3 Plans

> Once the European navigator has developed his operating plan and has available the appropriate technical resources, the implementation and monitoring of his navigation can be accomplished with a minimum of thought. He has simply to perform almost mechanically the steps dictated by his training and by his initial planning synthesis. (Gladwin 1964, p. 175)

Every account of communication involves assumptions about action, in particular about the bases for action's coherence and intelligibility. This chapter and the next discuss two alternative views of action. The first, adopted by most researchers in artificial intelligence, locates the organization and significance of human action in underlying plans. As old at least as the Occidental hills, this view of purposeful action is the basis for traditional philosophies of rational action and for much of the behavioral sciences. It is hardly surprising, therefore, that it should be embraced by those newer fields concerned with intelligent artifacts, particularly cognitive science and information-processing psychology.

On the planning view, plans are prerequisite to and prescribe action, at every level of detail. Mutual intelligibility is a matter of the reciprocal recognizability of our plans, enabled by common conventions for the expression of intent, and shared knowledge about typical situations and appropriate actions. The alternative view, developed here in chapter 4, is that while the course of action can always be projected or reconstructed in terms of prior intentions and typical situations, the prescriptive significance of intentions for situated action is inherently vague. The coherence of situated

action is tied in essential ways not to individual predispositions or conventional rules but to local interactions contingent on the actor's particular circumstances. A consequence of action's situated nature is that communication must incorporate both a sensitivity to local circumstances and resources for the remedy of troubles in understanding that inevitably arise.

This chapter reviews the planning model of purposeful action and shared understanding. Those who adopt the planning model as a basis for interaction between people and machines draw on three related theories about the mutual intelligibility of action: (1) the planning model itself, which takes the significance of action to be derived from plans, and identifies the problem for interaction as their recognition and coordination, (2) speech act theory, which accounts for the recognizability of plans or intentions by proposing conventional rules for their expression, and (3) the idea of shared background knowledge, as the common resource that stands behind individual action and gives it social meaning. Each of these theories promises to solve general problems in human communication, such as the relation of observable behavior to intent, the correspondence of intended and interpreted meaning, and the stability of meaning assignments across situations, in ways that are relevant to particular problems in people's interaction with machines.

3.1 The planning model

The planning model in cognitive science treats a plan as a sequence of actions designed to accomplish some preconceived end. The model posits that action is a form of problem solving, where the actor's problem is to find a path from some initial state to a desired goal state, given certain conditions along the way.[1] Actions are described, at whatever level of detail, by their preconditions and their consequences:

[1] See Newell and Simon 1972 for the seminal formulation of this view.

> In problem-solving systems, actions are described by prerequisites (i.e. what must be true to enable the action), effects (what must be true after the action has occurred), and decomposition (how the action is performed, which is typically a sequence of subactions). (Allen 1984, p. 126)

Goals define the actor's relationship to the situation of action, since the situation is just those conditions that obstruct or advance the actor's progress toward his or her goals. Advance planning is inversely related to prior knowledge of the environment of action, and of the conditions that the environment is likely to present. Unanticipated conditions will require re-planning. In every case, however, whether constructed entirely in advance, or completed and modified during the action's course, the plan is prerequisite to the action.

3.1.1 Plan generation and execution monitoring

One of the earliest attempts to implement the planning model on a machine occurred as part of a project at Stanford Research Institute, beginning in the mid 1960s. The project's goal was to build a robot that could navigate autonomously through a series of rooms, avoiding obstacles and moving specified objects from one room to another. The robot, named by its designers Shakey, was controlled by a problem-solving program called STRIPS, which employed a means–end analysis to determine the robot's path (Fikes and Nilsson 1971). The STRIPS program examined the stated goal, and then determined a subset of operators, or actions available to the robot that would produce that state. The preconditions of those actions in turn identified particular subgoal states, which could be examined in the same way. The system thus worked backward from the goal until a plan was defined from the initial state to the goal state, made up of actions that the robot could perform. Subsequent work on problem solving and plan synthesis consisted in large part in refinements to this basic means–ends strategy, toward the end of achieving greater efficiency by constraining the search through possible solution paths.[2]

[2] For a review of subsequent work, see Sacerdoti 1977, ch. 3.

Beyond the problem of constructing plans, artificial intelligence researchers have had to address problems of what Nilsson (1973) terms "failure and surprise" in the execution of their planning programs, due to the practical exigencies of action in an unpredictable environment. The objective that Shakey should actually be able to move autonomously through a real (albeit somewhat impoverished) environment added a new class of problems to those faced by mathematical or game-playing programs operating in an abstract formal domain:

> for a problem-solver in a formal domain is essentially done when it has constructed a plan for a solution; nothing can go wrong. A robot in the real world, however, must consider the execution of the plan as a major part of every task. Unexpected occurrences are not unusual, so that the use of sensory feedback and corrective action are crucial. (Raphael, cited in McCorduck 1979, p. 224)

In Shakey's case, execution of the plan generated by the STRIPS program was monitored by a program called PLANEX. The PLANEX program monitored not the actual moves of the robot, however, but the execution of the plan. The program simply assumed that the execution of the plan meant that the robot had taken the corresponding action in the real world. The program also made the assumption that every time the robot moved there was some normally distributed margin of error that would be added to a "model of the world," or representation of the robot's location. When the cumulative error in the representation got large enough, the plan monitor initiated another part of the program that triggered a camera which could, in turn, take a reading of Shakey's location in the actual world.

The uncertainty to which Shakey was to respond consisted in changes made to the objects in its environment. Another order of uncertainty was introduced with Sacerdoti's system NOAH (an acronym for Nets of Action Hierarchies). Also developed at the

Stanford Research Institute as part of the Computer-Based Consultant project, NOAH was designed to monitor and respond to the actions of a human user. With NOAH, Sacerdoti extended the techniques of problem-solving and execution monitoring developed in the planning domain to the problem of interactive instruction:

> NOAH is an integrated problem solving and execution monitoring system. Its major goal is to provide a framework for storing expertise about the actions of a particular task domain, and to impart that expertise to a human in the cooperative achievement of nontrivial tasks. (Sacerdoti 1977, p. 2)

The output of the planning portion of Sacerdoti's program is a "procedural net," or hierarchy of partially ordered actions, which becomes in turn the input to the execution-monitoring portion of the system. The execution monitor takes the topmost action in the hierarchy, provides the user with an instruction, and then queries the user regarding the action's completion. A principal objective of the innovations that Sacerdoti introduced for the representation of procedures in NOAH was to extend execution monitoring to include tracking and assessment of the user's actions in response to the instructions generated:

> The system will monitor the apprentice's work to ensure that the operation is proceeding normally. When the system becomes aware of an unexpected event, it will alter instructions to the apprentice to deal effectively with the new situation. (ibid., p. 3)

A positive response from the user to the system's query regarding the action is taken to mean that the user understood the instruction, and has successfully carried it out, while a negative response is taken as a request for a more detailed instruction. The system allows as well for a "motivation response," or query from the user as to why a certain task needs to be done (to which the system

responds by listing tasks to which the current task is related), and for an "error response," or indication from the user that the current instruction cannot be carried out.

Just as the accumulation of error in the PLANEX program required feedback from the world in order to re-establish the robot's location, the error response from the user in Sacerdoti's system requires that NOAH somehow repair its representation of the user's situation:

> PLANEX presumed that an adequate mechanism existed for accurately updating the world model. This was almost the case, since there were only a small number of actions that the robot vehicle could take, and the model of each action contained information about the uncertainty it would introduce in the world model. When uncertainties reached a threshold, the vision subsystem was used to restore the accuracy of the world model.
>
> For the domain of the Computer-based Consultant, or even for a richer robot domain, this approach will prove inadequate ... NOAH cannot treat the world model as a given. It must initiate interactions with the user at appropriate points to ensure that it is accurately monitoring the course of the execution ...
>
> [W]hen a serious error is discovered (requiring the system to be more thorough in its efforts to determine the state of the world), the system must determine what portions of its world model differ from the actual situation. (ibid., pp. 71–2)

The situation in which Shakey moved consisted of walls and boxes (albeit boxes that could be moved unexpectedly by a human hand). The problem in designing Shakey was to maintain consistency between the represented environment and the physical environment in which the robot moved. In introducing the actions of a user, the computer's environment becomes not only a physical but a social one, requiring the interpretation of the user's actions, and an assessment of the user's understanding of his or her situation. The difficulty of maintaining a shared understanding of a situation, as we will see more clearly in chapters 4 and 5, is not just a matter of monitoring the course of events, but of establishing their significance. Nonetheless, with Sacerdoti we have at least a preliminary recognition of the place of the situation in the intelligibility of action and communication.

3.1.2 *Interaction and plan recognition*

Adherents of the planning model in artificial intelligence research have taken the requirement of interaction as an injunction to extend the planning model from a single individual to two or more individuals acting in concert. The planning model attempts to bring concerted action under the jurisdiction of the individual actor by attaching to the others in the actor's world sufficient description, and granting to the actor sufficient knowledge, that he or she is able to respond to the actions of others as just another set of environmental conditions. The problem of social interaction, consequently, becomes an extension of the problem of the individual actor. The basic view of a single, goal-directed agent, acting in response to an environment of conditions, is complicated – the conditions now include the actions of other agents – but intact.

The problem for interaction, on this view, is to recognize the actions of others as the expression of their underlying plans. The complement to plan generation and execution in artificial intelligence research, therefore, is plan recognition, or the attribution of plans to others based on observation of their actions. The starting premise for a theory of plan recognition is that an observer takes some sequence of actions as evidence, and then forms hypotheses about the plans that could motivate and explain those actions. One persisting difficulty for action understanding in artificial intelligence research has been the uncertain relation between actions and intended effects. Allen (1984) illustrates this problem with the example of turning on a light:

> There are few physical activities that are a necessary part of performing the action of turning on a light. Depending on the context, vastly different patterns of behavior can be

> classified as the same action. For example, turning on a light usually involves flipping a light switch, but in some circumstances it may involve tightening the light bulb (in the basement) or hitting the wall (in an old house). Although we have knowledge about how the action can be performed, this does not define what the action is. The key defining characteristic of turning on the light seems to be that the agent is performing some activity which will cause the light, which was off when the action started, to become on when the action ends. An important side effect of this definition is that we could recognize an observed pattern of activity as "turning on the light" even if we had never seen or thought about that pattern previously. (p. 126)

Allen's point is two-fold. First, the "same" action as a matter of intended effect can be achieved in any number of ways, where the ways are contingent on circumstance rather than on definitional properties of the action. And secondly, while an action can be accounted for *post hoc* with reference to its intended effect, an action's course cannot be predicted from knowledge of the actor's intent, nor can the course be inferred from observation of the outcome. Allen identifies the indeterminate relationship of intended effect to method as a problem for planning or plan recognition systems: a problem that he attempts to resolve by constructing a logical language for action descriptions that handles the distinction between what he calls the "causal definition" of an action (i.e. the pre and post conditions that must hold in order to say that the action has occurred, independent of any method), and the action's characterization in terms of a particular method or procedure for its accomplishment.[3]

[3] Another, less problematic uncertainty that Allen attempts to capture is the observation that while some components of an action are sequentially ordered in a necessary way (i.e. one is prerequisite to the other), other components, while necessary to the action, have no necessary sequential relationship to each other. The incorporation of unordered actions into the structure of plans, pioneered by Sacerdoti (1975), was viewed as a substantial breakthrough in early planning research.

While Allen's approach to the problem of plan recognition is an attempt to reconstruct logically our vocabulary of purposeful action, a few more psychologically oriented researchers in artificial intelligence have undertaken experiments designed to reveal the process by which people bring the actions of others under the jurisdiction of an ascribed plan. Schmidt, Sridharan, and Goodson (1978) observe, for example, that plan attribution seems to require certain transformations of the sequential organization of the action described.[4] They report that throughout the process of plan attribution the problem to be solved by the subject remains "ill-formed," by which they mean that at any given time neither the range of possible plans that the other might be carrying out, nor the criteria for assessing just what plan is actually in effect, are clearly defined (p. 80). Nonetheless, they report that their subjects are able to posit an underlying plan. Their strategy appears to be to adopt tentatively a single hypothesis about the other's plan, rather than entertain all or even some number of logical possibilities simultaneously. The preferred hypothesis regarding the other's plan then affects what actions are noted and recalled in the subject's accounts of the action, and the temporal order of events is restructured into logical "in order to" or "because" relationships, such that relations among actions are not restricted to consecutive events in time. At the same time, the current hypothesis is always subject to elaboration or revision in light of subsequent events, to the extent that subjects are often required to suspend judgment on a given hypothesis, and to adopt a "wait and see" strategy. Wherever possible, actions that violate the structure of an attributed plan are explained away before the plan itself is reconsidered. Schmidt, Shridharan, and Goodson conclude that all of these observations "support the generalization that action understanding is simply a process of plan recognition" (p. 50). It is worth noting, however, that while these observations clearly point to a

[4] The empiricism of their study is unusual in artificial intelligence research, where work generally proceeds on the basis of imagination and introspection.

process of plan attribution by the observer, there is no independent evidence that the process of plan attribution is a process of recognizing the plan of the actor.

3.1.3 The status of plans

Assessment of the planning model is complicated by equivocation in the literature between plans as a conceptual framework for the analysis and simulation of action, and plans as a psychological mechanism for its actual production. When researchers describe human action in terms of plans, the discussion generally finesses the question of just how the formulations provided by the researcher are purported to relate to the actor's intent. The claim is at least that people analyze each other's actions into goals and plans in order to understand each other. But the suggestion that the plan is "recognized" implies that it has an existence prior to and independent of the attribution: that it actually determines the action.

The identification of the plan with the actor's intent is explicit in the writing of philosophers of action supportive of artificial intelligence research, like Margaret Boden, who writes:

> unless an intention is thought of as an action-plan that can draw upon background knowledge and utilize it in the guidance of behavior one cannot understand how intentions function in real life. (1973, pp. 27–8)

Intentions, in other words are realized as plans-for-action that directly guide behavior. A logical extension of Boden's view, particularly given an interest in rendering it more computable, is the view that plans actually are prescriptions or instructions for action. An early and seminal articulation of this view came from Miller, Galanter, and Pribram (1960), who define an intention as "the uncompleted parts of a Plan whose execution has already begun" (p. 61). With respect to the plan itself:

> Any complete description of behavior should be adequate to

> serve as a set of instructions, that is, it should have the characteristics of a plan that could guide the action described. When we speak of a plan . . . the term will refer to a *hierarchy* of instructions . . . *A plan is any hierarchical process in the organism that can control the order in which a sequence of operations is to be performed.*
>
> A Plan is, for an organism, essentially the same as a program for a computer . . . we regard a computer program that simulates certain features of an organism's behavior as a theory about the organismic Plan that generated the behavior.
>
> Moreover, we shall also use the term "Plan" to designate a rough sketch of some course of action . . . as well as the completely detailed specification of every detailed operation . . . We shall say that a creature is executing a particular Plan when in fact that Plan is controlling the sequence of operations he is carrying out. (p. 17, original emphasis)

With Miller, Galanter, and Pribram, the view that purposeful action is planned is put forth as a psychological "process theory," compatible with the interest in a mechanistic, computationally tractable account of intelligent action. By improving upon or completing our common-sense descriptions of the structure of action, the structure is now represented not only as a plausible sequence, but as an hierarchical plan. The plan reduces, moreover, to a detailed set of instructions that actually serves as the program that controls the action. At this point, the plan as stipulated becomes substitutable for the action, insofar as the action is viewed as derivative from the plan. And once this substitution is done, the theory is self-sustaining: the problem of action is assumed to be solved by the planning model, and the task that remains is the model's refinement.

While attributing the plan to the actor resolves the question of the plan's status, however, it introduces new problems with respect to

what we actually mean by "purposeful action." If plans are synonymous with purposeful action, how do we account, on the one hand, for a prior intent to act which may never be realized, and, on the other, for an intentional action for which we would ordinarily say no plan was formed ahead of time?[5] And if any plan of action can be analyzed at any level of detail, what level of description represents that which we would want to call purposeful action? If at every level, there is no reason in principle to distinguish, for example, between deliberate action and involuntary response, as the latter always can be ascribed to a process of planning unavailable to the actor. In fact, this is just what Boden would have us do. On her account, action can be reduced to basic units for which "no further procedural analysis could conceivably be given." Those units compose "complex procedural schemata or action-plans," which in turn produce "complex intentional effects" (1973, p. 36). Psychological processes at the level of intention, in other words, are reducible ultimately to bodily operations.

But while the planning model would have a statement of intent reflect an actual set of instructions for action, even casual observation indicates that our statements of intent generally do not address the question of situated action at any level of detail. In fact, because the relation of the intent to accomplish some goal to the actual course of situated action is enormously contingent, a statement of intent generally says very little about the action that follows. It is precisely because our plans are inherently vague – because we can state our intentions without having to describe the actual course that our actions will take – that an intentional vocabulary is so useful for our everyday affairs.

The confusion in the planning literature over the status of plans mirrors the fact that in our everyday action descriptions we do not normally distinguish between accounts of action provided before and after the fact, and action's actual course. As common-sense

[5] Davis (cited in Allen 1984) gives the example of a person driving who brakes when a small child runs in front of the car. See also Searle's distinction (1980) between "prior intentions" and "intentions-in-action."

constructs, plans are a constituent of practical action, but they are constituent as an artifact of our *reasoning about* action, not as the generative *mechanism of* action. Our imagined projections and our retrospective reconstructions are the principal means by which we catch hold of situated action and reason about it, while situated action itself, in contrast, is essentially transparent to us as actors.[6] The planning model, however, takes over our common-sense preoccupation with the anticipation of action, and the review of its outcomes, and attempts to systematize that reasoning as a model for action itself, while ignoring the actual stuff, the situated action, which is the reasoning's object.

3.2 *Speech acts*

A growing number of research efforts devoted to machine intelligence have as their objective, for both theoretical and practical reasons, human–machine communication using English, or "natural language" (for example, Brady and Berwick 1983; Bruce 1981; Joshi, Webber, and Sag 1981). Researchers in natural language understanding have embraced Austin's observation (1962) that language is a form of action, as a way of subsuming communication to the planning model. If language is a form of action, it follows that language understanding, like the interpretation of action generally, involves an analysis of a speaker's utterances in terms of the plans those utterances serve:

> Let us start with an intuitive description of what we think occurs when one agent A asks a question of another agent B which B then answers. A has some *goal*; s/he creates a plan (*plan construction*) that involves asking B a question whose answer will provide some information needed in order to achieve the goal. A then executes this plan, asking B the

[6] One result of the transparency of situated action is that we have little vocabulary with which to talk about it, though chapters 4 and 5 attempt to present some recent efforts in social science. For a treatment of the philosophical vocabulary proposed by Heidegger, see Dreyfus, in press, ch. 6.

> question. B interprets the question, and attempts to infer A's plan (*plan inference*). (Allen 1983, p. 110, original emphasis)

As with the interpretation of action, plans are the substrate on which the interpretation of natural language utterances rests, insofar as "[h]uman language behavior is part of a coherent plan of action directed toward satisfying a speaker's goals" (Appelt 1985, p. 1). We understand language, and action more generally, when we successfully infer the other's goals, and understand how the other's action furthers them. The appropriateness of a response turns on that analysis, from which, in turn:

> The hearer then adopts new goals (e.g., to respond to a request, to clarify the previous speaker's utterance or goal), and plans his own utterances to achieve those. A conversation ensues. (P. Cohen n.d., p. 24)

Given such an account of conversation, the research problem with respect to language understanding is essentially the same as that of the planning model more generally: that is, to characterize actions in terms of their preconditions and effects, and to formulate a set of inference rules for mapping between actions and underlying plans. Among researchers in the natural language area of artificial intelligence research, Searle's speech act theory (1969) is seen to offer some initial guidelines for computational models of communication:

> We hypothesize that people maintain, as part of their models of the world, symbolic descriptions of the world models of other people. Our plan-based approach will regard speech acts as operators whose effects are primarily on the models that speakers and hearers maintain of each other. (Cohen and Perrault 1979, p. 179)

Searle's conditions of satisfaction for the successful performance of speech acts are read as the speech act's "preconditions," while its illocutionary force is the desired "effect":

> Utterances are produced by actions (speech acts) that are executed in order to have some effect on the hearer. This effect typically involves modifying the hearer's beliefs or goals. A speech act, like any other action, may be observed by the hearer and may allow the hearer to infer what the speaker's plan is. (Allen 1983, p. 108)

In describing utterances by their preconditions and effects, speech acts seem to provide at least the framework within which computational mechanisms for engineering interaction between people and machines might emerge. But while Searle's "conditions of satisfaction" state conventions governing the illocutionary force of certain classes of utterance, he argues against the possibility of a rule-based semantics for construing the significance of any particular utterance. While the maxims that speech act theory proposes – for example, the felicity condition for a directive is that S wants H to do A – tell us something about the general conditions of satisfaction for a directive, they tell us nothing further about the significance of any particular directive. With respect to the problem of interpretation, Gumperz (1982b, p. 326) offers the following example from an exchange between two secretaries in a small office:

A: Are you going to be here for ten minutes?
B: Go ahead and take your break. Take longer if you want.
A: I'll just be outside on the porch. Call me if you need me.
B: OK. Don't worry.

Gumperz points out that B's response to A's question clearly indicates that B interprets the questions as an indirect request that B stay in the office while A takes a break, and, by her reply, A confirms that interpretation. B's interpretation accords with a categorization of A's question as an indirect speech act (Searle 1979), and with Grice's discussion of implicature (1975); i.e. B assumes that A is cooperating, and that her question must be relevant, therefore B searches her mind for some possible context or interpretive frame that would make sense of the question, and comes up with the break. But, Gumperz points out, *this analysis begs the question of how B arrives at the right inference:*

so on" (p. 423). Every situation, in other words, has its plan made up of ordered action sequences, each action producing the conditions that enable the next action to occur. Admittedly, the normative order of these action sequences can be thrown off course, by any one of what Schank and Abelson term "distractions," "obstacles," or "errors." Distractions, about which they have little to say, comprise the interruption of one script by another, while:

> An obstacle to the normal sequence occurs when someone or something prevents a normal action from occurring or some enabling condition for the action is absent. An error occurs when the action is completed in an inappropriate manner, so that the normal consequences of the action do not come about. (p. 426)

Not only does the typical script proceed according to a normal sequence of actions, in other words, but each script has its typical obstacles and errors that, like the script itself, are stored in memory along with their remedies, and retrieved and applied as needed.

While plans associate intentions with action sequences, scripts associate action sequences with typical situations. In practice, however, the stipulation of relevant background knowledge for typical situations always takes the form of a partial list, albeit one offered as if the author could complete the list, given the requisite time and space:

> If one intends to buy bread, for instance, the knowledge of which bakers are open and which are shut on that day of the week will enter into the generation of one's plan of action in a definite way; one's knowledge of local topography (and perhaps of map-reading) will guide one's locomotion to the selected shop; one's knowledge of linguistic grammar and of the reciprocal roles of shopkeeper and customer will be needed to generate that part of the action-plan concerned

> with speaking to the baker, and one's financial competence will guide and monitor the exchange of coins over the shop counter. (Boden 1973, p. 28)

Like Boden's story of the business of buying bread, attempts in artificial intelligence research to formalize common-sense knowledge rely upon an appeal to intuition that shows little sign of yielding to scientific methods. The difficulty is not just that every action presupposes a large quantity of background knowledge; though it would pose practical problems, such a difficulty would be tractable eventually. Just because "implicit knowledge" can in principle be enumerated indefinitely, deciding in practice about the enumeration of background knowledge remains a stubbornly *ad hoc* procedure, for which researchers have not succeeded in constructing rules that do not depend, in their turn, on some deeper *ad hoc* procedures.

Nevertheless, the image evoked by "shared knowledge" is a potentially enumerable body of implicit assumptions or presuppositions that stands behind every explicit action or utterance, and from which participants in interaction selectively draw in understanding each other's actions. This image suggests that what actually does get said on any occasion must reflect the application of a principle of communicative economy, which recommends roughly that to the extent that either the premises or rationale of an action can be assumed to be shared, they can be left unspoken. That means, in turn, that speakers must have procedures for deciding the extent of the listener's knowledge, and the commensurate requirements for explication. The listener, likewise, must make inferences regarding the speaker's assumptions about shared knowledge, on the basis of what he or she chooses explicitly to say. What is unspoken and relevant to what is said is assumed to reside in the speaker's and listener's common stock of background knowledge, the existence of which is proven by the fact that an account of what is said always requires reference to further facts that, though unspoken, are clearly relevant.

what is the background knowledge used in communication?

This image of communication is challenged, however, by the results of an exercise assigned by Garfinkel to his students (1972). Garfinkel's aim was to press the common-sense notion that background knowledge is a body of things thought but unsaid, that stands behind behavior and makes it intelligible. The request was that the students provide a complete description of what was communicated, in one particular conversation, as a matter of the participants' shared knowledge. Students were asked to report a simple conversation by writing on the left hand side of a piece of paper what was said, and on the right hand side what it was that they and their partners actually understood was being talked about. Garfinkel reports that when he made the assignment:

> many students asked how much I wanted them to write. As I progressively imposed accuracy, clarity, and distinctness, the task became increasingly laborious. Finally, when I required that they assume I would know what they had actually talked about only from reading literally what they wrote literally, they gave up with the complaint that the task was impossible. (p. 317)

The students' dilemma was not simply that they were being asked to write "everything" that was said, where that consisted of some bounded, albeit vast, content. It was rather that the task of enumerating what was talked about itself extended what was talked about, providing a continually receding horizon of understandings to be accounted for. The assignment, it turned out, was not to describe some existing content, but to generate it. As such, it was an endless task. The students' failure suggests not that they gave up too soon, but that what they were assigned to do was not what the participants in the conversation themselves did in order to achieve shared understanding.

While the notion of "background assumptions" connotes an actual collection of things that are there in the mind of the speaker – a body of knowledge that motivates a particular action or linguistic expression, and makes it interpretable – Garfinkel's exercise, as

well as the phenomenology of experience, suggest that there is reason to question the view that background assumptions are part of the actor's mental state prior to action:

> As I dash out the door of my office, for example, I do not consciously entertain the belief that the floor continues on the other side, but if you stop me and ask me whether, when I charged confidently through the door, I believed that the floor continued on the other side, I would have to respond that indeed, I did. (Dreyfus 1982, p. 25)

A background assumption, in other words, is generated by the activity of accounting for an action when the premise of the action is called into question. But there is no particular reason to believe that the assumption actually characterizes the actor's mental state prior to the act. In this respect, the "taken for granted" denotes not a mental state but something outside of our heads that, precisely because it is non-problematically there, we do not need to think about. By the same token, in whatever ways we do find action to be problematical, the world is there to be consulted should we choose to do so. Similarly, we can assume the intelligibility of our actions, and as long as the others with whom we interact present no evidence of failing to understand us, we do not need to explain ourselves, yet the grounds and significance of our actions can be explicated endlessly. The situation of action is thus an inexhaustibly rich resource, and the enormous problems of specification that arise in cognitive science's theorizing about intelligible action have less to do with action than with the project of substituting definite procedures for vague plans, and representations of the situation of action, for action's actual circumstances.

To characterize purposeful action as in accord with plans and goals is just to say again that it is purposeful and that *somehow*, in a way not addressed by the characterization itself, we constrain and direct our actions according to the significance that we assign to a particular context. How we do that is the outstanding problem. Plans and goals do not provide the solution for that problem, they

simply re-state it. The dependency of significance on a particular context, every particular context's open-endedness, and the essential *ad hoc*ness of contextual elaboration are resources for practical affairs, but perplexities for a science of human action. And, to anticipate the analysis in chapter 7, it is an intractable problem for projects that rest on providing in advance for the significance of canonical descriptions – such as instructions – for situated action.

4 Situated actions

> This total process [of Trukese navigation] goes forward without reference to any explicit principles and without any planning, unless the intention to proceed to a particular island can be considered a plan. It is nonverbal and does not follow a coherent set of logical steps. As such it does not represent what we tend to value in our culture as "intelligent" behavior.
>
> (Gladwin 1964, p. 175)

This chapter turns to recent efforts within anthropology and sociology to challenge traditional assumptions regarding purposeful action and shared understanding. A point of departure for the challenge is the idea that common-sense notions of planning are not inadequate versions of scientific models of action, but rather are resources for people's practical deliberations about action. As projective and retrospective accounts of action, plans are themselves located in the larger context of some ongoing practical activity. As common-sense notions about the structure of that activity, plans are part of the subject matter to be investigated in a study of purposeful action, not something to be improved upon, or transformed into axiomatic theories of action.

The premise that practical reasoning about action is properly part of the subject matter of social studies is due to the emergence of a branch of sociology named *ethnomethodology*. This chapter describes the inversion of traditional social theory recommended by ethnomethodology, and the implications of that inversion for the problem of purposeful action and shared understanding. To designate the alternative that ethnomethodology suggests – more a

reformulation of the problem of purposeful action, and a research programme, than an accomplished theory – I have introduced the term *situated action*. That term underscores the view that every course of action depends in essential ways upon its material and social circumstances. Rather than attempting to abstract action away from its circumstances and represent it as a rational plan, the approach is to study how people use their circumstances to achieve intelligent action. Rather than build a theory of action out of a theory of plans, the aim is to investigate how people produce and find evidence for plans in the course of situated action. More generally, rather than subsume the details of action under the study of plans, plans are subsumed by the larger problem of situated action.

The view of action that ethnomethodology recommends is neither behavioristic, in any narrow sense of that term, nor mentalistic. It is not behavioristic in that it assumes that the significance of action is not reducible to uninterpreted bodily movements. Nor is it mentalistic, however, in that the significance of action is taken to be based, in ways that are fundamental rather than secondary or epiphenomenal, in the physical and social world. The basic premise is twofold: first, that what traditional behavioral sciences take to be cognitive phenomena have an essential relationship to a publicly available, collaboratively organized world of artifacts and actions, and secondly, that the significance of artifacts and actions, and the methods by which their significance is conveyed, have an essential relationship to their particular, concrete circumstances.

The ethnomethodological view of purposeful action and shared understanding is outlined in this chapter under five propositions: (1) plans are representations of situated actions; (2) in the course of situated action, representation occurs when otherwise transparent activity becomes in some way problematic; (3) the objectivity of the situations of our action is achieved rather than given; (4) a central resource for achieving the objectivity of situations is language, which stands in a generally indexical relationship to the circumstances that it presupposes, produces, and describes; (5) as a consequence of the indexicality of language, mutual intelligibility is

achieved on each occasion of interaction with reference to situation particulars, rather than being discharged once and for all by a stable body of shared meanings.

4.1 Plans are representations of action

The pragmatist philosopher and social psychologist George Herbert Mead (1934) has argued for a view of meaningful, directed action as two integrally but problematically related kinds of activity. One kind of activity is an essentially situated and *ad hoc* improvisation – the part of us, so to speak, that actually acts. The other kind of activity is derived from the first, and includes our representations of action in the form of future plans and retrospective accounts. Plans and accounts are distinguished from action as such by the fact that, to represent our actions, we must in some way make an object of them. Consequently, our descriptions of our actions come always before or after the fact, in the form of imagined projections and recollected reconstructions.

Mead's treatment of the relation of deliberation and reflection to action is one of the more controversial, and in some ways incoherent, pieces of his theory. But his premise of a disjunction between our actions and our grasp of them at least raises the question for social science of the relationship between projected or reconstructed courses of action, and actions *in situ*. Most accounts of purposeful action have taken this relationship to be a directly causal one, at least in a logical sense (see chapter 3). Given a desired outcome, the actor is assumed to make a choice among alternative courses of action, based upon the anticipated consequences of each with respect to that outcome. Accounts of actions taken, by the same token, are just a report on the choices made. The student of purposeful action on this view need know only the predisposition of the actor and the alternative courses that are available in order to predict the action's course. The action's course is just the playing out of these antecedent factors, knowable in advance of, and standing in a determinate relationship to, the action itself.

The alternative view is that plans are resources for situated action, but do not in any strong sense determine its course. While plans presuppose the embodied practices and changing circumstances of situated action, the efficiency of plans as representations comes precisely from the fact that they do not represent those practices and circumstances in all of their concrete detail. So, for example, in planning to run a series of rapids in a canoe, one is very likely to sit for a while above the falls and plan one's descent.[1] The plan might go something like "I'll get as far over to the left as possible, try to make it between those two large rocks, then backferry hard to the right to make it around that next bunch." A great deal of deliberation, discussion, simulation, and reconstruction may go into such a plan. But, however detailed, the plan stops short of the actual business of getting your canoe through the falls. When it really comes down to the details of responding to currents and handling a canoe, you effectively abandon the plan and fall back on whatever embodied skills are available to you. The purpose of the plan in this case is not to get your canoe through the rapids, but rather to orient you in such a way that you can obtain the best possible position from which to use those embodied skills on which, in the final analysis, your success depends.

Even in the case of more deliberative, less highly skilled activities, we generally do not anticipate alternative courses of action, or their consequences, until *some* course of action is already under way. It is frequently only on acting in a present situation that its possibilities become clear, and we often do not know ahead of time, or at least not with any specificity, what future state we desire to bring about. Garfinkel (1967) points out that in many cases it is only after we encounter some state of affairs that we find to be desirable that we identify that state as the goal toward which our previous actions, in retrospect, were directed "all along" or "after all" (p. 98). The fact that we can always perform a *post hoc* analysis of situated action that will make it appear to have followed a rational plan

[1] This example was suggested to me by Randy Trigg, to whom I am indebted for the insight that plans orient us for situated action in this way.

says more about the nature of our analyses than it does about our situated actions. To return to Mead's point, rather than direct situated action, rationality anticipates action before the fact, and reconstructs it afterwards.

4.2 Representation and breakdown

While we can always construct rational accounts of situated action before and after the fact, when action is proceeding smoothly it is essentially transparent to us. Similarly, when we use what Heidegger terms equipment that is "ready-to-hand," the equipment "has a tendency to 'disappear'":

> Consider the example (used by Wittgenstein and Merleau-Ponty) of the blind man's cane. We can hand the man the cane and ask him to tell us what properties it has. After hefting and feeling it, he can tell us that it is light, smooth, about three feet long, and so on; it is present-at-hand for him. But when the man starts to use the cane (when he grasps it in that special mode of understanding that Heidegger calls "manipulation") he loses his awareness of the cane itself; he is aware only of the curb (or whatever object the cane touches); or, if all is going well, he is not even aware of that. Thus it is that equipment that is ready-to-hand is invisible just when it is most genuinely appropriated. (Dreyfus, in press, ch. 6)

In contrast, the "unready-to-hand," in Heidegger's phrase, comprises occasions wherein equipment that is involved in some practical activity becomes unwieldy, temporarily broken, or unavailable. At such times, inspection and practical problem-solving occur, aimed at repairing or eliminating the disturbance in order to "get going again." In such times of disturbance, our use of equipment becomes "explicitly manifest as a goal-oriented activity," and we may then try to formulate procedures or rules:

The scheme peculiar to [deliberating] is the "if-then"; if this

> or that, for instance, is to be produced, put to use, or averted, then some ways and means, circumstances, or opportunities will be needed (Heidegger, cited in Dreyfus, in press, ch. 6)

Another kind of breakdown, that arises when equipment to be used is unfamiliar, is discussed in chapter 6 in relation to the "expert help system" and the problem of instructing the novice user of a machine. The important point here is just that the rules and procedures that come into play when we deal with the "unready-to-hand" are not self-contained or foundational, but contingent on and derived from the situated action that the rules and procedures represent. The representations involved in managing problems in the use of equipment presuppose the very transparent practices that the problem renders noticeable or remarkable. Situated action, in other words, is not made explicit by rules and procedures. Rather, when situated action becomes in some way problematic, rules and procedures are explicated for purposes of deliberation and the action, which is otherwise neither rule-based nor procedural, is then made accountable to them.

4.3 *The practical objectivity of situations*

If we look at the world commonsensically, the environment of our actions is made up of a succession of situations that we walk in to, and to which we respond. As I noted in chapter 3, advocates of the planning model not only adopt this common-sense realist view with respect to the individual actor, but attempt to bring concerted action under the same account by treating the actions of others as just so many more conditions of the actor's situation. In the same tradition, normative sociology posits, and then attempts to describe, an objective world of social facts, or received norms, to which our attitudes and actions are a response. Emile Durkheim's famous maxim that "the objective reality of social facts is sociology's fundamental principle" (1938) has been the methodological

premise of social studies since early in this century. Recognizing the human environment to be constituted crucially by others, sociological norms comprise a set of environmental conditions beyond the material, to which human behavior is responsive: namely, the sanctions of institutionalized group life. Human action, the argument goes, cannot be adequately explained without reference to these "social facts," which are to be treated as antecedent, external, and coercive *vis-à-vis* the individual actor.

By adopting Durkheim's maxim, and assuming the individual's responsiveness to received social facts, social scientists hoped to gain respectability under the view that human responses to the facts of the social world should be discoverable by the same methods as are appropriate to studies of other organisms reacting to the natural world. A principal aim of normative sociology was to shift the focus of attention in studies of human behavior from the psychology of the individual to the conventions of the social group. But at the same time that normative sociology directed attention to the community or group, it maintained an image of the individual member rooted in behaviorist psychology and natural science – an image that has been dubbed by Garfinkel the "cultural dope":

> By "cultural dope" I refer to the man-in-the-sociologist's-society who produces the stable features of the society by acting in compliance with preestablished and legitimate alternatives of action that the common culture provides. (1967, p. 68)

Insofar as the alternatives of action that the culture provides are seen to be non-problematic and constraining on the individual, *their* enumeration is taken to constitute an account of situated human action. The social facts – that is to say, what actions typically come to – are used as a point of departure for retrospective theorizing about the "necessary character of the pathways whereby the end result is assembled" (p. 68).

In 1954, the sociologist Herbert Blumer published a critique of traditional sociology titled "What Is Wrong with Social Theory?" (see

Blumer 1969, pp. 140–52). Blumer argues that the social world is constituted by the local production of meaningful action, and that as such the social world has never been taken seriously by social scientists. Instead, Blumer says, investigations by social scientists have looked at meaningful action as the playing out of various determining factors, all antecedent and external to the action itself. Whether those factors are brought to the occasion in the form of individual predispositions, or are present in the situation as pre-existing environmental conditions or received social norms, the action itself is treated as epiphenomenal. As a consequence, Blumer argues, we have a social science that is about meaningful human action, but not a science of it.

For the foundations of a science of action, Blumer turns to Mead, who offers a metaphysics of action that is deeply sociological. Blumer points out that a central contribution of Mead's work is his challenge to traditional assumptions regarding the origins of the common-sense world, and of purposeful action:

> His treatment took the form of showing that human group life was the essential condition for the emergence of consciousness, the mind, a world of objects, human beings as organisms possessing selves, and human conduct in the form of constructed acts. He reversed the traditional assumptions underlying philosophical, psychological, and sociological thought to the effect that human beings possess minds and consciousness as original "givens," that they live in worlds of pre-existing and self-constituted objects, and that group life consists of the association of such reacting human organisms. (ibid., p. 61)

Mead's "reversal," in putting human interaction before the objectivity of the common-sense world, should not be read as an argument for metaphysical idealism; Mead does not deny the existence of constraints in the environment in which we act. What Mead is working toward is not a characterization of the natural world *simpliciter*, but of the natural world *under interpretation*, or the world as

construed by us through language. The latter is precisely what we mean by the *social* world and, on Mead's account, interaction is a condition for that world, while that world is a condition for intentional action.

More recently, ethnomethodology has turned Durkheim's maxim on its head with more profound theoretical and methodological consequences. Briefly, the standpoint of ethnomethodology is that what traditional sociology captures is precisely our common-sense view of the social world (see Sacks 1963; Garfinkel 1967; and Garfinkel and Sacks 1970). Following Durkheim, the argument goes, social studies have simply taken this common-sense view as foundational, and attempted to build a science of the social world by improving upon it. Social scientific theories, under this attempt, are considered to be scientific insofar as they remedy shortcomings in, and preferably quantify, the intuitions of everyday, practical sociological reasoning.

In contrast, ethnomethodology grants common-sense sociological reasoning a fundamentally different status than that of a defective approximation of an adequate scientific theory. Rather than being *resources* for social science to improve upon, the "all things being equal" typifications of common-sense reasoning are to be taken as social science's *topic*. The notion that we act in response to an objectively given social world is replaced by the assumption that our everyday social practices render the world publicly available and mutually intelligible. It is those practices that constitute ethnomethods. The methodology of interest to ethnomethodologists, in other words, is not their own, but that deployed by members of the society in coming to know, and making sense out of, the everyday world of talk and action.

The outstanding question for social science, therefore, is not whether social facts are objectively grounded, but how that objective grounding is accomplished. Objectivity is a product of systematic practices, or members' methods for rendering our unique experience and relative circumstances mutually intelligible. The source of mutual intelligibility is not a received conceptual scheme,

or a set of coercive rules or norms, but those common practices that produce the typifications of which schemes and rules are made. The task of social studies, then, is to describe the practices, not to enumerate their product in the form of a catalogue of commonsense beliefs about the social world. The interest of ethnomethodologists, in other words, is in how it is that the mutual intelligibility and objectivity of the social world is achieved. Ethnomethodology locates that achievement in our everyday situated actions, such that our common sense of the social world is not the precondition for our interaction, but its product. By the same token, the objective reality of social facts is not the fundamental *principle* of social studies, but social studies' fundamental *phenomenon*.