8

Communication

Agents communicate; this is one of the defining characteristics of a multiagent system. In traditional linguistic analysis, the communication is taken to have a certain form (syntax), to carry a certain meaning (semantics), and to be influenced by various circumstances of the communication (pragmatics). As we shall see, a closer look at communication adds to the complexity of the story. We can distinguish between purely *informational* theories of communication and *motivational* ones. In informational communication, agents simply inform each other of different facts. The theories of belief change, introduced in Chapter 14, look at ways in which beliefs change in the face of new information—depending on whether the beliefs are logical or probabilistic, consistent with prior beliefs or not. In this chapter we broaden the discussion and consider motivational theories of communication, involving agents with individual motivations and possible courses of actions.

We divide the discussion into three parts. The first concerns *cheap talk* and describes a situation in which self-motivated agents can engage in costless communication before taking action. As we see, in some situations this talk influences future behavior, and in some it does not. Cheap talk can be viewed as "doing by talking"; in contrast, *signaling games* can be viewed as "talking by doing." In signaling games an agent can take actions that, by virtue of the underlying incentives, communicate to the other agent something new. Since these theories draw on game theory, cheap talk and signaling both apply in cooperative as well as in competitive situations. In contrast, *speech-act theory*, which draws on philosophy and linguistics, applies in purely cooperative situations. It describes pragmatic ways in which language is used not only to convey information but to effect change; as such, it too has the flavor of "doing by talking."

8.1 "Doing by talking" I: cheap talk

Consider the Prisoner's Dilemma game, reproduced here in Figure 8.1. Recall that the game has a unique equilibrium in dominant strategies, the strategy profile (D, D), which is ironically also the only outcome that is not Pareto optimal; both players would do better if they both choose C instead. Suppose now that the prisoners are allowed to communicate before they play; will this change the outcome of the game? Intuitively, the answer is no. Regardless of the other agent's action, the given agent's best action is still D; the other agent's talk is indeed

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

	C	D
C	$-1, -1$	$-4, 0$
D	$0, -4$	$-3, -3$

Figure 8.1 The Prisoner's Dilemma game.

cheap. Furthermore, regardless of his true intention, it is the interest of a given agent to get the other agent to play C; his talk is not only cheap, but also not credible (or, as the saying goes, the talk is free—and worth every penny).

Contrast this with cheap talk prior to the Coordination game given in Figure 8.2.

	L	R
U	$1, 1$	$0, 0$
D	$0, 0$	$1, 1$

Figure 8.2 Coordination game.

self-committing utterance

self-revealing utterance

Here, if the row player declares "I will play U" prior to playing the game, the column player should take this seriously. Indeed, this utterance by the row player is both *self-committing* and *self-revealing*. These two notions are related but subtly different. A declaration of intent is self-committing if, once uttered, and assuming it is believed, the optimal course of action for the player is indeed to act as declared. In this example, if the column player believes the utterance "I will play U," then his best response is to play L. But then the row player's best response is indeed to play U. In contrast, an utterance is self-revealing if, assuming that it is uttered with the expectation that it will be believed, it is uttered only when indeed the intention was to act that way. In our case, a row player intending to play D will never announce the intention to play U, and so the utterance is self-revealing.

It must be mentioned that the precise analysis of this example, as well as the later examples, is subtle in a number of ways. In particular, the equilibrium analysis reveals other, less desirable equilibria than the ones in which a meaningful message is transmitted and received. For example, this example has another, less obvious equilibrium. The column player could ignore anything the row player says, allowing its beliefs to be unaffected by signals. In this case, the row player has no incentive to say anything in particular, and he might as well "babble," that is, send signals that are uncorrelated with his type. For this reason, we call this a *babbling equilibrium*. In theory, every cheap talk game has a babbling equilibrium; there is always an equilibrium in which one party sends a meaningless signal and the other party ignores it. An equilibrium that

babbling equilibrium

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

revealing equilibrium

is not a babbling equilibrium is called a *revealing equilibrium*. In a similar fashion one can have odd equilibria in which messages are not ignored but are used in a nonstandard way. For example, the row player might send the signal U when she means D and vice versa, so long as the column player adopts the same convention. However, going forward we will ignore these complications, and assume a meaningful and straightforward communication among the parties.

focal point

Stag Hunt game

It might seem that self-commitment and self-revelation are inseparable, but this is an artifact of the pure coordination nature of the game. In such games the utterance creates a so-called *focal point*, a signal on which the agents can coordinate their actions. But now consider the well-known *Stag Hunt game*, whose payoff matrix is shown in Figure 8.3.

	Stag	Hare
Stag	9, 9	0, 8
Hare	8, 0	7, 7

Figure 8.3 Payoff matrix for the Stag Hunt game.

In the story behind this game, Artemis and Calliope are about to go hunting, and are trying to decide whether they want to hunt stag or hare. If both hunt stag, they do very well; if one tries to hunt stag alone, she fails completely. On the other hand, if one hunts rabbits alone, she will do well, for there is no competition; if both hunt rabbits together, they only do OK, for they each have competition.

In each cell of the matrix, Artemis' payoff is listed first and Calliope's payoff is listed second. This game has a symmetric mixed-strategy equilibrium, in which each player hunts stag with probability $\frac{7}{8}$, yielding an expected utility of $7\frac{7}{8}$. But now suppose Artemis can speak to Calliope before the game; can he do any better? The answer is arguably yes. Consider the message "I plan to hunt stag." It is not self-revealing; Artemis would like Calliope to believe this, even if she does not actually plan to hunt stag. However, it *is* self-committing; if Artemis were to think that Calliope believes her, then Artemis would actually prefer to hunt stag. There is however the question of whether Calliope would believe the utterance, knowing that it is not self-revealing on the part of Artemis.

For this reason, some view self-commitment without self-revelation as a notion lacking force. To gain further insight into this issue, let us define the Stag Hunt game more generally. Consider the game in Figure 8.4. Here, if x is less than 7, then the message "I plan to hunt stag" is possibly credible. However, if x is greater than 7, then that message is not at all credible, because it is in Artemis' best interest to get Calliope to hunt stag, no matter what Artemis actually intends to play.

We have so far spoken about communication in the context of games of perfect information. In such games all that can possibly be revealed in the intention to

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

	Stag	Hare
Stag	9, 9	0, x
Hare	x, 0	7, 7

Figure 8.4 More general payoff matrix for the Stag Hunt game.

High-ability game

Signal high ability	3, 1
Signal low ability	0, 0

Low-ability game

Signal high ability	0, 0
Signal low ability	2, 1

Figure 8.5 Payoff matrix for the Job Hunt game.

act a certain way. In games of incomplete information, however, there is an opportunity to reveal one's own private information prior to acting.

Consider the following example. The Acme Corporation wants to hire Sally into one of two positions: a demanding and an undemanding position. Sally may have high or low ability. Sally prefers the demanding position if she has high ability (because of salary and intellectual challenge) and she prefers the undemanding positions if she instead has low ability (because it will be more manageable). Acme too prefers that Sally be in the demanding position if she has high ability, and that she be in the undemanding position if she is of low ability. The actual game being played is determined by Nature; for concreteness, let us assume that selection is done with uniform probability. Importantly, however, only Sally knows what her true ability level is. However, before they play the game, Sally can send Acme a signal about her ability level. Suppose for the sake of simplicity that Sally can only choose from two signals: "My ability is low," and "My ability is high." Note that Sally may choose to be either sincere or insincere. The situation is modeled by the two games in Figure 8.5; in each cell of the matrix, Sally's payoff is listed first, and Acme's payoff is listed second.

What signal should Sally send? It seems obvious that she should tell the truth. She has no incentive to lie about her ability. If she were to lie, and Acme were to believe her, then she would receive a lower payoff than if she had told the truth. Acme knows that she has no reason to lie and so will believe her. Thus there in an equilibrium in which when Sally has low ability she says so, and Acme gives her an undemanding job, and when Sally has high ability she also says so, and Acme gives her a demanding job. The message is therefore self-signaling; assuming she will be believed, Sally will send the message only if it is true.

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

8.2 "Talking by doing": signaling games

We have so far discussed the situation in which talk preceded action. But sometimes actions speak louder than words. In this section we consider a class of imperfect-information games called *signaling games*.

signaling game

Definition 8.2.1 (Signaling game) *A* signaling game *is a two-player game in which Nature selects a game to be played according to a commonly known distribution, player 1 is informed of that choice and chooses an action, and player 2 then chooses an action without knowing Nature's choice, but knowing player 1's choice.*

In other words, a signaling game is an extensive-form game in which player 2 has incomplete information.

It is tempting to model player 2's decision problem as follows. Since each of the possible games has a different set of payoffs, player 2 must first calculate the posterior probability distribution over possible games, given the message that she received from player 1. She can calculate this using Bayes rule with the prior distribution over games and the conditional probabilities of player 1's message given the game. More precisely, the expected payoff for each action is as follows.

$$\begin{aligned} u_2(a, m) &= \mathbb{E}(u_2(g, m, a)|m, a) \\ &= \sum_{g \in G} u_2(g, m, a) P(g|m, a) \\ &= \sum_{g \in G} u_2(g, m, a) P(g|m) \\ &= \sum_{g \in G} u_2(g, m, a) \frac{P(m|g)P(g)}{P(m)} \\ &= \sum_{g \in G} u_2(g, m, a) \frac{P(m|g)P(g)}{\sum_{g \in G} P(m|g)P(g)} \end{aligned}$$

One problem with this formulation is that the use of Bayes' rule requires that the probabilities involved be nonzero. But more acutely, how does player 2 calculate the probability of player 1's message given a certain game? This is not at all obvious in light of the fact that player 2 knows that player 1 knows that player 2 will go through such reasoning, et cetera. Indeed, even if player 1 has a dominant strategy in the game being played the situation is not straightforward. Consider the following signaling game. Nature chooses with equal probability one of the two zero-sum normal-form games given in Figure 8.6.

Recall that player 1 knows which game is being played, and will choose his message first (U or D), and then player 2, who does not know which game is being played, will choose his action (L or R). What should player 1 do?

Note that in the leftmost game (U, R) is an equilibrium in dominant strategies, and in rightmost game (D, L) is an equilibrium in dominant strategies. Since player 2's preferred action depends entirely on the game being played, and he is confident that player 1 will play his dominant strategy, his best response is R if

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

	L	R
U	$4, -4$	$1, -1$
D	$3, -3$	$0, 0$

	L	R
U	$1, -1$	$3, -3$
D	$2, -2$	$5, -5$

Figure 8.6 A signaling setting: Nature chooses randomly between the two games.

player 1 chooses U, and L if player 1 chooses D. If player 2 plays in this fashion, we can calculate the expected payoff to player 1 as

$$\mathbb{E}(u_1) = (0.5)1 + (0.5)2 = 1.5.$$

This seems like an optimal strategy. However, consider a different strategy for player 1. If player 1 always chooses D, regardless of what game he is playing, then his payoff is independent of player 2's action. We calculate the expected payoff to player 1 as follows, assuming that player 2 plays L with probability p and R with probability $(1 - p)$:

$$\mathbb{E}(u_1) = (0.5)(3p + 0(1 - p)) + (0.5)(2p + 5(1 - p)) = 2.5.$$

Thus, player 1 has a higher expected payoff if he always chooses the message D.

The example highlights an interesting property of signaling games. Although player 1 has privileged information, it may not always be to his advantage to exploit it. This is because by exploiting the advantage, he is effectively telling player 2 what game is being played and thereby losing his advantage. Thus, in some cases player 1 can receive a higher payoff by ignoring his information.

games of asymmetric information

Spence signaling game

Signaling games fall under the umbrella term of *games of asymmetric information*. One of the best-known examples is the so-called *Spence signaling game*, which offers a rationale for enrolling in a difficult academic program. Consider the situation in which an employer is trying to decide how much to pay a new worker. The worker may or may not be talented, and can signal this to the employer by choosing to get either a high or low level of education. Specifically, we can model the setting as a Bayesian game between an employer and a worker in which *Nature* first chooses the level of the worker's talent, θ, to be either θ_L or θ_H, such that $\theta_L < \theta_H$. This value of θ defines two different possible games. In each possible game, the worker's strategy space is the level of education e to get for both possible types, or level of talent. We use e_L to refer to the level of education chosen by the worker if his talent level is θ_L and e_H for the education chosen if his talent level is θ_H. We assume that the worker knows his talent.

Finally, the employer's strategy specifies two wages, w_H and w_L, to offer a worker based on whether his signal is e_H or e_L. We assume that the employer does not know the level of talent of the worker, but does get to observe his level of education. The employer is assumed to have two choices. One is to ignore the signal and set $w_H = w_L = p_H\theta_H + p_L\theta_L$, where $p_L + p_H = 1$ are the probabilities with which $Nature$ chooses a high and low talent for the worker.

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

The other is to pay a worker with a high education w_H and a worker with a low education w_L.

The payoff to the employer is $\theta - w$, the difference between the talent of the worker and the payment to him. The payoff to the worker is $w - e/\theta$, reflecting the assumption that education is easier when talent is higher.

pooling equilibrium

separating equilibrium

This game has two equilibria. The first is a *pooling equilibrium*, in which the worker will choose the same level of education regardless of his type ($e_L = e_H = e^*$), and the employer pays all workers the same amount. The other is a *separating equilibrium*, in which the worker will choose a different level of education depending on his type. In this case a low-talent worker will choose to get no education, $e_L = 0$, because the wage paid to this worker is w_L, independent of e_L. The education chosen by a high-talent worker is set in such a way as to make it unprofitable for either type of worker to mimic the other. This is the case only if the following two inequalities are satisfied.

$$\theta_L \geq \theta_H - e_H/\theta_L$$
$$\theta_L \leq \theta_H - e_H/\theta_H$$

These inequalities can be rewritten in terms of e_H as

$$\theta_L(\theta_H - \theta_L) \leq e_H \leq \theta_H(\theta_H - \theta_L).$$

Note that since $\theta_H > \theta_L$, a separating equilibrium always exists.

8.3 "Doing by talking" II: speech-act theory

Human communication is as rich and imprecise as natural language, tone, affect, and body language permit, and human motivations are similarly complex. It is not surprising that philosophers and linguists have attempted to model such communication. As mentioned at the very start of the chapter, human communication is analyzed on many different levels of abstraction, among them the *syntactic*, *semantic*, and *pragmatic* levels. The discussion of speech acts lies squarely within the pragmatic level, although it should be noted that there are legitimate arguments against a crisp separation among these layers.

8.3.1 *Speech acts*

The traditional view of communication is that it is the sharing of information. Speech-act theory, due to the philosopher J. L. Austin, embodies the insight that some communications can instead be viewed as actions, intended to achieve some goal.

locutionary act

Speech-act theory distinguishes between three different kinds of speech acts, or, if you wish, three levels at which an utterance can be analyzed. The *locutionary act* is merely the emission of a signal carrying a certain meaning. When I say "there's a car coming your way," the locution refers to the content transmitted. Locutions establish a proposition, which may be true or false. However, the

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

illocutionary act

utterance can also be viewed as an *illocutionary act*, which in this case is a *warning*. In general, an illocution is the invocation of a conventional force on the receiver through the utterances. Other illocutions can be making a request, telling a joke, or, indeed, simply informing.

perlocutionary act

Finally, if the illocution captures the intention of the speaker, the *perlocutionary act* is bringing about an effect on the hearer as a result of an utterance. Although the illocutionary and perlocutionary acts may seem similar, it is important to distinguish between an illocutionary act and its perlocutionary consequences. Illocutionary acts do something *in* saying something, while perlocutionary acts do something *by* saying something. Perlocutionary acts include scaring, convincing, and saddening. In our car example, the perlocution would be an understanding by the hearer of the imminent danger causing him to jump from in front of the car.

performative

Illocutions thus may or may not be successful. *Performatives* constitute a type of act that is inherently successful. Merely saying something achieves the desired effect. For example, the utterance "please get off my foot" (or, somewhat more stiffly, "I hereby request you to get off my foot") is a performative. The speaker asserts that the utterance is a request, and is thereby successful in communicating the request to the listener, because the listener assumes that the speaker is an expert on his own mental state. Some utterances are performatives only under some circumstances. For example, the statement "I hereby pronounce you man and wife" is a performative only if the speaker is empowered to conduct marriage ceremonies in that time and place, if the rest of the ceremony follows protocol, if the bride and groom are eligible for marriage, and so on.[1]

8.3.2 *Rules of conversation*

rules of conversation

Building on the notion of speech acts as a foundation, another important contribution to language pragmatics takes the form of *rules of conversation*, as developed by P. Grice, another philosopher. The simple observation is that humans seem to undertake the act of conversation cooperatively. Humans generally seek to understand and be understood when engaging in conversation, even when other motivations may be at odds. It is in both parties' best interest to communicate clearly and efficiently. This is called the *cooperative principle*.

cooperative principle

It is also the case that humans generally follow some basic rules when conversing, which presumably help them to achieve the larger shared goal of the Cooperative Principle. These rules have come to be known as the *Gricean maxims*. The four Gricean maxims are *quantity*, *quality*, *relation*, and *manner*. We discuss each one in turn.

Gricean maxims

The rule of *quantity* states that humans tend to provide listeners with exactly the amount of information required in the current conversation, even when they have access to more information. As an example, imagine that a waitress asks you, "how do you like your coffee?" You would probably answer, "Cream, no sugar, please," or something similar. You would probably not answer, "I like

1. It is however interesting to contemplate a world in which any such utterance results in a marriage.

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

arabica beans, grown in the mountains of Guatemala. I prefer the medium roast from Peet's Coffee. I like to buy whole beans, which I keep in the freezer, and grind them just before brewing. I like the coffee strong, and served with a dash of cream." The latter response clearly provides the waitress with much more information than she needs. You also probably would not respond, "no sugar," because this does not give the waitress enough information to do her job.

The rule of *quality* states that humans usually only say things that they actually believe. More specifically, humans do not say things they know to be false, and do not say things for which they lack adequate evidence. For example, if someone asks you about the weather outside, you respond that it is raining only if in fact you believe that it is raining, and if you have evidence to support that belief.

The rule of *relation* states that humans tend to say things that are relevant to the current conversation. If a stranger approaches you on the street to ask for directions to the nearest gas station, they would be quite surprised if you began to tell them a story about your grandmother's cooking.

Finally, the rule of *manner* states that humans generally say things in a manner that is brief and clear. When you are asked at the airport whether anyone unknown to you has asked you to carry something in your luggage, the appropriate answer is either "yes" or "no," not "many people assume that they know their family members, but what does that really mean?" In general, humans tend to avoid obscurity, ambiguity, prolixity, and disorganization.

These maxims help explain a surprising phenomenon about human speech, namely that we often succeed in communicating much more meaning than is contained directly in the words they say. This phenomenon is called *implicature*. For example, suppose that *A* and *B* are talking about a mutual friend, *C*, who is now working in a bank. *A* asks *B* how *C* is getting on in his job, and *B* replies, "Oh quite well, I think; he likes his colleagues, and he has not been to prison yet." Clearly, by stating the simple fact that *C* hasn't been to prison yet, which is a truism for most people, *B* is implying, suggesting, or meaning something else. He may mean that *C* is the kind of person who is likely to yield to temptation or that *C*'s colleagues are really very treacherous people, for example. In this case the implicature may be clear from the context of their conversation, or *A* may have to ask *B* what he means.

implicature

Grice distinguished between *conventional* and *nonconventional* implicature. The former refers to the case in which the conventional meaning of the words used determines what is implicated. In the latter, the implication does not follow directly from the conventional meaning of the words, but instead follows from context, or from the structure of the conversation, as is the case in *conversational implicatures*.

conversational implicature

In conversational implicatures, the implied meaning relies on the fact that the hearer assumes that the speaker is following the Gricean maxims. Let us begin with an example. *A* is standing by an immobilized car, and is approached by *B*. *A* says, "I am out of gas." *B* says, "There is a garage around the corner." Although *B* does not explicitly say it, she implicates effectively that she thinks that the garage is open and sells gasoline. This follows immediately from the assumption that *B* is following the Gricean maxims of relation and quality. If she were not following

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

the maxim of relation, her utterance about the garage could be a *non sequitur*; if she were not following the maxim of quality, she could be lying. In order for a conversational implicature to occur, (1) the hearer must assume that the speaker is following the maxims, (2) this assumption is necessary for the hearer to get the implied meaning, and (3) it is common knowledge that the hearer can work out the implication.

Grice offers three types of conversational implicature. In the first, no maxim is violated, as in the aforementioned example. In the second, a maxim is violated, but the hearer assumes that the violation is because of a clash with another maxim. For example, if A asks, "Where does C live?" and B responds, "Somewhere in the South of France," A can presume that B does not know more and thus violates the maxim of quantity in order to obey the maxim of quality. Finally, in the third type of conversational implicature, a maxim is flouted, and the hearer assumes that there must be another reason for it. For example, when a recommendation letter says very little about the candidate in question, the maxim of quantity is flouted, and the reader can safely assume that there is very little positive to say.

We give some examples of commonly-occurring conversational implicatures. Humans often use an *if* statement to implicate an *if and only if* statement. Suppose A says to B, "If you teach me speech act theory I'll kiss you." In this case, if A did not mean *if and only if*, then A might kiss B whether or not B teaches A speech act theory. Then A would have been violating the maxim of quantity, telling the B something that did not contain any useful information.

In another common case, people often make a direct statement as a way to implicate that they believe the statement. When A says to B, "Austin was right," B is meant to implicate, "A believes Austin was right." Otherwise, A would have been violating the maxim of quality.

Finally, humans use a presupposition to implicate that the presupposition is true. When A says to B, "Grice's maxims are incomplete," A intends B to assume that Grice has axioms. Otherwise, A would have been violating the maxim of quality.

indirect speech act

Note that conversational implicatures enable *indirect* speech acts. Consider the classic Eddie Murphy skit in which his mother says to him, "It's cold in here, Eddie." Although her utterance is on the surface merely an informational locution, it is in fact implicating a request for Eddie to do something to warm up the room.

8.3.3 *A game-theoretic view of speech acts*

The discussion of speech acts so far has clearly been relatively discursive and informal as compared to the discussion in the other sections, and indeed to most of the book. This reflects the nature of the work in the field. There are advantages to the relative laxness; it enables a very broad and multifaceted theory. Indeed, quite a number of researchers and practitioners in several disciplines have drawn inspiration from speech act theory. But it also comes at a price, as the theory can be pushed only so far before the missing details halt progress.

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

One could look in a number of directions for such formal foundations. Since the definition of speech acts appeals to the mental state of the speaker and hearer, one could plausibly try to apply the formal theories of mental state discussed later in the book, and in particular theories of attitudes such as belief, desire and intention. Section 14.4 outlines one such theory, but also makes it clear that so-called BDI theories are not yet fully developed. Here we will explore a different direction. Our starting point is the fact that there are at least two agents involved in communication, the speaker and the hearer. So why not model this as a game between them, in the sense of game theory, and analyze that game?

Although this direction too is not yet well developed, we shall see that some insights can be gleaned from the game-theoretic perspective. We illustrate this via the phenomenon of *disambiguation* in language. One of the factors that render natural language understanding so hard is that speech is rife with ambiguities at all levels, from the phonemic through the lexical to the sentence and whole text level. We will analyze the following sentence-level ambiguity:

disambiguation

Every ten minutes a person gets mugged in New York City.

The intended interpretation is of course that every ten minutes some different person gets mugged. The unintended, but still permissible, alternative interpretation that the same person gets mugged over and over again. (Indeed, if one adds the sentence "I feel very bad for him," the implausible interpretation becomes the only permissible one.) How do the hearer and speaker implicitly understand which interpretation is intended?

One way is to set this up as a common-payoff game of incomplete information between the speaker and hearer (indeed, as we shall see, in this example we end up with a signaling game as defined in Section 8.2, albeit a purely cooperative one). The game proceeds as follows:

1. There exist two situations:
   - s: Muggings of different people take place at ten-minute intervals in NYC.
   - t: The same person is repeatedly mugged every ten minutes in NYC.
2. Nature selects between s and t according to a distribution known commonly to A and B.
3. Nature's choice is revealed to A but not to B.
4. A decides between uttering one of three possible sentences:
   - p: "Every ten minutes a person gets mugged in New York City."
   - q: "Every ten minutes some person or another gets mugged in New York City."
   - r: "There is a person who gets mugged every ten minutes in New York City."
5. B hears A, and must decide whether s or t obtain.

This is a simplified view of the world (more on this shortly), but let us simplify it even further. Let us assume that A cannot utter r when t obtains, and cannot utter q when s obtains (i.e., he can be ambiguous, but not deceptive). Let us

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

furthermore assume that when B hears either r or q he has no interpretation decision, and knows exactly which situation obtains (s or t, respectively).

In order to analyze the game, we must supply some numbers. Let us assume that the probability of s is much higher that that of t. Say, $P(s) = .99$ and $P(t) = .01$. Finally, we need to decide on the payoffs. We assume that this is a game of pure coordination, that is a common-payoff game. A and B jointly have the goal that B correctly have the right interpretation. In addition, though, both A and B have a preference for simple sentences, since long sentences place a cognitive burden on them and waste time. And so the payoffs are as follows: If the sentence used is p and a correct interpretation is reached, the payoff is 10. If either q or r are uttered (after which by assumption a correct interpretation is reached), the payoff is 7; and if an incorrect interpretation is reached the payoff is -10.

The resulting game is depicted in Figure 8.7.

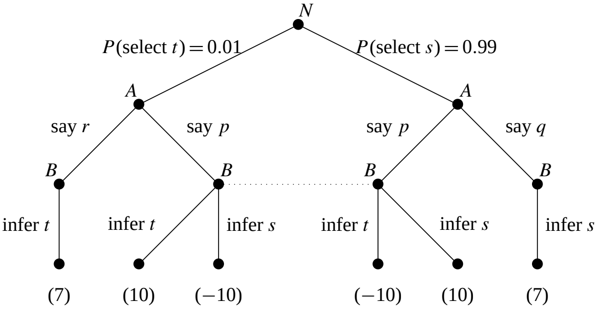

Figure 8.7 Communication as a signaling game.

What are the equilibria of this game? Here are two.

1. A's strategy: say q in s and r in t. B's strategy: When hearing p, select between the s and t interpretations with equal probability.
2. A's strategy: say p in s and r in t. B's strategy: When hearing p, select the s interpretation.

First, you should persuade yourself that these are in fact Nash equilibria (and even subgame-perfect ones, or, to be more precise since this is a game of imperfect information, sequential equilibria). But then we might ask, is there a reason to prefer one over the other? Well, one way to select is based on the expected payoff to the players. After all, this is a cooperative game, and it makes sense to expect the players to coordinate on the equilibrium with the highest payoff. Indeed, this would be one way to implement Grice's cooperative principle. Note that in the first equilibrium the (common) payoff is 7, while in the second equilibrium the expected payoff is $0.99 \cdot 10 + 0.01 \cdot 7 = 9.97$. And so it would seem that we have a winner on our hands, and a particularly pleasing one since this use of language accords well with real-life usage. Intuitively, to economize we use shorthand for commonly-occurring situations. This allows the hearer to make some default assumptions, but use more verbose language in the relatively rare situations in which those defaults are misleading.

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

This example can be extended in various ways. A can be given the freedom to say other sentences, and B can be given greater freedom to interpret them. Not only could A say q in s, but A could even say “I like cucumbers” in s. This is no less useful a sentence than p, so long as B conditions its interpretation correctly on it. The problem is of course that we end up with infinitely many good equilibria, and payoff maximization cannot distinguish between them. And so language can be seen to have evolved so as to provide *focal points* among these equilibria; the “straightforward interpretation” of the sentence is a device to coordinate on one of the optimal equilibria.

focal point

Although we are still far from being able to account for the entire pragmatics of language in this fashion, one can apply similar analysis to more complex linguistic phenomena, and it remains an interesting area of investigation.

8.3.4 *Applications*

The framework of speech-act theory has been put to practical use in a number of computer science and artificial intelligence applications. We give a brief description of some of these applications below.

Intelligent dialog systems

dialog system

One obvious application of speech act theory is a *dialog system*, which communicates with human users through a natural language dialog interface. In order to communicate efficiently and naturally with the user, dialog systems must obey the principles of human conversation, including those from Austin and Grice presented in this chapter.

TRAINS/TRIPS is a well-known dialog system, and is to assist the user in accomplishing tasks in a transportation domain. The system has access to information about the state of the transportation network, and the user makes decisions about what actions to take. The system maintains an ongoing conversation with the user about possible actions and the state of the network.

The TRAINS/TRIPS dialog system both uses and extends the principles of speech act theory. It incorporates a *Speech Act Interpreter*, which hypothesizes what speech acts the user is making, and a *Dialog Manager*, which uses knowledge of those acts to maintain the dialog. It extends speech act theory by creating a hierarchy of *conversation acts*, as shown in Table 8.1. As you can see, speech acts appear in this framework as the conversation acts that occur at the discourse level.

Workflow systems

Another useful application of speech act theory is in workflow software, software used to track and manage complex interactions within and between human organizations. These interactions range from simple business transactions to long-term collaborative projects, and each requires the involvement of many different human participants. To track and manage the interactions effectively, workflow software provides a medium for structured communications between all of the participants.

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

Discourse level	Act type	Sample acts
Multidiscourse	Argumentation acts	elaborate, summarize, clarify, convince
Discourse	Speech acts	inform, accept, request, suggest, offer, promise
Utterance	Grounding acts	initiate, continue, acknowledge, repair
Subutterance	Turn-taking acts	take-turn, keep-turn, release-turn, assign-turn

Table 8.1 Conversation acts used by the TRAINS/TRIPS system.

Many workflow applications are designed around an information processing framework, in which, for example, interactions may be modeled as assertions and queries to a database. This perspective is useful, but lacks an explicit understanding and representation of the pragmatic structure of human communications. An alternative is to view each communication as an illocutionary speech act, which states an intention on the part of the sender and places constraints on the possible responses of the recipient. Instead of generic messages, as in the case of email communications, users must choose from a set of communication types when composing messages to other participants. Within this framework, they can write freely. For example, when responding to a request, users might be given the following options.

- Acknowledge
- Promise
- Free form
- Counter offer
- Commit-to-commit
- Decline
- Interim report
- Report completion

The speech act framework confers a number of advantages to developers and users of workflow software. Because the basic unit of communication is a conversation, rather than a message, the organization of communications is straightforward, and retrieval simple. Furthermore, the status and urgency of messages is clear. Users can ask "In which conversations is someone waiting for me to do something?" or "In which conversations have I promised to do things?". Finally, access to messages can be organized and controlled easily, depending on project involvement and authorization levels. The downside is that it involves additional overhead in the communication, which may not be justified by the benefits, especially if the conversational structures implemented in the system do not capture well the rich set of communications that takes place in the workplace.

Agent communication languages

Perhaps the most widespread use of speech act theory within the field of computer science is for communication between software applications. Increasingly, computer systems are structured in such a way that individual applications can act

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

as agents (e.g., with the popularization of the Internet and electronic commerce), each with its own goals and planning mechanisms. In such a system, software applications must communicate with each other and with their human users to enlist the support of other agents to achieve goals, to commit to helping another agent, to report their own status, to request a status report from another, and so on.

Not surprisingly, several proposals have been made for artificial languages to serve as the medium for this interapplication communication. A relatively simple example is presented by *Knowledge Query and Manipulation Language (KQML)*, which was developed in the early 1990s. KQML incorporates some ideas from speech-act theory, especially the idea of performatives. It has a built-in set of performatives, such as ACHIEVE, ADVERTISE, BROKER, REGISTER, and TELL.

KQML

The following is an example of a KQML message, taken from a communication between two applications operating in the blocks world domain.

```
(tell
       :sender    Agent1
       :receiver  Agent2
       :language  KIF
       :ontology  Blocks-World
       :content   (AND (Block A) (Block B) (On A B)))
```

Note that the message is a performative. The content of the message uses blocks world semantics, which are completely independent of the semantics of the performative itself.

KQML is no longer an influential standard, but the ideas of structured interactions among software agents that are based in part on speech acts live on in more modern protocols defined on top of abstract markup languages such as *XML* and the so-called *Semantic Web*.

XML

Semantic Web

Rational programming

We have described how speech act theory can be used in communication between software applications. Some authors have also proposed to use it directly in the development of software applications, that is, as part of a programming language itself. This proposal is part of a more general effort to introduce elements of rationality into programming languages. This new programming paradigm has been termed *rational programming*. Just as object-oriented programming shifted the paradigm from writing procedures to creating objects, rational programming shifts the paradigm from creating informational objects to creating motivational agents.

rational programming

So where does communication come in? The motivational agents created by rational programming must act in the world, and because the agents are not likely to have a physical embodiment, their actions consist of sending and receiving signals; in other words, their actions will be speech acts. Of course, as shown in the previous section, it is possible to construct communicating agents within existing programming paradigms. However, by incorporating speech acts as primitives,

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

rational programming constructs make such programs more powerful, easier to create, and more readable.

We give a few examples for clarity. *Elephant2000* is a programming language described by McCarthy which explicitly incorporates speech acts. Thus, for example, an Elephant2000 program can make a promise to another and cannot renege on a promise. The following is an example statement from Elephant2000, taken from a hypothetical travel agency program:

```
if ¬ full (flight)
   then accept.request(
        make (commitment (admit (psgr, flight))))
```

The intuitive reading of this statement is “if a passenger has requested to reserve a seat on a given flight, and that flight is not full, then make the reservation.”

Agent-Oriented Programming (AOP)

Agent-Oriented Programming (AOP) is a separate proposal that is similar in several respects. It too embraces speech acts as the form and meaning of the communication among agents. The most significant difference from Elephant2000 is that AOP also embraces the notion of mental state, consisting of beliefs and commitments. Thus the result of an inform speech act is a new belief. AOP is not actually a single language, but a general design that allows multiple languages; one particular simple language, *Agent0* was defined and implemented. The following is an example statement in Agent0, taken from a print server application.

```
IF
   MSG COND: (?msgId ?someone REQUEST
              ((FREE-PRINTER 5min) ?time))
   MENTAL COND:
             ((NOT (CMT ?other (PRINT ?doc (?time+10min))))
              (B (FRIENDLY ?someone)))
THEN COMMIT
   ((?someone (FREE-PRINTER 5min) ?time)
    (myself (INFORM ?someone (ACCEPT ?msgId)) now))
```

The approximate reading of this statement is “if you get a request to free the printer for five minutes at a future time, if you are not committed to finishing a print job within ten minutes of that time, and if you believe the requester to be friendly, then accept the request and tell them that you did.”

8.4 History and references

The literature on language and natural language understanding is of course vast and we cannot do it justice here. We will focus on the part of the literature that bears most directly on the material presented in the chapter.

Two early seminal discussions on cheap talk are due to Crawford and Sobel [1982] and Farrell [1987]. Later references include Rabin [1990] and Farrell [1993]. Good overviews are given by Farrell [1995] and Farrell and Rabin [1996].

The literature on signaling games dates considerably farther back. The Stackelberg leadership model, later couched in game-theoretic terminology (as we do

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

in the book), was introduced by Heinrich von Stackelberg, a German economist, as a model of duopoly in economics [von Stackelberg, 1934]. The literature on information economics, and in particular on asymmetric information, continued to flourish, culminating in the 2001 Nobel Prize awarded to three pioneers in the area (Akerlof, Spence, and Stiglitz). The Spence signaling game, which we cover in the chapter, appeared in Spence [1973].

Austin's seminal book, *How to Do Things with Words*, was published in 1962 [Austin, 1962], but is also available in a more recent second edition [Austin, 2006]. Grice's ideas were developed in several publications starting in the late 1960s, for example, Grice [1969]. This and many other of his relevant publications were collected in Grice [1989]. Another important reference is Searl [1979]. The game-theoretic perspective on speech acts is more recent. The discussion here for the most part follows Parikh [2001]. Another recent reference covering a number of issues at the interface of language and economics is Rubinstein [2000].

The TRAINS dialog system is described by Allen et al. [1995], and the TRIPS system is described by Ferguson and Allen [1998]. The speech-act-based approach to workflow systems follows the ideas of Winograd and Flores [1986] and Flores et al. [1988]. The KQML language is described by Finin et al. [1997]. The term *rational programming* was coined by Shoham [1997]. Elements of the Elephant2000 programming language are described by McCarthy [1994]. The AOP framework was described by Shoham [1993].

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009

Downloaded from https://www.cambridge.org/core. University of Warwick, on 10 Jan 2019 at 08:28:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511811654.009