

Different faces of Risky Speech

Robert van Rooij and Merlijn Sevenster*

1 Communication as coordination problem

Suppose two individuals agreed to meet each other tonight at 10.00 o'clock in Amsterdam, but forgot to agree on a place (and don't have the chance anymore to make an agreement). The two are now facing a coordination problem: only if they make a 'correlated' decision, they will both end up at the same place and meet each other as desired.

Schelling (1960) distinguishes two ways to solve such coordination problems: convention and salience. A coordination problem is solved by *convention* if the participants were engaged in similar coordination problems before, and have formed the habit, or convention, to solve these problems in a particular way. Both participants see the overwhelming similarity between the previous coordination problems and the current one, and, either out of habit, or because they expect that the other participant will behave similarly as before, they behave similarly as they did in these previous encounters. A coordination problem is solved by *salience* if the participants do not expect that the problem can, or will, be solved by habit or convention, but have reason to assume that the other participant will behave in a certain way, because one kind of behavior is most 'obvious'.¹

David Lewis (1969) had the insight that we can think of successful communication as a way to successfully solve the coordination problem of how to transfer information. The problem involves both the speaker and the hearer: the speaker *S* has to decide which signal to send to transfer the intended information, and the hearer *H* has to interpret the signal in the way as intended by the speaker in order for the communicative act to be successful. As is the case of all coordination problems, *expectations* are crucial: the speaker's decision which signal to send will be based on how she expects the hearer *H* will interpret the signals she considers, and the hearer's decision will be based on what he expects the speaker could have meant.² So how can the participants in a conversation

*We would like to thank the reviewers for their critical comments and useful suggestions.

¹Of course, convention or precedence, gives rise to salience, and can be thought of as a special case. With Clark (1996) and others, however, we will here assume the intuitive distinction between solving a game by convention and by salience.

²Of course, these (first-order) expectations about what the other will do are based on the (higher-order) expectations of both participants of the conversation of the other's expectation about one's own behavior, and so on. In this paper we won't go through the way the expect-

have correct expectations about the communicative behavior of their partners?

Just as in the coordination games studied by Schelling, also here the two ways of solving the problem are either by convention, or by salience. For both ways, expectations are crucial. The communication problem is solved by *convention* if the speaker ‘encodes’ her communicative intention by using a signal which has been used many times before (or at least is composed out of signals used many times before) and which has received the interpretation the speaker now wants to communicate. The speaker uses the symbol on the expectation that the hearer will interpret it in the same way as before, while the hearer interprets it on the expectation that the speaker intended to communicate the same as on previous occasions when she used the signal. Of course, linguistic conventions are much more complicated than this picture suggests, but, essentially, this is the idea.

The communication problem is solved by *salience* if the conventional meaning (if any) of the signal used by the speaker *underspecifies* its actual intended interpretation, or in case the speaker wants to implicitly convey (by *conversational implicature*) something on top, or instead, of what is conventionally communicated by the use of the sentence. For such cases, expectations are even more important: speaker and hearer have to agree on what would be the most obvious interpretation of the signal in this context.

The traditional emphasis of linguistics has been on conventional or rule-governed communicative behavior: syntax and (lexical and compositional) semantics. However, for pragmatics, the theory of language use, it is the concept of salience that is of crucial importance.

To a large extent, the notion of salience is a psychological notion that largely has ‘escaped’ game theoretical analysis.³ It crucially involves expectations, and (at least traditional) game theory has nearly nothing to say about how these expectations are formed. However, we can abstract away from the particular expectations that participants of a conversation have, and use game theoretical reasoning to make predictions concerning their expected behavior in certain kinds of situations. That is what we will do in this paper.

2 Games, expectations, and communication

2.1 Expectations and equilibrium selection

Coordination problems can obviously be thought of in a game theoretical way.⁴ Suppose Row and Column have to make their respective decisions independently of one another. Row has to decide between performing R_1 or R_2 and Column has C_1 and C_2 as his alternative actions. In the simplest coordination games, both Row and Column are equally happy when they coordinate on either $\langle R_1, C_1 \rangle$ or on $\langle R_2, C_2 \rangle$. Such a game can be described in terms of the following payoff-table:

tations about the other’s behavior are formed, but just stick to the first-order expectations.

³But see Asher and William (this volume) for an interesting exception.

⁴Though it was the insight of Schelling (1960) that the analysis of how to solve such problems is more complicated, and thus more interesting, than previously assumed.

Game 1 :

	C_1	C_2
R_1	1, 1	0, 0
R_2	0, 0	1, 1

The action pairs on which they want to coordinate are both Nash-equilibria of the game, but their problem is on which one they should coordinate. Given that they have to decide independently of one another, their chosen action will depend on their expectations about what the other will do. In case the payoffs are equal, as in game 1, Row, for instance, will choose R_1 just in case she expects, or takes it to be more likely, that Column will play C_1 . For game 1, the choice of how to perform depends only on the players' expectations on what the other will do. But this is just because here both equilibria have the same payoff (for both players). In general, different equilibria can give rise to different payoffs, and both players will choose by maximizing their expected utilities. These expected utilities involve both payoffs and the probabilities that a player assigns to the different actions that the other player will perform. Suppose that the probability function P_R represents Row's expectations about what Column will perform, i.e., $P_R(C_i)$ will represent the probability with which Row thinks Column will perform action C_i . The expected utility held by Row to play R_1 , $EU_R(R_1)$, will then be $P_R(C_1) \times U_R(R_1, C_1) + P_R(C_2) \times U_R(R_1, C_2)$. It is easy to see that the expected utility of R_1 is higher than the expected utility of R_2 , $EU_R(R_1) > EU_R(R_2)$, just in case Row thinks it is more likely that Column will play C_1 than C_2 , i.e. when $P_R(C_1) > P_R(C_2)$. Obviously, something similar holds for Column.

Things are a little bit more complicated when the payoffs of the different equilibria are not the same. Consider, for instance, the following coordination problems:

Game 2 :	<table border="1"> <tr> <td></td> <td>C_1</td> <td>C_2</td> </tr> <tr> <td>R_1</td> <td>2, 2</td> <td>0, 0</td> </tr> <tr> <td>R_2</td> <td>0, 0</td> <td>1, 1</td> </tr> </table>		C_1	C_2	R_1	2, 2	0, 0	R_2	0, 0	1, 1	Game 3 :	<table border="1"> <tr> <td></td> <td>C_1</td> <td>C_2</td> </tr> <tr> <td>R_1</td> <td>8, 8</td> <td>0, 0</td> </tr> <tr> <td>R_2</td> <td>0, 0</td> <td>1, 1</td> </tr> </table>		C_1	C_2	R_1	8, 8	0, 0	R_2	0, 0	1, 1
	C_1	C_2																			
R_1	2, 2	0, 0																			
R_2	0, 0	1, 1																			
	C_1	C_2																			
R_1	8, 8	0, 0																			
R_2	0, 0	1, 1																			

Also in these games, both $\langle R_1, C_1 \rangle$ and $\langle R_2, C_2 \rangle$ are equilibria. However, now both would in principle prefer the former equilibrium to the latter. But this doesn't give them an automatic incentive to perform their part of equilibrium $\langle R_1, C_1 \rangle$: what one should do in order to maximize payoff depends also on one's expectations what the other will do. On the coordination problem of Game 2, Row should do her part of the coordination equilibrium $\langle R_1, C_1 \rangle$ only if she thinks (for whatever reason) that the probability that Column will choose C_1 is at least $\frac{1}{3}$, because only in that case Row's expected utility of playing R_1 , $EU_R(R_1)$ tops her expected utility of playing R_2 . Similarly for the coordination problem of Game 3. Now Row should choose R_2 instead of R_1 if she takes it to be at least eight times as probable that Column-chooser will choose C_2 than that he will choose C_1 . Although expectations always play a role when several equilibria are possible, the contrast between games 1, 2, and 3 shows that this

role increases if the expected utilities of the different equilibria become more alike.

Games 1, 2, and 3 each have two equilibria (in pure strategies), because the expectations that players have of other players' behavior were not supposed to play a role. We saw that these expectations are in fact crucial to predict what will be played, and we will show now that they even influence the equilibria of the game. Consider Game 3 again, and assume that before deliberation Row expects, for some reason, with a probability of 0.7 that Column plays C_2 and Column expects with the same probability that Row plays R_2 . Suppose, moreover, that these probabilities are common knowledge. Then, by taking these expectations into account as well, this gives rise to a new situation, described in Table 4, where the payoffs are now the expected utilities:

Table 4 :

	C_1	C_2
R_1	2.4, 2.4	2.4, 0.7
R_2	0.7, 2.4	0.7, 0.7

It is clear that in this situation we end up with play $\langle R_1, C_1 \rangle$, which is intuitively the correct equilibrium of Game 3. This discussion indicates how important expected utility theory is for game theory. Even if we start out with the strong assumption that the players have common knowledge of each others' initial expectations about one another's strategy choices,⁵ these expectations need not be their final expectations about these strategy choices. The reason is that these prior expectations don't take into account the reasoning of Column (Row) given that he (she) knows Row's (Column's) prior expectations, i.e., the process of deliberation.⁶ It is the final expectations (with probability 1, if we disregard mixed strategies) that count to determine what is the equilibrium that is being played. And this is what we saw in this situation: the initial expectations were overruled in the deliberation process that takes expected utility, and not just prior expectations into account. So, our discussion points out the relevance of individualistic Bayesian rationality.

The games we looked at so far had two features in common that singled them out as pure coordination games. First, the preference relations between action pairs were the same for both participants. Second, it was important for both Row and Column to coordinate: the payoff of the coordinating action pair was higher *for both agents* than the payoff of a non-coordinating action pair. In this paper we are interested in games where we give up one or both of these assumptions. First we will discuss (communication) games with two equilibria in which only one of these equilibria is a *strict* one: to receive the (lower) payoff of the other equilibrium, only one of the players (the speaker) has to perform his part of the equilibrium play. After that, we will discuss (communication) games

⁵Harsanyi & Selten (1988) discuss a technique in which, what they call, an 'objective prior' can be defined solely based on the structure of the game. But it is disputable whether players really use this technique to determine these prior expectations about what the other(s) will do. For instance, it is unclear why precedence and other notions are taken to be irrelevant.

⁶For some analyses of deliberation in games, see Harsanyi & Selten's (1988) *tracing procedure*, and Skyrms' (1990) various methods of rational deliberation.

where both participants of the conversation have an incentive to coordinate on an equilibrium (that is, the game has two strict equilibria), but where the payoffs vary on out of equilibria actions-pairs. As we will see, giving up looking only at pure coordination games introduces extra considerations concerning *risk*.

2.2 Risky versus safe play

In pure coordination games as discussed in the previous section, risk already plays an important role. If one does not know for sure which equilibrium strategy the other participant will play, it is possible that by maximizing one's expected utility one actually ends up empty handed in a non-coordinating play of the game. Because all non-equilibrium outcomes have the same zero-payoff for both participants, one strategy might be called more risky than another just in case its expected utility is lower. In the games we are going to discuss in this section, however, some strategies can be called more risky than others because of differences in the payoffs of non-equilibrium-plays of the game.

In the previous section we assumed that any non-equilibrium play of the game gives both participants a payoff of 0. In an appealing article, Sally (2003) observes, however, that in games that model communication (or many other types of situations), this assumption is wrong: some non-equilibria can be worse than others for one or both participants. If a speaker deliberates whether she should encode the information she wants to communicate in a funny, indirect way or not, for instance, Sally notes that she has to take into account that unsuccessful communication resulting from her (trying to) be(ing) funny is probably worse than unsuccessful communication without her being indirect. Consequently, Sally (2003) calls, for instance, ironical indirect speech *risky*.⁷ We think this is a very useful way of looking at communicative behavior. In contrast to Sally (2003), however, we will distinguish different types of games where the notion of 'risk' is involved. In doing so, we claim that Sally's notion of 'risky speech' is perhaps more widely applicable than suggested by Sally's own discussion.

Let us first discuss a game like the following:

Game 5 :	<table style="border-collapse: collapse; text-align: center;"> <tr> <td style="border: none;"></td> <td style="border: 1px solid black; padding: 2px 10px;">C_1</td> <td style="border: 1px solid black; padding: 2px 10px;">C_2</td> </tr> <tr> <td style="border: none; padding-right: 5px;">R_1</td> <td style="border: 1px solid black; padding: 2px 10px;">$1 + \epsilon, 1$</td> <td style="border: 1px solid black; padding: 2px 10px;">$1 - \epsilon', 0$</td> </tr> <tr> <td style="border: none; padding-right: 5px;">R_2</td> <td style="border: 1px solid black; padding: 2px 10px;">$1, 1$</td> <td style="border: 1px solid black; padding: 2px 10px;">$1, 1$</td> </tr> </table>		C_1	C_2	R_1	$1 + \epsilon, 1$	$1 - \epsilon', 0$	R_2	$1, 1$	$1, 1$
	C_1	C_2								
R_1	$1 + \epsilon, 1$	$1 - \epsilon', 0$								
R_2	$1, 1$	$1, 1$								

Although this type of game has two equilibria, $\langle R_1, C_1 \rangle$ and $\langle R_2, C_2 \rangle$, it is

⁷Parikh's (2001) compares direct and indirect speech as well in his analysis of miscommunication. If speaker and hearer make contrasting assumptions about the style used by the other conversational participant, miscommunication follows, because they modeled the conversational situation as different games which have different outcomes. Parikh proposes that in case style is involved, interlocutors first (should) play a metagame concerning the style (or the use of language), and only then one of interpretation. One might think of Sally (2003) and section 5 of this paper as an analysis of such a metagame.

not really one of coordination. The reason is that now the equilibrium $\langle R_2, C_2 \rangle$ is not a strict one: it doesn't matter what Column plays if Row plays R_2 . On the other hand, Row would benefit from the combination $\langle R_1, C_1 \rangle$ (if $\epsilon, \epsilon' > 0$). This is not only a *strict* equilibrium, but it is *payoff-dominant* as well. We call an equilibrium payoff-dominant if and only if there is no other equilibrium in the game that yields a strictly higher payoff for at least one player.

In case Row has no idea whether Column will play strategy C_1 or C_2 , we assume that Row takes both strategies to be equally likely.⁸ In that case, the expected utility of playing R_1 is higher/equal/lower than the expected utility of playing R_2 if and only if $\epsilon > / = / < \epsilon'$. For this reason, we will say that Row is *risk-loving* iff $\epsilon > \epsilon'$, he is *risk-neutral* iff $\epsilon = \epsilon'$, and he is *risk-averse* iff $\epsilon < \epsilon'$. Assuming by default that $\epsilon > \epsilon'$, we will denote strategy R_1 by Risky and strategy R_2 by Safe.

The other kind of games we are interested in is one in which both players can choose between risky and safe strategies. Also this type of game has two equilibria, but now it is important for both players to coordinate. This game differs from games 2 and 3 discussed in section 2.1 in that both equilibria have something distinctive to speak in their favor.

Consider Rousseau's (1755) famous Reindeer hunt game as described by Lewis (1969) and extensively discussed by Skyrms (2004); a simple two-player symmetric game with two strict equilibria: both hunting Reindeer, $\langle R, R \rangle$, or both hunting Squirrel, $\langle S, S \rangle$. Note that we slightly changed the story of Rousseau's game, but the contention of the game is intact.⁹ The first equilibrium gives the highest payoff to both, i.e., is payoff-dominant (or Pareto optimal), because it gives to both a utility of, let us say, 6, while the second equilibrium yields only one of 4. However, assume that if one hunts Reindeer but the other Squirrel, the payoff is (4,0) in 'favor' of the Squirrel-hunter. In that case, the payoff-dominated equilibrium where both are hunting Squirrel still has something to speak in its favor: if one player is equally likely to play either strategy, the expected utility of hunting Squirrel for the other is optimal.

Game 6: Reindeer hunt:	<table style="border-collapse: collapse; display: inline-table;"> <tr><td style="border: none;"></td><td style="border: none; padding: 0 5px;">R</td><td style="border: none; padding: 0 5px;">S</td></tr> <tr><td style="border: none; padding: 0 5px;">R</td><td style="border: 1px solid black; padding: 2px 5px; text-align: center;">6, 6</td><td style="border: 1px solid black; padding: 2px 5px; text-align: center;">0, 4</td></tr> <tr><td style="border: none; padding: 0 5px;">S</td><td style="border: 1px solid black; padding: 2px 5px; text-align: center;">4, 0</td><td style="border: 1px solid black; padding: 2px 5px; text-align: center;">4, 4</td></tr> </table>		R	S	R	6, 6	0, 4	S	4, 0	4, 4	<table style="border-collapse: collapse; display: inline-table;"> <tr><td style="border: none;"></td><td style="border: none; padding: 0 5px;">Risky</td><td style="border: none; padding: 0 5px;">Safe</td></tr> <tr><td style="border: none; padding: 0 5px;">Risky</td><td style="border: 1px solid black; padding: 2px 5px; text-align: center;">1 + ϵ, 1 + ϵ</td><td style="border: 1px solid black; padding: 2px 5px; text-align: center;">-ϵ', 0</td></tr> <tr><td style="border: none; padding: 0 5px;">Safe</td><td style="border: 1px solid black; padding: 2px 5px; text-align: center;">0, -ϵ'</td><td style="border: 1px solid black; padding: 2px 5px; text-align: center;">1, 1</td></tr> </table>		Risky	Safe	Risky	1 + ϵ , 1 + ϵ	- ϵ' , 0	Safe	0, - ϵ'	1, 1
	R	S																		
R	6, 6	0, 4																		
S	4, 0	4, 4																		
	Risky	Safe																		
Risky	1 + ϵ , 1 + ϵ	- ϵ' , 0																		
Safe	0, - ϵ'	1, 1																		

The more abstract right-hand example also has two (strict) Nash Equilibria (if both ϵ and ϵ' are higher than 0): both playing Risky, or both playing it Safe. It is obvious that equilibrium $\langle \text{Risky}, \text{Risky} \rangle$ is payoff-dominant. Following Harsanyi and Selten (1988), we will say that Nash equilibrium $\langle a^*, b^* \rangle$ is risk-dominant iff for all Nash equilibria $\langle a, b \rangle$ of the game,

⁸We realize that this is an unnatural assumption for Game 5, given that C_1 weakly dominates C_2 . It is discussed here only for illustrative purposes.

⁹The game is normally called a 'Stag hunt', and hunting Stag is normally contrasted with hunting Rabbit.

$$\frac{(U_{Row}(a^*, b^*) - U_{Row}(a, b^*)) \times (U_{Col}(a^*, b^*) - U_{Col}(a^*, b))}{(U_{Row}(a, b) - U_{Row}(a^*, b)) \times (U_{Col}(a, b) - U_{Col}(a, b^*))} \geq$$

In the above example $\langle \text{Safe}, \text{Safe} \rangle$ is risk-dominant exactly if $\epsilon' \geq \epsilon$. For this reason, we will call a player *risk-loving* iff $\epsilon\epsilon' > \epsilon^2$, she is *risk-neutral* iff $\epsilon = \epsilon'$, and she is *risk-averse* iff $\epsilon < \epsilon'$.

In contrast to the concept of payoff-dominance, the concept of risk-dominance is based on *individual* rationality. Think of the numerical version of the Reindeer hunt game, where it is common knowledge that the prior expectations before deliberation that the other will play R is 0.5. In that case the game gives rise to the following ‘expected utility’-table:

Expected utilities: Reindeer hunt:

	R	S
R	3, 3	3, 4
S	4, 3	4, 4

Obviously, in this situation the Nash equilibrium $\langle S, S \rangle$ will be played in which both are hunting Squirrel.

The preference for playing her part of the *risk-dominant* equilibrium in the Reindeer hunt game is closely related with the preference for playing Safe (or R_2) in Game 5 (if $\epsilon > \epsilon'$). Suppose that a player doesn’t know what the other player will do. In that case the speaker should choose the strategy that has the highest expected utility. Suppose that, for lack of reasons otherwise, a player takes both actions of the other player to be equally likely. One can show that in that case the action which has the highest expected utility in the Reindeer hunt game (or any other symmetric 2×2 game) is the strategy which is risk-dominant. And this will be the case for the Safe/Risky-strategy if and only if the player is risk-averse/risk-loving.

In the rest of this paper we will suggest that some decisions speakers and hearers have to make when they have to coordinate their *communicative* behavior by salience can be modeled by the decisions that have to be made by the players in the games discussed in this subsection. First, we will discuss an example that we suggest can be modeled analogously to (an incomplete information variant of) Game 5: implicit communication where the conventional meaning of the expression *underspecifies* what the speaker actually wants to communicate. Then we will discuss some examples that can best be modeled as ‘impure’ coordination games with rankable equilibria and varying off-diagonal payoffs as the Reindeer hunt. Following Sally (2003), we will suggest that this game models the communicative decisions involved in cases that the meaning intended to be communicated can ‘overturn’ the conventional meaning. Parikh’s (2001) game-theoretical analysis of miscommunication in terms of ‘metagames’ is closely related. Finally we will discuss an example that is somewhere in between games 5 and 6. Before we will come to these modelings, however, we first will shortly introduce ‘signaling games’. These signaling games will be extended in the following sections.

3 Games of communication

3.1 Standard signaling games

Lewis (1969) defined the notion of a signaling game in order to explain the conventionalization of meaning of language without assuming any pre-existing relation between messages and meanings.

A signaling game is a cooperative game amongst two players: a sender S and a receiver R , whose shared goal it is to let R perform an action that is appropriate with respect to the state S and R are in. This state is only observed by S though, and S communicates it by means of a meaningless message. Think of a couple of agents, the one looking out for hungry predators and the other searching for food on the ground. Both players are in the same state — there is a predator approaching or there is not —, but only S knows which state they are in. S signals the state by means of a message that has no pre-defined meaning — say, ‘buh’ or ‘bah’. In turn, R hears the message and is free to perform any action of her liking, but every state has a most appropriate action. E.g. if there is a hungry predator approaching, R should flee, otherwise R should keep on searching. The best thing for S to do is to say ‘buh’ if there is a hunting predator and ‘bah’ otherwise; and for R to flee when he hears ‘buh’ and not to in case he hears ‘bah’. (It is equally good of course to do the same, but with ‘buh’ and ‘bah’ interchanged.) If S and R play the game in such an optimal way, the meanings of the messages ‘buh’ and ‘bah’ are created in the play of the game. As we will see below, playing in a way that makes the messages meaningful amounts to coordinating on a Pareto-optimal Nash equilibrium.

For future reference, we formally define signaling games. Let T be the set of states, M be the set of messages and A be the set of actions such that $|T| = |A| \leq |M|$. Let $f : T \rightarrow A$ be the bijective function, that adds the appropriate action $f(t) \in A$ to every type $t \in T$. Then S (R) plays the signaling game following strategy s (r), that is a function from $T \rightarrow M$ ($M \rightarrow A$). In *cheap talk signaling games*, successful communication of state t (thus in case $f(t) = r(s(t))$) is rewarded with 1, whereas unsuccessful communication (thus in case $f(t) \neq r(s(t))$) is rewarded with 0, independent of the state t and the message $s(t)$:

$$u_S(f, t, s, r) = u_R(f, t, s, r) = \begin{cases} 1, & \text{if } f(t) = r(s(t)); \\ 0, & \text{if } f(t) \neq r(s(t)). \end{cases} \quad (1)$$

We assume that Nature picks the state according to some probability distribution P over T .¹⁰ The utility function for S and R is the expected utility relative to the probability distribution P over T :

$$U_S(s, r) = U_R(s, r) = \sum_{t \in T} P(t) \times u_S(f, t, s, r). \quad (2)$$

Finally, we define a *cheap talk signaling game* G as a tuple $\langle \{S, R\}, \mathbf{P}, \{\mathbf{S}, \mathbf{R}\}, \{u_S, u_R\} \rangle$, where \mathbf{P} is a probability distribution over T , \mathbf{S} is the set

¹⁰We assume that $P(t) > 0$, for every $t \in T$.

of strategies $s : T \rightarrow M$ for player S ; \mathbf{R} is the set of strategies $r : M \rightarrow A$ for player R ; and $\{u_S, u_R\}$ contains both players' utility functions. G is called 'cheap talk' because u_S and u_R , simultaneously defined in (1) are called this way.

As a small example, consider the signaling game with only two states t_1, t_2 , two messages m_1, m_2 and two actions a_1, a_2 , where $f(t_i) = a_i$ for $i \in \{1, 2\}$. Obviously, both players have four (pure) strategies each. Furthermore, let $x = P(t_1) > P(t_2) = y$. Then, we have the payoff matrix below.

	t_1	t_2		m_1	m_2				
s_1	m_1	m_1	r_1	a_1	a_1				
s_2	m_1	m_2	r_2	a_1	a_2				
s_3	m_2	m_1	r_3	a_2	a_1				
s_4	m_2	m_2	r_4	a_2	a_2				

	r_1	r_2	r_3	r_4
s_1	x, x	x, x	y, y	y, y
s_2	x, x	$1, 1$	$0, 0$	y, y
s_3	x, x	$0, 0$	$1, 1$	y, y
s_4	x, x	x, x	y, y	y, y

The resulting signaling game has four Nash equilibria: $\langle s_1, r_1 \rangle, \langle s_2, r_2 \rangle, \langle s_3, r_3 \rangle$ and $\langle s_4, r_1 \rangle$. As the reader can check, only in $\langle s_2, r_2 \rangle$ and $\langle s_3, r_3 \rangle$ communication takes place: these are precisely the payoff-dominant equilibria. Lewis calls such equilibria 'signaling systems'. Technically, $\langle s, r \rangle$ is a signaling system iff $f(t) = r(s(t))$, for every $t \in T$. Necessary condition for $\langle s, r \rangle$ to be a signaling system is that both s and r are injective or one-to-one functions.

3.2 Super conventional signaling games

Signaling games are used by Lewis (1969) to explain literal, or conventional, meaning in terms of the game-theoretic notion of 'stability'. This doesn't mean, however, that we cannot motivate the use of unconventional message-meaning combinations by making use of game-theoretical equilibrium as well. In this paper we study the risk of using non-conventional, non-explicit, or non-literal speech as a means of communication. In order to do so we introduce in this section signaling games where there already exists a convention of explicit literal meaning. We will call the resulting signaling games *super conventional*, and sometimes write SC signaling games.

The intuition underlying SC signaling games has it that S and R play a signaling game enjoying common knowledge of the fact that some strategy-pair $\langle s, r \rangle$ is the conventional signaling system. We denote the conventional sender and receiver strategy by means of cs and cr , respectively. It is this pair of strategies that model the literal or explicit meanings. While playing an SC signaling game, S and R have agreed on the conventional meaning of messages in

$$M' = \{m \in M \mid \text{there exists a } t \in T \text{ such that } cs(t) = m\},$$

the set that contains the messages that convey the to-be communicated types. Since s is an injective function, $|M'| = |T|$. We assume that only the messages in M' can have a non-literal meaning.

Typically, non-explicit or non-literal utterances have it that the sentence leave the actual interpretation underspecified, or, if taken literally, means something different than was intended by the speaker. In formal terms, although S is of type t she uses a message $m \neq cs(t)$, and S wants and expects R not to perform $cr(m)$, if that exists at all, but $f(t)$.

We will model the extra gain, in case of successful non-conventional communication by a parameter $\epsilon \geq 0$, whereas we punish the player (possibly both) who deviated from its conventional strategy with the parameter-value $\epsilon' \geq 0$, in case of unsuccessful communication.

This brings us to the main definition of this section. A *Super Conventional signaling game* $G_{\langle cs, cr \rangle}$ is a standard signaling game G equipped with a convention:

$$\langle \{S, R\}, \mathbf{P}, \{\mathbf{S}, \mathbf{R}\}, \{u_S, u_R\}, \langle cs, cr \rangle \rangle,$$

where u_S and u_R are the players' utility function, defined below. \mathbf{P} is the probability distribution over the set of types. For the utility functions of speaker and hearer in the signaling games discussed in the previous subsection it was taken to be irrelevant which message was being used; it only mattered whether communication was successful or not. In our superconventional signaling games we will assume that at least for the speaker, but perhaps also for the hearer, it is important which message is used for communication. In particular, it is taken to be advantageous for the speaker (and perhaps for the hearer) to communicate successfully with a non-explicit message, or with a message that should receive a non-literal interpretation. In the following two sections we are going to discuss examples where the speaker's utility function should not be defined as in (1) but rather as follows

$$u_S^{cs}(f, t, s, r) = \begin{cases} 1 + \epsilon, & \text{if } f(t) = r(s(t)) \text{ and } s(t) \neq cs(t); \\ 1, & \text{if } f(t) = r(s(t)) \text{ and } s(t) = cs(t); \\ 0, & \text{if } f(t) \neq r(s(t)) \text{ and } s(t) = cs(t); \\ -\epsilon', & \text{if } f(t) \neq r(s(t)) \text{ and } s(t) \neq cs(t). \end{cases} \quad (3)$$

Intuitively, u_S^{cs} hard-wires that the speaker is moderately rewarded or punished if she sticks to the conventional sender strategy (i.e. $s(t) = cs(t)$). Of course, still the utility-function tells that successful communication (i.e. $f(t) = r(s(t))$) is better rewarded than unsuccessful communication (i.e. $f(t) \neq r(s(t))$).

As for the hearer's utility function, we will discuss two special cases: In section 4, where we discuss the risk of non-explicit communication, we will take the hearer's utility function to be the same as in (1):

$$u_R(f, t, s, r) = \begin{cases} 1, & \text{if } f(t) = r(s(t)); \\ 0, & \text{if } f(t) \neq r(s(t)). \end{cases} \quad (4)$$

This means that only the speaker has to decide whether to play risky or not. The hearer just has to assign the correct meaning to the given message. In section 5, however, we will discuss linguistic phenomena where also the hearer

can play either risky or safe. In that section we will assume that the hearer’s utility function is given by the following function:

$$u_R^{cr}(f, t, s, r) = \begin{cases} 1 + \epsilon, & \text{if } f(t) = r(s(t)) \text{ and } r(s(t)) \neq cr(s(t)); \\ 1, & \text{if } f(t) = r(s(t)) \text{ and } r(s(t)) = cr(s(t)); \\ 0, & \text{if } f(t) \neq r(s(t)) \text{ and } r(s(t)) = cr(s(t)); \\ -\epsilon', & \text{if } f(t) \neq r(s(t)) \text{ and } r(s(t)) \neq cr(s(t)). \end{cases} \quad (5)$$

In section 6 we will discuss an example where the players’ utility functions are even more involved than in (4) and (5).

4 Risk of implicit communication

Let us assume a very simple signaling game, where we have two kinds of meanings, t_1 and t_2 , and expressions, m_1 and m_2 that conventionally denote t_1 and t_2 , respectively, in a context-independent way. Let us now assume that, in addition, we have an expression m_u that is lighter than either of m_1 and m_2 and that has an underspecified meaning: it can mean both t_1 or t_2 . Formally, let us fix that using m_u instead of m_1 or m_2 yields a bonus of $\epsilon > 0$. It is easy to see (e.g. Parikh, 2001, van Rooij, 2004) that if it is common knowledge that the hearer, R , takes t_1 to be more probable (salient) than t_2 , $\mathbf{P}(t_1) > \mathbf{P}(t_2)$, and m_1 is more costly than m_u , $C(m_1) > C(m_u)$, the ‘coding’ strategy that uses m_u to denote t_1 (and m_2 to denote t_2) is the most efficient, i.e., payoff-dominant, ‘coding’ strategy to denote t_1 and t_2 . In particular, it is more efficient than the coding strategy that uses m_1 to denote t_1 . However, when the relative probabilities of t_1 and t_2 are not shared between speaker S and hearer R and the latter has to guess (by tossing a coin) which meaning the speaker takes to be more salient, or probable, using a light message with an underspecified meaning is not going to have a positive payoff.¹¹ In the simple case above where the message with the underspecified meaning can have only two specific denotations, the benefit of communicating with a light expression must be very high in order to overcome the *risk* of miscommunication. We are going to discuss a case like that of Game 5 repeated below.

Game 5 :	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border: none;"></td> <td style="border: none; padding: 0 10px;">C_1</td> <td style="border: none; padding: 0 10px;">C_2</td> </tr> <tr> <td style="border: none; padding: 5px;">Risky</td> <td style="border: 1px solid black; padding: 5px; text-align: center;">$1 + \epsilon, 1$</td> <td style="border: 1px solid black; padding: 5px; text-align: center;">$1 - \epsilon', 0$</td> </tr> <tr> <td style="border: none; padding: 5px;">Safe</td> <td style="border: 1px solid black; padding: 5px; text-align: center;">$1, 1$</td> <td style="border: 1px solid black; padding: 5px; text-align: center;">$1, 1$</td> </tr> </table>		C_1	C_2	Risky	$1 + \epsilon, 1$	$1 - \epsilon', 0$	Safe	$1, 1$	$1, 1$
	C_1	C_2								
Risky	$1 + \epsilon, 1$	$1 - \epsilon', 0$								
Safe	$1, 1$	$1, 1$								

For the case at hand, thus where t_1 is the case, we assume that the Safe strategy is to send the correct explicit message in the relevant state, while the Risky strategy is to use the light message with the underspecified meaning. C_1 and C_2 are the strategies that always interpret the explicit messages in the expected way, and to interpret m_u as t_1 and as t_2 , respectively. Thus, successful

¹¹In case P_R is not commonly known, the situation cannot really be described as a (signaling) game. Indeed, the standard equilibrium reasoning is not appropriate anymore.

communication by context-independent expressions m_1 and m_2 — i.e., playing the Safe strategy — is 1 for both agents; (ii) unsuccessful communication has a payoff of 0, i.e., we assume that $\epsilon' = 1$; and (iii) the benefit of successful communication with the light underspecified expression m_u instead of the conventional explicit expression m_1 is ϵ , which is higher than 0.¹²

The hearer interprets the speaker’s message in the only appropriate way in case the message has a context-independent completely specified meaning. What the hearer does if he receives the underspecified message m_u depends on his beliefs: he interprets m_u as t_1 if he takes t_1 to be more likely, $P_R(t_1) > P_R(t_2)$, and he interprets m_u as t_2 if he takes t_2 to be more likely, $P_R(t_1) < P_R(t_2)$. Thus, the hearer has a choice between two strategies: C_1 and C_2 that reflect that any conventional message m_c is attached to its conventional meaning t_c and that C_i attaches t_i to m_u , for $i \in \{1, 2\}$. The speaker’s payoffs of these two strategies in the different situations are given by following tables:

t_1	C_1	C_2
Implicit	$1 + \epsilon$	0
Explicit	1	1

t_2	C_1	C_2
Implicit	0	$1 + \epsilon$
Explicit	1	1

The speaker doesn’t know how the hearer will interpret the underspecified message m_u because she does not know whether the hearer will take t_1 or t_2 to be more likely. We have seen above already that if the speaker takes C_1 and C_2 to be equally likely, i.e., if $P_S(C_1) = P_S(C_2)$, the benefit of using the underspecified message has to be at least 1, $\epsilon \geq 1$. But what if $P_S(C_1) \neq P_S(C_2)$? Let us assume that the speaker believes with probability n that $P_R(t_1) > P_R(t_2)$ (and thus with probability $1 - n$ that $P_R(t_1) \leq P_R(t_2)$). It is easy to see that the speaker takes implicit communication to be worthwhile in situation t_1 if and only if $n \times (1 + \epsilon) > 1$. That is, for the expected utility of being implicit to be higher than the expected utility of being explicit it has to be the case that $\epsilon > \frac{1-n}{n}$. The equality $\epsilon = \frac{1-n}{n}$ can be plotted as in figure 1.

Obviously, if n is very close to 0 the use of m_u will be a bad choice, but also for other choices of n , it probably won’t pay to be implicit: if n is $\frac{1}{3}$ or $\frac{1}{4}$, for instance, the value of ϵ has to be 2, or 3, respectively, which seems to be much too high.

Being explicit is a *safe* strategy. It is optimal under the maximin strategy and the minimax strategy. Things are more complicated when expected utility is at issue, for now it also depends on the relative weight of n and ϵ . But the main, and perhaps obvious, conclusion of these considerations, whether expected utility plays a role or not, is always that *it is safer to be explicit if you don’t know (for sure) what your conversational participants takes to be the most salient situation of T , and that it is risky to be implicit.*¹³

¹²Notice that S prefers to play Risky if she takes strategies C_1 and C_2 taken by R to be equally likely iff $\epsilon > 1$, (given that $\epsilon' = 1$).

¹³We have analyzed this situation with respect to a particular situation. We could have analyzed also the more general situation, where the strategies *implicit* and *explicit* stand for the strategies to use for all situations t_i messages m_u and m_i respectively, and where the actions of the hearer still depend crucially on the expectations. It is easy to see, however,

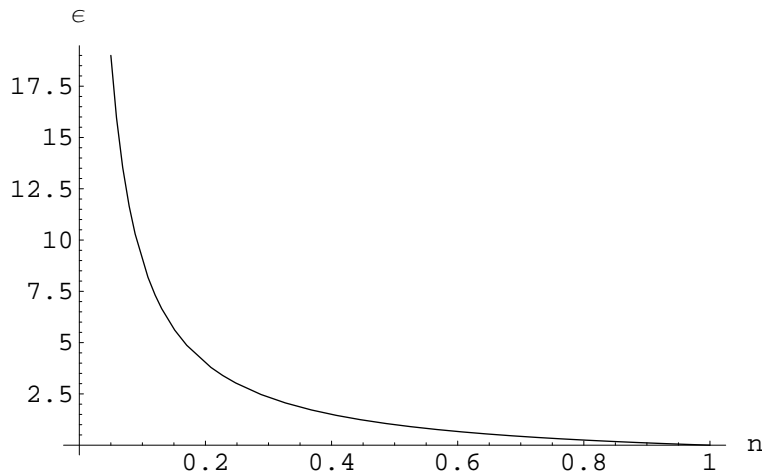


Figure 1: A depiction of the states in which the expected utility of being implicit equals the expected utility of being explicit. That is, $\epsilon = \frac{1-n}{n}$.

5 Risk-dominance versus payoff-dominance

In this section we study the use of non-literal speech as a means of communication. Following standard practice in philosophy of language, we distinguish between what a sentence means and what a speaker means by uttering this sentence. If the sentence meaning gives rise to a fully specified interpretation, the two will coincide in standard situations, i.e., when the speaker uses the sentence in the conventional way. In specific circumstances, however, it can be that even if the sentence has a fully specified meaning, the speaker means something quite different with her use of the sentence than the literal meaning of the sentence. This is the case when the sentence has besides a literal, also a non-literal meaning. Think of the use of indirect speech (acts), irony (such as over- and understatement), and metaphor. In the hearer's process of attaching the non-literal interpretation of a sentence, the hearer first has to recognize the defectiveness of the utterance's literal meaning. Following some suggestions of Sally (2003), we will argue that this type of speech can be successfully modelled as, what we called, a Reindeer hunt game, where both successful communication by the non-literal and the literal use of one's language are equilibrium plays of the game, but where the former is payoff-dominant, whereas the latter is risk-dominant.

Sally (2003) argues that “people play the language game in a way that

that this would not result in a different analysis.

is consistent with their play in all games.” Sally does so by fixing rules of thumb¹⁴ that describe people’s behavior while playing coordination games in a lab-setting. For instance, Sally considers the rules

- (A) In a game with one outcome risk-dominant and another “modestly” payoff-dominant, the former is more likely to be chosen.
- (B) As sympathy between the players increases, a payoff-dominant, risk dominated equilibrium is more likely to be realized.¹⁵

Concerning the status of these rules Sally says that “these empirical findings are clearly not hard and fast rules of coordination game play, but rather tendencies manifest in normal play.” And this is exactly the way we will treat them below.

To make these findings relevant for pragmatics, Sally introduces a “more complete coordination game of communication” that does not take states/meanings as primary objects but rather speech acts, as proposed by Austin (1962) and Searle (1969). Accordingly, the payoff-functions range over pairs of speech acts. Sally does not define them formally, however.

In contrast, we have strictly defined our super conventional signaling games in section 3. Recall that super conventional signaling games were defined in terms of states and actions — just like Lewis did —, and furthermore that rewarded payoff not only depends on the success of communication but also on whether or not the conventional strategies were respected. This conventional meaning was supposed to be a commonly known parameter of the super conventional game. As such, our model is not only more precisely defined than Sally’s but also requires less complicated notions. I.e. the only notion required besides the ones presupposed by Lewis is the notion of conventional meaning, which itself can be considered the result of a Lewisian game. We will see that this limitation does not limit Sally’s claim on language use resembling game playing.

Sally, namely, applies his game-theoretical rules of thumb to his signaling game to make predictions as to how people use language. He argues implicitly that the payoff-dominant equilibria are the signaling systems that communicate non-literally, whereas the risk-dominant equilibria communicate according to the convention. Then, for instance, rule (A) would predict that people speak literally by default; and rule (B) would predict that as sympathy between the players increases, people are more likely to communicate non-literally.

In Sevenster (2004) it is proved that

¹⁴Sally calls them “Wittgensteinian signposts”.

¹⁵This rule is the result of empirical research. We think that this rule also has a theoretical counterpart. As we saw above, different expectations about the opponent yield different Nash equilibria. By modelling ‘sympathy’ as having expectations that the other player behaves such as to maximize his actual utility (leaving out considerations of *expected* utility for the moment), we can theoretically enforce risk dominated equilibria. That is, the more sympathetic the players are towards each other, the more they will be tempted to play riskily. We believe that this might have interesting consequences for the analysis of language change, but will not speculate about this here.

- If $\langle s, r \rangle$ is a signaling system in $G_{\langle cs, cr \rangle}$, then $\langle s, r \rangle$ is a Nash equilibrium in $G_{\langle cs, cr \rangle}$.

This result corresponds to Lewis’ result and establishes that also in this model, signaling systems are the first-class citizens. Furthermore, characterizations of the payoff-dominant and risk-dominant equilibria are given:

- $\langle s, r \rangle$ is a payoff-dominant Nash equilibrium in $G_{\langle cs, cr \rangle}$ iff $\langle s, r \rangle$ is a signaling system and for every $t \in T$ it is the case that $s(t) \neq cs(t)$.
- If $\epsilon' > \epsilon$, then $\langle s, r \rangle$ risk-dominates all other signaling systems in $G_{\langle cs, cr \rangle}$ iff $s = cs$ and $r = cr$.

Taken strictly these formal characterizations do not teach us anything. In line with Sally, however, we can make a sketchy account as to how they should be interpreted. These characterizations, namely, enable us to apply rules (A) and (B) to our super conventional signaling games, and the same predictions can be made as Sally did.

To get a better understanding of these results let us consider the case of indirect requests and see what are the implications for the parameters ϵ , and ϵ' .¹⁶

As to indirect requests, think of a room containing a hearer having control over the open window and a speaker who is cold. The speaker wants the hearer to close the window and has two ways to communicate this. Either he uses the conventional message, such as “Could you close the window” or he makes an indirect request, such as “It’s cold in here”. The hearer on the other side has also two options, either interpret the message figuratively or literally. This simple game has two equilibria:

		<i>Figurative</i>	<i>Literal</i>
Game 7:	It’s cold in here	$1 + \epsilon, 1 + \epsilon$	$-\epsilon', 0$
	Could you close the window?	$0, -\epsilon'$	$1, 1$

That the correctly communicated “It’s cold in here” is more rewarding for both ($1 + \epsilon$ vs. 1) can be explained in terms of politeness: the speaker did not have to command the hearer and the hearer is not commanded.

That $\epsilon' > \epsilon$ means for the speaker that the benefit of being indirect is lower than the cost of being misunderstood. In case of misunderstanding by the use of a short message, the speaker would have to make a direct request, in order to accomplish her goal. Misunderstanding for the speaker is less bad if she is being literal, because she does not have to take the blame — I said so! On the other hand, misunderstanding for the hearer is less bad if interpreted literally — why didn’t you say so? Communicating literally is thus safe: the sentence meaning provides a face-saving excuse in the event of miscoordination.

¹⁶The same story can be told for ironical statements, i.e. understatements like “I wasn’t overimpressed by her speech.”

6 How to interpret answers exhaustively

Until now we have discussed situations where it was possible to decompose the utility function by answering two questions which were taken to be independent: (i) was the intended content successfully communicated? and (ii) did the agent use the conventional safe strategy or not? In the final substantial section of this paper we are going to discuss a somewhat more complicated case: we will give up the assumption that successful communication is a yes-or-no matter. In particular, we are going to discuss an example where it is intuitively the case that if the speaker adopts a risky strategy and the hearer a safe strategy there will be some useful transfer of communication, although this information transmission is not perfect. Thus, we will make a distinction in utilities when the speaker is adopting a risky strategy between the case where the hearer adopts an incorrect risky strategy, and a safe strategy.

In the previous sections we assumed that the speaker could choose to play risky or safe, and that the hearer will just interpret the underspecified message either correctly or incorrectly. Now we are going to look at a situation where also the hearer can interpret an underspecified message either in a risky way or in a safe way, and where the safe interpretation of an underspecified message is better than the incorrect risky interpretation, but worse than a correct risky interpretation. In abstract, these kind of situations give rise to a game like the following:

Game 8 :

t_1	C_1	C_2	C_3
Risky	$1 + \epsilon, 1$	$0, 0$	$1 - \epsilon', 1 - \epsilon'$
Safe	$1, 1$	$1, 1$	$1, 1$

where $\epsilon' < 1$. In the matrix of Game 8, C_1 denotes the risky strategy that is correct in this situation; C_2 the risky strategy that is incorrect in this situation, while C_3 stands for the safe strategy. Before we are going to analyze this kind of situations, let us first convince ourselves of their existence by looking at the interpretation of answers.¹⁷

Consider the following dialogue:

- (1) a. Bob: Who passed the examination?
- b. Ann: John and Mary.

What can Bob conclude from Ann's answer, besides the fact that John and Mary passed the examination? It seems only natural to assume that Ann mentioned all individuals of which she knows that they passed the examination. By making this assumption, Bob concludes that it is not the case that Ann knows that Sue, for instance, passed the examination. This seems a very reasonable inference to make.¹⁸

¹⁷This is just one of many examples that behave in this way. We believe, however, that it is an example that is more easy to explain than many of its alternatives.

¹⁸For a general formalization of this kind of this kind of reasoning, see van Rooij & Schulz (2004).

In many circumstances, however, Bob concludes something more from Ann's answer than just (i) the semantic meaning of the answer, and (ii) that it is not the case that Ann knows that Sue, for instance, passed the examination: Bob also concludes that Sue *did not* pass the examination. This extra inference comes about via Bob's extra assumption that Ann knows exactly who in fact passed the examination, i.e., the assumption that Ann is *competent* on the extension of the question-predicate.¹⁹ From the fact that Ann did not say that Sue passed the examination; the assumption that she mentioned all individuals of which she knows that they passed the examination; and the extra assumption that she knows who passed, Bob concludes that Ann knows that Sue did not pass. Due to the fact that knowledge entails truth, Bob concludes that Sue did, in fact, not pass.

We see that due to an assumption of competence, Bob can strengthen what she can infer from Ann's answer by taking Ann to obey the principle to mention all individuals you know to satisfy the question-predicate. There is, however, also another way to strengthen this inference: it can be that Bob assumes that Ann is *not competent* on the extension of the question-predicate. In that case, Bob can strengthen her inference that it is not the case that Ann knows *that* Sue passed the examination to the inference that it is not the case that Ann knows *whether* Sue passed the examination. Notice that this is indeed a strengthening, because the lack of knowledge *that* Sue passed is compatible with the knowledge that Sue did not pass, but this is not the case for the lack of knowledge *whether* Sue passed.

The above discussion shows that even if Bob assumes by Gricean reasoning that Ann mentioned all the individuals of which she knows that they passed the examination, this still leaves open three interpretations: one where Bob cannot infer any more than this; one where Bob can conclude that Sue did not pass; and one where Bob concludes that Ann doesn't know whether Sue passed. The latter two inferences are due to assumptions of competence and incompetence, respectively. Of course, Bob's (pragmatic) interpretation of Ann's answer by making these assumption is *risky*: his assumptions could turn out to be false, and, consequently, his additional inferences as well.

But for Ann to give an answer like (1b) without explicitly mentioning what more, if at all, she knows about the extension of the question-predicate is risky as well: Bob might adopt the wrong assumption concerning Ann's competence about the question-predicate and interprets the answer in a different way as intended. If Ann wants to be sure that Bob will understand the answer correctly, she has to play it *safe* and be very explicit about her knowledge.

We can think of the dialogue as a game between Ann and Bob where both either play risky or safe. Notice that this kind of game is not really a coordination game, because it seems natural to assume that in case Ann plays it safe and is completely explicit about her knowledge, Bob will always interpret the answer in the correct way. That is why also this game can most naturally be thought of as a game with alignment of preferences. Assume that the dialogue

¹⁹Again, see van Rooij & Schulz (2004) for one way to make this kind of reasoning precise.

takes place in a situation where Ann is in fact competent on the extension of the question-predicate. Then it gives rise to the following payoff-table at the left hand side. At the right hand side, we represent the expected values on the assumption that n denotes the probability that Ann is competent:

Game 9 :	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border: none;"></td> <td style="border: none;"><i>Comp</i></td> <td style="border: none;"><i>Incomp</i></td> <td style="border: none;"><i>Unkown</i></td> </tr> <tr> <td style="border: none;">Risky</td> <td style="border: 1px solid black; padding: 2px 5px;">$1 + \epsilon$</td> <td style="border: 1px solid black; padding: 2px 5px;">0</td> <td style="border: 1px solid black; padding: 2px 5px;">$1 - \epsilon'$</td> </tr> <tr> <td style="border: none;">Safe</td> <td style="border: 1px solid black; padding: 2px 5px;">1</td> <td style="border: 1px solid black; padding: 2px 5px;">1</td> <td style="border: 1px solid black; padding: 2px 5px;">1</td> </tr> </table>		<i>Comp</i>	<i>Incomp</i>	<i>Unkown</i>	Risky	$1 + \epsilon$	0	$1 - \epsilon'$	Safe	1	1	1	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border: none;"></td> <td style="border: none;">Risky</td> <td style="border: none;">Safe</td> </tr> <tr> <td style="border: none;">Risky</td> <td style="border: 1px solid black; padding: 2px 5px;">$n \times (1 + \epsilon)$</td> <td style="border: 1px solid black; padding: 2px 5px;">$1 - \epsilon'$</td> </tr> <tr> <td style="border: none;">Safe</td> <td style="border: 1px solid black; padding: 2px 5px;">1</td> <td style="border: 1px solid black; padding: 2px 5px;">1</td> </tr> </table>		Risky	Safe	Risky	$n \times (1 + \epsilon)$	$1 - \epsilon'$	Safe	1	1
	<i>Comp</i>	<i>Incomp</i>	<i>Unkown</i>																				
Risky	$1 + \epsilon$	0	$1 - \epsilon'$																				
Safe	1	1	1																				
	Risky	Safe																					
Risky	$n \times (1 + \epsilon)$	$1 - \epsilon'$																					
Safe	1	1																					

Suppose first that Bob, the column player, always plays risky. In that case we are back to our discussion in section 4, and it pays for Ann to play risky as well iff $\epsilon > \frac{1-n}{n}$, where n denotes the percentage by which Bob makes the correct prediction of competence. Thus, in case Bob is known to normally only ask questions to somebody who he thinks is competent (a not unreasonable procedure), the benefit of the short, but risky, answer for the answerer Ann doesn't have to be very large to still be worthwhile.

Now assume that Ann cannot assume that Bob always plays risky. Instead, Bob interprets answers only in 50 percent of the cases by making the assumption that the speaker was competent or incompetent. Denote the percentage by which Bob makes the correct assumption of competence, given that Ann is either competent or incompetent, to be n . In that case, it pays for speaker Ann to be risky if and only if $n(1 + \epsilon) \geq 1 + \epsilon'$.²⁰ If we now assume that (for some reason) the benefit of successful implicit communication is twice as high as when the hearer interprets things in a safe way: $\epsilon = 2\epsilon'$, the equality $n(1 + \epsilon) \geq 1 + \epsilon'$ can be plotted as in figure 2.

It should not be surprising that, indeed, for Ann to play risky if she is not sure whether Bob interprets in a risky way or not, she has to be even surer that Bob makes the correct assumption of competence than in case it is known that Bob plays risky.

7 Conclusion

Starting point of our paper is the insight that initial expectations that players have of each other's choice of action is important to solve a game with several equilibria. This important role of expectations — though crucial for Lewis's (1969) analysis of convention — only recently was given an interesting twist in game-theoretical analyses of conversation in the work of Sally (2003). He discusses how the notion of 'risk' might be important in conversational situations between speakers and hearers. In this paper we try to go beyond this work (i) by clarifying the connection of Sally's work with Lewisian signaling games, and (ii) by looking at some additional ways in which speech can be risky.

²⁰This can be shown by the following calculation:

$$\begin{aligned} \frac{[\frac{n}{2}(1 + \epsilon)] + [\frac{1}{2}(1 - \epsilon')]}{n(1 + \epsilon)} &\geq 1 \quad \text{iff} \\ &\geq 1 + \epsilon'. \end{aligned}$$

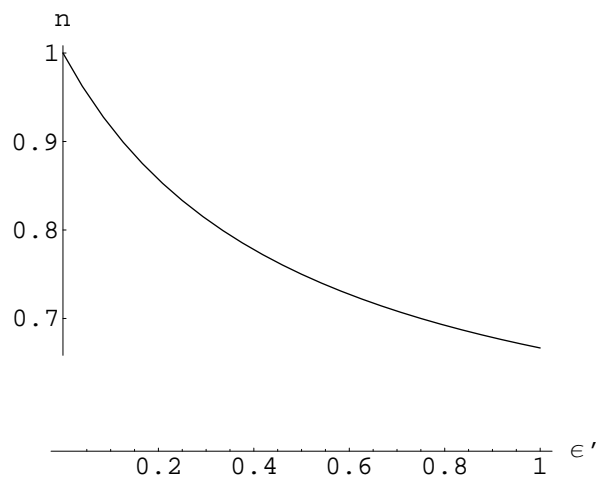


Figure 2: A depiction of the function $n = (1 + \epsilon') / (1 + 2\epsilon')$. It reflects the states in which the expected utility of playing risky, $n(1 + \epsilon)$ equals the expected utility of playing safe, $1 + \epsilon'$, where $\epsilon = 2\epsilon'$.

References

- [1] Austin, J.L. (1962), *How To Do Things With Words*, Oxford University Press, Oxford.
- [2] Clark, H. H. (1996), *Using Language*, Cambridge University Press, Cambridge.
- [3] Grice, H.P. (1967), ‘Logic and conversation’, *William James Lectures*, Harvard University, reprinted in *Studies in the Way of Words*, 1989, Harvard University Press, Cambridge, Massachusetts.
- [4] Harsanyi, J. C. and R. Selten (1988), *A General Theory of Equilibrium Selection in Games*, MIT Press, Cambridge, Massachusetts.
- [5] Lewis, D. (1969), *Convention: A Philosophical Study*, Harvard University Press, Cambridge, Massachusetts.
- [6] Parikh, P. (2001), *The Use of Language*, CSLI Publications, Stanford.
- [7] Rooij, R. van (2004), ‘Evolution of conventional meaning and conversational principles’, *Knowledge, Rationality & Action*, a special section of *Synthese*, **139**: 331-366.
- [8] Rooij, R. van and K. Schulz (2004), ‘Exhaustive interpretation of complex sentences’, *Journal of Logic, Language, and Information*, **13**: 491-519.

- [9] Rousseau, J.J. (1755), *Discours sur l'origine et les fondement de l'inégalité les hommes*, chez Marc Michel Rey, Amsterdam.
- [10] Sally, D. (2003), 'Risky speech: behavioral game theory and pragmatics', *Journal of Pragmatics*, **35**: 1223-1245.
- [11] Schelling, T. (1960), *The Strategy of Conflict*, Harvard University Press, Cambridge.
- [12] Searle, J.R. (1969), *Speech Acts. An essay in the Philosophy of Language*, Cambridge University Press, Cambridge.
- [13] Sevenster, M. (2004), 'Signaling games and non-literal meaning', In: P. Egré et al (eds), *Proceedings of the ESSLLI 2004 student session*.
- [14] Skyrms, B. (1990), *The Dynamics of Rational Deliberation*, Harvard University Press, Cambridge, Massachusetts.
- [15] Skyrms, B. (2004), *The Stag Hunt and the Evolution of Social Structure*, Cambridge University Press, Cambridge, Massachusetts.