3 | Joint actions

Joint activities advance mostly through joint actions. In buying items in a drugstore, a customer joins a server in opening the transaction, settling on the items wanted, establishing the price, exchanging money, and closing. In a chess game, the players join in specifying discrete moves from the opening of the game to the checkmate. Joint actions like these belong to an extended family of actions that also includes moving together in waltzing, playing notes together in a string quartet, paddling in unison in a canoe, and passing a ball in soccer or basketball. It also includes asking questions, making requests, making assertions, making references – much of what we think of as language use.

What makes an action a joint one, ultimately, is the coordination of individual actions by two or more people. There is coordination of both *content*, what the participants intend to do, and *processes*, the physical and mental systems they recruit in carrying out those intentions. When Ann and Ben paddle a canoe together, they coordinate on their plans – the content of what they do. Overall, they aim to reach the spit of land on the other side of the lake as efficiently as possible, with Ann in front and Ben in the rear. At any moment, they aim to stay on course, with Ann pulling on one side and Ben on the other. Ann and Ben also coordinate on their physical and mental processes. They pull their paddles in rhythm and with a force adjusted to keep them on course; if Ann changes sides, so does Ben; if Ann stops, so does Ben. In joint actions, the processes recruited depend on the plans, and the plans chosen depend on the processes available. If a log drifts in front of their canoe, Ann and Ben both adjust their processes to avoid it, then return to their course. Joint actions cannot be accounted for without understanding the interplay between content and process, and their place in overall joint activities.

Joint actions with language are no different. They too require the coordination of actions – with all that that requires. Although this may be a truism, it is a truism widely ignored. Some of the basic principles of language use are really general principles of joint action, and to understand language use, we must look to the broader principles.

Individual and joint actions

Joint actions pose a paradox. Recall that intentional actions divide into two types-individual and joint actions (Chapter 1). Ann is performing individual actions when she plays a flute, paddles a kayak, or shakes a stick. Ann and Ben are performing joint actions when they play a flute–piano duet, paddle a canoe together, or shake hands. Individual actions are performed by individual people, and joint actions, by ensembles of people. Clearly, there is no agent named Ann-and-Ben who decides "Ah, I am now going to play this duet" and then plays it. Ensembles of people don't intend to do things. Only individuals do (Clark and Carlson, 1982a, b). Yet ensembles of people play duets, paddle canoes, shake hands, and do other things individuals cannot do alone. The paradox is this: An ensemble can do things that it cannot intend to do.

The paradox dissolves once we see that joint actions have individual actions as parts. In Ann and Ben's flute and piano duet, there are three distinct actions:

0. the ensemble Ann-and-Ben plays the duet (a joint action)
1. Ann plays the flute part as part of 0 (an individual action by Ann)
2. Ben plays the piano part as part of 0 (an individual action by Ben)

The joint action in 0 is performed by means of the individual actions in 1 and 2. These individual actions are of a special type (Chapter 1). When Ann plays alone in the privacy of her living room, she doesn't coordinate her actions with anyone else. They are *autonomous actions*. But when she plays the flute part as part of the duet, as in 1, she performs actions as a means of participating with Ben in playing the duet. These I have called *participatory actions*. Joint actions can only be performed by means of participatory actions – by the individual participants each doing their parts. So we can denote a joint action by A and B as a joining of two participatory actions, $part_1(A)$ and $part_2(B)$, as here:

joint[$part_1$(A), $part_2$(B)]

Ann and Ben's duet becomes: joint[Ann plays flute part, Ben plays piano part].

Autonomous and participatory actions are distinguished by the intentions behind them. Here is one way to characterize individual actions:

Individual A is doing individual action *k* if and only if:
0. the action *k* includes 1;
1. A intends to be doing *k* and believes that 0.

For Ann to be playing a flute piece alone in her living room, she must intend to be playing that piece and believe she has those intentions. Joint actions look different:

Ensemble A-and-B is doing joint action *k* if and only if:
0. the action *k* includes 1 and 2;
1. A intends to be doing A's part of *k* and believes that 0;
2. B intends to be doing B's part of *k* and believes that 0.

For Ann to be playing her part of the flute–piano duet, she must intend to be playing her part, and believe she has these intentions and that Ben has the parallel intentions and beliefs. With participatory acts, Ann does what she does only in the continuing belief that Ben is intending to do his part.[1]

Joint actions must be distinguished from *adaptive* and *deceptive* actions. Consider this series of actions:

1. In a dart game, A throws a dart at a stationary dart board B.
2. In an arcade game, A shoots a pellet at a moving mechanical duck B.
3. As a spy, A shadows an unwary B through San Francisco.
4. In a game of catch, A throws a ball for B to catch.
5. In tennis, A tries to hit a ball past B.

In all five descriptions, A takes actions with respect to B based on where she predicts B will be. In 1 and 2, her prediction is based on the mechanical properties of dart boards and mechanical ducks. In 3, it is based on what B would do autonomously. In 4, it is based on what B would do in trying to coordinate with A. And in 5, it is based on what B would do believing he thought she was trying to deceive him. Of these, only 4 is a genuine joint action, in which A and B converge on a mutually desired outcome.

[1] For discussions of intentions in joint actions, see Grosz and Sidner (1990), Searle (1990), Tuomela (1996), Tuomela and Miller (1988).

In 3, spy A adapts unilaterally to B's actions, and in 5, tennis player A actively deceives her opponent B – a type of anti-coordination. Coordination is different from both adaptation and deception. Our primary concern is coordination.

Coordination

Joint actions are created when people coordinate with each other. Why should they coordinate? The reason, according to Thomas Schelling (1960), is to solve *coordination problems*. Two people have a coordination problem whenever they have common interests, or goals, and each person's actions depend on the actions of the other. To reach their goals, they have to coordinate their individual actions in a joint action. In this view, joint actions are created from the goal backward. Two people realize they have common goals, realize their actions are interdependent, and work backward to find a way of coordinating their actions in a joint action that will reach those goals. It was David Lewis' (1969) insight that language use is really people solving coordination problems. In our terms, it is a complex of joint activities. If Lewis is right, we should learn a great deal about language use from studying these problems. Let us begin with Schelling's analysis.

SCHELLING GAMES

There are many situations in which two people's actions are interdependent and their interests, or goals, are identical. Schelling studied these situations by devising a variety of one-shot problems I will call Schelling games. In each game, two people give their solutions to the same problem, but without consulting each other. Here are four Schelling games:

1. *Coin*. Name "heads" or "tails." If you and your partner name the same, you both win a prize.

2. *Numbers*. Circle one of the numbers listed here. You win if you both succeed in circling the same number:

7 100 13 261 99 555

3. *Meeting*. You are to meet somebody in New York City. You have not been instructed where to meet; you have no prior understanding with the person on where to meet; and you cannot communicate with each other. You are simply told that you will have to guess where to meet and that he is being told the same thing and that you will just have to try to make your guesses coincide. You were

told the date but not the hour of the meeting; the two of you must guess the exact minute of the day for the meeting. At what time will you appear at the meeting place that you elected?

4. *Money*. You are to divide $100 into two piles, labeled A and B. Your partner is to divide another $100 into two piles labeled A and B. If you allot the same amounts to A and B, respectively, that your partner does, each of you gets $100; if your amounts differ from his, neither of you gets anything.

When Schelling got about forty people to play these games, there was a surprising agreement in their responses. For the coin game, 86 percent of them said "heads." For the numbers game, 90 percent selected one of the first three numbers; 7 and 100 were the most popular. For the meeting game – the players were all from New Haven – "an absolute majority" suggested the information booth at Grand Central Station, and "virtually all" would go there at noon. For the money game, 88 percent of the players put $50 in pile A and $50 in pile B. As Schelling pointed out, the players in each game had little to go on: In principle, any solution was as good as any other. Still, they managed to win most of the time.

These are problems of *pure coordination*, where the two partners' interests coincide completely. But, as Schelling argued, the same factors apply even when the two partners' interests diverge, so long as they don't diverge too much. Here are two more Schelling games:

1'. *Unequal coin.* A and B are to choose "heads" or "tails" without communicating. If both choose "heads," A gets $3 and B gets $2; if both choose "tails," A gets $2 and B gets $3. If they choose differently, neither gets anything. You are A (or B); which do you choose?

4'. *Unequal money*. You and your partner are to be given $100 if you can agree on how to divide it without communicating. Each of you is to write the amount of his claim on a sheet of paper; and if the two claims add to no more than $100, each gets exactly what he claimed. If the two claims exceed $100, neither of you gets anything. How much do you claim?

For the unequal coin game, the two partners should still converge on "heads," since otherwise they both lose money. They did: 73 percent of the A's and 68 percent of the B's chose "heads" (compared to 86 percent in the original game). For the unequal money game, the goals shouldn't change either, and 90 percent of the players split the money fifty-fifty (compared to 88 percent in the original game). All it takes to be a coordination problem, as Lewis (p. 24) put it, is that "coincidence of interest predominates."

Everyday coordination problems are more varied than these examples suggest. They vary in number of possible solutions (from two to infinity), number of participants (from two to entire communities), what is at stake (from minor incivilities to nuclear war), and coincidence of interest (from partial to complete). They may be discrete, like the one-shot Schelling games, but more often they are continuous, like playing duets, paddling a canoe, or conversing, and that complicates matters immensely. Despite their differences, all these problems share certain characteristics. One of these is the coordination of expectations.

COORDINATION DEVICES

What does it take to solve coordination problems? It isn't enough, as Schelling noted, simply to predict what one's partner will do, since the partner will do what he or she predicts the first will do, which is whatever the first predicts that the partner predicts the first to do, and so on *ad infinitum*. Schelling argued:

> What is necessary is to coordinate predictions, to read the same message in the common situation, to identify the one course of action that their expectations of each other can converge on. They must "mutually recognize" some unique signal that coordinates their expectations of each other. (Schelling, p. 54)

Schelling went on:

> Most situations – perhaps every situation for people who are practiced at this kind of game – provide some clue for coordinating behavior, some focal point for each person's expectation of what the other expects him to expect to be expected to do. Finding *the* key, or rather finding *a* key – any key that is mutually recognized as the key becomes *the* key – may depend on imagination more than on logic; it may depend on analogy, precedent, accidental arrangement, symmetry, aesthetic or geometric configuration, casuistic reasoning, and who the parties are and what they know about each other. (p. 57)

With Lewis, I shall call such a focal point, or key, a *coordination device*.

The six Schelling games illustrate several devices. Of heads and tails, heads seems more prominent because one says "heads or tails" or perhaps because fronts are more salient than backs. In the number problem, 7 has a certain prominence (the only single digit, the smallest number), and so do 100 (a standard round number) and 13 (a common unlucky number), but the other numbers don't. In New York, Grand Central Station is a place many people outside New York pass through and even use as a meeting place; of the two most conspicuous times of the

day, noon is more sensible for meeting. And of various money splits, fifty-fifty is an obviously unique solution, since it is symmetrical for A and B.

Coordination devices range even more widely. When you and I want to meet, we can meet in Jordan Hall at eight on the basis of an *explicit agreement*, or on the basis of *precedent* – that's when and where we met last week. We can meet for a seminar in Room 100 at noon on the basis of a *convention* – that's when and where the seminar conventionally meets. If we lose each other wandering through the Tate Gallery, we could meet at van Gogh's *Self-portrait* on the basis of a prior conversation about your coming especially to see it, or at the entrance on the basis of its uniqueness as a location in the gallery. I can meet you getting off a plane, without knowing what you look like, by having you wear a carnation in your lapel. Or, as has happened to me, we could meet on the basis of signs of personal uncertainty: You and I look for a passenger and a reception party who are obviously looking for a reception party and passenger they don't know.

What does all this have to do with joint actions? When Ann and Ben pick "heads" in the coin game, they are performing a joint action. They are each performing individual actions as parts of an action by the pair of them, denoted as follows:

Joint [Ann picks "heads," Ben picks "heads"]

The coordination device – the prominence of "heads" – is what enables them to choose the right participatory actions to perform.

JOINT SALIENCE

What coordination devices do is give the participants a rationale, a *basis*, for believing they and their partners will converge on the same joint action. These rationales can, in principle, come from any source so long as they lead to a unique solution. What they must do, as Schelling put it, is enable the participants to form a "mutual expectation" about the individual actions each participant will take.[2]

What is a mutual expectation? Intuitively, it is a type of shared belief. To describe it, we need the technical notion of common ground I take up in Chapter 4. For now, I will define it this way:

[2] Kraus and Rosenschein (1992; Fenster, Kraus, and Rosenschein, 1995) have studied automated procedures for identifying focal points in a limited set of domains.

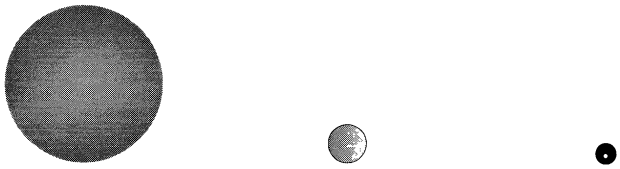
For two people *A* and *B*, it is common ground that *p* if and only if:

1. A and B have information that some basis *b* holds;
2. *b* indicates to A and B that A and B have information that *b* holds;
3. *b* indicates to A and B that *p*.

Suppose you and I agree to meet in Jordan Hall at eight. That agreement is a basis *b* for a certain piece of common ground. You and I each have information that the agreement holds. The agreement indicates to each of us that we both have information that it holds. Finally, it indicates to you and me that we each expect to go to Jordan Hall at eight. We can conclude: It is common ground for you and me that we each expect to go to Jordan Hall at eight. A mutual expectation is a mutual belief or supposition (a part of common ground) about what the participants will do.

It is mutual expectations like this that enable an ensemble of people to perform a joint action. Take meeting in Jordan Hall at eight. I won't believe I am taking part in that joint action unless I believe you are intending to go to Jordan Hall at eight too, and the same logic holds for you. But once we are armed with the mutual expectation, I can do my part (going to Jordan Hall at eight) in the belief that you are doing yours (going to Jordan Hall at eight), and you can do your part in the belief that I am doing mine. It is only with that mutual belief that we both believe we are taking part in the same joint action – meeting in Jordan Hall at eight.

Common ground, then, is a prerequisite for coordination – for joint actions.[3] The point is easy to demonstrate. Suppose I am asked, in a Schelling game, to choose one of these three balls:



basketball baseball squash ball

Which should I choose? That depends on my assumptions about my partner.

Case 1. I don't know who my partner is. I therefore choose the basketball, reasoning: "My partner and I can take for granted that the basketball is perceptually the most salient one for any two humans.

[3] As Schelling noted, an effective coordination device will have "some kind of prominence or conspicuousness." "But," he went on, "it is a prominence that depends on time and place and who the people are."

Since I know nothing else about my partner, and I suppose he or she knows that, we must rely on perceptual salience alone."

Case 2. My partner and I are old friends; indeed, we arrived together to play the Schelling game. Further, she and I play squash three times a week (and not basketball or baseball). I therefore choose the squash ball. I reason: "We mutually know we play squash, making the squash ball especially salient for the two of us. The basketball may be the most salient perceptually, but that salience isn't unique to the two of us. That makes the squash ball the solution of choice."

Case 3. My partner is the same as in 2, and I have been told that she is my partner, but that she doesn't know I am her partner. I therefore choose the basketball, reasoning: "Even though she and I mutually know we play squash, that cannot guide her choice since she has no idea I am her partner. The game reduces to case 1, hence the choice of the basketball."

It is easy to show that, with other configurations of common ground, I should always make my choice against what I take to be my partner's and my current common ground, and this is just what people do (Clark, Schreuder, and Buttrick, 1983; see also Clark and Marshall, 1981).

The ideal solution to a coordination problem, then, isn't the solution that is most salient *simpliciter*. It is the solution that is the most salient with respect to the participants' current common ground. The principle is this:

Principle of joint salience. The ideal solution to a coordination problem among two or more agents is the solution that is most salient, prominent, or conspicuous with respect to their current common ground.

(For short, I will use *joint salience* to mean "salience with respect to the participants' current common ground.") Not that two people will always agree on what is jointly the most salient. They may have discrepant conceptions of their current common ground, or of the most salient solution in it. But people are sensitive to potential discrepancies (Clark, Schreuder, and Buttrick, 1983) and adept at managing those that arise (Chapter 8). Still, they should strive for the ideal – within limits – for they can take for granted that their partners are striving for the ideal too. This way they reduce the possibility of miscoordination.

Participant coordination problems

The standard Schelling game ("Name heads or tails") is a *third-party* coordination problem. Two partners, say Ann and Ben, are given a

problem by a third party or by nature. It is never specified who the third party is, or what his or her motives are, but these can be critical. The third party may know Ann and Ben well and have given them a problem they would find easy. But for all they know, the third party may be diabolical and have given them an unsolvable problem ("Choose 59 or 83 or 71").

Participant coordination problems are fundamentally different, as when Ann poses a coordination problem for Ben and herself. For such a problem, Ben can reason: "Ann, being rational, must want to win and expect me to want to win too. Since she had leeway in her choice of problem, I assume she has chosen one she believes has a unique solution that we can converge on. Furthermore, she should think I will reason this way." And she should. If so, Ann and Ben have four additional premises they can use in solving the problem:

The solvability premises. In a coordination problem set by one of its participants, all of the participants can assume that the first party:

1. chose the problem,
2. designed its form,
3. has a particular solution in mind, and
4. believes the participants can converge on that solution.

These are premises Ann and Ben couldn't take for granted for third-party coordination problems.

Riddles and puzzles, for example, differ in solvability. Modern riddles have solutions that their creators don't expect solvers to discover, as here (from Augarde, 1986):

Ann: When is a thought like the sea?
Ben: (after thinking a bit) I don't know. When?
Ann: When it's a notion.

Riddles aren't participant coordination problems precisely because they violate the solvability premises. Puzzles, in contrast, have solutions their creators *do* expect solvers to discover, as here (from Smullyan, 1978):

Ann: Twenty-four red socks and twenty-four blue socks are lying in a drawer in a dark room. What is the minimum number of socks I must take out of the drawer that will guarantee that I have at least two socks of the same color?
Ben: (after working out the answer) Three.

So puzzles fulfill the solvability premises. And whereas riddles take three steps (Ann, Ben, Ann), puzzles take only two (Ann, Ben).

The solvability premises have an important corollary. When Ann presents her puzzle to Ben, she specifies twenty-four red socks, twenty-four blue socks, a dark room, and other information. The two of them assume this is all Ben needs to solve it. If they didn't, the puzzle wouldn't be solvable. Ben cannot add convenient assumptions: "Let me assume I can turn on the light. So the answer is two." Or: "Let me assume there is also a pair of Argyle socks in the drawer. So the answer is four." Nor should Ann assume Ben will do this. If he were allowed to, the puzzle would no longer have a unique solution. The only information Ann and Ben can add is information from their common ground that is consistent with the principle of joint salience.[4] The assumption they make is this:

The sufficiency premise. In a coordination problem set by one of its participants, the participants can assume that the first party has provided all the information they need (along with the rest of their common ground) for solving it.

The solvability and sufficiency premises are merely corollaries of the principle of joint salience as applied to participant coordination problems.

Some participant coordination problems have an added constraint: They come in sequences so that the participants have to coordinate not only on the solution, but also on *when* to present the solution. In such a situation, Ann won't give the socks puzzle, because she cannot know how long Ben will take – an hour, two minutes, thirty seconds. The completion time, to be predictable, must itself satisfy joint salience, solvability, and sufficiency. It cannot be twenty seconds, ten seconds, or five seconds, for these aren't unique solutions. It must be effectively zero, or immediate. If it weren't, there would be unpredictable delays that would get compounded on the next problem:

The immediacy premise. In a coordination problem set by one of its participants in a time-constrained sequence of problems, the participants can assume that they can solve it immediately – with effectively no delay.

So for a time-limited problem posed by Ann, Ben can assume she designed it so he would solve it immediately, readily, in no time at all.

It is crucial, therefore, who the coordination problem is set by and why. Joint salience applies whatever the problem. Solvability and sufficiency can be assumed for problems set by a participant. And immediacy can be assumed for problems that must be solved in a predictable

[4] See McCarthy's (1980, 1986) characterization of circumscription in so-called non-monotonic reasoning.

interval of time. By now, it should be clear why participant coordination problems are of such interest: They are the form most coordination takes in language use.

Conventions

Of all the coordination devices I have noted, two are uniquely suited for solving coordination problems. One is explicit agreement. When you and I agree to meet at Jordan Hall at eight, we do so to solve the problem of when and where to meet. Indeed, explicit agreements generally preempt other potential coordination devices. If you and I agree to choose tails on the next coin problem, our agreement takes precedence over the usual rationale for choosing heads. The other coordination device *par excellence* is convention.

Tom, Dick, and Harriet have a recurrent coordination problem: They want to meet for lunch every Tuesday – a joint action. Week after week, they agree to meet at the faculty club at 12:15. After a while, they no longer have to say when and where they are to meet. They each simply go to the faculty club at 12:15 – their participatory actions – because that is what they mutually expect each other to do based on the regularity in their recent behavior. What they have evolved is a convention, and that is now the device by which they coordinate their meeting – by which they carry out their joint action.

A convention, according to Lewis (1969), is a community's solution to a recurrent coordination problem. In some societies, bowing is a solution to the recurrent problem of how to greet one other; in others, it is shaking hands. In America and Europe, placing knives, forks, and spoons on the table is a solution to the recurrent problem of what utensils to use in eating. In China and Japan, it is to place chopsticks. In North America, leaving a tip at the table in a restaurant is a solution to the recurrent problem of how to help pay the waiter or waitress. In Europe, it is to include the tip in the bill. Conventions come in many forms – for large and small communities, for simple and complex problems.

What makes something a convention? According to Lewis, it has these five properties:[5]

[5] I have updated Lewis' (1969) account slightly to deal with minor problems noted by Burge (1975), Gilbert (1981, 1983), and others. For consistency, I have also changed Lewis' "population P" to "community C" and "common knowledge" to "common ground" and simplified his formulation in other ways. See Lewis for the full story.

A convention is:

1. a regularity *r* in behavior
2. partly arbitrary
3. that is common ground in a given community C
4. as a coordination device
5. for a recurrent coordination problem *s*.

Take greeting. When any two old friends meet, they have a recurrent coordination problem of how to greet. In some American communities, the solution is for two men to shake hands and for a man and woman, or two women, to kiss each other once on the cheek. These actions constitute a regularity *r* in behavior. They are a coordination device that solves the recurrent coordination problem of how to greet. The regularity is common ground for the members of those communities. And it is partly arbitrary, for it could have been different; in other communities, two men hug; in still others, two people kiss two, or three, times.[6] I say "partly" because the options available may be constrained. In greetings, the available options may exclude slapping or kicking, actions that hurt or injure.

Most conventions don't evolve as Tom, Dick, and Harriet's did. Shaking hands with the right hand, for example, didn't evolve for just me and the people I met. It was already in use in my culture when I learned it. Most conventions are arbitrary in being accidents of history: It is an accident of history that we shake hands with the right hand. If history had been different, we could be using the left.[7] Becoming a member of a community means in part acquiring the conventions in that community that were already in place.

Most conventions belong to systems. In every culture, for example, the problem of greeting people face to face has evolved a system of solutions. Here is a fragment of one system:

[6] Still, as Tyler Burge (1975) argued, it needn't be common ground in a community that a convention has alternatives. If people thought *cockadoodledoo* was the only way one could express a rooster's crow, the word would be no less conventional for that.

[7] It is really an empirical question for each regularity in behavior whether it could have been different. Right-handedness, for example, may be so strong that it dominates all other interests in coordinating on shaking hands, making the choice of the right hand not a true convention. Most conventions vary across cultures, thereby demonstrating their historical arbitrariness.

Situation *s*	**Joint action** *r*
gender: man and woman; two women *relationship*: intimates	A and B hug
gender: man and woman; two women *relationship*: acquainted equals	A and B kiss each other once on the right cheek
gender: two men *relationship*: unacquainted equals *introduction*: by oneself or third party	A and B shake hands
gender: man and woman; two women *relationship*: unacquainted *introduction*: formal, by third party	A and B exchange "How do you do?"
gender: man and woman; two women *relationship*: unacquainted *introduction*: informal, by third party	A and B exchange "Hello"

In this example, the recurrent coordination problem – the situation *s* – is partitioned into five mutually exclusive classes, each with a different solution – a different joint action *r*. The system is so tightly constrained that it may be impossible to change one convention without changing others. If hugging were broadened to new situations, kissing, shaking hands, and the rest would have to be narrowed. And there are probably links between related conventions. Is it accidental that we shake the right hand and kiss the right cheek? A change in one might induce a change in the other.

Conventions, Lewis argued, aren't habits or practices. All the same, they seem to be maintained in part by habits and practices. Shaking hands with the right hand remains intact partly because it has become habitual for people to extend their right hand when shaking hands. And when the practice of men wearing hats disappeared, so did the convention of men greeting women by tipping their hats. How are conventions maintained? This is surely related to the processes by which people coordinate with each other, an issue we will return to.

Coordination in language use

In discourses, as in other joint activities, the participants advance their interests by creating joint actions as solutions to coordination problems. They create entire joint activities when faced with such coordination problems as how to plan a party, complete a business transaction, get a

story told, or exchange gossip. At each step in these activities, they create smaller joint actions as solutions to smaller coordination problems, such as how to make and accept offers, how to speak and be understood, and who is to speak when. These problems in turn divide into smaller coordination problems, leading to more local joint actions. Discourses emerge as solutions to hierarchies of coordination problems. If this is right, people should exploit the same coordination devices inside discourses as outside them, and they do.

In language use, a central problem is coordinating what speakers mean and what their addressees understand them to mean. These are really participant coordination problems – Schelling games set by speakers for their addressees and themselves to solve. Their solutions should therefore reflect joint salience, solvability, and sufficiency: Speakers and addressees should take for granted, within limits, that speakers have in mind unique solutions they believe their addressees will converge on. To see this, let us examine an analysis of signaling systems by David Lewis (1969).

SIGNALING SYSTEMS

As a model situation, Lewis drew from a legend of the American Revolutionary War about Paul Revere riding through the Massachusetts countryside to warn everyone that the redcoats – the British – were coming.[8] The scene Lewis chose has two participants, the sexton of the Old North Church and Paul Revere – a speaker and an addressee. The sexton acts according to one contingency plan:

If the redcoats are observed staying home, hang no lantern in the belfry.
If the redcoats are observed setting out by land, hang one lantern in the belfry.
If the redcoats are observed setting out by sea, hang two lanterns in the belfry.

And Revere acts according to another:

If no lantern is observed hanging in the belfry, go home.
If one lantern is observed hanging in the belfry, warn the countryside that the redcoats are coming by land.
If two lanterns are observed hanging in the belfry, warn the countryside that the redcoats are coming by sea.

[8] The legend is best known from Henry Wadsworth Longfellow's poem "Paul Revere's Ride" (1861), which every American schoolchild used to know by heart.

The sexton's contingency plan is a function *Fs* from states of affairs to signals – observable actions – and Revere's is a function *Fr* from observable signals to responses he could take. For Revere and the sexton to succeed, they need to coordinate contingency plans, and they do just that with their choice of plans.

Revere and the sexton have created a *signaling system*. They begin with a coincidence of goals: Both want Revere's response to be appropriate to the state of the British army as the sexton sees it. As Lewis put it, "Each agent will be acting according to the contingency plan that is best given the other's contingency plans and any state of affairs." A signaling system is a combination <*Fs, Fr*> that achieves "the preferred dependence of the audience's response upon the state of affairs."

Signaling systems are ideal for coordinating what speakers mean with what their addressees understand them to mean. By hanging one lantern in the belfry, the sexton *meant* that Revere was to warn the countryside about the redcoats coming by land. Since he believed the signaling system to be common ground for the two of them, he could use one lantern and count on Revere to recognize what he meant. As Lewis pointed out, all this can be said without any mention of the meaning of the signals themselves – for example, that one lantern meant that the redcoats were coming by land. "But nothing important seems to have been left unsaid, so what has been said must somehow imply that the signals have their meanings." What one lantern means is a consequence of the pairing of the sexton's and Revere's contingency plans. This anticipates a point I will return to in Chapter 5: Speaker's meaning is primary, and signal meanings derivative.

When the sexton hangs out a single lantern, he is posing a *participant* coordination problem. Revere can assume the sexton (1) chose the problem, (2) designed its form of presentation, (3) had a particular solution in mind, and (4) believed he and Revere would converge on that solution. He didn't design it to be solvable by just anyone. He might even have devised it to confound British spies. Coordination in language use is like this. When Ann tells Ben "Bob went out with Monique last night," she expects to be understood by Ben, but not by just any overhearer. Most overhearers wouldn't know who Bob and Monique were. If Ann said "You-know-who did you-know-what with you-know-who last night," she would be posing a coordination problem unsolvable by anyone not privy to the special common ground she shares with Ben (Clark and Carlson, 1982a; Clark and Schaefer, 1987b, 1992).

Signaling systems are therefore bases for joint actions. Revere's and the sexton's contingency plan gives them a rationale for this joint action:

Joint [the sexton hangs one lantern in belfry, Revere takes the sexton to mean that the redcoats are coming by land]

As in any joint action, Revere and the sexton each take individual actions in the belief that each of them is doing so as part of a joint action by the pair of them. So what for Lewis is an account of coordination in language use is for us also an account of joint actions in language use. One is the basis for the other.

"It is not at all necessary," Lewis noted, "to confine ourselves to conventional signaling systems in defining meaning for signals." It didn't matter that Revere and the sexton came to their signaling system by explicit agreement. One lantern in the belfry still meant that the redcoats were coming by land. Signaling systems can be based on explicit agreement, precedent, salience, convention – on any coordination device that works. Naturally occurring signaling systems exploit all types of coordination devices.

CONVENTIONS AND LANGUAGE

Languages like English are conventional signaling systems *par excellence*. Most English speakers, for example, have contingency plans that include this pairing of conditionals, which I will call a *signaling doublet*:

Speaker: If you intend to denote the cipher naught, you can utter the word *zero*.
Addressee: If a speaker utters the word *zero*, he or she can be denoting the cipher naught.

This doublet happens to be conventional. It is a regularity in behavior – when people want to denote naught, they can use *zero*, and others can understand them to be denoting naught. It is a coordination device for a recurrent coordination problem – speakers wanting to denote naught and their addressees wanting to recognize this. As a coordination device, it is common ground in the community of English speakers (not Japanese or Navaho speakers). And it is arbitrary – another doublet (like *null* for "naught") might have evolved instead if the history of English had been different. In de Saussure's classic *Cours de linguistique générale* (1916), he called such a doublet a *linguistic sign* and argued that "the lin-

guistic sign is arbitrary."[9] So just as the Old North Church signaling system has doublets, so does English. It is just that English has many more, organized in a complex system (Lewis, 1969).

Conventional doublets in language use come in many guises. Here are four broad categories:

Lexical entries. Many doublets are treated as lexical entries linking forms and meanings.[10] There is a lexical entry in English, for example, that pairs the signal type *zero* – its phonetic shape – with the signal meaning "naught." Construction types that have lexical entries include:

1. elementary words (e.g., *dog*, *zero*, *from*)
2. inflectional morphemes (e.g., *-s*, *-ed*, *-est*)
3. productive derivational morphemes (e.g., *-able*, *-er*, *un-*)
4. lexicalized complex words (e.g., *business*, whose meaning is not entirely derivable from the meanings of *busy* and *-ness*)
5. idioms (e.g., *by and large*, whose meaning is also not entirely derivable from the meanings of its parts)

Together these entries make up a complex signaling system. Not only is *zero* paired with "naught," but *one* is paired with "one," *two* with "two," etc., in a set of contrasting doublets for numbers. These, in turn, contrast with other quantifiers, such as *none*, *some*, and *all*, and eventually with all other lexical entries. How this is to be represented is one of the basic questions in linguistics.

Grammatical rules. Other doublets are expressed as grammatical rules that describe the composition of these basic forms. These include:

1. phonological rules (e.g., for what is a possible phonetic sequence in English)
2. morphological rules (e.g., for deriving adjectives like *shippable* from *to ship* and *-able*)
3. syntactic rules (e.g., for how a noun phrase may consist of an article plus a noun)
4. semantic rules (e.g., for how the meaning of a noun phrase is a composition of the meanings of its parts).

[9] "Le lien unissant le signifiant au signifié est arbitraire, ou encore, puisque nous entendons par signe le total résultant de l'association d'un signifiant à un signifié, nous pouvons dire plus simplement: *le signe linguistique est arbitraire*" (1916/1968, p. 100).

[10] For a related idea, see the notion of *lemma* (Levelt, 1989).

Conventions of use. Other doublets have been studied as conventions of use. In many cultures, you greet people by asking about their health, e.g., "How are you?" and in others, by asking where they are going. In some cultures, when a person sneezes, you say "Bless you," and to wish someone luck on stage, you say "Break a leg" (Morgan, 1978).

Conventions of perspective. Other doublets are really conventions about how one is to view certain entities. In Britain, a street is conceived of as an area that includes the roadway and the adjacent land on which the houses sit. So the British say, "My house is *in* Maiden Lane." In North America, a street is conceived of as a one-dimensional roadway that the adjacent land and houses touch. So North Americans say, "My house is *on* Maiden Lane." In Britain (and the rest of Europe), the "first" floor of a building is one story above the ground floor, but in North America, it *is* the ground floor. It isn't that the two communities have different meanings for *in*, *on*, and *first*. What differs are their conventional perspectives on streets and floors (Clark, 1996). Differences in conventional perspective are easy to confuse with differences in word or construction meaning.

As Lewis argued, the phonological, lexical, morphological, syntactic, and semantic rules of a language – its grammar – constitute a conventional signaling system. They describe regularities of behavior – what English speakers regularly do, and expect others to do, to achieve part of what they intend to do in using sounds, words, constructions, and sentences for communication.

NONCONVENTIONAL COORDINATION

The conventions of English are hardly enough to make communication work. They specify only the *potential* uses of a word or construction – and only some of these. They never specify the *actual* uses. The doublet for *zero* says how the word *can* be used. It doesn't say how it actually *is* used on some particular occasion. Every use of language raises non-conventional coordination problems, which depend for their solution on joint salience, solvability, and sufficiency. Here are four classes of problems that require non-conventional solutions.

Ambiguity. Almost every expression has more than one conventional meaning. Suppose *zero* has four conventional senses – "cipher naught," "nil," "freezing temperature," and "nonentity." The traditional idea is that when we are told, "I met a zero," or "It's zero outside," or "Write down zero," we select the lexical entry that "best fits" the utterance in

context. But what "best fit" comes down to really is joint salience – which sense is the most salient solution given our current common ground. We tend to underestimate the coordination problems created by ambiguity, which arise not only for ambiguous words like *zero*, but for ambiguous constructions like *criminal lawyer* and *I discovered the guy with my binoculars*.

Contextuality. In San Francisco in 1980 , a woman telephoned directory assistance to ask about toll charges, and the operator told her, "I don't know – you'll have to ask a zero."[11] If the caller had selected one of the conventional senses for *zero*, she might have chosen "nonentity" ("I don't know – you'll have to ask a nonentity"). Yet she reportedly interpreted the operator as meaning "person one can reach on a telephone by dialing the cipher naught." The operator used *zero* with a novel, non-conventional interpretation, and the caller interpreted it on the spot. How did they manage? The operator created a participant coordination problem that they solved on the basis of solvability, sufficiency, immediacy, and joint salience.

The operator's use of *zero* is a type of *contextual construction* (Clark and Clark, 1979; Clark, 1983). Contextual constructions aren't merely ambiguous, having a small fixed set of conventional meanings. They have in principle an infinity of potential non-conventional interpretations, each built around a conventional meaning of the word or words it is derived from. The operator's use of *zero* was built around "naught." In other circumstances, *zero* could have been used with an infinity of other interpretations. Contextual constructions rely on an appeal to context – to the participants' current common ground. They always require non-conventional coordination for their interpretation.

Contextual constructions are ubiquitous. In English, they include such types as these (Clark, 1983):[12]

[11] *San Francisco Chronicle*, November 24, 1980

[12] For discussions of these constructions, see Clark and Clark (1979), Clark (1978, 1983), Clark and Gerrig (1984), Downing (1977), Gleitman and Gleitman (1970), Kay and Zimmer (1976), Levi (1978), Nunberg (1979), and Sag (1981), though Levi assumes, contrary to the conclusion here, that nonpredicating adjectives have entirely conventional interpretations (see Clark, 1983).

Contextual construction	Examples
indirect description	You'll have to call a *zero*. I bought a *Henry Moore.*
compound noun	Sit on the *apple-juice chair*. I want a *finger cup.*
denominal noun	He's a *waller*. She's a *cupper*.
denominal verb	She *Houdini'd* her way out of the closet. My friend *teapotted* a policeman.
denominal adjective	She's very *San Francisco*. He's *Churchillian*.
nonpredicating adjective	That's an *atomic* clock, not a *manual* one.
possessive	That's *Calvin's* side of the room. Let's take *my* route.
main verb *do*	He *did* the street. He *did* a Nixon.
pronoun *one*	He has *one*.
pro-adjective *such*	He has just *such* a car.

This list also includes the main verb *do*, the indefinite pronoun *one*, and the pro-adjective *such*, which work like contextual constructions. When a friend tells you, "George *did* all three roofs," you understand what George did by assuming solvability and sufficiency and by appealing to joint salience.

The common ground needed for contextual constructions often lies far outside language. For Ann to tell Ben "I Houdini'd my way out of the closet," she must suppose they share salient biographical facts about Harry Houdini, the great escape artist (Clark and Gerrig, 1984). For her to say "Max went too far this time and teapotted a policeman" and by "teapot" mean "rub the back of with a teapot," she must suppose she and Ben share knowledge of Max's peculiar penchant for sneaking up behind people and rubbing them with a teapot (Clark and Clark, 1979). And for satirist Erma Bombeck to write "Stereos are a dime a dozen" and by "stereos" to mean "potential roommates who own a stereo," she must suppose she and her readers understand she is writing about difficulties in finding a roommate (Clark, 1983). Contextual constructions offer a convincing demonstration of the cumulative view of discourse: They can only be understood against the current state of the discourse.

Indexicality. Most references to particular objects, events, states, and processes are indexical: The referents cannot generally be identified without knowledge of the participants' current common ground. When I tell you, "That man is my cousin," I rely on conventions about the meanings of *that*, *man*, and noun phrases, but there is no convention linking the expression *that man* to my actual cousin. That

link we have to coordinate by non-conventional means. Perhaps you have just mentioned an infamous criminal, or we have just seen a man fall on an icy sidewalk, or I have pointed at a book about cars. We hit on the same referent by appealing to solvability, sufficiency, and joint salience (Clark, Schreuder, and Buttrick, 1983; Nunberg, 1979; see Chapter 6). Similar principles apply to definite descriptions (like *the man in the poster*), definite pronouns (*I*, *she*, *here*), and even proper names (*George*, *Connie*).

Indexicality poses even more of a problem in *indirect reference*. When Jack tells Connie, "Our house celebrates birthdays with strawberries and champagne," he is using *our house* to refer directly to his house, but only as a means for referring indirectly to its inhabitants. The link from Jack's house to its inhabitants is not conventional and has to be coordinated by Jack and Connie. The principle is, once again, joint salience.

Layering. Suppose Jack utters "Frankly, I don't give a damn." In talking to Connie, he could be speaking seriously and mean what he says. In other circumstances, he could be speaking nonseriously at another layer of action. He might be practicing the line for a play, demonstrating someone's tone of voice, offering a linguistic example, or citing Rhett Butler's line from the movie *Gone with the Wind*. Whether he is speaking seriously or nonseriously isn't a matter of convention, but of nonconventional coordination (Chapter 12).

NONCONVENTIONAL COORDINATION DEVICES

If convention isn't the only coordination device we exploit in language use, what are the others? The answer is, almost any device we can appeal to successfully. The ultimate criterion is, as before, joint salience. Three such devices are explicit agreement, precedent, and perceptual salience.

Take explicit agreement. In scholarly writing, the meaning of a term is often stipulated. When Peter Strawson (1974, p. 75) says: "I begin by introducing the notion of a perspicuous grammar. A perspicuous grammar is...," he is making an explicit agreement with his readers about what he will mean by *perspicuous grammar* for the rest of that article. Its very purpose is to preempt conventions that would otherwise apply. Stipulations can be made on the spot with locutions like *what I shall call*, *let us call this*, *hereafter*, *for short*, *termed*, *named*, and *abbreviated*, but they can also be established through more

elaborate codes. In principle, any convention of language can be preempted by stipulation.[13]

Explicit agreement is also found in baptismal dubbing and its secular counterparts. When a child is born, the parents explicitly agree on its name and then call it by that name or a derivative. The name then ordinarily becomes conventional, though it is assumed to have originated in an explicit agreement. All types of proper nouns and technical terms have similar origins, though the origins are generally much less formal (Ziff, 1977).

Precedent is another important coordination device in using language. Picture Helen and Sam each looking at the diagram of a maze and talking about it on the telephone. The horizontal passages in this maze can be described as rows, lines, columns, or paths, and so can the vertical passages. But once Helen has described the horizontal passages as *rows*, that sets a precedent. From then on, Sam must use *rows* for the horizontal passages and some other term – say, *columns* or *lines* – for the vertical ones. The reason: Helen's precedent becomes the jointly most salient solution to Sam's next reference to the passages, and Sam must conform or risk misunderstanding (Garrod and Anderson, 1987). Entrainment of terms like this is ubiquitous in conversation – powerful evidence for precedent as a major source of coordination in language use.

Perceptual salience is all too often ignored as an essential coordination device in language use. When I tell you, "Please stand by that tree," I may be pointing at a clump of ten trees. Still, you take the one I am referring to to be the biggest, nearest, or most unusual tree, the one that is jointly the most salient perceptually. Or I can say, "What was that?" and refer to a sudden explosion, flash of light, or eerie creak based on the jointly most salient perceptual event at the moment. Perceptual salience can be brought about by gestures, by third parties, by acts of nature, by almost anything. The sources of perceptual salience are limitless (Clark, Schreuder, and Buttrick, 1983).

[13] This is the basis for private codes among spies, and even between husbands and wives, or lovers. In Noel Coward's *Private Lives*, Amanda proposes to Elyot that the moment either one notices the two of them bickering, he or she should utter *Solomon Isaacs* (later shortened to *Solomon*) as a signal to stop all talk for five minutes (later shortened to two). Elyot agrees, and the signal works, for a bit. Similarly, Mad Margaret and Sir Despard Murgatroyd, in Gilbert and Sullivan's *Yeomen of the Guard*, agree that when he says *Basingstoke*, she will try to pull herself together, and that works too.

Processes in coordinating

In Schelling's and Lewis' schemes, coordination problems are treated as discrete events. The tacit assumption is that all coordination is achieved via such events. Nature serves people with one-shot Schelling games in which they make distinct choices, and that is that. The processes inside the event – setting the problem, understanding it, deciding on solutions, specifying the solutions – are irrelevant. It doesn't matter whether the players ruminate over their choices, as in diplomacy, bargaining, and chess, or make split-second decisions, as in canoeing, dancing, and shaking hands.

Most everyday coordination, however, is continuous, demanding adaptive moment-by-moment decisions that don't readily divide into discrete coordination problems. The difference is between a joint *act* and a joint *action*. When we view shaking hands as a joint act, we are treating it as a one-shot coordination problem, an event occurring at a single moment in time. But when we view it as a joint action, we are treating it as a process that unfolds in time. We might see it as a sequence of joint acts that are coordinated in time, or as a process of another kind. For continuous coordination, we must think of actions not acts. The added element is timing.

CONTINUOUS COORDINATION

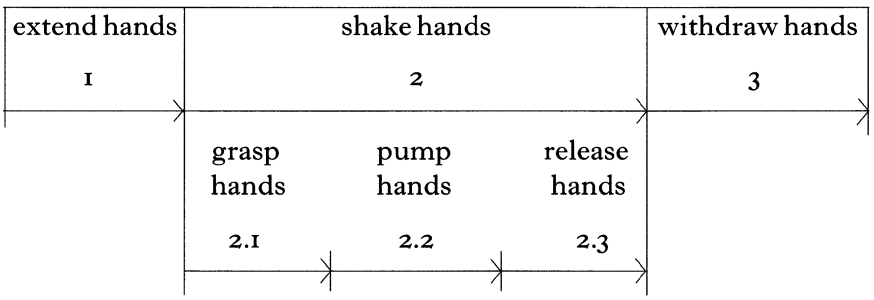
All coordination, even in one-shot problems, is at least quasi-continuous. In the coin game, players A and B are asked to name "heads" or "tails." Whether they are allowed two seconds or two years to respond is left unspecified. When there are twenty Schelling games in a row, timing cannot be left unspecified. A's choice in game 6 must be paired with B's choice in game 6, not in game 5 or 7. A and B really need to coordinate on three things: (1) the current coordination problem; (2) their solution to it; and (3) the moment of response. In truly continuous problems, A and B coordinate (1), (2), and (3) moment by moment. In the general case, joint activities are continuous.

Continuous coordination is *periodic* whenever the actions are synchronized mainly by a cadence or rhythm – waltzing, playing a duet, paddling a canoe, marching in step. More often, it is *aperiodic* – two people shaking hands, eating dinner together, helping each other on with their coats, waving good-bye, negotiating a doorway without bumping. Joint actions can also be mixtures of the two. Conversation is aperiodic.

Coordination can also be *balanced* or *unbalanced*. In some joint actions, the participants take similar actions with no one in the lead. Hand shaking, duet playing, and team juggling may be initiated by one person, but are otherwise balanced. Most joint actions, however, are unbalanced. At any moment, they are led, or directed, by one of the participants, and the rest follow. In waltzing it is the man who leads, in orchestras the conductor, in canoeing the fore paddler, and in conversation the speaker. Not, of course, that these leaders have *carte blanche* to go any direction they want. But their actions are the main basis for synchrony and for the actions taken by the other participants. Most aperiodic unbalanced activities alternate in who takes the lead. In playing catch, it is largely the thrower, not the catcher, who leads, but who is thrower and who is catcher alternates. Conversation is unbalanced.

PHASES AND SYNCHRONY

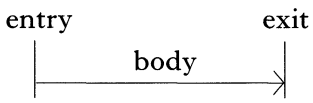
Joint actions can be coordinated, I suggest, because they divide into *phases*. By phase, I mean a stretch of joint action with a unified function and identifiable entry and exit times. Playing a Mozart string quartet has four obvious phases – the first, second, third and fourth movements. Shaking hands has three – extending the hands, shaking hands proper, and withdrawing the hands. Most phases are hierarchical, dividing into subphases, which divide into further subphases, and so on. In music, phase hierarchies are represented directly in the notation: Entire pieces divide into sections, which divide into phrases, which divide into measures, which divide into beats. And in shaking hands, the second phase seems to divide into three subphases – grasping, pumping, and releasing – as diagrammed here:



With time running from left to right, the overall handshake divides into phases 1, 2, and 3, and phase 2 divides into subphases 2.1, 2.2, and 2.3.

Phases are what actually get coordinated. A phase is really a joint

Phases are what actually get coordinated. A phase is really a joint action with an entry, a body, and an exit (see Chapter 2). It can be diagrammed this way:



The entry is the moment the participants believe they have entered the action – the tail of the arrow – and the exit is the moment they believe they have left it – the head of the arrow. The body is what they do between the entry and the exit – the shaft of the arrow. The participants have to coordinate on all three features.

Synchrony of action requires coordination on the entry and exit times to each phase. To achieve synchrony, the participants must be able to project both times from what went before. They should be helped whenever the times are: (1) good reference points – jointly salient moments in time; and (2) easy to project from the previous phases. The participants achieve continuous synchrony, I suggest, by means of three main *coordination strategies*.

The *cadence strategy* is limited to periodic activities. In these, entry times are highly salient, and the duration of a phase is entirely predictable from the cadence. So the participants can coordinate by reaching agreement on three features:

1. an entry time *t*
2. a duration *d*
3. for all participants *i*, the participatory action *p(i)* that *i* is to perform in *d*

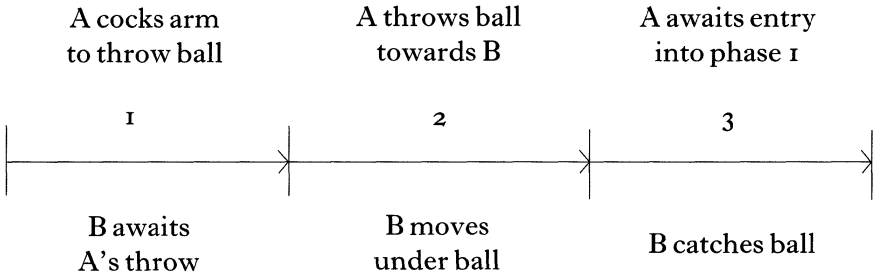
In music, entry times are marked by heavy beats for phases and by lighter beats for subphases; it is significant that in musical notation the salient beats mark entry times to a measure, not exit times. In marching, entry times are marked by footfalls, and in canoe paddling, by the starts of paddle pulls, which also occur in rhythm. In rhythmic activities, the duration of a phase is a fixed number of beats long and depends little on what the participants do during each phase.

The *entry strategy* is more general than the first strategy. In continuous actions, the exit from one phase coincides with the entry into the next. In shaking hands, you know you have left the "extending hands" phase the moment you have entered the "grasping hands" subphase. When this holds, the participants only have to coordinate on two features:

1. an entry time *t*
2. for all participants *i*, the participatory action *p(i)* that *i* is to perform in the phase

For this strategy to work, the entry times must be salient and projectable from the participatory actions of the previous phase. These conditions hold for many unbalanced aperiodic activities.

Most aperiodic activities have jointly salient entry times. Playing catch – tossing a ball back and forth – might have three main phases:



These phases define a cycle – a superphase in playing catch – and each time it is repeated, A's and B's roles are reversed. And these three phases themselves have subphases. The entry times into phases 1, 2, and 3 are as follows: the moment A begins to cock his or her arm; the moment of A's release of the ball; and the moment of B's contact with the ball. As the boundary strategy requires, these are major landmarks visible to both players.

The problem in aperiodic actions is projecting the entry times. Without a cadence, the participants need other devices, and the main device is the leader's actions. In playing catch, the entry time to phase 2 (the ball's leaving A's hand) can be projected by estimating how long A will take in throwing the ball. That can be projected more precisely from the subphases of 1 – say, bringing the arm back and thrusting it forward. The entry time to phase 3 (B's catching the ball) can be projected from the subphases of 2 – say, the ball rising to its apex, and the ball falling from its apex. So, to synchronize their actions, the participants track the subphases, and the easier they are to track, the more accurate the synchrony.

Aperiodic phases are usually *extendible*. Suppose that B in phase 3 goes to catch the ball, drops it, and has to pick it up again. The extra time he takes is added to phase 3 – or rather only one subphase of phase 3 – and doesn't affect phases 1 or 2 or any other subphases of 3. To keep in synchrony, all A and B have to do is extend the one subphase by the

right amount and continue. Extendibility is useful because it allows for local repairs, for inserting other joint actions – like time-outs – and for accommodation to temporary lapses from synchrony.

The third strategy is the *boundary strategy*. In continuous actions, the exit from one phase sometimes doesn't coincide with the entry into the next, and there is no cadence to help out. In these, the participants must coordinate on three features:

1. an entry time *t*
2. an exit time *u*
3. for all participants *i*, the participatory action *p(i)* that *i* is to perform in the phase

In the final phase of shaking hands, the entry time is projected from the participatory actions of the previous phase. But the exit time must be projected from the actions of the current phase since there is no following phase to mark it. In a handshake, the two people withdraw their hands together to end at the same time. It would be unseemly for one person to withdraw the hand too quickly.

People trying to coordinate need to estimate time accurately. When I throw a ball, I need to throw it to where my partner can catch it, and he needs to go to where I have thrown it. On my part, that takes estimates of how far and how fast he can run, and these will depend on the situation – the terrain, the type of ball, my partner's skill. If I overestimate, he won't catch the ball, and if I underestimate, the catch will be too easy, and he will get bored. The same goes for my partners. They must estimate how hard, how high, and in which direction I have thrown the ball, or they will miss it. Making moment-by-moment estimates like this is one of the great feats of joint actions.

In all three strategies, synchrony is achieved by the participants projecting entry times and participatory actions for each phase. The principle I suggest is this:

The synchrony principle. In joint actions, the participants synchronize their processes mainly by coordinating on the entry times and participatory actions for each new phase.

Put simply, joint actions are largely organized around entries and expected participatory actions.

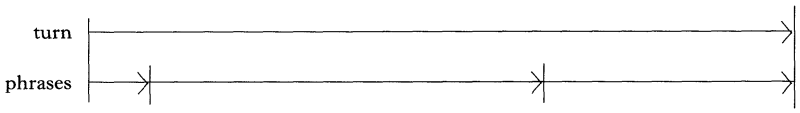
Language processes

Conversation is an example *par excellence* of a joint activity in which the joint actions are aperiodic, unbalanced, and alternating. It is aperiodic because it has no cadence, unbalanced because it is led largely by the speaker, and alternating because who speaks alternates turn by turn. Not that language use is always this way. It can be balanced, as when parishioners recite prayers in unison, and periodic, as when football fans, picketers, and opera choruses chant or sing in rhythm. Yet its primary form is aperiodic and alternating.

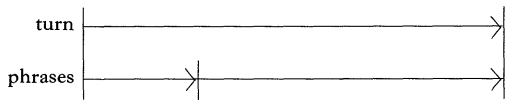
PHASES IN CONVERSATION

If conversation does consist of joint actions, it should divide into phases that have jointly salient entry times, and it does. Conversations divide into a well-known hierarchy of phases – from broad sections to phonetic segments. Some of these phases are illustrated here (1.3.986):

Cal: well what was the ^OUTcome of all this, what . trans^PIRED,



Viv: -- ^NOTHing, I haven't heard a ^WORD,-



Each line denotes a turn, and turns divide into intonation units, the ends marked here by commas (Chapter 9). Intonation units, which are themselves phrases, divide into smaller phrases (e.g., "what | was | the outcome | of all this"), which divide into words, syllables, then segments.

The entries and exits of these phases are marked in the syntax, morphology, and intonation. Intonation units, for example, tend to begin on a high pitch, drop gradually in pitch over the unit, and end with a distinctive fall or rise. They also tend to have a focal accented syllable at or near the end that allows listeners to project the exit time with great accuracy. Moreover, they are thought to be found in all languages.

They are precisely what speakers and listeners need for synchronizing their processing.

For the entry and boundary strategies to work, the participants must be able to project the entry times for the next phase with accuracy. And to do that, they must coordinate on the time the current phase consumes. Enter the immediacy premise. With the intonation unit "What . transpired?" Cal is posing a participant coordination problem – he is asking Viv a question. Viv cannot initiate the next phase – her answer – until she has solved that problem, until she has understood his question. By the immediacy premise, she can assume that Cal expected her to be able to grasp what he meant on completion of that phase.

In conversation, then, addressees are expected to have completed their processing of a phase roughly by the time speakers finish that phase. The immediacy premise should hold for phases of all sizes. At the level of single words, addressees should have completed hearing, identifying, and grasping a word by the time speakers go on to the next word. At the level of intonation units, they should have understood what was meant in the current unit before speakers initiate the next one. If processing weren't roughly immediate, delays in one phase would accumulate with delays in the next, making synchrony even more difficult down the line.

PRECISION OF TIMING

People are able to project entry and exit times in conversation with surprising precision (Jefferson, 1972, 1973; Sacks, Schegloff, and Jefferson, 1974; Chapter 9). For a preview of the issue, consider the coordination problem of how to enter the next turn as illustrated in this actual bit of conversation (1.3.215):

Kate:	how did you get on at your interview, . do tell us,
Nancy:	. oh - - god, what an experience, - - I don't know where to start, you know, it was just such a nightmare - - I mean this whole system, of being invited somewhere for lunch, and then for dinner, - and overnight, . *and breakfast*
Nigel:	*oh you st-* you you did stay

Speakers often try to initiate a new turn precisely as the previous turn ends. When they cannot, they create problems that have to be resolved. There are two such problems in this example – Nancy's and Nigel's.

Nancy's problem is that she doesn't immediately know what she wants to say. She has been selected to start speaking precisely at the end

of "your interview." Because she doesn't, Kate prompts her after a brief pause with, "Do tell us." Nancy knows that, if she doesn't start soon, she may be taken as not having heard or understood, or as opting out. So she commits herself with "oh - - god" and then hesitates to plan her answer in earnest. What Nancy and Kate do, then, is shaped by their mutual expectation that Nancy should initiate her turn immediately. Nigel's problem is different. He incorrectly projects the end of Nancy's turn after "and overnight," so his speech overlaps with Nancy's. He repairs the problem by stopping, making a new projection, and beginning again after "and breakfast." So Nigel's overlap and restart are also a result of a mutual expectation of immediate entry into the next turn.[14]

Entry times, as a result, carry evidence about the participants' mental states – their understanding, readiness, plans. Nancy's delayed entry showed her uncertainty about what to say next. Nigel's premature entry revealed his belief about when Nancy had completed her turn. Mistiming can also be used as a deliberate tactic, as when speakers time their turns to overlap with the end of a previous turn to show that they already recognize what is being said (Chapter 8). Entry times are useful both as evidence and as instruments of communication.

What sort of information do entry times provide? The principle that applies is quite straightforward:

Principle of processing time. People take it as common ground that mental processes take time, and that extra processes may delay entry into the next phase.

The principle is useful because we have surprisingly accurate heuristics for estimating processing difficulty. Here are a few. In speaking, processing should take longer, all else being equal, (a) the rarer the expression; (b) the longer the expression; (c) the more complex the syntax or morphology; (d) the more precise the message; and (e) the more uncertain a speaker is about what he or she wants to say. And in understanding, processing should take longer, all else being equal, (a) the rarer the expression; (b) the longer the expression; (c) the more complex the syntax or morphology; (d) the more precise the message; (e) the more extensive the implications; and (f) the less salient the referents. These are only some of the heuristics we use.

[14] That is, the current speaker provides evidence about when the next speaker can or should begin, and potential next speakers are expected to use this evidence to enter their turns at precisely those moments. This goes for all entry times. See Chapter 8.

So content and process are interdependent: The more complicated the content, generally, the longer the process. This helps us discover what our interlocutors are thinking, and reveal to them what we are thinking. Processing time is a resource we make exquisite use of (Chapters 7, 8, and 9).

ASYNCHRONOUS JOINT ACTIONS

Synchrony is required in conversation because speech is evanescent. If addressees are ever to recover an utterance, they must attend to the speech while it is being produced, and that requires speakers and addressees to synchronize their processes. Written language, however, is not evanescent, and writers' and readers' processes are asynchronous. When I write my sister a letter, I may take half an hour, pausing halfway through for coffee and revising it several times. She may read it in thirty seconds and reread it. Not only are her actions and mine not synchronized. There may be no point-by-point correspondence between them at all.

Writing and reading are no less joint actions for the lack of synchrony. My actions depend on what I expect my sister to do, and her actions depend on what she thinks I would expect her to do. We still coordinate on content. I use English, refer to people we mutually know, and allude to family matters all on the assumption she will recognize the coordination devices I am using – conventions, joint salience, precedent, and all the rest. But I will also design – and redesign, edit, and reedit – my sentences to match the processes I judge she will read them by. I expect her to scan the sentences in order at a certain pace and to do so optimally when I pack information at the right density. Even though our processes are not synchronous, she and I coordinate on them.

Joint actions are required in language use regardless of setting. The coordination of content required is much the same across settings, but the coordination of processes is not. In conversation, speakers and addressees synchronize the phases of their actions. In asynchronous settings, speakers try to make processing optimal for their addressees.

Summary

When two people talk, they coordinate on both content and process. They have to do this in performing any joint action – playing a duet, paddling a canoe together, or shaking hands. Many properties of language use are common to all joint actions.

Joint actions require the participants to coordinate on their individual actions. In each joint act, the participants face a coordination problem: What participatory actions do they expect each other to take? To solve this problem, they need a coordination device – something to tell them which actions are expected. Now, according to the principle of joint salience, the ideal coordination device for any such problem is the solution that is most salient, prominent, or conspicuous with respect to the common ground of the participants. The device may be a convention, a precedent, an explicit agreement, a jointly salient perceptual event – any device, really, that satisfies the principle. In language use, coordination problems have additional properties because they are devised by one of the participants. Two of these are solvability and sufficiency: The participants can assume that each coordination problem has a unique solution they can figure out with the available information. Joint salience, solvability, and sufficiency already allow us to account for many properties of language use. Later, we will see how they account for even more.

But language use requires continuous coordination. The participants have to coordinate not only on *what* they do but on *when* they do what they do. They accomplish that, I have suggested, by coordinating on the entry times, content, and exit times of each phase of their actions on the assumption that the addressees' processing of the current phase is expected to be complete roughly by the initiation of the next phase. Yet they also realize that additional mental processes may delay entry into the next phase. Later, we will see how these properties are put to good use.