Journal of Economic Perspectives—Volume 10, Number 3—Summer 1996—Pages 103–118

Cheap Talk

Joseph Farrell and Matthew Rabin

"Simply by making noises with our mouths, we can reliably cause precise new combinations of ideas to arise in each other's minds."
—Steven Pinker, *The Language Instinct* (1994), p. 1

"A verbal contract isn't worth the paper it's written on."
—attributed to Yogi Berra

"Your grandmother was making noises like she was going into town."
—Emily Ann Cramer, circa 1965, verbal communication to J. F.

"Won't that kid ever shut up?"
—W. R., circa 1965, verbal communication to A. R.

Much of modern microeconomics asks how private information is shared through market and other mechanisms. Hayek (1945), in a classic essay, even identified information sharing as the chief efficiency of competitive markets. Subsequent work by Hurwicz (1973) and others has pursued Hayek's intuitions and evaluated economic institutions in terms of their effects on information

■ *Joseph Farrell is Professor of Economics, University of California, Berkeley, California. He is also serving as Chief Economist at the Federal Communications Commission, Washington, D.C., where he is trying to make talk as cheap as possible; however, his utterances should not be interpreted as the views of the Commission or of any Commissioner. Matthew Rabin is Assistant Professor of Economics, University of California, Berkeley, California. Their e-mail addresses are farrell@econ.berkeley.edu and rabin@econ.berkeley.edu, respectively.*

transmission. On a smaller scale of analysis, game theorists and economists have found many instances of "signaling," as discussed by Spence (1974), where an informed agent takes costly actions that (according to a somewhat complex but inexorable logic) reveal the agent's information or "type." Yet, we suspect, most information sharing is not done through Spence-style signaling, through the price system, nor through carefully crafted Hurwicz-style incentive-compatible mechanisms: it is done through ordinary, informal talk.

How much can such "cheap talk" accomplish? Economists are inconsistent about this question. Some scathingly ask, when talk is cheap (meaning that it does not directly affect payoffs), what incentive is there to tell the truth? Others assume that "communication" will lead players to Nash equilibria: indeed, to equilibria that are Pareto-efficient within the set of equilibria. Both of these extreme views, we think, are wrong.

The scathing view, we believe, is a verbal confusion: talk is cheap (it does not *directly* affect payoffs), but, given that people respond to it, talk definitely affects payoffs. A misinformed listener will do something that is not optimal for himself and, if their interests are sufficiently aligned, this is bad for the speaker too. In a nutshell, this is how cheap talk can be informative in games, even if players ruthlessly lie when it suits them.[1]

Consider an example. The Rayco Corporation wants to hire Sally in one of two positions. One position demands high ability; the other is better done by someone of low ability (an able person will get bored and perform badly). Rayco does not know Sally's ability (which may be either "high" or "low"); it begins with 50–50 beliefs. Sally knows her own ability.

Rayco wants to give Sally the demanding job if she has high ability and the undemanding job if she has low ability. For this example, we suppose that if Sally has high ability she will enjoy the demanding job and will find the undemanding job boring, whereas if she has low ability she will be stressed out by the demanding job and find the undemanding job manageable. We can represent the situation as Example 1:

		Job Rayco gives Sally	
		Demanding	Undemanding
Sally's ability	High	2,1	0,0
	Low	0,0	1,3

[1] Valley, Thompson, Gibbons and Bazerman (1995) found that, in bargaining with cheap talk, experimental subjects outperformed the Myerson-Satterthwaite theoretical upper bound on gains from trade with private information. We interpret this, as well as the fact that people typically say what they want to have been believed even when the incentives clearly imply that cheap talk should not be believed, as suggesting that some people tell the truth despite incentives to lie.

There are four possible outcomes, corresponding to Sally's ability (high or low) and Rayco's choice of job offer (demanding or undemanding). In each cell, we write Sally's payoff and then Rayco's. Formally, this is an incomplete-information game, where Sally can be one of two "types" and where only Rayco has a (payoff-relevant) move.

Readers steeped in the signaling literature may already be thinking about how a high-ability Sally can engage in costly education beyond what a low-ability Sally would be willing to imitate . . . but this is a red herring. Costly signaling is surely important in economics (though we hope education is nobler than that), but in this example there is no need for it. Sally will just tell Rayco her ability, and Rayco will believe her: she has no reason to lie. This private-information problem is resolved through cheap talk.

In this paper we try to convey, without too much in the way of notation or theorems, a broader sense of when cheap talk can communicate private information in equilibrium, and of whether those equilibria or others are likely to arise. We argue that cheap talk can and often does matter, but it does not generally lead to efficiency.

Cheap Talk About Private Information

To formalize our intuitive claim about what will happen in Example 1, we view it as a two-stage game. In the first stage, Sally can say something to Rayco; in the second stage, Rayco decides which job to give her. Because "say something" represents a choice from a large set of actions (all reasonably short utterances in English, for a start), this game is very large. Although we learn a lot from a simplified version in which we actively consider only the possible messages "My ability is high" and "My ability is low," we should be aware that Sally *could* say other things.

In this two-stage game, the following strategies constitute an equilibrium.[2] Sally says "my ability is high" if her ability is high and "my ability is low" if her ability is low. Rayco infers that Sally's ability is high, and gives her the demanding job, if she says "my ability is high"; it infers that her ability is low, and gives her the undemanding job, if she says "my ability is low." If Rayco responds to her cheap talk in this way, Sally has no incentive to lie, because that would induce Rayco to make a mistake, and in this example a mistake for Rayco is bad for Sally.

A useful (if stilted) alternative way to say this is that Sally has preferences over Rayco's beliefs about her ability, because Rayco relies on those beliefs when it does something—assign her to a job—that she cares about. Each of Sally's possible types

[2] To be precise, they constitute a perfect Bayesian equilibrium, which is the commonly used solution concept for multistage games with private information. "Sequential equilibrium" (Kreps and Wilson, 1982) is often equivalent, including here. These equilibrium concepts are in effect generalizations of subgame-perfect equilibrium to games with private information: see, for instance, Gibbons (1992) for details.

(high ability and low ability) is "self-signaling": that is, she wants Rayco to think she has high ability if and only if she has, and similarly for low ability. Knowing this, Rayco is happy to use a strategy that effectively delegates its action to Sally, by making its job assignment depend on her cheap talk. In this example, there are no incentive problems and cheap talk can convey all the information.

Unfortunately, things are not always so nice. Consider next a modified problem, reminiscent of those discussed by Spence (1974). Here, Rayco still wants to put Sally in the demanding job if her ability is high and in the undemanding job if her ability is low; but now, perhaps because performance is hard to monitor and the demanding job is better paid, Sally wants the demanding job whatever her true ability. The bimatrix for Example 2 is:

		Job Rayco gives Sally	
		Demanding	Undemanding
Sally's ability	High	2,1	0,0
	Low	2,0	1,3

Sally's preferences over Rayco's beliefs are no longer correlated with the truth: instead, whatever the truth, she wants Rayco to think she has high ability. Neither type, therefore, is self-signaling: high ability is not because a low-ability Sally would also claim to have high ability, and low ability is not because a low-ability Sally does not want to be known as such. Because of this lack of self-signaling or lack of correlation, cheap talk cannot now convey Sally's private information about her ability level; either no information will be conveyed, or Sally will engage in costly signaling.

In the two examples given thus far, correlation between the "sender's" (Sally's) true type and her preference over the "receiver's" (Rayco's) beliefs either is perfect or fails completely.[3] We now describe, in Example 3, a more complicated setting with some correlation but also some conflict. This is a variant of a classic model developed by Crawford and Sobel (1982).

Now, Sally's ability lies on a continuum, rather than being just "high" or "low." Based on Rayco's beliefs about her ability, it will set her wage and make workplace demands on her: if it believes she has higher ability, it will demand more and will pay more. Sally knows her ability level, but Rayco only has its initial beliefs and what Sally says.

Because she likes money, Sally may well want to persuade Rayco that her ability is somewhat higher than it actually is. Extreme exaggeration may not be tempting,

[3] The correlation may fail even more dramatically—though with no more devastating results—if Sally's preference over Rayco's belief is *negatively* related to the truth. For example, suppose that high-ability types like to goof off in the undemanding job, but low-ability types like the higher salary associated with the demanding job.

however: even at a high salary, Sally does not want to talk herself into a demanding job she cannot handle. Depending on the details of the problem, it may turn out that if her true ability is t, she would most like Rayco to believe that her ability is $t + b$, where b is a positive parameter that is known to both players.

Can cheap talk be credible in such a problem? It's tempting to think not, because if Rayco "discounts" Sally's claimed ability by some amount, c, then Sally can just claim her ability is $t + b + c$, and thus fool Rayco. The flaw in this argument is the assumption that if cheap talk is meaningful, it must convey a *precise* meaning. If, instead, Rayco always believes one of a limited set of things about Sally's ability, then she might find that the only available exaggerations are too large to be tempting. Crawford and Sobel (1982) showed how, in such a model, *imprecise* cheap talk can indeed be an equilibrium, provided that Sally does not want to exaggerate too much (b is not too large). Larger values of b force cheap talk to take on less precise meaning: the maximum amount of communication, in some sense, depends on the extent to which the parties' interests are aligned.

These three examples suggest some general principles. Sometimes there is no incentive to lie, and cheap talk will fully convey private information. If there is too strong an incentive to lie, cheap talk becomes meaningless. However, even if there is some limited incentive to lie, cheap talk can convey some meaning in equilibrium.

As these principles suggest, cheap talk matters in a variety of mixed-motive economic interactions involving private information. For instance, Farrell and Gibbons (1989) and Matthews and Postlewaite (1989) show that cheap talk can matter in bargaining; Austen-Smith (1990, 1993) and Matthews (1989), among others, have shown how cheap talk matters in political contexts; and there is a growing literature of similar applications elucidating how limited common interest may lead to meaningful cheap talk.

If Cheap Talk Can Matter, Does It?

We have argued that cheap talk can convey information, and, following the usual game-theoretic approach, we backed up that claim by displaying *an* equilibrium in which it does. We did not claim that the equilibrium was unique; indeed, it is not. In any game, however "obvious" it may be that cheap talk will convey information, game theorists traditionally argue that it need not. It is consistent with common knowledge of rationality, and with equilibrium, for cheap talk to be completely ignored. This section addressed this uncomfortable state of affairs.

In Example 1, it was a natural equilibrium for Sally to describe her ability as "high" or "low," and for Rayco to give her the corresponding job.[4] It is also an equilibrium for her to say "swordfish" if her ability is high and "sauerkraut" if her

[4] Again, the idea of equilibrium should be understood in a technical sense as a perfect Bayesian equilibrium. See note 2.

ability is low—as long as Rayco understands this language. It is even an equilibrium for Sally to say "my ability is low" if her ability is high, and "my ability is high" if her ability is low, as long as Rayco gives her the more demanding job when she claims low ability, and the less demanding job when she claims higher ability.[5] Obviously, zillions of these equilibria exist: they differ in the assignment of messages to meanings—that is, in the language they use—but the differences are "inessential": in each, Sally ends up in the right job through cheap talk.

But there is another, essentially different, equilibrium. In this equilibrium, Rayco's 50–50 beliefs are unaffected by what Sally says. Given that Rayco ignores Sally's utterances, she might as well "babble"—that is, make noises that are uncorrelated with her type. In turn, her babbling justifies Rayco's strategy of ignoring cheap talk (and thus always assigning the undemanding job, which is its best move in expected value given its 50–50 prior). The point is quite general: it is always consistent with rationality to treat cheap talk as meaningless.

Such a "babbling" equilibrium always exists.[6] Since there is no intrinsic reason why Sally's noises would be correlated at all (let alone in any particular way) with her private information, one might think that the most natural, focal, or symmetry-preserving approach is to suppose that there is no such correlation and to ignore cheap talk—implementing the babbling equilibrium. If we didn't know better, we might think that this was the most reasonable outcome.

But we *do* know better. People don't usually take the destructively agnostic attitude that "I won't presume that the words mean what they have always meant." Rather, people take the usual or literal meaning seriously. This doesn't mean they believe whatever they hear; rather, they use the usual meaning as a starting point and then assess credibility, which involves asking questions such as, "Why would she want me to think that?"[7]

No such skeptical respect for language appears in the babbling equilibrium in Example 1. Instead, language such as "my ability is high" is gratuitously treated as meaningless—meaningless not because incentive problems imply that cheap talk cannot be credible in equilibrium, as in Example 2, but just because all talk is treated as babbling in the equilibrium! In this sense, the babbling equilibrium is absurd. When players share a language, they will not behave like that in Example 1.

Using confident introspection—with which we believe almost everyone would agree on reflection—we have disparaged one equilibrium in that example and favor

[5] The modern Greek for "yes" is "ne" and for "no" is "okeh," so an American visitor to Greece sometimes feels as if this experiment is being performed.

[6] Seidman (1992) gave an example of a game without a cheap-talk stage that has no sequential equilibrium, but in which, if a cheap-talk stage is added, there is a unique equilibrium, which involves meaningful communication. We believe (for different reasons) that this result does not contradict our key point.

[7] People in reality do not seem to lie as much, or question each other's statements as much, as game theory suggests they should. For the purposes of this work, we'll model ruthless economic Sally who tells the truth only whenever she finds it pays, and we'll suppose that Rayco expects that.

another one.[8] How can we extend this to cheap-talk games in general as a "refinement" of equilibrium—that is, a rule for picking out some equilibria that are more plausible than others?[9] Our predictions in Examples 1 and 2 might suggest one possible refinement: that cheap talk will communicate as much as is consistent with avoiding incentives to lie. But this would not capture the reasoning behind our intuitions in those examples, and in general it gives different and, we think, wrong results.

To illustrate, consider Example 4, the right-to-silence game. Sally knows which one of two tasks is efficient to perform. Rayco could hire Sally specifically to perform Job 1, specifically to perform Job 2, or as a highly paid manager who will choose which job to perform. If Rayco knew which task is efficient, it would still hire her to perform the task, but at a lower salary, because she has lost her informational advantage. Sally wants to be hired as manager, but prefers to be hired to do the right task and be more productive rather than to do the wrong task and be less productive. The payoffs to Example 4 are:

		Job1	Job2	Manager
Sally's knowledge	Task 1 efficient	2,5	1,−2	3,3
	Task 2 efficient	1,−2	2,5	3,3

Just as in Example 1, there are two essentially different equilibria: there is (as always) the uninformative "babbling" equilibrium, and there is a fully revealing equilibrium, where Sally reveals what is efficient before being hired. In contrast to Example 1, we will argue that the revealing equilibrium is implausible here.

In the revealing equilibrium, why doesn't Sally deviate and refuse to reveal her private information? The only possible answer is that refusing to reveal her information is not an option given that Rayco takes *every* possible message as meaning either that Job 1 is efficient or that Job 2 is. Even if Sally makes such sounds as "I refuse to tell you what's efficient before you hire me as manager," or "Notice that my refusal to reveal which task is efficient signals nothing, since whatever my private

[8] To some extent, recent experiments test and, we think, broadly confirm our arguments. Crawford (1995) provides an excellent brief review of experiments on cheap talk. So far as we are aware, the only careful experiment on the role of a common language in revealing private information is Sopher and Zapater (1993), who provide mixed support for the types of restrictions implied by Rabin (1990). More common are experiments on preplay communication regarding intentions (see below). Cooper, DeJong, Forsythe and Ross (1989, 1994), for instance, provide support for Farrell's (1987) notion of communication increasing coordination in the battle of the sexes—as well as the hypothesis that more rounds of talk will yield more coordination than one round.

[9] Some readers may be confused at the statement that the "equilibrium"—often in economics a term for what we believe will happen—is not plausible. Indeed, researchers in game theory sometimes write things like "the perfect Bayesian equilibrium is not really an equilibrium." Just to clarify: to say that a set of strategies constitutes a perfect Bayesian equilibrium means that they satisfy certain technical conditions; in many contexts that makes the perfect Bayesian equilibrium a reasonable outcome, but (as here) it does not always do so.

information is, I prefer that you not know it," or "la di dah," Rayco must (for the equilibrium to work) infer either "ah, she's revealing that Job 1 is efficient," or "ah, she's revealing that Job 2 is efficient." Otherwise, Sally would indeed say some such thing, so revealing would not be equilibrium. This is entirely consistent in the weird world of this equilibrium, but ludicrous from a broader viewpoint. In other words, the stability of the revealing equilibrium here depends on interpreting certain messages in ways that, while logically consistent and not inviting deception, are weird.[10]

Another way to say this is that the revealing equilibrium in this game can only happen if the players use a language that contains ways to say "Job 1" and "Job 2" but no way to say "no comment." If, in fact, they share a rich language such as English, then the "implausible equilibria" can survive only if *language is gratuitously misunderstood.* For the revealing equilibrium in Example 4, Rayco has to misunderstand "I won't tell you"; for the babbling equilibrium in Example 1, Rayco has to misunderstand "I have high ability." Contrary to what these equilibria specify, we think that in reality Sally surely can make such messages *understood*; and, once understood, they will be *believed*, for she has no incentive to lie about them.

The view that cheap talk may be blocked by incredulity (as in Example 2) but not by incomprehension has been called the "rich-language assumption," because it really assumes that players share a pre-existing rich common language in which a desired message can be expressed and understood. In the context of simple "sender-receiver" games such as Examples 1 to 4, where Sally sends a cheap-talk message and Rayco receives it, Farrell (1993) pursues the rich-language assumption and its implications. The approach generalizes our discussion of Examples 1, 2 and 4.[11]

Talking About Intentions

Often when people talk (and listen), the topic is not exogenous private information, as above, but rather what we plan to do. Cheap talk is used to coordinate, as when we agree on where to meet for dinner. The same problems arise as in talking about private information, but with some added issues.

Consider Example 5: pure coordination. Susan and Roberto work at the same office building in a small town and are both going to eat out tonight. They want to eat together. There are four eateries in town; three are of equal quality, but the

[10] We describe these interpretations as "weird," but some readers might ask whether they are not natural if Rayco can commit itself to, for example, hiring Sally for Job 1 unless she announces that she'd prefer Job 2. We agree that Rayco would like to make such a commitment and may try to do so, but in a one-shot setting *without* commitment, ex post, it will not be a best response after one of the "I won't tell you" messages.

[11] While we believe this approach is broadly correct, the exact way to implement these intuitions is more problematic. Modifications of Farrell's "neologism-proofness" have been proposed by Rabin (1990), Matthews, Okuno-Fujiwara and Postlewaite (1991), and others.

fourth—down by the town's (sole) railroad station, Grand Empire Station—is of lower quality. The game can be represented as follows, where the choice for each player is where to dine, and in each cell we show (in the conventional game-theory fashion) Susan's payoff followed by Roberto's:

		Roberto			
	Eatery:	1	2	3	Empire
Susan	Eatery 1	3,3	0,0	0,0	0,−2
	Eatery 2	0,0	3,3	0,0	0,−2
	Eatery 3	0,0	0,0	3,3	0,−2
	Empire	−2,0	−2,0	−2,0	1,1

Susan and Roberto are leaving from the same parking garage, using separate cars parked next to each other. What will happen? Intuitively, like Sally and Rayco in Example 1, they surely will talk and thus solve their problem: they will eat together at one of the good restaurants. How does this work?

Susan can say, "I'm off to Eatery 2." As in the analysis of cheap talk about private information, one could take a destructively agnostic attitude and note that there is a "babbling" equilibrium in which this statement means nothing. But people who think like that tend to eat alone. Empirically, it makes more sense to recognize that Roberto knows the literal meaning of Susan's remark, and he will evaluate it to see whether it is credible; if it is, he will believe it.

How does Roberto evaluate the credibility of such a remark? If Susan's destination were exogenous private information, like Sally's ability, then Roberto would ask two questions. First, if Susan is really going to Eatery 2, would she want me to believe she is? Here, she would. Second, if Susan is really going somewhere else, might she want me to believe she's going to Eatery 2? Here, she would not. So the message "I'm off to Eatery 2" is self-signaling: Susan wants to say it if and only if it is true.

Roberto should also ask a third question, which has no analogy in talking about exogenous private information. If Susan thought she had persuaded him that she's going to Eatery 2, would she have an incentive to go to Eatery 2? This test, too, is passed here: if she thinks he believes she is going to Eatery 2, she expects him to go there, and so she wants to go there. We can describe this by saying the message "I'm going to Eatery 2" is self-committing: if believed, it creates incentives for the speaker to fulfill it.[12]

[12] This is the same as saying that Susan's going to Eatery 2, along with Roberto's best response (here, going to Eatery 2), constitute a Nash equilibrium in the game without cheap talk. We introduce the term

A message that is both self-signaling and self-committing seems highly credible. Accordingly, provided there is no problem of making her messages understood, Susan can ensure a payoff of 3 by using this remark (or, similarly, by naming another of the good places), or a payoff of 1 by naming the Grand Empire Diner; or she can refrain from saying anything and take her chances. Since taking her chances cannot get her more than 3, and can easily do worse, it seems clear that she will name one of the good places. Formally, if it is common knowledge that self-signaling, self-committing messages are believed, then in every subgame-perfect equilibrium each player gets a payoff of 3. Cheap talk efficiently resolves this coordination problem.

Cheap Talk versus Convention

A popular alternative to talk in resolving coordination problems is convention. Many informal introductions to game theory stress Schelling's (1960) examples of how focal points develop and help people to coordinate. In one famous example, Schelling discussed (theoretically and experimentally) the problem facing people who had planned to meet in New York but forgotten to say where. The leading focal point at the time was Grand Central Station.

Although Schelling's work remains fascinating half a life span later, we find this emphasis on tacit coordination surprising, because we think that people in small-numbers coordination problems usually can and will talk. Moreover, in pure coordination games, cheap talk yields more efficient coordination than do Schelling-style focal points. In Example 5, suppose that the eatery at Grand Empire Station is a natural focal point for people who have agreed to meet but have forgotten to say where. Susan and Roberto, having grown up in this town, may both know that is where they should meet. This will be better than not coordinating, but worse than what they can get by speaking. More generally, pre-existing focal points are seldom optimal, because they are not tailored to the situation at hand. How convenient will Schelling's New York visitors find it to go to Grand Central Station at noon? Cheap talk is an excellent solution to such problems, and situations where people cannot talk are the exception, not the rule.[13]

Coordination Under Conflict

What can talk achieve in more challenging problems, where players' interests are not well aligned? Because of conflict, messages are less likely to be self-signaling or self-committing, and cheap talk will be less successful or less informative. An extreme example is Example 6, the prisoner's dilemma, which we present as the choice of high or low effort in a joint enterprise:

"self-committing" because we want to stress that this is a two-stage process in which talk not only reveals information but also itself changes the speaker's incentives (through changing what she expects the listener to do).

[13] But this does provide one reason to work toward cheap wireless telephony.

		Roberto	
		high effort	low effort
Susan	high effort	7,7	5,8
	low effort	8,5	6,6

Each player reasons that whether the other player uses high effort or low effort, her/his own selfishly best response is low effort. But this leads to the outcome (6, 6), which is Pareto-dominated by the outcome (7, 7), where both players offer high effort. Can Susan and Roberto talk their way out of this? Clearly not. To do so, they would have to coordinate on high effort. But suppose Susan says, "I will put in high effort; I expect you to do the same." This message is not self-signaling (Susan would like Roberto to put in high effort whatever she plans to do), and it is not self-committing (Susan will have no incentive to follow through on her promise, whether or not she expects Roberto to believe her); moreover, even if Roberto believes Susan plans to put in high effort, he will have no incentive to do so. Whatever they say, "low effort" remains a strictly dominant strategy.[14] As this example suggests, credible messages must be self-committing.

On the other hand, self-committing (Nash equilibrium) messages need not be entirely credible, as Example 7 suggests (Aumann, 1990). Each of two hunters must decide what to hunt one morning. If both hunt stag, both do very well; if both hunt rabbits, both do OK; but if Artemis alone hunts stag while Calliope hunts rabbits, Artemis will fail and Calliope will have no competition in catching rabbits. The payoffs can be represented as:

		Calliope	
		Stag	Rabbits
Artemis	Stag	9,9	0,8
	Rabbits	8,0	7,7

Artemis would like to persuade Calliope that she will hunt stag. It is a Nash equilibrium for them both to do so, so the message "I plan to hunt stag" is self-committing: if Artemis thinks Calliope believes it, she will indeed hunt stag. But, as

[14] So why do so many authors introducing the prisoners' dilemma stress that the "prisoners" are held in separate cells and cannot talk to each other? First, real prisoners might be able to *change the game*—Susan might say to Roberto "if you fink on me, I'll kill you" and perhaps be believed. Second, empirical studies suggest that the prisoners' dilemma, played for moderate stakes, is really a coordination game! That is, people behave as if psychological payoffs from "both cooperating" make each player want to cooperate if the other does, and not otherwise. See Rabin (1993) for discussion of reciprocity in games and Sally (1994) for a review of how communication affects play in games like this.

Aumann stresses, the message is *not* self-signaling: Artemis would like Calliope to believe it even if Artemis in fact plans to hunt rabbit. Aumann argues that this makes the message incredible (and therefore that cheap talk does nothing this game, and in particular will not lead the players to the "good" stag-hunting Nash equilibrium); although we see the force of Aumann's argument, we suspect that cheap talk will do a good deal to bring Artemis and Calliope to the stag hunt.

In another application of the same idea, consider an aspect of the theory of collusion. Suppose two duopolists are engaged in repeated Cournot competition. They have, let's suppose, been playing myopically, producing the Cournot equilibrium output each period. Now one duopolist says to the other, "You cut your output, and I'll cut mine." Will this cheap talk make collusion more likely?

Let's interpret the talk as suggesting that they both go to trigger strategies—where each firm restricts its output to half the monopoly output unless one of them has previously "cheated" by producing more, in which case they revert to myopic Cournot play. As is well known, if there is not too much discounting, these trigger strategies constitute an equilibrium in which the firms split monopoly profits. As it is an equilibrium, it is self-committing: a best response to the trigger strategy is the trigger strategy. But the proposal is not self-signaling: duopolist A would like its rival B to adopt this trigger strategy even if A has no intention of doing so! After all, having duopolist B restrict output for one period helps A earn higher profits in that period, even though B will then revert to myopic Cournot behavior. Just as in Aumann's stag hunt example, each player wants the other to play for the (privately) Pareto-superior equilibrium, even if he himself does not plan to do it.[15]

If cheap-talk invitations to collude are not self-signaling, are they ineffectual and socially harmless? Not necessarily: antitrust law does not approve of competitors engaging in cheap talk about pricing, and experimental evidence, as well as common sense, generally supports this position (Holt, 1995, pp. 409–411; Carlton and Perloff, 1994, ch. 7). We cannot be confident that cheap talk won't help players reach a privately Pareto-superior equilibrium even if such talk is not self-signaling.

Conflict in Talk

So far we have considered what happens when just one player talks and the other assesses the credibility of the message. This focuses the discussion on whether messages are credible, which is a major determinant of how much information will be shared by cheap talk. Yet often it is important that more than one player can talk. For instance, consider Example 8: the battle of the sexes. As the classic story is retold by Gibbons (1992):

> A man and a woman are trying to decide on an evening's entertainment; we analyze a gender-neutral version of the game. While at separate workplaces,

[15] Invitations to collude would be self-signaling if the punishment phase were sufficiently worse for the "defector" than never having left simple noncooperative play; this may be the case in practice.

> Pat and Chris must choose to attend either the opera or a prize fight. Both players would rather spend the evening together than apart, but Pat would rather they be together at the prize fight, while Chris would rather they be together at the opera.

We represent these preferences in a conventional game-theory bimatrix:

		Chris: Opera	Chris: Fight
Pat	Opera	2,1	0,0
	Fight	0,0	1,2

While this is usually analyzed as a simple simultaneous-move game, we suggest that this assumes the couple is rather odd. Normal couples, in a situation like this, talk and try to agree on where to go. Although one can forcibly assume that cheap talk is impossible (they must decide right away, are at separate workplaces and can't phone), doing so strips the game of most of what happens in most strategic interactions whose eventual payoffs look like this. Thus, while the (directly) payoff-relevant moves in the battle of the sexes can be represented as this bimatrix, the strategic interaction consists of a negotiation: in other words, this game is preceded by talking about what they will do.

This talk is not tightly structured, so it is not clear how we should model it, but it seems important that both players talk; for simplicity let us assume that there is one round of cheap talk, in which each player simultaneously announces a plan for what he or she might do in the second stage, the actual play. We might call this two-stage game "the battle of the sexes with one round of talk."

This two-stage game has many equilibria, including normatively appealing ones in which the outcome is (2, 1) or (1, 2) with equal probability.[16] But these "fair" equilibria do not seem likely to emerge. More plausibly, each player will argue for his or her preferred equilibrium. That is, this two-stage game is often played as a negotiation game in which cheap-talk "agreements" are sometimes reached; such agreements are generally followed. If Pat says, "I'm going to the opera," and Chris says, "I'm going to the opera," these messages, each of which is self-signaling and self-committing, reinforce each other, so we think it is extremely likely that they will regard themselves as having "agreed" on something self-enforcing, and they will carry out their announcements, just as in Example 5, the game of pure coordination. However, if Pat says, "I'm going to the opera," while Chris says, "I'm

[16] Even if somehow the players cannot toss a coin to decide which of these Nash equilibria to play, they could do something equivalent: for instance, each player can name "odd" or "even," and if both name the same they play (2, 1) and if they differ they play (1, 2); of course, there are many other equilibria with the same distribution of outcomes.

going to the fight," each message individually is self-signaling and self-enforcing but they conflict and cancel: unless it's common knowledge between them who's in charge (in which case probably only that person will speak), they have achieved no coordination but only a "sidewalk shuffle."[17]

This problem arises even in the pure coordination game, as the sidewalk shuffle illustrates. But in a pure coordination game, the problem disappears if the players can talk long enough. When there is some conflict, negotiation about on what to coordinate can dissipate much of the potential gains from successful coordination as in the battle of the sexes, as discussed by Farrell (1987). Indeed, when it is unclear who controls the communication, a rigid "social convention" may do better for the players than does cheap talk, in contrast to our earlier lesson. However, the losses from this negotiation are bounded: Rabin (1994) shows that in any game, when there are enough rounds of talk, in every plausible equilibrium each player gets an expected payoff at least as great as he would get in the other player's favorite Nash equilibrium. Effectively, the worst a player can do is to give up and say to the other, "You choose."

Lessons From All This

We see no simple general lessons. Indeed, we reject several enticingly simple (and popular) ideas: that cheap talk ensures that players will play a Nash equilibrium;[18] that an efficient outcome will emerge (even if this merely means efficient among Nash equilibria); or that all information will be shared (or even as much information as is consistent with incentives in the sense of equilibrium).

We particularly stress that cheap talk does not in any sense guarantee efficiency in games. Because talk can help avoid misunderstandings and coordination failures, it often improves outcomes (for the players), but even unlimited cheap talk does not reliably lead to a Pareto-efficient outcome. For instance, when players have divergent preferences across equilibria, much of the gain from coordination may be lost to dispute and bargaining problems. Moreover, in cheap talk about private information, the logic of equilibrium selection described above can readily be shown to lead to inefficient equilibrium selection in reasonable examples.

We close, as we began, by relating cheap talk to the Hurwicz-style theory of *mechanism design.* Cheap talk consists of costless, nonbinding, nonverifiable messages that may affect the listener's beliefs. Such messages are exactly what players

[17] Cooper, DeJong, Forsythe and Ross (1989) report experiments with a somewhat similar game. Two-way communication seemed much less effective in coordinating players than one-way communication. In their game, however, many players make announcements that were neither self-signaling nor self-committing (apparently hoping for an attractive but nonequilibrium payoff); this problem would not arise in the game described here.

[18] For reasons of space we have not discussed this underresearched topic here: interested readers should consult Farrell (1988), Rabin (1990) and Aumann (1990).

convey in the theory of mechanism design (where the messages are often called "reports of players' types"). How is cheap talk different from that theory? Clearly, every cheap-talk equilibrium is an incentive-compatible mechanism (this just means that it's an equilibrium). We think there are two differences between the approaches.

First, there is typically no mediator in cheap talk. As a result, players may find it hard to release just the information that they would like to release. A mediator can ask each player for his type and promise to reveal only part of the information conditional on the other player's type; this may be impossible in informal cheap talk. Many failures of communication occur because each player thinks secrecy is desirable, but would not think that if the player knew the other's information—as in O. Henry's short story, "The Gift of the Magi."

More fundamentally, in mechanism design, people choose how to structure communication for a specific problem (within the constraints of incentives). In Example 4, for instance, there is an incentive-compatible mechanism in which Sally reveals her information, but we do not think it is a reasonable prediction under informal cheap talk. Similarly, in mechanism design, the battle of the sexes has an efficient symmetric solution: there are correlated equilibria that equitably achieve efficiency. But cheap talk typically does not work that way. Rather, the study of cheap talk is based on how people skeptically, but reasonably and mostly conventionally, interpret language. It is the study of rational people who know how to communicate in the ordinary way.

■ *Both authors thank, without blaming, Vincent Crawford, Miguel Costa Gomes, Steven Matthews and Joel Sobel for helpful comments, as well as the helpful and patient editorial team of the journal.*

References

Aumann, Robert, "Nash Equilibria are not Self-Enforcing." In Gabszewicz, J. J., J.-F. Richard, and L. A. Wolsey, eds., *Economic Decision-Making: Games, Econometrics and Optimisation.* Amsterdam: Elsevier, 1990, pp. 201–6.

Austen-Smith, D., "Information Transmission in Debate," *American Journal of Political Science,* February 1990, *34*:1, 124–52.

Austen-Smith, D., "Interested Experts and Policy Advice: Multiple Referrals under Open Rule," *Games and Economic Behavior,* January 1993, *5*:1, 3–43.

Carlton, D., and J. Perloff, *Modern Industrial Organization.* New York: HarperCollins College Publishers, 1994.

Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross, "Communication in the Battle of the Sexes Game: Some Experimental Results," *Rand Journal of Economics,* Winter 1989, *20*:4, 568–87.

Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross, "Alternative Institutions for Resolving Coordination Problems: Experimental Evidence on Forward Induction and Preplay Communication." In Friedman, James, ed., *Problems*

of Coordination in Economic Activity. Boston: Kluwer, 1994, pp. 29–146.

Crawford, V., "Theory and Experiment in the Analysis of Strategic Interaction," invited symposium on Experiment in the Analysis of Strategic Interaction, Econometric Society Seventh World Congress, Tokyo, August 1995.

Crawford, V., and J. Sobel, "Strategic Information Transmission," *Econometrica,* November 1982, *50,* 1431–51.

Farrell, Joseph, "Cheap Talk, Coordination, and Entry," *Rand Journal of Economics,* Spring 1987, *18*:1, 34–39.

Farrell, Joseph, "Communication, Coordination and Nash Equilibrium," *Economic Letters,* 1988, *27*:3, 209–14.

Farrell, Joseph, "Meaning and Credibility in Cheap Talk Games," *Games and Economics Behavior,* 1993, *5*:4, 514–31.

Farrell, Joseph, and Robert Gibbons, "Cheap Talk Can Matter in Bargaining," *Journal of Economic Theory,* June 1989, *48*:1, 221–37.

Gibbons, Robert, *Game Theory for Applied Economists.* Princeton: Princeton University Press, 1992.

Hayek, F., "The Use of Knowledge in Society," *American Economic Review,* September 1945, *35,* 519–30.

Holt, Charles, "Industrial Organization: A Survey of Laboratory Research." In Kagel, John H., and Alvin E. Roth, eds., *Handbook of Experimental Economics.* Princeton: Princeton University Press, 1995, pp. 349–444.

Hurwicz, L., "The Design of Mechanisms for Resource Allocation," *American Economic Review,* May 1973, *63,* 1–30.

Kreps, David, and Robert Wilson, "Sequential Equilibrium," *Econometrica,* July 1982, *50,* 863–94.

Matthews, Steven, "Veto Threats: Rhetoric in a Bargaining Game," *Quarterly Journal of Economics,* May 1989, *104*:2, 347–70.

Matthews, Steven, and Andrew Postlewaite, "Pre-Play Communication in Two-Person Sealed-Bid Auctions," *Journal of Economic Theory,* June 1989, *48,* 245–52.

Matthews, Steven, Masahiro Okunō-Fujiwara, and Andrew Postlewaite, "Refining Cheap-Talk Equilibria," *Journal of Economic Theory,* December 1991, *55,* 247–73.

Pearce, D., "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica,* July 1984, *52*:4, 1029–50.

Pinker, S., *The Language Instinct.* New York: W. Morrow and Co., 1994.

Rabin, Matthew, "Communication Between Rational Agents," *Journal of Economic Theory,* June 1990, *51,* 144–70.

Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics," *American Economic Review,* December 1993, *83*:5, 1281–302.

Rabin, Matthew, "A Model of Pre-Game Communication," *Journal of Economic Theory,* August 1994, *63*:2, 370–91.

Sally, D., "Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1992," *Rationality and Society,* January 1994, *7*:1, 58–92.

Schelling, T. C., *The Strategy of Conflict.* Cambridge: Harvard University Press, 1960.

Seidmann, D. J., "Cheap Talk Games May Have Unique, Informative Equilibrium Outcomes," *Games and Economic Behavior,* 1992, *4*:3, 422–25.

Sopher, Barry, and Iñigo Zapater, "Communication and Coordination in Signalling Games: An Experimental Study," mimeo, June 1993.

Spence, A. M., *Market Signaling: Informational Transfer in Hiring and Related Screening Processes.* Cambridge: Harvard University Press, 1974.

Valley, K., L. Thompson, Robert Gibbons, and M. Bazerman, "Is Talk Really Cheap? Outperforming Equilibrium Models of Communication in Bargaining Games," mimeo, August 1995.