Part I

Assembling Intentionality

1

Evolutionary Naturalism

Successful farmers have social relations with one another, while hunter-gatherers have ecological relations with hazelnuts.

Quoted in C. Gamble, The Palaeolithic Societies of Europe *(1999), p. 2.*

1.1 Two Projects of Evolutionary Naturalism

Philosophy is not a natural science, but philosophy is intimately connected with the natural sciences, for one of its roles is integrative. All animals are biochemical machines with complex sensing and control systems. These systems enable those agents to adapt their behavior to the contingencies of their environment. To understand the distinctive features of our own control systems, we need to integrate the neurosciences and cognitive psychology with evolutionary biology, especially human behavioral ecology. For to understand the operation of complex systems, it is usually necessary to understand their function (Marr 1980). That task is hard, for historical processes destroy evidence about their own causes. Understanding the history of human cognition is especially difficult, for humans are most unusual primates. We are technically proficient, physically modifying our habitat in innumerable ways. Humans are encultured primates, and have been encultured as long as we have been human. As such, we are extraordinarily cooperative. For hundreds of thousands of years humans have lived by collective action. Moreover, we do not just live in groups, we are marinated in the material, behavioral, and informational products of our culture. Those social products profoundly influence our actions (Mithen 1996b; Tattersall 1998; Gamble 1999). We are sexually unusual, for we combine social life with paternal investment in children and fairly stable bonds between mated pairs. We use language, an extraordinarily complex and subtle communication system. We have invaded almost every terrestrial habitat, and almost every region of the earth.

Our behavior is strikingly variable over space and time. Thus human behavior is plastic and adaptable, and we have complex and subtle systems through which we register features of our environment and control our actions. I shall argue that we are unusual not just in our behavior and distribution; we have been built by unusual evolutionary mechanisms. For both group selection and nongenetic inheritance have been of profound importance in human evolution.

Thus one integrative project is *internal* to the sciences: it is the project of assembling a coherent theory of human agency and human evolutionary history from the fragments provided by the natural and the social sciences. This project will be one theme of this work. I shall call this set of facts, the ones that explain our behavioral plasticity and adaptability, the "wiring-and-connection" facts about human agency (for this terminology, see Godfrey-Smith 2002). This set includes facts about our internal organization (the wiring facts) and the facts about how that organization registers, reflects, or tracks external circumstances (the connection facts). But it also includes the evolution and development of our wiring and our connections to our world.

In pursuing this integrative project in part I, I shall sketch some ideas on the cognitive foundations of human uniqueness. I shall focus on the evolution of decoupled representational capacities: that is, the evolution of capacities to track features of the environment where that tracking capacity does not drive a specific behavioral response. And I shall discuss the evolution of motivational mechanisms that are not based on drives and sensations. In part II, I'll discuss the evolutionary mechanisms responsible for the transition to human cognitive capacities, and in part III the aim is to develop an alternative to currently influential "massively modular" theories of human cognitive architecture. To put it mildly, this whole discussion is both tentative and gappy: the main aim is to foreground issues and ideas that have been underplayed in current debates.

This internal integrative project coexists with an *external project* that will be rather more familiar to philosophers. Economics, anthropology, archaeology, and the other social sciences see humans as essentially social and encultured agents. These social sciences and the information they provide are central to the internal project. But it is also important that those social sciences typically depend on ideas of human agency that are refined versions of our folk self-conception. For science does not have to construct from scratch a theory of human agency. We inherit a picture of that agency as part of our common culture. Our

inherited picture portrays us as self-aware, more or less rational agents. We are intentional agents, and our actions are a reflection of our thoughts and preferences. It is this picture that the social sciences have inherited and modified.

Since we act, and are acted on, in both cultural and ecological communities, there must be some way of synthesizing human behavioral ecology with the social sciences. We are social as well as ecological agents, and a coherent account of human nature must weld together evolutionary-scientific and social-scientific conceptions of human agency. But it does not follow from the fact that we are both social and ecological beings that an integrated biocultural theory of human agency will vindicate anything like our folk self-conception. Our folk conception may be a self-deceiving view of human agency, so the external project is to explore the extent to which our folk conception, both on its own and as it has been integrated within the social sciences, can be integrated with a scientific conception of human agency. This external project presupposes the internal one. We cannot make much progress in answering the question "How do our folk conceptions relate to a scientific conception of agency?" unless we have a sketch of that scientific conception in play.

1.2 The Simple Coordination Thesis

Folk psychology is both rich and varied. It provides us with ways of thinking and talking about sensations, emotions, character traits, perceptions, and thoughts. But it is often supposed that the core of folk psychology is intentional psychology: that is, the prediction and explanation of the actions of agents in terms of their conception of the world (how they believe their world to be) and their preferences (how they want their world to be). Belief/preference explanation is apparently ubiquitous in human social life. We use it to understand our own behavior: I hung around after that appalling talk because I thought the department chair would have free drinks in his office. And we interpret the actions of others that way: he went to the party because he wanted to ask the new girl out.

The concepts of belief and preference, feelings and emotions, moods and character states, may describe our cognitive architecture well, badly, or not at all. However that turns out, we do use these conceptual categories to interpret others. These interpretative practices play an absolutely central role in human social life. Let's call the set of facts

about our folk concepts "the interpretation facts" (Godfrey-Smith 2002). The external integrative project, then, is to understand the relationship between the wiring-and-connection facts and the interpretation facts. Since the publication of Jerry Fodor's *The Language of Thought* in 1975, contemporary philosophy of psychology has been dominated by a bold and optimistic hypothesis about this relationship: the *Simple Coordination Thesis*. This hypothesis is motivated by three thoughts. First, we are very complex agents with subtle mechanisms of behavioral control. Second, despite our complexity, we are apparently quite good at predicting and explaining one another's actions. Third, as just noted, all normal humans inherit as part of their common cognitive stock a set of interpretative concepts. Perhaps the third of these facts explains the second despite the first. This is the hunch of the Simple Coordination Thesis, for it claims that: (a) Our interpretative concepts constitute something like a theory of human cognitive organization: they are a putative description of the wiring-and-connection facts; (b) Our interpretative skills depend on this theory, and our ability to deploy it on particular occasions; (c) We are often able to successfully explain or anticipate behaviour because this theory is largely true.

The Simple Coordination Thesis makes three crucial bets. The first is on a picture of the internal architecture of the human mind. There is a good deal of controversy about its commitments, but at a minimum the Simple Coordination Thesis is committed to the idea that the cognitive system of human agents has two subsystems. One has the function of registering states of the world as it is and as it might be. Another functions to register and rank the way the world could be changed. Moreover, these registrations of possible changes motivate the agent; they are goals. One system is described by our belief attributions and the other by our goal attributions.[1] The thesis is also committed to a second bet: the idea that the contents of these subsystems, and their consequences for behavior, are partially identifiable by other agents. In general, the registration-and-goal states of other agents are not inscrutable, and deducing their consequences for others' actions

[1] More controversial is the idea that the Simple Coordination Thesis is committed to a classical version of the computational theory of mind. When we attribute beliefs and preferences to other agents, we are taking them to have particular symbol structures in the belief and preference systems of their executive control structures. When we attribute a particular piece of reasoning ("She twigged the fact that they were having an affair once she noticed that they began coming into the office at the same time in the afternoon, and with wet hair") we are supposing that specific computational operations in the belief systems have transformed one symbol structure into another.

is not computationally intractable. Hence particular applications of intentional psychology – particular interpretations – are often correct. Though fallible, we are good at specifying the particular states of that architecture in ourselves and others. Third, it is committed to a particular view of the relationship between internal states of the agent and the world. It is part of the folk picture that thoughts have content. A preference is satisfied or not satisfied. A belief is true or false. A belief is about something; Peter, for example, believes that spiders are dangerous, and that is a belief about spiders. The Simple Coordination Thesis needs a view of how this aspect of our interpretative practices – talk of meaning or content – is related to the wiring and connection facts. According to the Simple Coordination Thesis, meaning is a *specific connection property* of the wiring-and-connection facts, though different versions of the hypothesis give different accounts of the nature of that connection.

The Simple Coordination Thesis has not generated a consensus in its favor. Instead it has spawned many versions and much skepticism. The Churchlands argue that though the interpretation facts purport to describe the wiring-and-connection facts they do a horrible job. Dan Dennett argues that the interpretation facts do not have the function of describing the internal organization of agents. Instead, in a rough but useful way, they specify behavioral dispositions of agents while being neutral on the categorical basis of those dispositions. More recently, defenders of simulation theory have attempted to drive a wedge between our interpretative conceptual apparatus and our actual skills of action anticipation. So sceptics abounded, and friends of the hypothesis did not talk with one voice. Even so, the idea in some form or other has dominated philosophy of psychology for the last quarter-century, and with important results.

Sad to say, those results have not yet included a vindication of their crucial bets. One possibility, of course, is that we have not been trying hard enough for long enough. But we need to take seriously the possibility that the relationship between the two sets of facts is much less clean than the Simple Coordination Thesis supposed. One way of evaluating this possibility is to place the thesis in an evolutionary context. That is my strategy, despite the fact that the Simple Coordination Thesis is not in itself an evolutionary hypothesis. It is primarily a hypothesis about the proximal mechanisms of human action. Nevertheless, evolutionary considerations can be part of a total package of evidence in favor of a proximate hypothesis (see, for example, the final chapter of Sober and Wilson 1998).

Moreover, evolutionary considerations are of especial relevance to hypotheses about our abilities to interpret each other's actions. For there is an important contrast between the set of conceptual tools that we use both to understand and to manipulate causal processes in our physical environment (folk physics) and those we use to interpret others. We use our folk physics whenever we make a simple tool. We use it when something is blocking a pipe, and we pick and strip a branch to break the obstruction; or when we make a simple walking stick from a branch while out bush-walking. In such physical interventions, we are unreflectively attuned to such factors as the length, rigidity, weight, and thickness of the branch; of the size, strength, and depth of the obstruction, and so forth. Our awareness of such properties, of their significance and interactions, is crucial to our capacities to intervene in, and remake, our physical environment. Folk physics, like folk psychology, may be a unique feature of hominid cognition (Tomasello 2000; Povinelli et al. 2000). And it is clearly of great importance in human lifeways. But the domain of folk physics – the macroscopic physical properties of objects – is unaffected by our having a theory of those properties. The trajectories of thrown rocks will not change, however well or ill we understand those trajectories.

That is not true of the relationship between folk interpretation and the wiring-and-connection facts. On operational, developmental, and evolutionary time scales there are interactions between the cognitive organization of agents, and how others interpret, respond to, and predict their actions. In particular, there are coevolutionary interactions between agents and interpreters; between the actual cognitive organization of an agent, and others' pictures of that organization. If human social organization were to have a fundamentally cooperative dynamic, we would expect that interaction to increase the accuracy of agents' interpretative practices (chapter 11.4). Indeed, the very fact that others interpret our actions and act on those interpretations may have shaped our cognitive systems in ways that make those interpretations more apt to be right (Mameli 2001). Other coevolutionary scenarios have contrasting implications for the chances that our intentional interpretation is a reasonably faithful picture of the facts of wiring-and-connection. Thus our picture of the fundamental dynamic of human social evolution has direct implications for our assessment of the reliability of interpretation.

The evolutionary perspective developed in this book has been an increasingly prominent element of the internal project, but it has been largely missing from the external project. This claim might seem

surprising to those familiar with the last 20 years or so of debate about meaning: debate about the relationship between the folk notion of meaning and the wiring-and-connection facts. For much of this debate has been funneled through a few specific animal examples. It is hard to find a paper on representation that is not ostensibly about frogs and their thoughts about flies. Moreover, on one prominent account, frogs' thoughts are about flies, because it is the biological function of those thoughts to enable frogs to catch flies (Millikan 1989; Papineau 1987; Neander 1995). The facts which make it true that frogs' internal states are about flies are facts about the history of selection on frog ancestors for fly-catching.

This work is important, but it actually illustrates the absence of an evolutionary perspective from the external project. For the evolutionary-historical theory of content is not an evolutionary perspective on *human cognition*. Very few of Millikan's ideas depend on specific empirical claims about primate evolution. For Fodor, Millikan, Neander, and others, frogs and their fly-hunting form a model system rather than part of a theory of the evolution of human cognitive capacities. These theorists deploy the obvious strategy of thinking about a system that is as simple as possible while still manifesting the phenomenon to be explained. Like other model systems, frogs are discussed precisely in the belief that it is possible to abstract away from the differences between the model system and other systems, including the differences in their evolutionary histories. Frogs cannot illuminate the evolutionary history of human cognition. At most, they illustrate very general features of cognition and its evolution in all animals. Frogs are well-chosen as model systems only if (a) the folk notions of meaning, content, and aboutness correctly describe certain connection features of human minds; and (b) those connection features are characteristic of a very large class of animal control systems. In chapter 11.5 I shall explain why this is a bet I am not taking.

In sum, then, this book shares with many others a focus on central themes in contemporary philosophy of psychology: the nature and status of intentional psychology and its relation to a scientific understanding of cognition. It contrasts with most other works in examining intentional psychology through an evolutionary and comparative lens. My focus is not on the neural or the computational mechanisms that realize interpretative capacities. It is on the evolutionary and adaptive mechanisms that assemble intentional agents, and on the specific evolutionary dynamic that built the special kind of intentional agents that we are. I begin by first discussing the evolution of the building-blocks

of intentional systems, and that is the aim of part I of this book. Parts II and III develop my account of the specific features of hominid cognitive evolution.

In the next chapter, I introduce a general framework for thinking about the evolution of cognition, and over the next three chapters I'll use that framework to develop some suggestions about the evolution of belief-like and preference-like states. There are, I shall suggest, reasons to think that we have evolved wiring-and-connection features that are something like, but not perfectly like, beliefs and preferences as portrayed by intentional psychology.