

Chapter 1

The Project

This essay is about commonsense psychology, also known in philosophy as folk psychology and in psychology as theory of mind or mind reading. I opt for the term *interpretation* as the attribution relation between an *interpreter* and a *subject*. I motivate this choice later in this chapter. First, I outline the background of the main topic in section 1, distinguish several stances on it, and extricate the one that matters to interpretation. With the territory so delineated, I preview in section 2 the main argument of the essay and conclude in section 3 with a word about terminological policies.

1 Background

The stances Interpretation is a competence that allows primates to make sense spontaneously and effectively of each other in terms of behavioral dispositions and psychological attributes, such as character traits, emotions, feelings, and attitudes. This essay seeks to contribute to an explanation of the nature of the competence, its history, and its job. Because of its evolutionary rationale, reflected in its job, interpretation has a privileged relation to the mind interpreted. I phrase the point diplomatically because the nature of the relation is much contested. This is why it is helpful from the outset to separate interpretation from other stances on the mind.

There are four stances worth distinguishing: naive, folkloric, ideological, and scientific (Dennett 1987, chapter 3; 1991b; also Bogdan 1991b, 1993a). Interpretation is a sort of *naive psychology*, programmed innately and exercised spontaneously (Humphrey 1988). The elucidation of its nature is the object of this essay. There is also a *naive phenomenology*, which provides unreflective and immediate access to the phenomenal data of mental life. The relation of this phenomenology to interpretation is still subject to inquiry and debate.

The *folklore of the mind* is a collection of popular notions and conceptions, some universal, others specific to a culture. This psychofolklore has many sources: interpretation and naive phenomenology, myth and religion, philosophical and artistic views that spread in a population and survive across generations, linguistic practices and social expectations, and ordinary observations about people and their behaviors that find their way into the general mentality. The notion of soul is folkloric (sources: myth, religion, philosophy), and so are the ordinary observations that people tend to act on their desires and beliefs (source: interpretation), that pain is unpleasant (source: naive phenomenology) and that people tend to believe what they say (sources: linguistic practice, social expectation, interpretation).

The *ideology of the mind* originates in myth, religion, folklore, and philosophy, and encompasses notions and conceptions that construe the mind not as a result of a search for the truth (science) or on the basis of ordinary observation (folklore) but according to religious, moral, social, or political views, implications, or prejudices. An ideological conception comes close to being wishful thinking, though it is not always detrimental. Freud suggested that the notion of an immortal soul may have been a clever concoction to maintain the mental health and motivation of the tribe confronted with the realization of impending death. Such a notion would be ideological. Equally ideological are the a priori notions that the human mind is unique, special, with no animal pedigree. That ideology is independent of interpretation is often seen in the fact that individuals may hold the silliest views about the mind (as the history of folklore and philosophy abundantly testify) while managing quite well in interpreting their conspecifics. Interpretation is one thing; false consciousness about the mind another thing.

The last stance is that of *science*. There are many sciences of the mind, nowadays collected under the label of ‘cognitive science’. Their aim is to figure out and explain the design and operation of the mind, from its neural mechanisms (neuroscience) to its various programs (psychology, linguistics, philosophy of mind). Whatever their angle on the mind, it is their methodology and objectives that radically distinguish the sciences of the mind from each of the other stances.

This essay is exclusively about interpretation as a naive psychology—although the word ‘psychology’ is merely a concession to current usage, for the argument developed below is that interpretation is neither a (naive) *logos* nor literally about the *psyche*. The notion that it is literally

though naively a logos of the psyche draws on preconceptions nurtured by the other stances. This is why keeping these intrusive stances at bay requires constant vigilance.

Motivation There are several reasons why I thought this project worth pursuing. If I were to single out the most motivating ones, I would choose three. The first is the desire to get *to the roots* of interpretation. At least since the emergence of philosophical behaviorism a few decades ago, there has been much talk about interpretation but relatively little effort to uncover its roots and reasons. Some views seemed to get things right, others didn't, but there wasn't a firm sense that one knew why. Dominating the debate were linguistic and introspective intuitions, views about the nature of psychology, semantic analyses of attributional sentences, and the metaphysics of mind. Even if these were the right levels of analysis (which I doubt), we were not told why. Turning to the psychology of interpretation, which rapidly became a new and dynamic research program, one discovers a larger theoretical horizon, robust data, and impressive experiments. And yet one retains the sense that the roots of interpretation remain elusive, not so much for lack of data but for lack of further theoretical probing.

Hence the second reason for this project: the prospect of *proximate impasse*. Current views tend to talk past each other, assigning to interpretation different and often conflicting domains, aims, and skills, or else they debate the same issues with no resolution in sight. Here is a sample of what I mean. Interpretation is (a) innate, no, learned; (b) modular and specialized, no, general-purpose; (c.1) procedural, no, more like an explicit theory; (c.2) neither procedural nor theorylike but rather projection or simulation; (d) naively true of the mind, no, totally false; (e) uniquely human, no, with animal precedents; (f) about the mind, no, about behavior only; and so on. There will be occasions later to label and explain these views. Now I stress their inconclusive oppositions. What is wrong with that, you would ask. Isn't science progressing in this way? Is there a better formula, you ask. The problem (I reply) is not with disagreement and debate but with the *level* of theorizing. The oppositions just noted are all proximate, that is, concerned with programs and their use, not with why the programs are there in the first place, which would be an explanation of the ultimate or evolutionary sort. Some of the questions cited (early in the list) may be settled by further data and insightful proximate research. Other questions (down the list) might not be so easy.

Thus I come to the third reason for this project, the *need for evolution*. I argue in chapter 3 that the domain and job of interpretation cannot be easily figured out without evolution. The competencies (so far) best understood in cognitive science without appeal to evolution, namely, vision and grammar, are *producers* of representations. Interpretation, however, is one level up on the processing ladder, being mostly a *categorization* capability. Such capabilities—from concepts to meaning representations—are notoriously hard to fathom. In interpretation, they are even harder to pin down without asking why they evolved, under what pressures, and in what contexts of selection. Once evolution is brought into the picture, it emerges that the domain and job of interpretation are different from what the mainstream positions hold. Or so I will argue.

Yet before embarking on an evolutionary venture, I acknowledge the formidable obstacles facing it. Antievolutionism, still rampant in philosophy and cognitive science, is animated by two chief prejudices. One is that complex mental abilities—and interpretation is one—did not evolve by natural selection, or even if they did, this fact need not illuminate their design. The other prejudice is that we will never know whether such abilities evolved and, even if they did, we would not know how. Brains do not fossilize; new functions can emerge without precedents; and the primate mind, the only candidate for interpretation, is simply too recent and complex to allow a coherent and verifiable evolutionary story.

Neither prejudice is as substantial as it sounds. But I find their methodological message depressing and rather reactionary, for they tread on mysterianism and lack of firm evidence. Neither is a good reason not to probe and speculate. One learns from mistakes but not from mysteries. This is when a handicap may turn into an asset. My handicap is that I am a philosopher of mind, not a psychologist or evolutionary theorist. My evolutionary ideas and arguments are those of an amateur who took some time to appreciate their force and relevance. Philosophers can be good at brainstorming and integrative speculation—an institutional adaptedness, if you like. The study of mind and interpretation needs that, and this is something appreciated nowadays in neighboring disciplines. Recent avenues of research on interpretation (e.g., its form of theory, innateness, false-belief recognition, simulation) have been suggested or even initiated by philosophers. Good omens to continue in the same spirit. Even if the case for the evolution of interpretation is not as tight or plausible as one may wish, attempting to make the case and show what this reveals about the design of interpretation is a heuristic journey worth making in the

quest for new clues and insights. There is nothing to lose, except some reading time. My sense, though, is that the journey is more than heuristic. The following, very briefly, explains why.

2 The Argument

From why to what At the heart of this essay is the notion that the *why* of interpretation illuminates its *what*, meaning that its *evolutionary reasons* can constrain and even specify its *job*. The quest for evolutionary reasons is a quest for the forces that shaped the goals and tasks of interpretation, or the design of its job. The why-to-what, or rationale-to-job, syllogism goes like this. If interpretation first evolved by natural selection, then it is an adaptation. If an adaptation, then it is a competence or family of skills good at doing something that serves the interpreter's goals and ultimately promotes her biological prosperity. Good-at-doing-something indicates the goals and tasks, or job design, of the competence. The job design of a competence is in turn the best avenue to understanding its operation.

On this view, identifying the selective pressures under which interpretation evolved provides the constraints on what interpretation is good at and hence on what its adaptedness consists of. I distinguish three families of such pressures—*communal* (cooperation, family life), *epistemic* (education, communication) and *political* (manipulation, deception)—and show that they are at work throughout the evolution of interpretation. At each evolutionary turn, these pressures correlate systematically with key properties of the respective form of interpretation. Such correlations not only show that interpretation is the outcome of evolution, which is one major aim of this essay, but also pave the ground for applying evolutionary insights to the analysis of the job of interpretation, which is the other major aim.

For this explanatory strategy to work, I need to identify the evolutionary *environments of selection* as the contexts where the forces of selection produce interpretive skills as adaptations. This identification yields clues to the specificity of the forces involved and of the skills they produce, and also to the common pattern of adaptedness found in the evolution of interpretation as a distinct competence. This pattern is displayed by (what I call) the *goal setting for interpretation*, where the satisfaction of the interpreter's goals depends on figuring out and doing something about the goals of the interpreted; that is what selects for interpretation.

Practice What goal settings make manifest and motivate is that interpretive skills were selected as *effective strategies* to do something about the subject's goals and behaviors when they interfered with or had consequences for those of the interpreter. Interpretation thus evolved as a cognitive instrument with practical import. The most salient insight provided by evolution is that interpretation is practical in two senses: it operates as a practice and has practical aims. It operates as a practice in that it conceives of and handles the subject's psychological attributes and behaviors in terms of effective strategies of engagement, intervention, and reaction, or in short, *involvement*. And it has practical aims because it picks up and categorizes those subject-world relations that allow such effective strategies.

The practical effectiveness of interpretation depends on the interpreter's correctly parsing the subject's relations to and actions on the world into units that afford opportunities for involvement, whether by incremental interception or alteration or influence or reaction, and that also afford opportunities for control of the whole process by fine-tuned expectations, predictions, and feedback. Thus emerges the leading theme of this essay, according to which the interpretive categories of psychological attributes are practical in nature, for they operate as parsers of the sort just sketched. Put simply, to interpret a subject's attributes is to know what to do about them.

To sum up, there are communal, epistemic, and political pressures for practical involvement in the subject's affairs. This translates into selection for skills to influence causally or take advantage of a subject by executing tasks that parse his relations to the world in a practical format. On this line of analysis I propose to explain the phylogenetic and ontogenetic differences in the job design of interpretation in terms of the accord between the epistemic, communal, and political demands of social life and the potential of the interpretive skills to affect the subject practically at joints where effective strategies of involvement are possible and successful. When that accord snaps, evolution ends up changing the rules of the game. The resulting historical pattern would look as follows.

Evolutionary turns I distinguish two major phases in the evolution of interpretation, one called *early* or *situated* interpretation, the other *advanced* or *unsituated* interpretation. The former is a reflex or instinctive interpretation, steeped in the immediacy of perception and behavioral response; the latter a thoughtful or reflective interpretation that ranges

across space, time, and imagined possibilities. The former is shaped mostly by natural selection, the latter mostly by culture. Within each major phase I distinguish several evolutionary turns, sufficiently distinct in terms of epistemic, political, and communal pressures and the skills they pressured for.

The first turn is *natural teleology*. It is the earliest and simplest form of interpretation. It responds to the crude agency of subjects and their basic goals. When pressured to identify specific goals, basic and nonbasic, by clues that reflect more than physiological propensities for fixed behavioral patterns, interpretation turns *psychobehavioral*. When pressured to share information and experiences, interpretation takes a uniquely human *psychosocial* turn, first in an infant version. Although different in terms of evolutionary pressures-skills packages, all these forms of interpretation are situated and reflex in manner. It is only at the *metarepresentational* turn, in early childhood, that interpretation moves beyond its perceptual and behavioral immediacy and becomes unsituated and able to attribute basic propositional attitudes, such as desire, belief, and intention. In disagreement with prevailing views, I think that evolution does not stop at this ontogenetic phase and has one more turn to take. *Reconstructive* interpretation becomes fully reflective, since it is based, on thinking, and takes the form of adult commonsense psychology.

The mind deal Interpretation is a practice evolved in response to distinct selective pressures, first natural and later cultural. It is a practice in the service of the interpreter's goals and ultimately of her biological fitness. As a result, interpretation is not a probe or reader or theory of minds but a pragmatic policy of tracking the subject's relations to the world to the extent that they affect the interpreter. Yet to be effective along these lines, interpretation evolves to factor in and exploit the subject's mental and behavioral propensities and their affordances for practical involvement. This need not yield an explicit knowledge of minds. At no evolutionary stage do interpretive skills generate explicit representations about the subject's mental architectures (functional resources and programs), data structures (in the case of attitudes), or other functional configurations (in the case of affects, feelings, and traits). Nor are any of the latter tracked implicitly but systematically. What interpretation tracks systematically, whether explicitly or implicitly, are the subject's *relations* to the world and their *external relata*. This is the only intentionality that interpretation cares about. Many theorists may agree with this position, in which case

the debate is about how to analyze the mind-world relations picked out in interpretation. My proposal is that the systematic tracking of subject-world relations is motivated and formatted by the practical interests of the interpreter, even in adult human interpretation, where there is much greater freedom to interpret beyond immediate needs. As a result, the interpreter has a practical knowledge of minds, mostly implicit and limited to the aspects that afford practical involvement. The success of such knowledge in various applications (prediction, control, manipulation, explanation) is secured by evolutionary arrangements, natural and cultural, and not by the interpreter's knowledge of minds.

To sum up, then, there are three main and tightly interwoven lines of argument, with ample polemical fuse, to be developed in this essay. One is the evolutionary hypothesis—the selection of interpretation. Another is its practicality—a tool to secure the interests of the interpreter and ultimately her biological prosperity. The third is the subject-world-interpreter relation at the heart of the domain of interpretation.

The plan Part I makes the general case for the evolution of interpretation. Chapter 2 argues that interpretation is a competence selected epistemically, communally, and politically. Chapter 3 holds that interpretation is distinct and, being in the categorization business, needs an evolutionary reading of its goals and tasks. The second part of the book focuses on situated interpretation, whose forms are in the direct grip of natural selection. Chapter 4 surveys the main forms and inquires into the selective pressures responsible for them. Chapter 5 takes its clues from a comparison between selective pressures and the skills they called for, and ventures a hypothesis about the design of the job of early interpretation. This is where I articulate the notion of interpretation as a practice consisting of effective strategies of involvement in the subject's affairs. Against the current consensus, chapter 6 suggests that the metarepresentational turn, when unsituated attitudes come under interpretation, is a transient phase that grounds what evolves next. Part III turns to reconstruction, the adult form of human interpretation. Chapter 7 retraces the chief evolutionary sources of reconstruction: the development of thinking and its cultural regimentation. Drawing on these evolutionary data, chapter 8 explores the job design of reconstruction. In the process, it argues that reconstruction remains practically motivated and operated, a feature reflected in its categorizations and attributions; that content ascriptions map aspects of practical interest to the interpreter into functional roles those aspects play

in the interpreter's thinking and acting; that intentional explanation is a creature of culture, attuned mostly to its norms, not to psychological regularities; and that the evolutionary plausibility of simulation may be more modest than thought by both its proponents and detractors. The final chapter 9 puts two and two together to reach a few conclusions.

Evolutionist credits Given the evolutionist thrust of this work, I salute at this point fellow travelers who have often guided my journey. (Other intellectual debts and points of reference are noted throughout this essay and in further detail in chapter 9, section 3.) My work stands and builds on many evolutionists' shoulders. Following the pioneering insights of Humphrey (1988) and the work of Premack and Woodruff (1978), many psychologists have turned in recent years to animal and child interpretation. It is an indication of the explosion of creative work in the last two decades that it is hard to list all those whose data and hypotheses provided grounds and guidelines for many ideas developed in these pages. Several excellent collections of papers sample these contributions and get due recognition later (Baron-Cohen, Tager-Flusberg, and Cohen 1993; Byrne and Whiten 1988; Carruthers and Smith 1996; Lewis and Mitchell 1994; Moore and Dunham 1995; Whiten 1991a). Although many students of interpretation have gestured toward evolution, few have adopted an explicit evolutionary stance, and even fewer have systematically explored interpretation with the tools of evolutionary analysis. Just doing animal or developmental psychology, even comparatively, is not yet doing evolutionary psychology. The latter requires explaining systematically the history and design of abilities in terms of selective forces or other evolutionary causes. Nothing less will do.

Yet there are notable exceptions. In philosophy, Dan Dennett has not only been the evolutionist par excellence but has also contributed significant evolutionary analyses of interpretation. Besides developing the concept of the intentional stance, so central to interpretation, and boldly defending adaptationism as a method of studying the mind, Dennett has allied the two in a new and influential look at animal interpretation (Dennett 1987). There are a few other (alas, too few) philosophical attempts to inject evolution more thoroughly into accounts of interpretation. I cite in particular those of Bennett (1976, 1991a, 1991b), Clark (1989), Graham (1987), and Levin (1984).

In animal and child psychology, several authors have turned more than occasionally to evolution. I am thinking of Baron-Cohen (1995b), Cheyney

and Seyfarth (1990), Humphrey (1988), Mitchell (1994), de Waal (1982, 1989), Whiten (1993), Whiten and Byrne (1988a, 1988c), and the more programmatic Daniel Povinelli (1993, 1996, Povinelli and Eddy 1996). Last but not least, I acknowledge the influence of the methodological program for evolutionary psychology articulated by Leda Cosmides and John Tooby (Cosmides and Tooby 1987; Tooby and Cosmides 1990b, 1992) and the conceptual tools with which they link the notions of selection and adaptation with the top-down analysis in cognitive science. Much as I have learned from these illustrious precedents, I found some room left for further probing, particularly in the direction of a systematic and comprehensive argument going from the evolutionary rationale of interpretation to the design of its job. I hope this essay will show why this argument is necessary and useful, and why interpretation may have evolved as a practice of involvement, not as mind reading.

3 Words

The term ‘interpretation’ is employed in many fields, from literary theory to hermeneutics, logic, and semantics. Yet the notion I so label is antecedent to all others in the order of history and explanation. Without organisms making sense of each other in contexts where this matters biologically—the Ursense of interpretation, if you like—there would be no other kind of interpretation. If the coevolutionary link between interpretation and primate cognition is validated, as I expect it to be, then the other senses of interpretation can be seen to depend historically on the one deemed the most basic in this work.

My choice also has terminological motivation, for I need a compact and convenient word that takes verb, adjective, and adverb forms easily and naturally; the other candidates don’t. I also get to name the termini of an interpretation relation. The interpreted organism is *subject*, and the one who interprets is *interpreter*. To keep track pronominally of who is who, across species, I stipulate that *she* is the interpreter and *he* the interpreted, unless the context says otherwise. I use the adjective ‘interpretive’ for what pertains to the use or functioning of interpretation, as in ‘interpretive skills or concepts’, and (unless I get confused) reserve the adjective ‘interpretational’ for what pertains to a theory, analysis, or language of interpretation.

Words matter in another crucial respect. The ordinary words ‘belief’, ‘desire’, ‘intention’ (and such like) belong to two equally busy and valid

vocabularies, which I call ‘ambivalent’ and ‘dedicated’ respectively. These vocabularies must be distinguished and the distinction respected, lest confusion engulf us. The *ambivalent* vocabulary does not discriminate between naively cognitivist concepts of forms and outputs of information processing (studied by cognitive science) and purely interpretive concepts. The latter belong to a vocabulary *dedicated* to interpretation. Throughout this essay I stick to the dedicated vocabulary and, whenever possible, replace the ambivalent vocabulary with cognitive-scientific paraphrases. To illustrate, ‘belief’ used dedicatedly means a subject-world relation as categorized by an interpreter; ‘belief’ used ambivalently may mean either the former or (as so often) an information state or process of the sort investigated by cognitive science.

All this may look pedantic, but it pays off methodologically. Mighty philosophical views and many intense debates have suffered from the confusion and begged many questions. Perhaps the worst question begging is to assume that interpretation is about the mind. It could be, but that determination ought to be the conclusion of an arduous inquiry, not its unmotivated starting point. Equally counterproductive is the use of the ambivalent vocabulary—with meanings imported from metaphysics, mental causation, physicalist reduction, and relation to various sciences—to make points about interpretation. The trouble is further compounded when one ambivalently identifies both naively cognitivist and interpretive concepts from the viewpoint of introspection and naive phenomenology—a misguided tactic that goes back to Descartes and the British empiricists and from which philosophy of mind has not yet fully recovered.

The dangers just signaled explain not only a terminological policy but also a methodological policy. Philosophers initiate and conduct arguments from inside a shared though contested territory of assumptions, vocabularies, and problems. Given the way I intend to map out the territory of interpretation, there is not much common ground around which to place the usual philosophical fencing. Few philosophical inquiries have taken the evolutionary route as far as I intend to, and few begin by separating what is going on in the mind (the cognitivist reading) from how the interpreter relates to a subject (the dedicated reading). As a result, playing the good old philosophical game would not be very productive, which is why I opt for a fresh start.