

Contents lists available at [ScienceDirect](#)

## Research in Economics

journal homepage: [www.elsevier.com/locate/rie](http://www.elsevier.com/locate/rie)

# Let's talk it over: Coordination via preplay communication with level- $k$ thinking<sup>☆</sup>

Vincent P. Crawford<sup>a,b,c</sup><sup>a</sup> All Souls College, Oxford OX1 4AL, UK<sup>b</sup> Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK<sup>c</sup> Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA

## ARTICLE INFO

## Article history:

Received 23 September 2016

Accepted 2 October 2016

Available online 11 October 2016

## Keywords:

Preplay communication of intentions

Coordination

Battle of the Sexes

Behavioral game theory

Noncooperative games

## ABSTRACT

This paper reconsiders Farrell's (1987) and Rabin's (1994) analyses of coordination via preplay communication, focusing on Farrell's analysis of Battle of the Sexes. Replacing their equilibrium and rationalizability assumptions with a structural non-equilibrium model based on level- $k$  thinking, I reevaluate Farrell and Rabin's assumptions on how players use language and their conclusions on the limits of communication in bringing about coordination. The analysis partly supports their assumptions about how players use language, but suggests that their "agreements" do not reflect a full meeting of the minds. A level- $k$  analysis also yields very different conclusions about the effectiveness of communication.

© 2016 The Authors. Published by Elsevier Ltd. on behalf of University of Venice. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Tacit coordination is ubiquitous in the animal kingdom, but explicit coordination—the use of preplay communication to structure relationships via non-binding agreements—may be fully realized only in human societies. Although explicit coordination is an essential part of our lives, and there is now a substantial body of experimental evidence on it (Crawford (1998) surveys early work), our theoretical understanding remains incomplete. This paper proposes and analyzes a model of coordination via pre-play communication that seeks to narrow the gaps between theory, evidence, and intuition, building on the work of Farrell (1987) and Rabin (1994) (see also Farrell, 1988 and Rabin, 1991), henceforth collectively "FR".

FR's analyses address two conjectures regarding complete-information games that are still widely held despite FR's partly negative conclusions: that preplay communication will yield an effective agreement to play an equilibrium in the underlying game; and that the agreed-upon equilibrium will be Pareto-efficient within that game's set of equilibria (henceforth "efficient"). FR assume that communication takes the form of one or more two-sided, simultaneous exchanges of messages about players' intended actions in the underlying game. The messages are in a pre-existing common language and they are nonbinding and costless. FR also assume equilibrium, sometimes weakened to rationalizability. They further restrict attention to outcomes that satisfy plausible behavioral restrictions defining which combinations of messages create

<sup>☆</sup> Prepared for a keynote address presented at the 26th Arne Ryde Symposium, "Communication in Games and Experiments," 24–25 August 2007, University of Lund, Lund, Sweden. I am grateful to Nageeb Ali, Tore Ellingsen, Nagore Iriberrri, Navin Kartik, Robert Östling, Adam Sanjurjo, and Joel Sobel for extremely helpful comments and suggestions. All Souls College and the University of California, San Diego provided research support. The research leading to these results also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 339179. The contents reflect only the author's views and not the views of the ERC or the European Commission, and the European Union is not liable for any use that may be made of the information contained therein.

E-mail address: [vincent.crawford@economics.ox.ac.uk](mailto:vincent.crawford@economics.ox.ac.uk)

<http://dx.doi.org/10.1016/j.rie.2016.10.001>

1090-9443/© 2016 The Authors. Published by Elsevier Ltd. on behalf of University of Venice. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

agreements, and whether and how agreements can be changed.<sup>1</sup> Under these assumptions FR show that rationalizable preplay communication need not assure equilibrium; and that, although communication enhances coordination, even equilibrium with “abundant” (Rabin’s term for “unbounded”) communication does not assure that the outcome will be Pareto-efficient.

In the part of FR’s analyses that is most important for this paper, Farrell (1987) uses Battle of the Sexes to study symmetry-breaking via one or more rounds of two-sided preplay communication with conflicting preferences about how to coordinate.<sup>2</sup> He focuses on the symmetric mixed-strategy equilibrium in the entire game, including the communication phase, in which the first pair of messages in the same communication round that identify a pure-strategy equilibrium in Battle of the Sexes are treated as an agreement to play that equilibrium, ignoring all previous messages. He calculates the equilibrium rate of efficient coordination with one or more rounds of communication, showing that the rate increases steadily with the number of rounds but converges to a limit less than one even with abundant communication.<sup>3</sup>

Because Farrell’s analysis is specific to Battle of the Sexes and assumes equilibrium, it is reasonable to ask how general his insights are. Rabin (1994) extends Farrell’s analysis to a wide class of underlying games while dropping his symmetry restriction; augmenting his restrictions on how players use language to allow them to make interim agreements, which can be improved upon in subsequent agreements; and considering the implications of rationalizability as well as equilibrium.<sup>4</sup> Rabin defines notions called *negotiated equilibrium* and *negotiated rationalizability* that combine the standard notions with his restrictions on how players use language. He shows that with abundant communication, each player’s negotiated equilibrium expected payoff is at least as high as in his worst efficient equilibrium in the underlying game. He then shows, replacing negotiated equilibrium by negotiated rationalizability, that even without equilibrium, each player expects (perhaps incorrectly) a payoff at least as high as in his worst efficient equilibrium. Thus Farrell’s insights are quite general: “...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.” (Rabin, p. 373).

Although equilibrium and rationalizability are the natural places to start in analyses like FR’s, recent experiments suggest that in settings without clear precedents people often deviate systematically from equilibrium, especially when the reasoning behind it is not straightforward. The evidence also suggests that in such settings a structural non-equilibrium model can often out-predict equilibrium.<sup>5</sup> While the existence of an empirically successful alternative to treating deviations as errors makes equilibrium seem too strong an assumption, rationalizability may be too weak. This paper takes a middle course, reconsidering FR’s analyses with particular attention to Farrell’s analysis of Battle of the Sexes, but replacing equilibrium or rationalizability with a non-equilibrium model based on level- $k$  thinking. Although level- $k$  models have not yet been thoroughly tested in this kind of setting, they explain much of the predictable part of subjects’ deviations from equilibrium in experiments that elicit initial responses to games in other settings, and their strong experimental support makes them a natural candidate.<sup>6</sup>

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. The 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.<sup>7</sup>

Turning to abundant communication, I assume in the spirit (but not the letter) of Rabin’s analysis that players always have the option of an additional round of communication by mutual consent, but that in any round either player can

<sup>1</sup> These restrictions make subgame-perfection superfluous. Rabin (1994, pp. 389–390) discusses the rationale for studying models in which two-sided messages are simultaneous rather than sequential. As he notes (and as Schelling (1960) noted), if there are no delay costs, as in FR’s and my analyses, with sequential messages the outcome may be arbitrarily determined by assumptions about who can speak last or how players form their beliefs.

<sup>2</sup> Farrell’s analysis also sheds light on the symmetry-breaking role of communication in the pure coordination games studied by Schelling (1960) and others.

<sup>3</sup> Symmetry is a natural restriction when players cannot distinguish their roles, and avoids begging the question of symmetry-breaking. Crawford and Haller (1990, p. 580) provide a justification for the symmetry assumption.

<sup>4</sup> Rabin’s model of communication is similar to Kalai and Samet’s (1985). They assume agreements are binding, though renegotiable; but this difference is unimportant here because in coordination games the potential agreements are equilibria, and Rabin’s assumptions make agreements to play them effectively binding, though renegotiable. Costa-Gomes (2002) extends Rabin’s theory and tests it with experimental data.

<sup>5</sup> With enough clear precedents, equilibrium is more reliable; but explicit agreements may then be unnecessary.

<sup>6</sup> Level- $k$  models, described in Section 1, also tend to out-predict equilibrium models with payoff-sensitive error distributions such as quantal response equilibrium. They were introduced to explain experimental data by Stahl and Wilson (1994) and Nagel (1995) and further developed by Ho et al. (1998); Costa-Gomes et al. (2001; “CGCB”); Camerer et al. (2004; “CHC”); Costa-Gomes and Weizsäcker (2008); Costa-Gomes and Crawford (2006; “CGC”); Crawford and Iriberrí (2007a, 2007b); and Crawford et al. (2008). Crawford et al. (2013, Section 3) review the evidence.

<sup>7</sup> Negotiated rationalizability is potentially relevant here because level- $k$  types choose  $k$ -rationalizable strategies (Bernheim (1984); CGC, Section 1) and  $k$ -rationalizability, even for moderate  $k$ , is close to rationalizability in this setting. However, Rabin’s analysis is not conclusive here because it requires levels of  $k$  higher than those that are realistic, and negotiated rationalizability builds the assumption that agreements are effective into players’ beliefs, but in the level- $k$  analysis agreements reflect weaker restrictions that may not always satisfy Rabin’s assumption.

unilaterally shut off communication and force play of the underlying game. As Rabin's analysis of negotiated rationalizability suggests, level- $k$  players need not keep communicating until an agreement is reached as in Farrell's equilibrium.

Finally, a level- $k$  analysis implies very different conclusions about the effectiveness of communication than Farrell's equilibrium analysis. A level- $k$  analysis suggests that coordination rates in Battle of the Sexes, with or without communication, will be largely independent of the difference in players' preferences, while in Farrell's equilibrium analysis coordination rates are highly sensitive to this difference. A level- $k$  analysis already has surprising implications for tacit coordination: Even with moderate differences in preferences, the level- $k$  coordination rate without communication is likely to be higher, for empirically plausible type distributions, than the mixed-strategy equilibrium rate. Further, with one round of communication, the level- $k$  rate is well above the rate without communication, and is likely to be higher than the equilibrium rate with one round of communication unless preferences are very close. Finally, with abundant communication, the level- $k$  coordination rate is likely to be higher than the equilibrium rate unless preferences are moderately close. The model's predictions with abundant communication are consistent with Rabin's bounds based on negotiated rationalizability, but their precision yields additional insight into the causes and consequences of breakdowns in negotiations.

This paper's closest relatives other than FR are Crawford (2003) and Ellingsen and Östling (2010; henceforth "EÖ"). Crawford (2003) introduces a level- $k$  model of one-sided communication of intentions and uses it to study deception in zero-sum games. EÖ generalize Crawford's model to allow two-sided communication and use it to study two central issues in coordination: symmetry-breaking in games like Battle of the Sexes; and reassurance in games like Stag Hunt, where there is a tension between the higher payoffs and greater fragility of the Pareto-dominant equilibrium. They show that in a level- $k$  model, as in equilibrium with suitable refinements, one-sided communication almost trivially solves the coordination problem in Battle of the Sexes, and is therefore more effective than two-sided communication, as is usually found in experiments (Crawford, 1998, Section 3). They also show that, unlike equilibrium with suitable refinements, a level- $k$  model can also explain why two-sided communication is more effective than one-sided in Stag Hunt, as is also found in experiments. EÖ focus on the implications of these results for organizational design. This paper adapts EÖ's generalized model of two-sided communication to a different purpose: reevaluating FR's assumptions about how players use language and providing a more realistic characterization of the effectiveness of communication in bringing about coordination via symmetry-breaking.<sup>8</sup>

The paper is organized as follows. Section 1 introduces the level- $k$  model by using it to analyze Battle of the Sexes without preplay communication, following CHC's (Section 3.3) level- $k$  (or as they call it, "cognitive hierarchy") analysis of closely related market-entry games. It has long been noted that subjects in market-entry experiments (Rapoport et al., 1998 and Rapoport and Seale, 2002) regularly achieve better ex post coordination (number of entrants closer to market capacity) than in the symmetric mixed-strategy equilibrium, the natural equilibrium benchmark. Earlier versions of this result led Daniel Kahneman (1988) to remark, "...to a psychologist, it looks like magic." CHC show that Kahneman's "magic" can be explained by a level- $k$  model, in which the predictable heterogeneity of strategic thinking allows some players to mentally simulate others' entry decisions and accommodate them. The more sophisticated players become like Stackelberg followers, with coordination benefits for all.

Section 1's analysis adapts CHC's analysis to Battle of the Sexes, showing that level- $k$  thinking yields similar symmetry-breaking benefits there. The analysis suggests a view of tacit coordination profoundly different from the traditional view: With level- $k$  thinking, equilibrium and, a fortiori, selection principles such as risk- or payoff-dominance (Harsanyi and Selten, 1988) play no direct role in players' strategic thinking. Coordination, when it occurs, is an almost accidental (though predictable) by-product of the use of non-equilibrium decision rules. These striking differences motivate a level- $k$  analysis of explicit coordination: At the very least, a level- $k$  analysis will shift the equilibrium benchmarks in Farrell's analysis.

Section 2 reviews Farrell's equilibrium analysis of communication in Battle of the Sexes and the implications of Rabin's analysis of negotiated rationalizability in this setting.

Section 3 presents a level- $k$  analysis of Battle of the Sexes with one round of communication. It then compares the resulting coordination outcomes with Section 1's level- $k$  outcomes for Battle of the Sexes without communication, and with Section 2's equilibrium outcomes with one round. Finally, it uses the level- $k$  model to reevaluate Farrell's assumptions regarding which combinations of messages create agreements.

Section 4 extends Section 3's analysis to allow abundant communication, modeled as allowing players the option, at the end of any communication round, of an additional round by mutual consent. It then compares the resulting coordination outcomes with the level- $k$  outcomes with one round of communication, and with the outcomes in Farrell's equilibrium characterization of the limits of abundant communication.

Section 5 is the conclusion.

<sup>8</sup> I do not consider one-sided communication because it begs the question of symmetry-breaking that is at the heart of the coordination problem in Battle of the Sexes. Nonetheless, as EÖ show, the model used here has the "right" implications to explain experimental results with one-sided as well as two-sided communication. Kartik et al. (2007) introduce level- $k$  models of one-sided strategic information transmission, in the limited sense of credulous receivers; see also Kawagoe and Tazikawa (2009).

## 1. A level- $k$ model of tacit coordination

This section introduces the level- $k$  model by using it to analyze Battle of the Sexes without communication, following CHC's (Section 3.3) analysis of market-entry games.

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*. 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.

As implemented here, 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.  $L1$  and higher types have accurate models of the game and are rational in that they choose best responses to beliefs. Their only departure from equilibrium is in replacing its assumed perfect model of others with simplified models that avoid the complexity of equilibrium analysis.

In applications the type frequencies are treated as behavioral parameters (or in CHC's cognitive hierarchy model, a parameterized distribution) to be 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 very 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; but the definition of higher types via iterated best responses allows an empirically plausible explanation of behavior in most settings.

In CHC's market-entry games,  $n$  risk-neutral firms simultaneously decide whether to enter a market with capacity  $m < n$ . If  $m$  or fewer firms enter, the entrants all earn a profit; but if more than  $m$  enter they all earn a loss. Staying out yields zero. Like Battle of the Sexes, this game has a unique symmetric equilibrium in mixed strategies, in which the expected number of entrants is approximately  $m$ , but there are significant probabilities of over- or under-entry. Yet in Rapoport et al.'s (1998) and Rapoport and Seale's (2002) experiments with closely related games, the numbers of entrants ex post were systematically closer to  $m$  than in the symmetric equilibrium.

How can subjects do systematically better than in the symmetric equilibrium? CHC show that this can be explained by a level- $k$  model with an empirically plausible type distribution. In their model,  $L0$  is uniformly random, the usual assumption for normal-form games.  $L1$ s mentally simulate  $L0$ s' random entry decisions and accommodate them, entering only if they expect enough  $L0$ s to stay out.  $L2$ s accommodate  $L1$ s' (and in CHC's model, unlike in mine,  $L0$ s') entry decisions; and so on. Even though players' decisions are simultaneous and there is no communication, the heterogeneity of strategic thinking allows more sophisticated types to accommodate less sophisticated types' decisions, just as (noisy) Stackelberg followers would.

Now consider the closely related Battle of the Sexes game in Fig. 1, where  $a > 1$  without loss of generality. Two players choose simultaneously between two pure actions, H for Hawk or D for Dove, using the standard labeling of the strategies from evolutionary game theory to emphasize the symmetry of actions and payoffs across player roles. 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' expected payoffs are  $a/(1+a) < 1$ , worse for each player 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  $L0$  chooses its action randomly, with  $\Pr\{H\} = \Pr\{D\} = 1/2$ . Higher types' best responses are easily calculated:  $L1$  chooses H,  $L2$  chooses D,  $L3$  chooses H, and  $L4$  chooses D (Table 1). Although  $L3$  behaves like  $L1$  here, and  $L4$  behaves like  $L2$ , I retain all four for comparability with the analysis below. But I assume for simplicity, from now on, that the frequency of  $L0$  is 0.

The model's predicted outcome distribution is determined by the outcomes of the possible type pairings in Table 1 and the type frequencies. The type frequencies are assumed to be independent of payoffs, in keeping with the fact that, like equilibrium, they are intended as general models of strategic behavior. Because in Battle of the Sexes, the outcomes of the possible type pairings are independent of  $a$  as long as  $a > 1$ , the payoff-independence of the type frequencies implies that the model's predicted outcome distribution is independent of  $a$ . By contrast,  $a$  has a strong influence on the equilibrium coordination rate, so this independence is important in the comparison between level- $k$  and equilibrium rates.

		Column	
		H	D
Row	H	0	1
	D	1	0

Fig. 1. Battle of the sexes ( $a > 1$ ).

**Table 1**  
Level- $k$  outcomes without communication.

Types	L1	L2	L3	L4
L1	H, H	H, D	H, H	H, D
L2	D, H	D, D	D, H	D, D
L3	H, H	H, D	H, H	H, D
L4	D, H	D, D	D, H	D, D

With symmetry, players have equal ex ante payoffs, which are proportional to the expected coordination rate, so little is lost by focusing on the coordination rate. Lumping L1 and L3 together and letting  $\nu$  denote their total probability, and lumping L2 and L4 together and letting  $(1-\nu)$  denote their total probability, the coordination rate is  $2\nu(1-\nu)$ , which is maximized at  $\nu=1/2$ , where it takes the value  $1/2$ . Thus for  $\nu$  near  $1/2$ , which is empirically plausible in this setting, the coordination rate is close to  $1/2$ . (However, for more extreme values of  $\nu$  the rate is worse, falling to 0 as  $\nu \rightarrow 0$  or 1.) By contrast, the mixed-strategy equilibrium coordination rate,  $2a/(1+a)^2$ , is maximized when  $a=1$  where it takes the value  $1/2$ , and equals  $4/9$  when  $a=2$  and  $3/8$  when  $a=3$ , converging 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.

A player's ex ante (not conditioned on his type) expected payoff is  $(1+a)\nu(1-\nu)$ . This too is maximized at  $\nu=1/2$ , where it equals  $(1+a)/4$ , which is always greater than the  $a/(1+a)$  expected payoff of the mixed-strategy equilibrium. Thus for type frequencies near  $\nu=1/2$ , the level- $k$  model yields players greater ex ante expected payoffs than the mixed-strategy equilibrium; and for  $a > 3$ , greater even than in players' worst pure-strategy equilibria.<sup>9</sup>

From a mechanism-design point of view, 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  $\nu \equiv \Pr\{H\}=1/2$ . When  $\nu=1/2$ , the level- $k$  model can be viewed as using the heterogeneity of players' strategic thinking to purify this best symmetric rationalizable outcome. This is not to suggest that level- $k$  thinking always makes this ideal outcome attainable: the type frequencies are behavioral parameters, not choice variables.<sup>10</sup>

As noted in the Introduction, the level- $k$  model suggests a view of tacit coordination profoundly different from the traditional view: Equilibrium and selection principles such as risk- or payoff-dominance play no direct role in players' strategic thinking; and coordination, when it occurs, is an almost accidental by-product of how paired players' types interact.

## 2. Farrell's equilibrium analysis of communication

This section reviews Farrell's (1987) analysis of one- and multi-round communication in Battle of the Sexes and the implications of Rabin's (1994) analysis in this setting.

Farrell's underlying game has a richer payoff parameterization than the Battle of the Sexes game in Fig. 1, but the added richness is not relevant here, so I use Fig. 1's game. In Farrell's model, 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".

Recall that 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 either some round yields an equilibrium pair of messages, in which case they play that equilibrium; or the communication phase ends without an agreement, in which case they revert to the symmetric mixed-strategy equilibrium in Battle of the Sexes. I will describe his equilibrium, which is subgame-perfect, by players' common values of  $q \equiv \Pr\{h\}$  in each round and their common value of  $p \equiv \Pr\{H\}$  if there is no agreement.

Farrell calculates the equilibrium rate of coordination failure (which is more convenient to work with than the rate of coordination) and studies how it depends on the number of rounds of communication. 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  $a/(1+a)$ . Of course,  $1 - [p^2 + (1-p)^2] = 2p(1-p) = 2a/(1+a)^2$ , the equilibrium coordination rate calculated in Section 1.

<sup>9</sup> It is also interesting to evaluate players' welfare type by type. The expected payoff of an H player (L1 or L3 here) is  $a(1-\nu)$  and of a D player (L2 or L4) is  $\nu$ . Thus types differ on the ideal  $\nu$ , each preferring to be on the "rare" side.

<sup>10</sup> The level- $k$  approach has other implications for mechanism design, not developed here. For instance, because level- $k$  types (above  $k=0$ ) all respect simple dominance, mechanisms that implement desired outcomes in dominant strategies may, depending on the mix of types in the population, have an advantage over mechanisms that implement superior outcomes but only in equilibrium, especially if the latter use non-transparent devices like integer games. See for example Crawford et al. (2009).

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 without an agreement. Because the second event is conditionally independent of the first, the equilibrium failure rate is  $[q^2 + (1-q)^2][p^2 + (1-p)^2]$ , which because  $p$  is the same in both cases is always less than the rate without communication of  $[p^2 + (1-p)^2]$ . To see how much one round of communication reduces the failure rate, it is necessary to calculate the equilibrium  $q$ . This can be done by reducing the game to a simultaneous-move message game by plugging in the payoffs from the possible message pairs. The message game is qualitatively like Battle of the Sexes, but with different payoffs because it is not the last chance to coordinate. The equilibrium  $q = a^2/(1+a^2)$ , and the equilibrium failure rate is therefore  $(1+a^4)/[(1+a^2)(1+a)^2]$ . The corresponding coordination rate is  $1 - (1+a^4)/[(1+a^2)(1+a)^2] = 2(a+a^2+a^3)/[(1+a^2)(1+a)^2]$ , which is greater than the equilibrium coordination rate without communication,  $2a/(1+a)^2$ .

With abundant communication, the equilibrium failure rate is a product that generalizes  $[q^2 + (1-q)^2][p^2 + (1-p)^2]$ , with a separate  $q$  for each round (Farrell's (7), p. 38). If the  $q$ s were independent of the number of rounds and bounded between 0 and 1, then the failure rate would approach 0 as the number grew without limit. But each  $q$  must be in equilibrium in its round's message game, and although the failure rate declines with the number of rounds, the equilibrium  $q$ s converge to 1 so quickly that the failure rate converges to a limit above 0 even with abundant communication. 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,  $2(a+a^2+a^3)/[(1+a^2)(1+a)^2]$ . The limiting expected payoff with abundant communication is  $[(1+a)/2] \times [2/(1+a)] = 1$ , well above the mixed-strategy equilibrium payoff  $a/(1+a)$ . Thus Farrell's equilibrium with abundant communication exactly realizes the bound given by Rabin's results for negotiated equilibrium, whereby each player expects a payoff of at least the 1 of his worst pure-strategy equilibrium in Battle of the Sexes.

Farrell shows, more generally, that the equilibrium coordination rate is everywhere increasing in the number of rounds. When  $a=1$ , the coordination rate is  $1/2$  without communication,  $3/4$  with one round, and 1 with abundant communication. But as  $a \rightarrow \infty$ , even with abundant communication the coordination rate approaches 0.

The sensitivity of coordination rates to payoff differences highlights an incentive problem. In Farrell's model, players use mixed communication strategies to try to create a correlating agreement within their relationship to break the symmetry as required for efficient coordination.<sup>11</sup> Under his assumptions about how players use language, even proposing a non-binding agreement has real consequences, and this creates an incentive for players to negotiate more aggressively, the more rounds of communication there are. It is this incentive that drives their probabilities of sending message  $h$  to 1 too quickly for them to reach an efficient agreement with probability 1, creating an efficiency gap that increases with  $a$ . As a result, unless  $a$  is near 1, the benefits of abundant communication are limited and most of the gains from communication would be realized with only one round: With abundant communication, the coordination rate is 1 when  $a=1$ ,  $2/3$  when  $a=2$ , and  $1/2$  when  $a=3$ , while with one round the rate is  $3/4$  when  $a=1$ ,  $28/45 \approx 0.62$  when  $a=2$ , and  $39/80 \approx 0.49$  when  $a=3$ . Even so, Farrell's analysis shows that proposing and making non-binding agreements allows players to realize some of the benefits of an ideal binding agreement to play the best symmetric correlated equilibrium, which would yield expected payoff  $(1+a)/2$  instead of the 1 they obtain in Farrell's equilibrium.

### 3. A level- $k$ model with one round of communication

This section introduces a level- $k$  model with one round of two-sided communication and uses it to analyze Battle of the Sexes. I focus on two-sided communication because one-sided communication begs the question of symmetry-breaking that is at the heart of the coordination problem in Battle of the Sexes. The section then compares the level- $k$  coordination outcomes with one round of communication with those without communication, and with Section 2's equilibrium outcomes for the game with one round. Finally, it uses the level- $k$  model to reevaluate Farrell's assumptions regarding which combinations of messages create agreements.

#### 3.1. Modeling two-sided level- $k$ communication

The key difficulty in analyzing two-sided level- $k$  communication is extending level- $k$  types from normal-form games to extensive-form types that determine both messages and actions. I do this, following EÖ, by adapting the types in Crawford's (2003) model of one-sided communication.<sup>12</sup> As EÖ note (p. 1701), a player's beliefs and best responses as a credible sender

<sup>11</sup> Crawford and Haller (1990) provide an analogous analysis, in which players repeatedly play a tacit coordination game, using costly real-time play to generate precedents within their relationship that will eventually allow them to break the symmetry as needed; or, in other settings, to find more efficient ways to coordinate.

<sup>12</sup> EÖ's assumptions here are closely related (though not identical) to Crawford's (2003)  $L0$  specifications for senders and receivers: a "credible" sender, which tells the truth; and a "credulous" receiver, which believes whatever it is told. Given these  $L0$ s, in Battle of the Sexes 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  $D$ . Therefore, even one round of one-sided communication almost trivially solves the coordination problem.

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 EÖ, I do this by giving priority to the credible sender type and dispensing with the credulous receiver type. Thus I assume that an *L0* player uniformly randomizes its action, without regard to its partner's message, and sends a truthful message.<sup>13</sup> This *L0* is intuitively plausible—bearing in mind that it is only the starting point for players' strategic thinking—with fairly strong experimental support.<sup>14</sup> It is a generalization of Section 1's uniform random *L0*, the usual specification for games without communication. Because Crawford's one-sided model did not need to specify a priority between sent and received messages, it is also a generalization of his credible sender type.

In deriving the behavior of *L1* and higher types, I assume that a type always chooses an action with the highest expected payoff, given its beliefs. As in previous applications (e.g. CHC, EÖ, Crawford and Iriberry, 2007b), 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, as in Demichelis and Jörgen (2008) and (in a somewhat different form) in Kartik (2009), 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.<sup>15</sup> I also assume, on the grounds that message preferences are weaker than action preferences, 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.

### 3.2. Types' strategies

I now characterize the behavior of *L1* through *L4* in Battle of the Sexes with one round of communication. 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. It therefore accommodates by choosing action D if it receives message h from its partner, and choosing action H if it receives message d. Because *L1* expects its own message to be ignored, truthful and untruthful messages would yield it the same payoffs, and it would therefore prefer to be truthful. However, at the time it chooses its own message it has not yet received its partner's message, and so it cannot predict its own action. Further, 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 them 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. It 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 like *L1*, *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, on the grounds that message preferences are weaker than action preferences, 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.)

<sup>13</sup> If it is assumed instead that *L0* uniformly and independently randomizes its message as well as its action, then communication is completely ineffective and the model reduces to Section 1's model without communication. The credulous receiver type, because it deals with beliefs about another player's communication strategy, is arguably less fundamental than the credible sender type. Crawford and Iriberry (2007b) argue that *L0* should be as nonstrategic as possible, and show (in a completely different context) that this tends to yield a more useful model.

<sup>14</sup> See the experiments reported in Blume et al. (2001), Cai and Wang (2006), and Wang et al. (2010). Truthful *L0*s also play important roles in Crawford and Iriberry's (2007a) analysis of auctions and in the classical literature on deception; Crawford (2003) gives further references.

<sup>15</sup> In Crawford and Iriberry's (2007a) analysis of common-value auctions, they assume that level-*k* types can draw inferences about others' private information from their bids, but not inferences about others' types. In Crawford (2003) inferences from others' messages about their types are drawn by the *Sophisticated* type (whose decisions are in equilibrium, taking non-equilibrium players' decisions into account), but not by the level-*k* types. EÖ (online Appendix 3) assume that even level-*k* types draw such inferences in their analysis of CHC's Poisson cognitive hierarchy model, where types above *L1* have priors with positive weights on all lower types. Adding CHC's cognitive-hierarchy types or Crawford's *Sophisticated* type would cloud the waters here without adding insight.

**Table 2**  
Level- $k$  messages and outcomes with one round of communication.

Type (message)	$L1$ (random)	$L2$ (h)	$L3$ (d)	$L4$ (h)
$L1$ (random)	$\frac{1}{2}H + \frac{1}{2}D, \frac{1}{2}H + \frac{1}{2}D$	D, H	H, D	D, H
$L2$ (h)	H, D	H, H	H, D	H, H
$L3$ (d)	D, H	D, H	D, D	D, H
$L4$ (h)	H, D	H, H	H, D	H, H

### 3.3. Coordination outcomes

Table 2 gives the messages for all types and the coordination outcomes on the non-equilibrium path for all type pairings. “ $\frac{1}{2}H + \frac{1}{2}D, \frac{1}{2}H + \frac{1}{2}D$ ” refers to players’ independently random choices in  $L1$  versus  $L1$ , which make all four possible outcomes equally likely.

There are three notable differences from Table 1’s coordination outcomes for the level- $k$  model without communication. 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:  $L2$  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 accommodated by choosing D. To put it another way, returning to Section 1’s Stackelberg analogy, without communication  $L1$  is effectively committed (in  $L2$ ’s mind) to choosing H; but with communication  $L1$  is not committed not to listen, and this allows  $L2$  to use its message to take over the leadership role.  $L3$  expects its partner to choose H without regard to the messages and so accommodates, sending d and choosing D; but without communication  $L3$  expected  $L2$  to choose D, and so accommodated by choosing H.  $L4$  expects its partner to choose D without regard to the messages and so accommodates, sending h and choosing H; but without communication  $L4$  expected  $L3$  to choose H, and so accommodated by choosing D.

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. It 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 admittedly 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 Section 1’s level- $k$  outcomes without communication, or Section 2’s equilibrium outcomes with one round? With symmetry across player roles ex ante, little is lost by focusing 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), respectively, 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.<sup>16</sup> Suppose for definiteness that the frequencies of  $L1$ ,  $L2$ ,  $L3$ , and  $L4$  are  $r \approx 0.4$ ,  $s \approx 0.3$ ,  $t \approx 0.2$ , and  $u \approx 0.1$  respectively, which are probably reasonable estimates (CGCB, CGC). Then the overall coordination rate without communication is  $2(r+t)(s+u) \approx 0.48$ , while with communication the overall coordination 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 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.

### 3.4. Reevaluating Farrell’s assumptions about which message combinations create agreements

Now recall Farrell’s assumptions about which message combinations create agreements. Focusing on the model with one round of communication, he assumes that a message pair that identifies a pure-strategy equilibrium in Battle of the Sexes is treated as an agreement to play that equilibrium, and that players otherwise play the mixed-strategy equilibrium in that game.

As indicated in Table 2, on the non-equilibrium path  $L1$  sends a random message,  $L2$  and  $L4$  send h, and  $L3$  sends d. 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 that a message pair that identifies an equilibrium is treated as an agreement to play that equilibrium.

<sup>16</sup> Thus, as EÖ conclude, one round of two-sided communication yields “little overall coordination” relative to the great success of one-way communication, although more than in the mixed Nash equilibrium or the Farrell outcome.



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.<sup>17</sup> 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, and of the "right" equilibrium; but for these pairings communication never enhances coordination.)  $L1$ 's listening skills bring about coordination often enough to 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.

#### 4. A level- $k$ model with abundant communication

Although Section 1's level- $k$  model of tacit coordination improves upon the mixed-strategy equilibrium in Battle of the Sexes, it is not yet clear how close a level- $k$  model with abundant communication will come to the outcome distribution of Farrell's equilibrium with abundant communication. This section extends Section 3's analysis to allow abundant communication. It then compares level- $k$  coordination outcomes with abundant communication to Section 3's outcomes with one round, and to the outcomes in Farrell's equilibrium analysis of abundant communication.

##### 4.1. Modeling abundant level- $k$ communication

Recall that Farrell's equilibrium analysis of abundant communication assumes that players continue exchanging messages indefinitely until an agreement is reached. I assume instead, in the spirit of Rabin's analysis, that players can always agree to continue for an additional round of (two-sided) communication by mutual consent, but that in any round either player can unilaterally cut off communication and force play of the underlying game. Finally, I assume that players have a slight preference for avoiding additional rounds, all else equal.

The model adds players' options to request to continue communication to a multistage version of Section 3's model, as a pair of simultaneous decisions in each round following the exchange of messages. If both players request to continue, then communication continues for (at least) one more round. Otherwise the communication phase ends and players play the underlying game. As is usual in unanimity games, 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, in the spirit of Section 3's model, 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 equilibrium players do in Feddersen and Pesendorfer, 1996 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.

##### 4.2. Types' communication strategies with abundant communication

The analysis of types' communication strategies with abundant communication builds on Section 3's analysis to determine which type pairs, following which realized message pairs in the current round, decide to exercise the option to extend communication.

Note first that both players requesting to continue communication can never be better for both players if their current messages already identify a pure-strategy equilibrium in Battle of the Sexes. If communication is cut off they will play that equilibrium, which is fully Pareto-efficient (not just efficient in the set of equilibria, which is not all that is relevant for level- $k$  types). By continuing they incur the slight cost of an additional round of communication, and no deviation from the Pareto-efficient current agreement 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:  $L1$  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; and  $L2$  or  $L4$  versus  $L2$  or  $L4$  following their normal message pair h,h.

<sup>17</sup>  $L1$  can be described as a good listener but a bad talker.  $L2$ , by contrast, is a good talker but a bad listener; and  $L3$  and  $L4$  are good talkers but mediocre listeners—mediocre because they choose the right action on the non-equilibrium path, but they are too sure of their beliefs to respond to their partners' messages when the messages contradict their beliefs. No 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-level types have communication skills no better than  $L1$ 's through  $L4$ 's. As Rabin notes, an equilibrium analysis also fails to explain a meeting of the minds, as opposed to assuming one. It is possible that a full meeting of the minds requires more than mechanical decision rules, something like a Gricean leap of the imagination (Grice, 1975). Compare the notion of "team reasoning" in the experimental coordination literature (e.g. Crawford et al., 2008 and the references cited there).

**Table 3**  
Level- $k$  outcomes with abundant communication.

Type	$L1$	$L2$	$L3$	$L4$
$L1$	$\frac{1}{2}H + \frac{1}{2}D, \frac{1}{2}H + \frac{1}{2}D$ 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
$L2$	H, D	H, H	H, D	H, H
$L3$	D, H	D, H	D, D (?)	D, H
$L4$	H, D	H, H	H, D	H, H

First consider  $L1$  versus  $L1$  following message pair d,d. Each expects to play H against its 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 cut off communication, and so play H, H in the underlying game.<sup>18</sup>

$L1$  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 they know there is potential for improvement. Even so, it may seem pointless to continue to communicate, because they know they will still be the same people who have just failed to reach an agreement in a round exactly like the one that would ensue. Recall however that  $L1$ 's message is random because  $L1$  cannot predict its own action before receiving its partner's message, so that both messages have equal expected payoffs and are equally likely to be true. If the randomness of  $L1$ 's message is an unstudied response to those indifferences—as for example in a logit error distribution—then the random outcomes need not be correlated each round, even though the setting is the same. 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 below that if they 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 for them to continue if and only if  $(1+a)/3 > 1$ , or equivalently if  $a > 2$ .

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 the pair d,d, then they cut off communication and play H, H. When  $a < 2$ , if they send the pair h,h they cut off communication and play D, D. When  $a > 2$ , if they send the pair h,h they continue communicating for (at least) one more round. In that case, under my assumption that the types draw no inferences from the history of their interactions, the process is a Markov chain, with all states but h,h absorbing. Letting  $x, y$ , and  $z$  be the probabilities that the process converges to H, H; D, H; or H, D respectively, the transition probabilities imply  $x = \frac{1}{4} + \frac{1}{4}x$ ,  $y = \frac{1}{4} + \frac{1}{4}y$ , and  $z = \frac{1}{4} + \frac{1}{4}z$ . Thus  $x = y = z = 1/3$ , and the ex ante coordination rate of  $L1$  versus  $L1$  following message pair h,h is  $2/3$ .

For  $L1$  versus  $L1$  following message pair h,h when  $a > 2$ , the definition of  $L1$  gracefully overcomes what might appear an insurmountable problem in extending Farrell's equilibrium analysis of abundant communication to a level- $k$  model: These models concern repeated interaction in fixed pairs, and Farrell's use of communication to solve the coordination problem 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.

Now consider  $L2$  or  $L4$  versus  $L2$  or  $L4$ . Like  $L1$  versus  $L1$  following message pair d,d, these types are too optimistic to continue communicating. Instead they too cut off communication after the first round, and so play H, H in the underlying game.

Finally, consider  $L3$  versus  $L3$ . Like  $L1$  versus  $L1$  following message pair h,h, their beliefs are too pessimistic. But unlike  $L1$ 's messages,  $L3$ 's are deterministic, so they may conclude that it is pointless to continue communicating anyway. If they do continue, they are doomed to repeat d,d forever and never reach an efficient agreement. The only ray of hope is that, if they do continue communicating, and there is some exogenous randomness in how messages are sent or received,  $L3$  versus  $L3$  might eventually reach an efficient agreement by accident.<sup>19</sup>

#### 4.3. Coordination outcomes

Table 3 gives the coordination outcomes on the non-equilibrium path for all type pairings with abundant communication. As in Table 2, “ $\frac{1}{2}H + \frac{1}{2}D, \frac{1}{2}H + \frac{1}{2}D$ ” refers to the uniform distribution over the four possible coordination outcomes for  $L1$  versus  $L1$  following message pair h,h when  $a < 2$ . The outcomes with abundant communication are the same as with

<sup>18</sup> I am grateful to Navin Kartik for correcting an error in my initial analysis of this case.

<sup>19</sup> By contrast, such randomness in communication is superfluous for  $L1$  versus  $L1$  following message pair h,h and won't significantly alter their chances of agreeing on an efficient equilibrium. And it won't help  $L1$  versus  $L1$  following message pair d, d or  $L2$  or  $L4$  versus  $L2$  or  $L4$  escape their traps, because their optimistic beliefs will still make them cut off communication after the first round.

one round, except that if  $a > 2$ ,  $L1$  versus  $L1$  now have coordination rate  $2/3$  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. Table 3 reflects the first change in that the  $L1$  versus  $L1$  cell now contains “ $1/3H, H+1/3D, H+1/3H, D$  if  $a > 2$ ,” reflecting players’  $1/3$  limiting probabilities of  $H, H; D, H$ ; or  $H, D$ ; and the second change in that the  $D, D$  outcome in the  $L3$  versus  $L3$  cell now has a question mark after it.

Updating Section 3’s calibration, 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  $t^2(1-0) = 0.06$ , raising the coordination rate to approximately 0.77 or 0.74. Thus, with abundant communication the level- $k$  coordination rate is greater than the equilibrium coordination rate,  $2/(1+a)$ , which equals 1 when  $a=1$ ,  $2/3$  when  $a=2$ , and converges to 0 like  $1/a$  as  $a \rightarrow \infty$ , whenever  $a > 1.94$  and possibly for lower values of  $a$ . To the extent that level- $k$  types do better than in Farrell’s equilibrium, they do so because, as in Section 1’s analysis, the level- $k$  model relaxes the equilibrium incentive constraints.

Just as for equilibrium, the benefits of abundant communication are limited and most of the gains from communication would be 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, reducing  $L1$  versus  $L1$ ’s coordination rate.)

## 5. Conclusion

This paper has reconsidered FR’s analyses of coordination via preplay communication, focusing on Farrell’s analysis of Battle of the Sexes and replacing FR’s equilibrium and rationalizability assumptions with a structural non-equilibrium model based on level- $k$  thinking. The analysis gives a unified treatment of players’ messages and actions and how messages create agreements, and allows a reevaluation of FR’s assumptions 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. Further, with abundant communication, as Rabin’s analysis of negotiated rationalizability suggests, level- $k$  players need not keep communicating until an agreement is reached as in Farrell’s equilibrium.

Finally, a level- $k$  analysis implies very different conclusions than Farrell’s equilibrium analysis about the effectiveness of communication in Battle of the Sexes. The level- $k$  coordination rate in that game, unlike the equilibrium rate, is largely independent of the difference in players’ preferences. Even with moderate differences in preferences, for plausible type distributions the level- $k$  coordination rate is likely to be higher than the equilibrium rate, with or without communication. The level- $k$  model’s predictions with abundant communication are consistent with Rabin’s bounds based on negotiated rationalizability, but their precision yields additional insight into the causes and consequences of breakdowns in negotiations.

## References

- Bernheim, B. Douglas, 1984. Rationalizable strategic behavior. *Econometrica* 52 (4), 1007–1028.
- Blume, Andreas, DeJong, Douglas, Kim, Yong-Gwan, Sprinkle, Geoffrey, 2001. Evolution of communication with partial common interest. *Games Econ. Behav.* 37 (1), 79–120.
- Cai, Hongbin, Wang, Joseph Tao-Yi, 2006. Overcommunication in strategic information transmission games. *Games Econ. Behav.* 56 (1), 7–36.
- Camerer, Colin, Ho, Teck-Hua, Chong, Juin Kuan, 2004. A cognitive hierarchy model of games. *Q. J. Econ.* 119 (3), 861–898.
- Costa-Gomes, Miguel A., 2002. A suggested interpretation of some experimental results on preplay communication. *J. Econ. Theory* 104 (1), 104–136.
- Costa-Gomes, Miguel A., Crawford, Vincent P., 2006. Cognition and behavior in two-person guessing games: an experimental study. *Am. Econ. Rev.* 96 (5), 1737–1768.
- Costa-Gomes, Miguel A., Crawford, Vincent P., Broseta, Bruno, 2001. Cognition and behavior in normal-form games: an experimental study. *Econometrica* 69 (5), 1193–1235.
- Costa-Gomes, Miguel A., Weizsäcker, Georg, 2008. Stated beliefs and play in normal-form games. *Rev. Econ. Stud.* 75 (3), 729–762.
- Crawford, Vincent P., 1998. A survey of experiments on communication via cheap talk. *J. Econ. Theory* 78 (2), 286–298.
- Crawford, Vincent P., Costa-Gomes, Miguel A., Iriberrí, Nagore, 2013. Structural models of nonequilibrium strategic thinking: theory, evidence, and applications. *J. Econ. Lit.* 51 (1), 5–62.
- Crawford, Vincent P., Gneezy, Uri, Rottenstreich, Yuval, 2008. The power of focal points is limited: even minute payoff asymmetry may yield large coordination failures. *Am. Econ. Rev.* 98 (4), 1443–1458.
- Crawford, Vincent P., Haller, Hans, 1990. Learning how to cooperate: optimal play in repeated coordination games. *Econometrica* 58 (3), 571–595.
- Crawford, Vincent P., Iriberrí, Nagore, 2007a. Level- $k$  auctions: can a non-equilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions? *Econometrica* 75 (6), 1721–1770.
- Crawford, Vincent P., Iriberrí, Nagore, 2007b. Fatal attraction: salience, naivete, and sophistication in experimental ‘hide-and-seek’ games. *Am. Econ. Rev.* 97 (5), 1731–1750.
- Crawford, Vincent P., Kugler, Tamar, Neeman, Zvika, Pauzner, Ady, 2009. Behaviorally optimal auction design: an example and some observations. *J. Eur. Econ. Assoc.* 7 (2–3), 365–376.

- Demichelis, Stefano, Jörgen, W. Weibull, 2008. Language, meaning, and games: a model of communication, coordination, and evolution. *Am. Econ. Rev.* 98 (4), 1292–1311.
- Ellingsen, Tore, Östling, Robert, 2010. When does communication improve coordination. *Am. Econ. Rev.* 100 (4), 1695–1724.
- Farrell, Joseph, 1987. Cheap talk, coordination, and entry. *Rand J. Econ.* 18 (1), 34–39.
- Farrell, Joseph, 1988. Communication, coordination and nash equilibrium. *Econ. Lett.* 27 (3), 209–214. [and 1990. "Erratum," 33(3): 299].
- Feddersen, Timothy J., Pesendorfer, Wolfgang, 1996. The swing voter's curse. *Am. Econ. Rev.* 86 (3), 408–424.
- Grice, H. Paul, 1975. *Logic and conversation* [New York: Academic Press; reprinted in *Studies in the Way of Words*]. Cole, P., Morgan, J. (Eds.), In: Grice, H.P. (Ed.), *Syntax and Semantics, Volume 3: Speech Acts*, 1991. , Harvard University Press, Cambridge, pp. 22–40.
- Harsanyi, John, Selten, Reinhard, 1988. In: *A General Theory of Equilibrium Selection in Games* MIT Press, Cambridge, MA.
- Ho, Teck Hua, Camerer, Colin, Weigelt, Keith, 1998. Iterated dominance and iterated best response in experimental 'p-Beauty Contests'. *Am. Econ. Rev.* 88 (4), 947–969.
- Kahneman, Daniel, 1988. Experimental economics: a psychological perspective. In: Tietz, R., Albers, W., Selten, R. (Eds.), *Bounded Rational Behavior in Experimental Games and Markets*, Springer-Verlag, New York, pp. 11–18.
- Kalai, Ehud, Samet, Dov, 1985. Unanimity games and pareto optimality. *Int. J. Game Theory* 14 (1), 41–50.
- Kartik, Navin, 2009. Strategic communication with lying costs. *Rev. Econ. Stud.* 76 (4), 1359–1395.
- Kartik, Navin, Ottaviani, Marco, Squintani, Francesco, 2007. Credulity, lies, and costly talk. *J. Econ. Theory* 134 (1), 93–116.
- Kawagoe, Toshiji, Tazikawa, Hirokazu, 2009. Equilibrium refinement vs. level-k analysis: an experimental study of cheap-talk games with private information. *Games Econ. Behav.* 66 (1), 238–255.
- Nagel, Rosemarie, 1995. Unraveling in guessing games: an experimental study. *Am. Econ. Rev.* 85 (5), 1313–1326.
- Rabin, Matthew, 1991. *Focal Points in Pre-Game Communication*. UC Berkeley Department of Economics Working Paper 91–179.
- Rabin, Matthew, 1994. A model of pre-game communication. *J. Econ. Theory* 63 (2), 370–391.
- Rapoport, Amnon, Seale, Darryl A., 2002. Coordination success in noncooperative large group market entry games. In: Zwick, Rami, Rapoport, Amnon (Eds.), *Experimental Business Research*, Kluwer Academic, Dordrecht, The Netherlands.
- Rapoport, Amnon, Seale, Darryl A., Erev, Ido, Sundali, James A., 1998. Equilibrium play in large group market entry games. *Manag. Sci.* 44 (1), 119–141.
- Schelling, Thomas C., 1960. In: *The Strategy of Conflict* Harvard University Press, Cambridge.
- Stahl II, Dale O., Paul, W. Wilson, 1994. Experimental evidence on players' models of other players. *J. Econ. Behav. Organ.* 25 (3), 309–327.
- Wang, Joseph Tao-Yi, Spezio, Michael, Camerer, Colin F., 2010. Pinocchio's pupil: using eyetracking and pupil dilation to understand truth-telling and deception in games. *Am. Econ. Rev.* 100 (3), 984–1007.