

PRELUDE AND AN OUTLINE

Inside Man is a live exercise in modeling human ways of being—thinking, feeling, acting. “Live” means that the book does not merely introduce *models*; it also attempts to teach—impart, imprint—*modeling* and to produce within the reader the abilities, predispositions, and attitudes of the modeler: a *distancing* from the individual whose behavior is modeled, an *engineering* approach to the model-building process, a (self-)critical approach to the process of model testing and elaboration, and a pedagogical and therapeutic approach to communicating and enacting models—to using them to transform the humans whom they model. This book seeks to make the process and the phenomenon of modeling transparent and explicit, and clarifies why modeling human behavior, thought, and feeling should be an interactive process between the modeler and the modeled. As such, it is situated at the intersection of analytical and computational thinking about rationality, reasoning, and choice with the more recent traditions of action science and action research.

Inside Man is a different sort of book from the kind one typically finds on “how to build a model of humans.” It introduces the reader not only to the nuts and bolts of *models* but also to the life-world, activities, and mental habits of the *modeler*—and it does so without oversimplifying the logical

structures of models themselves. It is different from other books in the field (e.g., Lave and March [1975]; Kreps [1988]; Elster [1989]; Casti [1991]; Gershenfeld [1998]) in that it treats the modeling languages it utilizes as *fundamentally unfinished* and the modeling enterprise as an iteratively and recursively self-correcting activity.

In what follows, the modeler and the modeled are real protagonists—with real wishes, aims, and temptations. The modeled sometimes models the modeler back, and this is something the modeler must keep in mind when building and deploying his models. Here, the modeler's usually privileged epistemic position is never quite secure: Alternative explanations are always possibly and sometimes probably better or more valid than those arising from the model currently being tested. The process of testing a model is not the tidy one whereby “data” supplied by the behavior of the modeled provide dispositive verifications or refutations of hypotheses, but rather one by which the status of data as data must be secured by a process of interaction with the modeled. Moreover, the individual whose behavior or thought is being modeled is available to the modeler for questioning, deliberation, debate, and confrontation; and the modeler will often *need* to avail himself of this opportunity for dialogue.

This is a messy situation. Messy situations can lead to messy outcomes, and it is the *métier* of the modeler to stare the mess in the face without flinching, averting his gaze, or seeking the comfort of “disciplinary” legitimacy and “foundations.” It is just such a “mess” that this book intentionally wades into as a way of highlighting the thoughts, behaviors, and ways of being of the modeler—rather than focusing merely on the cognitive by-products of the modeling activity (i.e., the models themselves).

This does not mean that the models themselves will receive short shrift. They are on prominent display throughout; without them, where would “modeling,” as a life form, be? But the models here do not “steal the show” from the activity of modeling, for one *lives out* models of behavior as much as one *knows of*—or can see the world in terms of—such models. It is this “living out” of the models, and of the activity of modeling itself, to which I want to draw the attention of those who can spare it.

I begin by placing the project of this book in the context of the current discursive landscape of normative, descriptive, and prescriptive “science” and of the wider aims of a phenomenology of being-in-the-world (Chapter 1).

My protagonist modeler is heedful of the cleanliness and precision of axiomatic models of choice but is also attuned to both experimental investigations that can be interpreted as placing the validity of the axiomatic bases in peril and to intuitions and insights that arise from his own first-person access to the life-world of the object of his models—another individual—with the result that the modeler will attempt to forge a connected and responsive dialogue among phenomenology and the descriptive, normative approaches to “modeling human behavior.”

I then proceed along a path best summarized as “differentiation by parts”: Some part of the total function must be kept constant while taking the derivative of the part that is currently in focus. I first consider models of the behavior of a human individual as the outcome of a choice or a set of choices that express or instantiate a set of preferences via more or less complicated deliberation and maximization exercises (Chapter 2). I show how models of behavior as choice can be productively modified and augmented to deal with concerns raised by experimentalists about the putative coherence and invariance of preferences and the consciousness of the chooser. I inquire into ways by which the modeler can justify (to the modeled) the representational choices he makes—his choice of models—and show the importance of the justifiability of a modeler’s model to the modeled in situations where the model’s object is able to field and respond to reasoned arguments.

I shift the focus of attention next to modeling the epistemic states—or states of mind that are about propositions describing states of the world—and consider probabilistic models of the personal degrees of belief that an individual holds regarding the truth value of various propositions (Chapter 3). The modeler inquires into the epistemic states of the subject of his inquiry by first positing a model for the structure of these states and then offering the subject bets on the truth value of propositions that may be true or false depending on states of the world and ways in which these states are rendered into language. The modeler can justify these models to his subject by positing scenarios in which violating their structure places the subject, with high probability, on the losing side of a bet or a sequence of bets. I show how the modeler can use the basic probabilistic framework for measuring personal degrees of belief to model and inquire into the meaning, kinematics, and dynamics of the epistemic states of his subject and into the dependence of these epistemic

states on the subject's visceral and emotional states and the subject's purpose in holding the beliefs that she does.

Chapter 4 introduces models of the processes (e.g., inference) by which the subject of inquiry comes to hold certain beliefs. These models are based on the canonical structure of the problems that a subject can attempt to solve and on a "pico-economics" of thinking, which aims to measure the average, fixed, marginal, and opportunity costs of thinking. The resulting economics of mental behavior allows the modeler to deploy decision-theoretic representations of behavior to represent the dynamics of the subject's "states of mind." The modeler uses this approach to represent the basic processes by which the subject reasons, perceives, and infers (i.e., thinks) and to make inferences and measurements of the states that a thinking subject may be said to undergo, "along the way," while thinking. The model is then extended to include states of being of the subject that, although "intelligent," are nonrepresentational in the following sense: They do not rely on the *explicit* representation by the subject of a specific problem, of the procedures that may be used to solve it, of the objects and entities that together are used to describe the current and desired conditions of the problem solver, or of the space of possible solutions to the problem the subject is attempting to solve.

Although learning can be understood as a problem-directed inferential process, it is treated separately in Chapter 5, wherein the modeler attempts to come to grips with what it means for a subject to "learn" and what it means for the modeler to learn about the ways in which the subject learns. A particular model of learning is proposed—one that navigates between the Scylla of unbounded inductivism and the Charybdis of untethered fallibilism—and is shown to accommodate basic insights and intuitions about learning. This model attempts to salvage both the "betting odds" interpretation of personal degrees of belief and the fallibilist approach to the lawlike generalizations that undergird inference from the known to the unknown.

Because the modeler is in frequent and crucial *dialogue* with the subject of his inquiry, the final chapter (Chapter 6) sets forth a model of dialogical communication that constrains both the modeler and the subject to produce oral behaviors that are coherent, connected, responsive, relevant, and informative contributions to a dialogue. Participation in a true dialogue commits the discussants to certain obligations to justify and explain their utterances

lest the latter should fail to count as “speech acts.” This model of communication enjoins participants in a genuine dialogue from transgressing against a set of constitutive rules about what it means for an exchange of oral noises to constitute a dialogue—on pain of excluding themselves from the dialogue and losing their status as communicators.