Vincent P. Crawford

26th Arne Ryde Symposium, "Communication in Games and Experiments"

24-25 August 2007, Lund, Sweden



“What we got here...is a failure to communicate.”

—Paul Newman as the title character in *Cool Hand Luke*

## Introduction

The paper reconsiders two classic analyses of explicit coordination—the use of preplay communication of intentions to structure relationships via non-binding agreements—in the light of recent work on structural non-equilibrium models of initial responses to games based on level- $k$  thinking

- Joseph Farrell (1987), “Cheap Talk, Coordination, and Entry,” *Rand Journal of Economics* (see also his “Communication, Coordination and Nash Equilibrium,” *Economics Letters* 1988)
- Matthew Rabin (1994), “A Model of Pre-Game Communication,” *Journal of Economic Theory* (see also his “Focal Points in Pre-Game Communication,” Berkeley 1991)

Henceforth collectively “FR”

## FR's analyses

FR's models consist of a preplay communication phase followed by play of an underlying game; together I call them the "entire game"

Farrell studies symmetry-breaking with conflicting preferences about how to coordinate as in Battle of the Sexes (or in pure coordination games)

Rabin studies coordination more generally, in a more general class of games

Their analyses address two conjectures

- That preplay communication will yield an effective agreement to play an equilibrium in the underlying game
- That the agreed-upon equilibrium will be Pareto-efficient within the underlying game's set of equilibria (henceforth "efficient" for short)

Regarding the structure of the communication phase, FR assume

- Communication takes the form of one or more two-sided, simultaneous exchanges of messages about players' intentions in the underlying game
- The messages are in a pre-existing common language, hence understood
- The messages are nonbinding and costless (“cheap talk”)

(Rabin (1994, pp. 389-390) discusses the rationale for studying simultaneous two-sided messages rather than one-sided or sequential messages)

Regarding players' behavior, FR assume

- Equilibrium, sometimes weakened to rationalizability
- Plausible behavioral restrictions defining which combinations of messages create agreements, and whether and how agreements can be changed

Under these assumptions, they show that

- Rationalizable preplay communication need not assure equilibrium
- Although communication enhances coordination, even equilibrium with “abundant” (Rabin’s term for “unbounded”) communication does not assure that the outcome will be Pareto-efficient

## Motivation

Equilibrium and rationalizability are natural places to start in analyses like FR's

But recent experiments suggest that in settings without clear precedents people often deviate systematically from equilibrium (with clear precedents equilibrium is more reliable, but explicit agreements may be unnecessary)

The experimental evidence also suggests that in such settings a structural non-equilibrium model based on level- $k$  thinking (or “cognitive hierarchy,” as Camerer, Ho, and Chong (2004 *QJE*) call it) can often out-predict equilibrium

Level- $k$  models also tend to out-predict “equilibrium with noise” models with payoff-sensitive error distributions such as quantal response equilibrium

## Level- $k$ models

Level- $k$  models allow behavior to be heterogeneous, but they assume that each player follows a rule drawn from a common distribution over a particular hierarchy of decision rules or *types* (as they are called in this literature)

Type  $Lk$  anchors its beliefs in a nonstrategic  $L0$  type and adjusts them via thought-experiments with iterated best responses:  $L1$  best responds to  $L0$ ,  $L2$  to  $L1$ , and so on

In applications the type frequencies are treated as behavioral parameters (or a parameterized distribution), estimated or translated from previous analyses

The estimated distribution is fairly stable across games, with most weight on  $L1$ ,  $L2$ , and  $L3$

The estimated frequency of the anchoring  $L0$  type is usually 0 or small; thus  $L0$  exists mainly as  $L1$ 's model of others,  $L2$ 's model of  $L1$ 's model, and so on

Even so, the specification of  $L0$  is the main issue in defining a level- $k$  model and the key to its explanatory power

$L0$  often needs to be adapted to the setting, as it does here; but the definition of higher types via iterated best responses allows a simple, empirically plausible explanation of behavior in many different settings

$L_1$  and higher types have accurate models of the game and are rational in that they choose best responses to beliefs ( $L_k$  makes  $k$ -rationalizable decisions)

$L_k$ 's only departure from equilibrium is replacing its assumed perfect model of others with simplified models that avoid the complexity of equilibrium analysis

In the words of Selten (*EER* '98):

“Basic concepts in game theory are often circular in the sense that they are based on definitions by implicit properties.... Boundedly rational strategic reasoning seems to avoid circular concepts. It directly results in a procedure by which a problem solution is found. Each step of the procedure is simple, even if many case distinctions by simple criteria may have to be made.”

## Experimental evidence

Camerer, Ho, and Chong (2004 *QJE*, Section IV) and Costa-Gomes and Crawford (2006 *AER*, Introduction and Section II.D) summarize the experimental evidence for level-k thinking/cognitive hierarchy models, which includes experiments on games with a variety of different structures

- Stahl and Wilson (1994 *JEBO*, 1995 *GEB*)
- Nagel (1995 *AER*)
- Ho, Camerer, and Weigelt (1998 *AER*)
- Costa-Gomes, Crawford, and Broseta (2001 *ECMA*)
- Costa-Gomes and Weizsäcker (2005)
- Cai and Wang (2006 *GEB*)
- Costa-Gomes and Crawford (2006 *AER*)

## Applications

Level- $k$  models have also been used to resolve a variety of puzzles

- Crawford (2003 *AER*): preplay communication in zero-sum games
- Camerer, Ho, and Chong (2004 *QJE*): coordination via structure, market-entry games, speculation and zero-sum betting, money illusion
- Ellingsen and Östling (2006): organizational design, Aumann's (1990) critique and why one-sided communication works better in games like Battle of the Sexes but two-sided communication works better in Stag Hunt
- Crawford and Iriberri (2007 *ECMA*): overbidding in independent-private-value and common-value auctions
- Crawford and Iriberri (2007 *AER*): systematic deviations from unique mixed-strategy equilibrium in zero-sum two-person hide-and-seek games with non-neutrally framed locations, as in Rubinstein and Tversky's experiments
- Crawford, Gneezy, and Rottenstreich (2007): coordination via structure and framing, label salience, payoff salience, and focal points

## Motivation continued

The existence of an empirically successful alternative to treating deviations from equilibrium as errors makes equilibrium (or quantal response equilibrium) seem too strong, but rationalizability may be too weak

This paper takes a middle course, reconsidering FR's analyses but replacing equilibrium or rationalizability with a level- $k$  model

Although level- $k$  models have not yet been thoroughly tested in this setting, their strong experimental support elsewhere makes them a natural candidate

This paper focuses on Farrell's analysis of Battle of the Sexes, but with some attention to Rabin's more general analysis

As suggested by Camerer, Ho, and Chong's (2004 *QJE*, Section III.C) analysis of market-entry games, a level- $k$  analysis already has surprising implications for tacit coordination in Battle of the Sexes

Subjects in market-entry experiments regularly achieve better ex post coordination (number of entrants closer to market capacity) than in the symmetric mixed-strategy equilibrium, the natural equilibrium benchmark; this led Kahneman (1988) to remark, "...to a psychologist, it looks like magic"

Camerer, Ho, and Chong show that Kahneman's "magic" can be explained by a level- $k$  model of closely related market-entry games: The predictable heterogeneity of strategic thinking allows some players to mentally simulate others' entry decisions and accommodate them, breaking the symmetry as required for coordination; the more sophisticated players become like Stackelberg followers, with coordination benefits for all

In Battle of the Sexes without communication, even with moderate differences in preferences, the level- $k$  coordination rate, for empirically plausible type distributions, is likely to be higher than the mixed-strategy equilibrium rate

In Battle of the Sexes with communication, a level- $k$  analysis allows a unified treatment of players' messages and actions and how messages create agreements, deriving all three from simple assumptions that explain behavior in other settings

A level- $k$  analysis also allows a reevaluation of FR's plausible but ad hoc restrictions on how players use language

With one round of communication, the analysis justifies FR's assumption that a message pair that identifies an equilibrium leads to that equilibrium

However, the resulting "agreements" do not fully reflect the meeting of the minds that FR sought to model

Instead they reflect either one player's perceived credibility as a sender or the other's perceived credulity as a receiver, never both at the same time

As a result, a level- $k$  analysis may not fully support the assumptions about agreements in Rabin's analysis of negotiated rationalizability

With abundant communication, as Rabin's analysis of negotiated rationalizability suggests, level- $k$  players need not keep communicating until an agreement is reached as they do in Farrell's equilibrium

A level- $k$  analysis also yields very different conclusions about the effectiveness of communication than Farrell's equilibrium analysis

With or without communication, level- $k$  coordination rates in Battle of the Sexes are largely independent of the difference in players' preferences

By contrast, in Farrell's equilibrium analysis coordination rates are highly sensitive to the difference in players' preferences

Unless the difference in preferences is very small, coordination rates are likely to be higher with level- $k$  thinking than in Farrell's equilibria

## Outline

I. A Level- $k$  analysis of tacit coordination in Battle of the Sexes

II. Farrell's equilibrium analysis of Battle of the Sexes without and with communication, and Rabin's generalizations

III. A Level- $k$  analysis of Battle of the Sexes with one round of communication

IV. A Level- $k$  analysis of Battle of the Sexes with abundant communication

## A Level- $k$ analysis of tacit coordination in Battle of the Sexes

I adapt Camerer, Ho, and Chong's analysis of market-entry games to Battle of the Sexes, showing that level- $k$  thinking yields similar benefits there

		Column	
		H	D
Row	H	0      0	1 $a$
	D	1 $a$	0      0

**Battle of the Sexes ( $a > 1$ )**

The unique symmetric equilibrium is in mixed strategies, with  $p \equiv \Pr\{H\} = a/(1+a)$  for both players

The expected coordination rate is  $2p(1-p) = 2a/(1+a)^2$ ; and players' payoffs are  $a/(1+a) < 1$ , worse for each than his worst pure-strategy equilibrium

In the level- $k$  model, each player follows one of four types,  $L1$ ,  $L2$ ,  $L3$ , or  $L4$ , with each player role filled by a draw from the same distribution

I assume, as in most previous analyses, that without communication  $L0$  chooses its action randomly, with  $\Pr\{H\} = \Pr\{D\} = \frac{1}{2}$

Higher types' best responses are easily calculated:

$L1$ s mentally simulate  $L0$ s' random decisions and best respond to them, choosing H; similarly,  $L2$ s choose D,  $L3$ s choose H, and  $L4$ s choose D

Although  $L3$  behaves like  $L1$  here, and  $L4$  like  $L2$ , I retain all four for comparability with the analysis below

But I assume for simplicity that the frequency of  $L0$  is 0

I assume throughout that both player roles are filled from the same distribution of types, which restricts attention to symmetric outcome distributions, paralleling Farrell's restriction to the symmetric mixed-strategy equilibrium

The model's predicted outcome distribution is determined by the outcomes of the possible type pairings in Table 1 and the type frequencies

<b>Types</b>	<b><i>L1</i></b>	<b><i>L2</i></b>	<b><i>L3</i></b>	<b><i>L4</i></b>
<b><i>L1</i></b>	H, H	H, D	H, H	H, D
<b><i>L2</i></b>	D, H	D, D	D, H	D, D
<b><i>L3</i></b>	H, H	H, D	H, H	H, D
<b><i>L4</i></b>	D, H	D, D	D, H	D, D
<b>Table 1. Level-<i>k</i> Outcomes without Communication</b>				

The type frequencies are assumed independent of payoffs, in keeping with the fact that the types are intended as general models of strategic behavior; thus the model's predicted outcome distribution is independent of a

With symmetry, players have equal ex ante payoffs, which are proportional to the expected coordination rate

Lumping  $L1$  and  $L3$  together and letting  $v$  denote their total probability, and lumping  $L2$  and  $L4$  together and letting  $(1-v)$  denote their probability, the coordination rate is  $2v(1-v)$ , maximized at  $v = 1/2$  where it takes the value  $1/2$

Thus for  $v$  near  $1/2$ , which is empirically plausible, the coordination rate is close to  $1/2$  (for more extreme values of  $v$  the rate is worse, falling to 0 as  $v \rightarrow 0$  or 1)

The mixed-strategy equilibrium coordination rate,  $2a/(1+a)^2$ , is maximized when  $a = 1$  where it takes the value  $1/2$ , but converges to 0 like  $1/a$  as  $a \rightarrow \infty$

Thus even for moderate values of  $a$ , the level- $k$  coordination rate is quite likely to be higher than the equilibrium rate

The level- $k$  model improves upon the symmetric equilibrium by “relaxing” the incentive constraints requiring players’ responses to be in equilibrium

Because level- $k$  types best respond to non-equilibrium beliefs, it is natural to compare the level- $k$  outcome to the best symmetric rationalizable outcome, in which each player plays a non-equilibrium mixed strategy with  $v \equiv \Pr\{H\} = \frac{1}{2}$

When  $v = \frac{1}{2}$ , the level- $k$  model can be viewed as using the heterogeneity of strategic thinking to purify this best symmetric rationalizable outcome

This is not to suggest that level- $k$  thinking always makes this ideal outcome attainable: type frequencies are behavioral parameters, not choice variables

The analysis suggests a view of tacit coordination profoundly different from the traditional view

Even though decisions are simultaneous and there is no communication, the predictable heterogeneity of strategic thinking allows some players to mentally simulate others' entry decisions and accommodate them, just as (noisy) Stackelberg followers would, with coordination benefits for all players

Equilibrium and selection principles such as risk- or payoff-dominance play no direct role in level- $k$  players' strategic thinking

Coordination, when it occurs, is an almost accidental (though statistically predictable) by-product of players' use of non-equilibrium decision rules

## **Farrell's equilibrium analysis of Battle of the Sexes with communication, and Rabin's generalizations**

In Farrell's model of Battle of the Sexes with communication, the underlying game is preceded by one or more communication rounds in which players send simultaneous messages regarding their pure-strategy intentions

The messages are in a pre-existing common language and they are nonbinding and costless

I denote the possible messages "h" meaning "I intend to play H" and "d" meaning "I intend to play D"

Farrell studies the symmetric mixed-strategy equilibrium in the entire game, including the communication phase, in which players take the first pair of messages that identify a pure-strategy equilibrium in the underlying game as an agreement to play that equilibrium, ignoring all previous messages

In Farrell's equilibrium, players randomize their messages in each round until some round yields an equilibrium pair of messages, in which case they play that equilibrium; or the communication phase ends, in which case they revert to the symmetric mixed-strategy equilibrium in Battle of the Sexes

Farrell calculates the equilibrium coordination rate with one or more rounds of communication and studies how it depends on the number of rounds

I will describe his equilibrium by players' common values of  $q \equiv \Pr\{h\}$  in each communication round and  $p \equiv \Pr\{H\}$  if the communication phase ends and they play Battle of the Sexes without an agreement

Without communication, the equilibrium failure rate is  $[p^2 + (1-p)^2]$ , which equals  $(1+a^2)/(1+a)^2$  when  $p$  takes its equilibrium value of  $a/(1+a)$

$(1 - [p^2 + (1-p)^2]) = 2p(1-p) = 2a/(1+a)^2$ , the equilibrium coordination rate)

With one round of communication, coordination fails if and only if players' message pair does not specify an equilibrium *and* players' pure actions are not in equilibrium when they then play the underlying game

The equilibrium failure rate is  $[q^2 + (1-q)^2][p^2 + (1-p)^2]$ , less than the rate without communication of  $[p^2 + (1-p)^2]$

The equilibrium  $q$  can be calculated by reducing the game to a simultaneous-move message game by plugging in the payoffs from message pairs

$q = a^2/(1+a^2)$ , so the equilibrium failure rate is  $(1+a^4)/[(1+a^2)(1+a)^2]$

The corresponding coordination rate is  $1 - (1+a^4)/[(1+a^2)(1+a)^2] = 2(a+a^2+a^3)/[(1+a^2)(1+a)^2]$ , greater than the equilibrium coordination rate without communication,  $2a/(1+a)^2$

With abundant communication, the equilibrium failure rate is a product like  $[q^2 + (1-q)^2][p^2 + (1-p)^2]$ , but with a separate  $q$  for each round

If the  $q$ s were bounded between 0 and 1, the failure rate would approach 0 as the number of rounds grew; but each  $q$  must be in equilibrium in its round's message game, and the equilibrium  $q$ s converge to 1 so quickly that the failure rate converges to a limit above 0 even with abundant communication

Farrell shows that the limiting failure rate is  $(a-1)/(a+1)$ , and the corresponding coordination rate is  $1 - [(a-1)/(a+1)] = 2/(1+a)$ , which is greater than the equilibrium coordination rate with one round of communication; the limiting expected payoff is  $[(1+a)/2] \times [2/(1+a)] = 1$ , above the mixed-strategy equilibrium payoff  $a/(1+a)$  and exactly realized Rabin's bound

But even with abundant communication, the coordination rate  $\rightarrow 0$  as  $a \rightarrow \infty$

Rabin (1994) evaluates the generality of Farrell's analysis

- A much wider class of underlying games
- No symmetry restriction
- Richer characterization of how players use language, allowing interim agreements
- Considering the implications of rationalizability as well as equilibrium

Rabin defines notions called *negotiated equilibrium* and *negotiated rationalizability* that combine the standard notions of equilibrium and rationalizability with his restrictions on how players use language

With abundant communication, each player's negotiated equilibrium expected payoff is at least his worst efficient equilibrium payoff in the underlying game

Replacing negotiated equilibrium by negotiated rationalizability, each player expects (perhaps wrongly) at least the payoff of his worst efficient equilibrium

Thus Farrell's insights are quite general (Rabin, p. 373):

"...the potential efficiency gains from communication illustrated by [Farrell (1987)] do not rely on ad hoc assumptions of symmetry or on selecting a particular type of mixed-strategy equilibrium. Rather, the efficiency gains...inhere in the basic assumptions about how players use language."

Costa-Gomes (2002 *JET*) extends Rabin's theory and tests it with the experimental data of Cooper, DeJong, Forsythe, and Ross (1989 *Rand*) and the data from Roth and collaborators' experiments on unstructured bargaining

## A Level- $k$ analysis of Battle of the Sexes with one round of communication

The key difficulty in analyzing two-sided level- $k$  communication is extending normal-form level- $k$  types to types that determine both messages and actions

I do this, following Ellingsen and Östling (2006, Section 3.2), by adapting the  $L0$  sender type in Crawford's (2003) model of one-sided communication

(Crawford's (2003) type hierarchy is built upon a "credible" sender type, which tells the truth (there called  $W0$  but here called  $L0$ ; Crawford's "credulous" receiver type  $S0$  is a best response to  $W0$ , hence analogous to an  $L1$ ))

But with two-sided communication, as Ellingsen and Östling note (footnote 8), a player's beliefs and best responses as a credible sender and a credulous receiver are inconsistent for sent and received messages that do not specify an equilibrium action pair, so the analysis must reconcile them in some way

Like Ellingsen and Östling, I do this by giving priority to the credible sender type and dispensing (with regard to  $L0$ ) with the credulous receiver type

(The credulous receiver type, because it deals with beliefs about another player's communication strategy, is less fundamental than the credible sender type; Crawford and Iriberri (2007 *AER*) argue that making  $L0$  as nonstrategic as possible tends to yield more useful level- $k$  models)

Thus I assume that  $L0$  uniformly randomizes its action, without regard to its partner's message, and sends a truthful message

This truthful  $LO$  is intuitively plausible—bearing in mind that it is only the starting point for players' strategic thinking—with some experimental support

- Blume, DeJong, Kim, and Sprinkle (2001 *GEB*)
- Crawford (2003) gives references to the classical literature on deception
- Kawagoe and Tazikawa (2005)
- Cai and Wang (2006 *GEB*)
- Wang, Spezio, and Camerer (2006)
- Sánchez-Pagés and Vorsatz (2007 *GEB*, 2007)
- Crawford and Iriberri (2007 *ECMA*) (truthful  $LO$  bidder types in auctions)

It generalizes the uniform random  $LO$  used for games without communication

It also generalizes Crawford's (2003) truthful  $WO$  sender type (which made a more precise assumption about  $WO$ 's action, but only for definiteness)

(If it were assumed instead that  $L0$  uniformly and independently randomizes its message as well as its action, then communication would be completely ineffective and the model would reduce to the model without communication)

(Ellingsen and Östling (Section 3.2) show that their truthful specification of  $L0$  gets the (empirically) right answer with one-sided communication as well:

In Battle of the Sexes with one-sided communication, an  $L1$  receiver will believe the message it receives and accommodate; an  $L1$  sender will expect its message to be believed, and will therefore send message  $h$  and choose action  $H$

$L2$  and higher senders will also send  $h$  and choose  $H$ ; thus  $L1$ ,  $L2$  and higher receivers will all choose  $D$ )

## Types' strategies

In deriving types' strategies in Battle of the Sexes with two-sided communication, I assume that a type always chooses an action with the highest expected payoff, given its beliefs

As in previous applications, I assume that payoff ties are broken randomly, so that a type chooses equally desirable actions with equal probabilities

I also assume that the types have a slight preference for truthfulness, so that if telling the truth and lying have exactly equal payoffs, a type tells the truth

If, in addition, both messages have equal probabilities of being true, I assume that a type sends them with equal probabilities

With regard to types' beliefs, I assume that, because each type has a unitary model of others ( $L2$  believing others are  $L1$ , and so on), it does not draw sophisticated inferences about others' types from their messages

(In Crawford and Iriberri's (2007 *ECMA*) analysis of common-value auctions, level- $k$  types can draw inferences about others' private information from their bids but not inferences about others' types; in Crawford's (2003) analysis the *Sophisticated* type but not the level- $k$  types draw inferences from others' messages about their types; Ellingsen and Östling assume that level- $k$  types draw such inferences in their analysis of the "Poisson cognitive hierarchy" model, where types above  $L1$  have positive weights on all lower types)

I also assume that if a type receives a message that contradicts its beliefs regarding its partner's action, it disregards the message and maintains its beliefs about the action, on the grounds that action preferences are stronger

Given  $L0$ 's strategy of uniformly randomizing its action and sending a truthful message,  $L1$  expects its partner's message to be truthful and its own message to be ignored

$L1$  therefore accommodates by choosing action D if it receives message h from its partner, and choosing action H if it receives message d

At the time  $L1$  chooses its own message it has not yet received its partner's message, and so it cannot predict its own action; and because  $L1$  expects its partner's message to be h and d with equal probabilities, both of its own messages have equal probabilities of being true

$L1$  therefore sends h and d with equal probabilities, independent of its action

Given  $L1$ 's strategy,  $L2$  expects its partner's message to be uninformative and its own message to be believed and accommodated

$L2$  therefore chooses action H and sends message h, in each case without regard to its own or its partner's message (but if for some reason it had chosen action D instead, it would have sent message d)

Given  $L2$ 's strategy,  $L3$  expects its partner's action to be H, its partner's message to be truthful, and its own message to be ignored

If  $L3$  receives message h, reinforcing its belief that its partner's action will be H, then it accommodates, choosing action D

Because  $L3$  expects its own message to be ignored, but unlike  $L1$  it expects its partner to choose action H, it sends the message it expects to be true, d

If  $L3$  instead receives message d, contradicting its belief that its partner's action will be H, then I assume that  $L3$  still expects its partner to choose H and still sends the message it expects to be true, d

Thus  $L3$  always chooses action D and sends message d (but if it had chosen action H instead, it would have sent message h)

Given  $L3$ 's strategy,  $L4$  expects its partner's message to be truthful and its own message to be ignored

If  $L4$  receives message  $d$ , reinforcing its belief that its partner's action will be  $D$ , then it accommodates, choosing action  $H$

Because  $L4$  expects its own message to be ignored and expects its partner to choose action  $D$ , it sends the message it expects to be true,  $h$

If  $L4$  instead receives message  $h$ , contradicting its belief that its partner's action will be  $D$ ,  $L4$  still expects its partner to choose  $D$  and still sends the message it expects to be true,  $h$

Thus  $L4$  always chooses action  $H$  and sends message  $h$  (but if it had chosen action  $D$ , it would have sent message  $d$ )

## Coordination outcomes

Table 2 gives the messages for all types and the coordination outcomes on the non-equilibrium path for all type pairings

<b>Type</b> (message)	<b>L1</b> (random)	<b>L2 (h)</b>	<b>L3 (d)</b>	<b>L4 (h)</b>
<b>L1</b> (random)	$\frac{1}{2}H + \frac{1}{2}D,$ $\frac{1}{2}H + \frac{1}{2}D$	D, H	H, D	D, H
<b>L2 (h)</b>	H, D	H, H	H, D	H, H
<b>L3 (d)</b>	D, H	D, H	D, D	D, H
<b>L4 (h)</b>	H, D	H, H	H, D	H, H
<b>Table 2. Level-<i>k</i> Messages and Outcomes with One Round of Communication</b>				

“ $\frac{1}{2}H + \frac{1}{2}D, \frac{1}{2}H + \frac{1}{2}D$ ” refers to players’ independently random choices in *L1* versus *L1*, which make all four possible outcomes equally likely

Repeat Table 1 for the level- $k$  model without communication for comparison:

<b>Types</b>	<b><i>L1</i></b>	<b><i>L2</i></b>	<b><i>L3</i></b>	<b><i>L4</i></b>
<b><i>L1</i></b>	H, H	H, D	H, H	H, D
<b><i>L2</i></b>	D, H	D, D	D, H	D, D
<b><i>L3</i></b>	H, H	H, D	H, H	H, D
<b><i>L4</i></b>	D, H	D, D	D, H	D, D
<b>Table 1. Level-<math>k</math> Outcomes without Communication</b>				

<b>Type</b> (message)	<b><i>L1</i></b> (random)	<b><i>L2</i></b> (h)	<b><i>L3</i></b> (d)	<b><i>L4</i></b> (h)
<b><i>L1</i></b> (random)	$\frac{1}{2}H + \frac{1}{2}D,$ $\frac{1}{2}H + \frac{1}{2}D$	D, H	H, D	D, H
<b><i>L2</i></b> (h)	H, D	H, H	H, D	H, H
<b><i>L3</i></b> (d)	D, H	D, H	D, D	D, H
<b><i>L4</i></b> (h)	H, D	H, H	H, D	H, H
<b>Table 2. Level-<math>k</math> Messages and Outcomes with One Round of Communication</b>				

There are three notable differences between Table 1 and Table 2

First, with one round of communication types other than *L1* always (without regard to the message sent or received) choose the action opposite to the one they choose without communication

For example, *L2* expects its messages to be believed and accommodated, and so sends *h* and chooses *H*; but without communication *L2* expected *L1* to choose *H*, and so accommodates by choosing *D*

Returning to the Stackelberg analogy used for tacit coordination, without communication *L1* is effectively committed (in *L2*'s mind) to choosing *H*; but with communication *L1* is *not* committed not to listen (because its *L0* is truthful), and this allows *L2* to use its message to take over the leadership role

Second, in the pairing  $L1$  versus  $L1$ , there are now equal probabilities of all four {H, D} combinations, instead of the H, H outcome without communication

This is because  $L1$  expects its partner's message to be truthful and its own message to be ignored

$L1$  therefore believes and accommodates its partner's message but (unable to predict which message will be true) chooses its own message randomly, so that both  $L1$ s end up playing H and D with equal probabilities

$L1$ 's communication skills here leave something to be desired, but its listening skills still yield a large improvement over the  $L1$  versus  $L1$  outcome without communication

Third, in the pairing  $L1$  versus  $L3$ ,  $L1$  still chooses H but  $L3$  now accommodates by choosing D

This is because  $L3$  expects its partner to choose H, and so chooses D and sends d, while  $L1$  sends a random message but expects its partner's message to be truthful, and so ends up choosing H

Although  $L1$  is not good at talking, it doesn't matter because  $L3$  is not listening

The improvement here is entirely due to  $L1$ 's listening skills, which suffice for coordination with  $L3$

How much does one round of level- $k$  communication improve coordination over level- $k$  outcomes without communication or equilibrium outcomes with one round?

Focus again on the coordination rate (ignoring changes from H, D to D, H, or vice versa)

Comparing the level- $k$  outcomes without communication (Table 1) and with one round (Table 2), the rate goes up from 0 to  $\frac{1}{2}$  for the pairing  $L1$  versus  $L1$ , from 0 to 1 for the pairings  $L1$  versus  $L3$ , and is otherwise unchanged

If the frequencies of  $L1$ ,  $L2$ ,  $L3$ , and  $L4$  are  $r \approx 0.4$ ,  $s \approx 0.3$ ,  $t \approx 0.2$ , and  $u \approx 0.1$  then the overall coordination rate without communication is  $2(r+t)(s+u) \approx 0.48$ , while with communication the overall rate goes up by  $\frac{1}{2}r^2 + 2rt$ , to 0.68

Comparing the level- $k$  and equilibrium coordination rates with one round of communication, the equilibrium rate is  $2(a+a^2+a^3)/[(1+a^2)(1+a)^2]$ , which equals  $3/4$  when  $a = 1$ ,  $28/45$  when  $a = 2$ , and converges to 0 like  $1/a$  as  $a \rightarrow \infty$

Thus when  $a \approx 1$  the coordination rate with one round of communication is likely to be somewhat higher for equilibrium than for a level- $k$  model (0.75 versus 0.68)

But even for moderate values of  $a$ , the level- $k$  coordination rate is likely to be higher than the equilibrium coordination rate

## Reevaluating Farrell's assumptions about which message combinations create agreements

Focusing on Farrell's assumptions as they apply with one round of communication, he assumes a message pair that identifies a pure-strategy equilibrium in Battle of the Sexes is treated as an agreement to play that equilibrium

On the non-equilibrium path in Table 2,  $L1$  sends a random message,  $L2$  and  $L4$  send  $h$ , and  $L3$  sends  $d$ ; and in all twelve possible pairings from  $\{L1, L2, L3, L4\}$ , message pairs that identify an equilibrium in Battle of the Sexes always lead to both players playing that equilibrium

Thus, taken literally, the analysis justifies Farrell's assumption for one round

However, the resulting agreements do not reflect the meeting of the minds that FR sought to model

Instead they reflect either one player's perceived credibility as a sender or the other's perceived credulity as a receiver, but never both at the same time

The problem is that no level- $k$  type is both a good talker and a good listener, as would be required (at the least) for a full meeting of the minds (higher types have communication skills no better than  $L1$ 's through  $L4$ 's)

As a result, pairings of  $L1$  versus  $L2$ ,  $L3$ , or  $L4$  always lead to equilibrium play, without regard to whether or not the message pair identifies an equilibrium; and pairings of  $L1$  versus  $L1$  sometimes lead to equilibrium play, again without regard to whether or not the messages identify an equilibrium (for pairings from  $\{L2, L3, L4\}$  only agreements lead to equilibrium play; but for these pairings communication never enhances coordination)

$L1$ 's listening skills raise the coordination rate well above the rate without communication, but a level- $k$  analysis may not fully support the assumptions about agreements in Rabin's analysis of negotiated rationalizability

As Rabin notes, an equilibrium analysis also fails to explain a meeting of the minds: perhaps a full meeting of the minds requires more than mechanical decision rules, a Gricean leap of the imagination or "team reasoning" (e.g. Crawford, Gneezy, and Rottenstreich (2007) and the references cited there)

## **A Level-k analysis of Battle of the Sexes with abundant communication**

Farrell's equilibrium analysis of abundant communication assumes that players continue exchanging messages until an agreement is reached

I assume, in the spirit of Rabin's analysis, that players can always agree to continue for an additional round of communication by mutual consent

I also assume players have a slight preference for avoiding additional rounds

The model adds players' options to request to continue communication as simultaneous decisions each round following the exchange of messages

As usual, there is always an equilibrium in the request game in which neither player requests to continue; I simply assume that if continuing is better for both players, given their beliefs, then they both request to continue

I also assume that players draw no inferences about their partners' types from the history of their interactions; and that in their request decisions they draw no conditional inferences about their partners' types (as in Feddersen and Pesendorfer's (1996 *AER*) equilibrium analysis of the "swing voter's curse")

The assumption that players draw no inferences from history is obviously strained for some outcome paths; I maintain it anyway to make the most important points as simply as possible

## Types' strategies

Note first that both players requesting to continue communication can never be better for both players if their current messages already identify one of the Pareto-efficient pure-strategy equilibria in Battle of the Sexes

By continuing they incur the slight cost of an additional round of communication, and no deviation could make that worthwhile for both of them

This implies (finding Table 2's inefficient outcomes) that there are three kinds of type pair and realized message pair that might continue communication

<b>Type</b> (message)	<b>L1</b> (random)	<b>L2 (h)</b>	<b>L3 (d)</b>	<b>L4 (h)</b>
<b>L1</b> (random)	$\frac{1}{2}H + \frac{1}{2}D,$ $\frac{1}{2}H + \frac{1}{2}D$	D, H	H, D	D, H
<b>L2 (h)</b>	H, D	H, H	H, D	H, H
<b>L3 (d)</b>	D, H	D, H	D, D	D, H
<b>L4 (h)</b>	H, D	H, H	H, D	H, H
<b>Table 2. Level-<i>k</i> Messages and Outcomes with One Round of Communication</b>				

- *L1* versus *L1* following one of the message pairs, d,d or h,h, that don't identify an equilibrium
- *L3* versus *L3* following its normal message pair d,d
- *L2* or *L4* versus *L2* or *L4* following their normal message pair h,h

*L1* versus *L1* following message pair  $d,d$  both expect to play  $H$  against their partner's  $D$  if communication is cut off, because each expects its partner's message to be truthful and its own to be ignored

Given this, each is too sure of its optimistic beliefs to continue communicating

Instead, as Rabin's analysis of negotiated rationalizability suggests is possible out of equilibrium, *L1* versus *L1* following message pair  $d,d$  both cut off communication, and so play  $H, H$  in the underlying game

*L1* versus *L1* following message pair  $h,h$  both expect to play  $D$  against their partner's  $H$  if communication is cut off

These beliefs are too pessimistic, so there is potential for improvement; but it may seem pointless to continue because they will be the same people who have just failed to reach agreement in a round like the one that would ensue

But both of *L1*'s messages have equal expected payoffs and are equally likely to be true; if *L1*'s randomness is an unstudied response to those indifferences, then the random outcomes need not be perfectly correlated each round

Given this, the outcome if *L1* versus *L1* following message pair  $h,h$  continue will be a new random pair of messages, with a new, positive probability of identifying an efficient equilibrium (compare Costa-Gomes's (2002) "mutual grain of agreement" assumption)

It is shown in the paper that if  $L1$  versus  $L1$  continue, the eventual outcome will be H, H; D, H; or H, D, each with probability  $1/3$ , with expected payoff  $(1+a)/3$

If they cut off communication, they expect to play D against H, with payoff 1

Thus it is better to continue if and only if  $(1+a)/3 > 1$ , or equivalently if  $a > 2$

In this case the definition of  $L1$  gracefully overcomes what might appear an insurmountable problem in extending Farrell's equilibrium analysis of the effectiveness of abundant communication to a level- $k$  model

These models concern repeated interaction in fixed pairs, and Farrell's analysis of abundant communication inherently relies on randomness

We are socialized to think that equilibrium players can and do consciously randomize; but it is conventional to assume (and I think empirically plausible) that level- $k$  players cannot, or at least do not, consciously randomize

Fortunately, level- $k$  players can *unconsciously* randomize, and the definition of  $L1$  creates just the indifferences needed to make this work for  $L1$  versus  $L1$  following message pair  $h,h$

Summing up for  $L1$  versus  $L1$ , in the first round each of the four possible message pairs is equally likely

If players send one of the pairs,  $d,h$  or  $h,d$ , that identify an equilibrium, then they cut off communication and play that equilibrium

If they send  $d,d$ , then they cut off communication and play  $H, H$

When  $a < 2$ , if they send  $h,h$ , they cut off communication and play  $D, D$

When  $a > 2$ , if they send  $h,h$  they continue communicating for (at least) one more round; in that case, if as I have assumed the types draw no inferences from history, the process is a Markov chain, with all states but  $h,h$  absorbing; the eventual outcome is either  $H, H$ ;  $D, H$ ; or  $H, D$ , each with probability  $1/3$

Like  $L1$  versus  $L1$  following message pair  $d,d$ ,  $L2$  or  $L4$  versus  $L2$  or  $L4$  are too optimistic to continue communicating

They too cut off communication after the first round, and so play  $H, H$  in the underlying game

Finally, like  $L1$  versus  $L1$  following message pair  $h,h$ ,  $L3$  versus  $L3$  are too pessimistic

But unlike  $L1$ 's messages,  $L3$ 's are deterministic, so  $L3$  versus  $L3$  may conclude that it is pointless to continue communicating anyway

If they do continue, they are possibly doomed to repeat  $d,d$  forever and never reach an efficient agreement

The only ray of hope is that, if  $L3$  versus  $L3$  do continue and there is some exogenous randomness in how messages are sent or received, or some random variation in how they learn from experience, they might eventually reach an efficient agreement by accident (such randomness is superfluous for  $L1$  versus  $L1$  following  $h,h$ ; and it won't stop  $L1$  versus  $L1$  following  $d,d$  or  $L2$  or  $L4$  versus  $L2$  or  $L4$  from cutting off communication after the first round)

## Coordination outcomes

Table 3 gives the coordination outcomes on the non-equilibrium path

Type	<i>L1</i>	<i>L2</i>	<i>L3</i>	<i>L4</i>
<i>L1</i>	$\frac{1}{2}H + \frac{1}{2}D$ , $\frac{1}{2}H + \frac{1}{2}D$ if $a < 2$ ; $\frac{1}{3}H, H + \frac{1}{3}D, H + \frac{1}{3}H, D$ if $a > 2$	D, H	H, D	D, H
<i>L2</i>	H, D	H, H	H, D	H, H
<i>L3</i>	D, H	D, H	D, D (?)	D, H
<i>L4</i>	H, D	H, H	H, D	H, H

**Table 3. Level-*k* Outcomes with Abundant Communication**

“ $\frac{1}{2}H + \frac{1}{2}D$ ,  $\frac{1}{2}H + \frac{1}{2}D$ ” refers to the uniform distribution over the four possible outcomes for *L1* versus *L1* following message pair h,h when  $a < 2$

Repeat Table 2 for the level- $k$  model with one round of communication for comparison

<b>Type</b> (message)	<b>L1</b> (random)	<b>L2</b> (h)	<b>L3</b> (d)	<b>L4</b> (h)
<b>L1</b> (random)	$\frac{1}{2}H + \frac{1}{2}D,$ $\frac{1}{2}H + \frac{1}{2}D$	D, H	H, D	D, H
<b>L2</b> (h)	H, D	H, H	H, D	H, H
<b>L3</b> (d)	D, H	D, H	D, D	D, H
<b>L4</b> (h)	H, D	H, H	H, D	H, H
<b>Table 2. Level-<math>k</math> Messages and Outcomes with One Round of Communication</b>				

<b>Type</b>	<b>L1</b>	<b>L2</b>	<b>L3</b>	<b>L4</b>
<b>L1</b>	$\frac{1}{2}H + \frac{1}{2}D,$ $\frac{1}{2}H + \frac{1}{2}D$ if $a < 2$ ; $\frac{1}{3}H, H + \frac{1}{3}D, H + \frac{1}{3}H, D$ if $a > 2$	D, H	H, D	D, H
<b>L2</b>	H, D	H, H	H, D	H, H
<b>L3</b>	D, H	D, H	D, D (?)	D, H
<b>L4</b>	H, D	H, H	H, D	H, H
<b>Table 3. Level-<math>k</math> Outcomes with Abundant Communication</b>				

The outcomes with abundant communication are the same as with one round of communication, except that if  $a > 2$ ,  $L1$  versus  $L1$  now have a coordination rate of  $2/3$  instead of  $1/2$ ; and some exogenous randomness might allow  $L3$  versus  $L3$  to raise its coordination rate above its rate of 0 with one round (“?”)

Updating the calibration for one round of communication, with frequencies of  $L1$ ,  $L2$ ,  $L3$ , and  $L4$   $r \approx 0.4$ ,  $s \approx 0.3$ ,  $t \approx 0.2$ , and  $u \approx 0.1$ , if  $a > 2$  the first change adds another  $r^2/6 \approx 0.03$  to the overall level- $k$  coordination rate with abundant communication, raising it to approximately 0.71 from the rate of 0.68 with one round and of 0.48 without communication (if  $a < 2$  the rate stays at 0.68)

The second change could conceivably add as much as  $t^2(1-0) = 0.06$  more, raising the coordination rate to approximately 0.77 or, if  $a < 2$ , 0.74

With abundant communication, when  $a > 1.94$  and possibly for lower values, the level- $k$  coordination rate is greater than the equilibrium rate,  $2/(1+a)$ , which equals 1 when  $a = 1$ ,  $2/3$  when  $a = 2$ , and  $\rightarrow 0$  like  $1/a$  as  $a \rightarrow \infty$

To the extent that level- $k$  types do better than in Farrell's equilibrium analysis, they do so because, as in the level- $k$  analysis of tacit coordination, the level- $k$  model relaxes the equilibrium incentive constraints

As in Farrell's analysis, the benefits of abundant communication are limited and, unless players' preferences are fairly close, most of the gains from communication are realized with only one round (here, oddly, the benefits of abundant communication are more limited when  $a$  is small, because  $L1$  versus  $L1$  following message pair  $h,h$  then cut off communication)

The level- $k$  model's predictions are consistent with Rabin's bounds based on negotiated rationalizability, but their precision yields additional insight