

4. *Modeling Human Actors*

It is not possible yet to point to a single theory of human behavior that has been successfully formulated and tested in a variety of settings.

—Elinor Ostrom

In this chapter, we address a question that lies at the core of this book: How do we model people? In many of the models that follow, people will be the fundamental unit of analysis. We will construct models of people who vote, cooperate, start uprisings, participate in fads, invest in retirement accounts, and become addicted to drugs. Within each model, we will have to make assumptions about people. What are their objectives? Do they only care about themselves or are they altruistic? What are their potential actions? How do they choose what action to take, or do they not even have a choice?

We could make ad hoc assumptions for each model. But to do so would invite confusion and miss an opportunity. We would be left with an idiosyncratic set of constructions. Each new model would require new thinking about how people act. The resulting heterogeneity would limit our ability to think across and combine models. We could not be effective many-model thinkers.

The approach we follow stresses coherence along with variety. We will model people as either *rule-based actors* or *rational actors*. Within the set of rule-based actors, we consider those who act based on *simple fixed rules* and those who act based on *adaptive rules*. Someone acting based on an adaptive rule can change her behavior based upon information or past success or because she watches others. As we shall discuss, no bright lines distinguish these cases from one another; an adaptive rule can sometimes be interpreted as a fixed rule, and sometimes rational actions will take the form of simple rules.

How we choose to model people will depend on the context and on our goals. Are we predicting or explaining? Are we evaluating policy actions? Are we trying to design an institution? Or are we exploring? In low-stakes environments, such as

modeling what color coat people buy or whether they stand for an ovation after a show, we will most often assume that people apply fixed rules. When people decide whether or not to cooperate in a venture or to trust another person, we will assume that people learn and adapt. Finally, in high-stakes environments, we will assume that informed, sophisticated people make optimal choices.

Before describing our approach in more detail, we first address some common misconceptions. Many people first encounter formal models of social phenomena in introductory economics courses. Those models often rely on a rudimentary rational-actor model in which everyone is self-interested and capable of optimizing. The model may also assume everyone has the same preferences and income level. Economists then solve for equilibria within these models, enabling them to evaluate the effects of shocks to a market or policy changes. These models, though based on incorrect assumptions, are useful. They help economists to communicate and students to understand.

Based on this experience, many people infer that formal modeling requires a narrow, unrealistic view of human nature, in which people are self-interested and never make mistakes. That is not the case. In fact, not even economists think that to be true. The frontiers of economics consist of models with imperfectly informed, heterogeneous actors who adapt in response to what they learn, and who sometimes, though not always, care about the payoffs to others. The extent to which people exhibit other-regarding preferences also depends on the situation; when donating to charity or volunteering, for example, a person may be more other-regarding than when buying a house.

Nevertheless, the unfortunate impression persists that modeling assumes selfish, unrealistically rational people. We must disabuse ourselves of that view. As an analogy, if you only wade a few steps into the ocean, you might infer that it is shallow. As you swim out farther, you begin to sense the depth. Here, we start near the shore. At times we venture further out and show how models can accommodate other-focused, boundedly rational people.

Whatever assumptions we do make, we cannot escape their implications. We are tied to the mast of logical coherence. We cannot manufacture implications. If we assume strong social influences in consumer choices, our model will produce a handful of products with large market shares. If we assume people obtain information through networks, then people who fill structural holes will possess power.

The remainder of the chapter begins with an overview of some challenges of modeling people: we are diverse, socially influenced, mistake-prone, purposive, adaptive, and possessed of agency. We cannot include all of these characteristics in a single model without creating a complicated mess, so we must pick and choose.

If heterogeneity matters little, perhaps we assume identical agents. If the problem is simple or the people sophisticated, perhaps we assume people do not make mistakes.

We next describe the rational-actor model and discuss its theoretical underpinnings and the justifications for its use despite its descriptive inaccuracy. We conclude that whether the rational-actor model functions as a gold standard, a straw man, or somewhere in between depends on our model's purpose. The rational actor will be less successful at predicting human behavior than as a tool for communicating, evaluating actions, and designing policies.

We then show how we can add psychological biases and altruistic preferences onto the standard rational-actor model. The choice of whether to include a bias or a concern for others rests again on what we are studying. Some human biases such as loss aversion and presentist bias—caring more about delays today than in the future—may be necessary to include in some instances. For example, those assumptions may be important for models of retirement savings or riots. The assumptions may be less important for models of driving behavior or disease transmission.

In the fourth section, we describe rule-based behavior. This category of models has the advantage of being both flexible—any behavior we can write down as a rule is fair game—and tractable. We need only encode that behavior in a computer program, an agent-based model, and watch what unfolds. That freedom comes with responsibility. As we can choose any behavioral rule, we must guard against ad hoc assumptions. In some cases, behavioral rules can be justified as optimal behavior given an objective function, though this will not always be the case.

The chapter concludes with a reconsideration of the value of rationality as a benchmark behavior. Even if people do not optimize, they do adapt to changing circumstances and new knowledge. That observation produces a conundrum of sorts. If we design an institution or policy based on the assumption that people have a bias or that they act in ways not in their self-interest, we run the risk that people will change their behaviors. People may be fooled once, but they may be harder to fool two or three times. We need not conclude that rationality is the only plausible assumption, but the logic does argue for rationality as a relevant benchmark. Logic also supports considering simple rules of behavior as well, as a lower bound on rationality. And, in modeling any given situation, we might apply any number of adaptive and psychological rules as a way for us to explore the giant space in between those extremes.

The Challenge of Modeling People

Modeling people presents a challenge because while models require low-dimensional representations, people defy simple characterizations. People are *diverse*, we are *socially influenced*, we are *error-prone*, we are *purposive*, and we *learn*. In addition, people *possess agency*—we have the capacity to act.

By way of contrast, physical objects, such as carbon atoms and billiard balls, exhibit none of these six properties. Carbon atoms lack diversity (though they can occupy heterogeneous positions within compounds, such as in propane). Carbon atoms never violate the laws of physics nor do the lead purposive lives. They do not change their behaviors based on past experiences. They lack agency; they do not decide to lead uprisings or switch careers. Hence the oft-repeated quip by social scientists: how difficult physics would be if electrons could think. Physics would be even harder if electrons could write models.

We can start with the problems created by diversity. People differ in our preferences, in our capacity to act, in the social networks we form, in our levels of altruism, and in the level of cognitive attention we allocate to actions. Modeling would be easier if everyone were the same. Sometimes we rely on statistical logic and assume that the behavioral diversity cancels. For example, we might construct a model that predicts charitable donations as a function of income. For a given income level and tax rate, some people may be more altruistic than we assume and others may be less. If the deviations from the model average out (and in Chapter 5 we cover models of distributions that explain why they might), then our model may be accurate. This canceling out of diversity will not occur unless actions are independent. When behavior is socially influenced, extreme actions can create spillovers. This occurs when political activists energize voters. We will encounter this effect of diversity when we model riots.

Whether or not mistakes cancel in the aggregate depends on the context. Errors that result from a lack of cognitive attachment may be random and independent. Errors that arise from cognitive biases may be systematic and correlated. For example, people may overweight recent events and recall narratives better than statistics. A shared bias like this will not cancel out.

The next challenge relates to what people desire. A central challenge in writing models of people will be making an accurate assessment of their goals and objectives. Some people desire wealth and fame. Others want to contribute to the betterment of their communities and the world. In the rational-actor model, we represent a person's payoff directly in the form of a function. In rule-based models, purposes are more implicit. A behavioral rule in which people seek to live in an in-

tegrated neighborhood but move out of a neighborhood if the percentage of people who share their racial identity falls below 10% embeds certain beliefs about what people desire.

The final challenge to modeling people results from the fact that people have agency: the ability to take action, to change what we do, and to learn. That said, in some contexts, people may be better characterized as creatures of habit. Actions may be outside our control. Few people choose to be addicted to opioids or to be poor. Yet people take actions that produce those outcomes.

Often, when people take actions that produce bad outcomes, they adapt their behavior. We can capture this by including learning in our models. How people learn varies by context. When learning how many hours they need to study for an exam in order to get a good grade, or how many times a week they need to exercise, people may learn based on individual experiences and introspection. When learning what grocery store to visit or whether to contribute to a charity, people may learn through observing others. In Chapter 26, we show how in non-strategic contexts, learning generally works. People learn the best action. We also show that in strategic contexts, which we model as games, all bets are off. Neither individual nor social learning necessarily produces good outcomes.

Each of these six characteristics are potential model features. If we include a feature, we must decide how much of it to include. How diverse do we make our actors? How much social influence do we include? Do people learn from others? How do we define objectives? How much agency do people possess? We may possess less agency than we believe. Jonathan Haidt describes our lack of agency with his metaphor of the rider and the elephant. “The image I came up with for myself, as I marveled at my weakness, was that I was a rider on the back of an elephant. I’m holding the reins in my hands, and by pulling one way or the other I can tell the elephant to turn, to stop, or to go. I can direct things, but only when the elephant doesn’t have desires of his own. When the elephant really wants to do something, I’m no match for him.”¹ Sometimes we do ride the elephant. Sometimes we do not. No single approach to modeling humans will be appropriate in all settings, so we model humans in a variety of ways.

The Rational-Actor Model

The *rational-actor model* assumes that people make optimal choices given a payoff or utility function. These actions can be *decisions*, where the payoff depends only on the individual’s own action, or they can take place within a *game*, where payoffs depend on what others do. In a game with simultaneous choices or with incomplete

information, the rational-actor model also specifies *beliefs* about what the other actors will do.

Rational-Actor Model

An individual's preferences are represented by a mathematical utility or pay-off function defined over a set of possible actions. The individual chooses the action that maximizes the function's value. In a game, that choice may require beliefs about the actions of other players.

As an example, we construct a primitive rational-actor model of an individual's decision for how much income to allocate to housing. The model characterizes her utility as a function of housing and all other consumption, with the latter including food, clothing, and entertainment (see box). The model assumes a price for housing and a price for all other goods. The model is far from realistic. It treats all housing the same. And, it lumps all other goods into one category called consumption and prices them all the same. We can set those inaccuracies aside for the moment, as the purpose of the model is to explain the proportion of income spent on housing.

A Rational-Actor Model of Consumption

Assumption: An individual's utility from general consumption, C , and housing, H , can be written as follows:

$$U(C, H) = C^{\frac{2}{3}} H^{\frac{1}{3}}$$

Result: A utility-maximizing individual (a rational actor) spends exactly one-third of her income on housing.²

In the model, the proportion of income a person spends on housing does not depend on the price of housing or on income. Both results are reasonable approximations of the data.³ Other than people at the extremes of the income distribution, most people spend about a third of their income on housing. The finding has policy implications: if housing prices fall by 10%, people will buy 10% more housing. The finding also provides a justification for assuming identical agents. If people spend a fixed percentage of income on housing, total spending on housing depends only on average income.

Using a utility function makes our models analyzable, testable, and tractable. We can estimate the functions with data, we can derive optimal actions, and we

can ask “what if” questions by changing parameter values. In assuming a utility function, we imply a coherence to preferences that may not exist. For preferences to be representable by a utility function, they must satisfy certain axioms. Theorems that prove the existence of utility functions assume a set of alternatives along with a *preference ordering*. Imagine that we can list all possible bundles of goods a person might buy. A preference ordering ranks these bundles from most to least favored. A person might prefer coffee with milk to tea with lemon; if so, she ranks the bundle {coffee, milk} above the bundle {tea, lemon}.

A utility function represents preferences if it assigns a higher value to bundle *A* than bundle *B* if and only if the preference ordering ranks *A* above *B*. For preferences to be consistent with a utility function they must satisfy completeness, transitivity, independence, and continuity. *Completeness* requires that the preference ordering is defined over all pairs of alternatives. *Transitivity* rules out preference cycles. If someone prefers bundle *A* to bundle *B* and bundle *B* to bundle *C*, she must also prefer *A* to *C*. In other words, if a person prefers apples to bananas and bananas to cheese, then she must also prefer apples to cheese. This condition rules out inconsistent preferences.

Independence requires that people evaluate the outcomes in a lottery separately. A lottery is a probability distribution over alternatives, such as a 60% probability of *A* and a 40% probability of *B*. Preferences satisfy independence if when *A* ranks above *B*, then in any lottery that includes *B* as an outcome, the person prefers an alternative lottery in which we replace *B* with *A*. Independence rules out strong risk aversion. A risk-averse person might rank a trip to New Orleans over a trip to Disney World but prefers knowing for certain that he will be going to Disney World over entering a lottery that sends him to New Orleans with probability $\frac{1}{2}$ and to Disney World otherwise. The final condition, *continuity*, requires that if a person ranks *A* above *B* and *B* above *C*, then there exists a lottery in which she gets *A* with probability *p* and *C* with probability $(1-p)$ that she likes exactly as much as *B*. This assumption also rules out strong preferences for certain outcomes.⁴

The assumptions of independence and transitivity, which people violate, on top of the dubious claim that people optimize leads many to question the widespread use of the rational-actor model, particularly by economists. Yet there exist good reasons to assume rationality. First, people may act “as if” they optimize. They may apply rules that produce nearly optimal behavior. When people play pool, catch a Frisbee, or drive a car, they do not write down mathematical equations. The mathematics required to time a leap to catch a Frisbee would overwhelm almost anyone. Yet people catch Frisbees. So, by the way, do dogs. Thus, both people and dogs act as if we solve a difficult optimization problem.

This same logic extends to high-dimensional problems. An analysis of the actions of Harold Zurcher, the superintendent of maintenance for the Metropolitan Bus Company in Madison, Wisconsin, found that he made near optimal decisions about when and whether to replace bus engines.⁵ Though Zurcher did not write down any mathematics, he relied on heuristics. Those heuristics, informed by experience, meant that he acted (almost) as if he were a rational actor.

Second, even if people do make mistakes, in cases where a situation is repeated, our capacity to learn should push us toward optimal actions. Third, in cases where the stakes are large, people should put in the time and energy to make near-optimal choices. People may overpay 30% for coffee or AAA batteries, but they do not overpay 30% for cars or houses. The claim that learning and higher stakes increase rationality has ample empirical and experimental support.⁶

A fourth reason for adopting the rational-actor model, paradoxically, is that it simplifies the analysis. Most utility functions will have a unique optimal action. A person can behave suboptimally in thousands of ways. Saying that people do not optimize opens an enormous box of possibilities. If we assume people make choices to maintain their identities or to enforce cultural norms, we may lack a single clear answer. Rational choice is not realistic, but realism comes at the cost of messiness. An answer, even if it is known to be wrong, can be more useful than having no answer at all, as it allows us to bring the model to data and to work through the effects of changes in variables.⁷

Arguments for Rational Choice

“As if”: Intelligent rule-based behavior may be indistinguishable from optimal or near-optimal behavior.

Learning: In situations that are repeated, people should approach optimal behavior.

Large stakes: On important decisions, people gather information and think slowly.

Uniqueness: Optimal behavior is often unique, making the model testable.

Consistency: Optimal behavior creates a consistent model. If people learn the model, they will not change their behavior.

Benchmark: Optimal behavior provides a benchmark as an upper bound on people’s cognitive abilities.

Fifth, the rational-actor assumption guarantees internal consistency. If a model assumes suboptimal behavior and the model is in the public domain, the model can be learned. People can change their behavior. They might not optimize, but any assumption other than optimality is subject to the criticism that it is not consistent. We return to this point in the discussion at the end of the chapter.

Last, and some would argue most important, rationality can function as a benchmark.⁸ When designing a policy, making a prediction, or choosing an action, we should consider what would happen if people had rational preferences and optimized. That exercise may point to flaws in our thinking. We should also be open to the possibility that the exercise will lead us to conclude that the rational-actor model does not apply and that we should privilege other models instead. To this list we might add a seventh reason: many-model thinking. If people apply many models, they are less likely to make mistakes.

Psychological Biases

The rational-actor model has been challenged by psychologists, economists, and neuroscientists, who note that it does not match up with how humans behave. Empirical findings from laboratory and natural experiments show that people suffer a variety of biases, including a status quo bias. We ignore base rates when making probability calculations, we attach too much significance to sure things, and we are loss-averse.

As researchers begin to link behavior and beliefs to processes within the brain, evidence of hardwired biases becomes more compelling. For example, neuroeconomics uses brain imaging studies to study economically relevant behaviors such as attitudes toward risk, levels of confidence, and responses to information.⁹ Kahneman argues that what we know so far supports making a distinction between two types of thinking: quick, intuitive rules (*fast thinking*) and deliberate contemplation (*slow thinking*). Fast thinking is more likely to be subject to the aforementioned biases.¹⁰ In the long run, we may be able to infer some behavioral patterns from brain structures, but we should keep in mind that the brain has tremendous plasticity. It is capable of overcoming biases by thinking slowly.

Further, we should be cautious about accepting as universal any finding documented in just a handful of studies. Many psychological findings have not proven robust. A recent study failed to replicate half of one hundred findings published in leading psychology journals.¹¹ Furthermore, replicability need not imply universality. Subject pools for many studies lack economic and cultural diversity.¹² We might expect that more diverse subject pools would produce fewer behavioral regularities, providing even greater reason to avoid generalizations about behavior.

Last, in attempting to make more realistic models, we must keep tractability in mind. More-realistic models may require more-sophisticated mathematics.¹³ None of these concerns is so persuasive to suggest abandoning models with psychologically realistic behaviors, but collectively they imply that we proceed with caution and emphasize well-documented behavioral regularities.

Two deviations that have been replicated many times are loss aversion and hyperbolic discounting. *Loss aversion* states that people are risk-averse over gains and risk-loving over losses. Kahneman and Tversky refer to this general theory of behavior as *prospect theory*.¹⁴ Loss aversion does not at first appear irrational, but it implies that people choose different actions when an identical scenario is presented as a potential loss as opposed to a potential gain.

For example, people prefer winning \$400 for certain rather than entering a lottery with an even chance of winning \$1,000. Yet they will enter a lottery with an even chance of losing \$1,000 rather than lose \$600 for certain. This same inconsistency extends to nonmonetary domains. Doctors given choices framed as gains are risk-averse. When choices are presented as losses, doctors take more risks.¹⁵

Prospect Theory: Example

Gain Framing: You have two options:

Option A: Win \$400 for certain.

Option B: Win \$1,000 if a fair coin comes up heads, and \$0 if tails.

Loss Framing: You are given \$1,000 and have two options:

Option $\hat{A}$: Lose \$600 for certain.

Option $\hat{B}$: Lose \$0 if a fair coin comes up heads, and lose \$1,000 if tails.

A and $\hat{A}$ are equivalent as are B and $\hat{B}$. According to prospect theory, more people choose A and $\hat{B}$.

Hyperbolic discounting implies stronger discounting of the immediate future. Standard economic models assume *exponential discounting*, a constant rate at which people discount the future. A person with an annual discount rate of 10% values \$1,000 next year as equal to \$900 today. She would discount next year's money by 10% for every year into the future. Evidence shows that most people do not discount the future at a fixed discount rate. Instead, they suffer from an *immediacy bias*: they discount the near future far more than the later future.¹⁶ For example, if you ask people whether they would prefer \$9,500 twenty years from now or \$10,000 twenty years and one day from now, almost everyone will wait one more

day for the extra \$500. If you ask those same people if they would prefer \$9,500 today or \$10,000 tomorrow, many will take the \$9,500 now. This is an example of immediacy bias.¹⁷

That bias produces time-inconsistent behavior. One year from now, most people prefer waiting one more day and taking the \$10,000. Such preferences are not logically consistent. Hyperbolic discounting has been put forward as a reason why people run up credit card debts, eat unhealthy foods, have unprotected sexual relations, and fail to save for retirement.

In summary, depending on how we will use our model, we may choose to assume loss aversion and hyperbolic discounting given that these assumptions appear to better match behavior for most people. The main reason not to do so would be if they complicate the model without qualitatively changing what we find, or if by assuming hyperbolic discounting our model produces unrealistic behavior.

Rule-Based Models

We now turn to rule-based models.¹⁸ While optimization-based models assume an underlying utility or payoff function that people maximize, rule-based models assume specific behaviors. A rule-based model might assume that in an auction, a person will bid 10% less than her true value for an item, or that a person will copy a friend's action if that friend consistently receives higher payoffs. Many people equate optimization-based models with mathematics and rule-based models with computation. The distinction between optimization-based models and rule-based models is not as clean as might be thought. Think back to our model of housing consumption. Optimal behavior took the form of a simple rule: spend one-third of income on housing. The key difference between the two approaches lies in their foundational assumptions. In an optimization-based model, preferences or payoffs are fundamental. In a rule-based model, the behavior is fundamental.

Behavioral rules can be fixed or adapt. A fixed rule applies the same algorithm at all times. Just as rational-choice models provide an upper bound on people's cognitive abilities, fixed-rule models provide a lower bound. A common fixed rule in markets, the *zero intelligence rule*, accepts any offer that produces a higher payoff. It never takes a stupid (i.e., utility-reducing) action. Suppose we want to gauge the efficiency of a one-sided market design in which sellers post take-it-or-leave-it offers for some good. A seller following a zero intelligence rule would randomly pick a price above her value. A buyer would purchase any good with a price below her value. When we encode those behaviors in a computer model, we find that in markets zero-intelligence traders produce nearly efficient outcomes. Thus, exchange markets do not need rational buyers and sellers to function well.¹⁹

An adaptive rule switches among a set of behaviors, evolves new behaviors, or copies the behaviors of others. It takes these actions in order to improve the payoff. Therefore, unlike fixed rules, adaptive rules require a utility or payoff function. Advocates of this approach argue that within any situation people tend toward simple and effective rules, and that if this is what people do, we should model them in that way.²⁰ Though rule-based models make no explicit assumption about rationality, adaptive-rule models exhibit ecological rationality—better rules come to predominate.²¹

To explain how adaptive-rule models operate, we describe the *El Farol model* of self-organized coordination.²² El Farol is a nightclub in Santa Fe, New Mexico, that features dancing every Tuesday night. Each week, a population of 100 potential dancers decide whether to go dance at El Farol or stay home. All 100 people like to dance, but they do not want to go if the club is too crowded. The model assumes a stark form of preferences. A person earns a payoff of zero from staying home, a payoff of 1 from attending if 60 or fewer people attend, and a payoff of -1 from attending when more than 60 people attend.

If we construct a fixed-rule model, anything might result. For example, if we assigned everyone the rule “go the first week; if more than 60 people attend, do not go the next week; and then go the following week,” the El Farol would have 100 attendees the first week, zero attendees the second week, and then 100 attendees the third week. The El Farol model creates adaptive rules by endowing each person with an ensemble of rules. Each rule tells the individual whether or not to go to El Farol. The rules take several forms. Some are fixed rules: go every other week. Others are based on trends in the number of people who attended El Farol in recent weeks. One rule might predict that the number of people that show up this week will be the same as last week. If fewer than 60 people attended last week, that rule would tell the person to go this week.

An adaptive behavioral rule model might assign a score to each rule equal to the percentage of weeks for which it gave correct advice. Each individual might then use the model in her ensemble with the highest score. The best rule will vary over the course of the weeks. Simulations of this type of model find that if individuals possess a large ensemble of rules, then approximately 60 people attend each week: coordination emerges without any central planner. In other words, the system of adaptive rules self-organizes into nearly efficient outcomes.

El Farol Model: Adaptive Rules

Each of 100 individuals decides independently whether or not to go to El Farol every week for a year. An individual who goes to El Farol earns a payoff of 1 if 60 or fewer people attend and a payoff of -1 otherwise. An individual who does not go to El Farol earns a payoff of zero.

Each individual has an ensemble of rules to decide whether to attend. These rules can be fixed or contingent on recent past attendances. Each week, each individual follows the rule in his ensemble that, if followed, would have produced the highest payoff in the past.

We can interpret behavior within adaptive-rule models, like the El Farol model, within the *micro-macro loop* (see figure 4.1). At the micro-level, a set of individuals take actions (denoted by the a_i 's) according to rules. These rules create macro-level phenomena (denoted by Macro_1 and Macro_2), as represented by the upward arrows. In the El Farol problem, the macro-level phenomena are the sequences of past attendances. The downward arrows represent how these macro-level phenomena feed back into the behaviors of the individuals. In the El Farol model, each person may be applying a different rule. If the rules people apply produce a crowded El Farol four weeks in a row, then rules that tell people to attend less often will produce higher payoffs. As people switch to those rules, fewer people will attend. The micro-level rules produce a macro-level phenomenon (over-attendance) that feeds back to the micro-level rules.

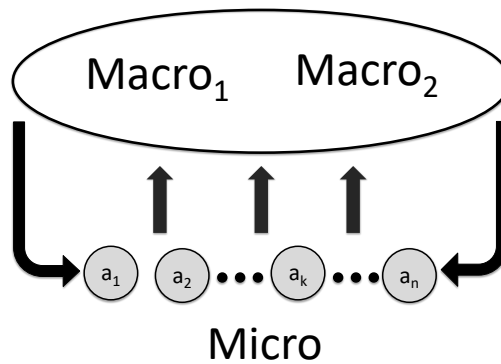


Figure 4.1: The Micro-Macro Loop

Cognitive Closure, a Big Question, and Many Models

The micro-macro loop elucidates a central tension as to how smart to make our agents. Should people infer all consequences of their actions? The loop also hints at a larger question that we encounter throughout the book as to what class of outcome a model produces: Does it go to equilibrium, produce randomness, create a cycle, or generate a complex series of outcomes?

We start with the question of how smart to make our agents. Suppose that we believe that individuals possess only modest cognitive abilities, so we build a model with *zero intelligence agents*. Their actions aggregate to produce aggregate macro-level phenomena. If the macro level produces efficient or nearly efficient outcomes, as we noted was the case with a one-sided market of buyers and sellers, then we may be justified in our assumption. An easy-to-follow fixed rule produces good outcomes. People would have little incentive to expend effort developing more sophisticated rules.

The tension arises when our model produces inefficient or even lousy macro-level outcomes. Such could be the case in the El Farol model, where a common fixed rule could lead to a cycle in which El Farol was overcrowded with dancers one week and empty the next. Confronted with an inefficient outcome, we might think that people would adapt. They might experiment. They might think through the logic of the situation to formulate a new action. If we follow that logic to its extreme and assume a low cost of thinking, then we find ourselves advocating the rational-actor model. Any person not behaving optimally could do better. While that is true, people also have to be able to formulate that better action.

This leads to a big question: What class of outcomes does the model produce? We have four options: equilibrium, cycles, randomness, or complexity. The class of outcome will matter for deciding how seriously we take the argument that people should learn their way to equilibrium. First, if the model produces randomness at the macro level, the individuals probably cannot learn anything. Our model is fine. A similar logic applies to models that produce complex patterns. In these cases, we would assume that people continue to adapt new rules, but we would not necessarily assume that they can choose optimally. To the contrary, the complexity of the macro-level phenomena makes optimal responses implausible. People would be more likely, as in the El Farol model, to confront complexity with an ensemble of simple rules.

The models that produce cycles or equilibria create a stationary environment. We therefore might expect that people can learn—that no one would continually take a suboptimal action. As an example, suppose we have a traffic model in which

everyone chooses a route to work using a fixed rule. In our model, the traffic system settles into an equilibrium. In that equilibrium, one of the individuals, Layne, spends 75 minutes each morning traveling from Calabasas to downtown Los Angeles. Given the equilibrium, if Layne took side streets through Topanga Canyon, her trip would take only 45 minutes. Given the value of an extra 30 minutes per day and the frequency with which people in Los Angeles talk about traffic, Layne would likely find the shorter route. She has no shortage of methods for finding it. She might use a route recommender, talk to a neighbor, or experiment.

Thus, if our model produces an equilibrium (or a simple cycle) and that equilibrium is not consistent with optimizing behavior, then our model suffers a logical flaw. If people have a better action available to them, they should figure it out. They should learn. Notice that we need not assume optimal behavior in order to reach the equilibrium. People could follow simple rules and produce an equilibrium in which no one can benefit by changing her action. At that equilibrium, it would look “as if” people are optimizing, because they are. Again, that logic need not apply for complex or random outcomes. If traffic patterns in Los Angeles produce a complex sequence of traffic slowdowns and jams, we have little reason to believe that Layne selects an optimal route each day. She almost surely cannot.

If adaptive rules that can adopt any action produce an equilibrium, then the equilibrium must be consistent with behavior by optimizing agents. If those same adaptive rules produce complexity, the agents’ behavior need not be optimal. We can restate this idea as follows: optimal behavior may be an unrealistic assumption, particularly in complex situations. On the other hand, if a system produces a stable outcome in which a person has better actions, she will probably figure out a better action to take.

An extension of this logic applies to policy interventions. Suppose that we use data to estimate people’s behavioral rule—say, the likelihood a person shows up at a hospital’s emergency room during lunch hour for minor health issues. If we assume a fixed rule, we might enlarge the size of our facility so that people do not have to wait. If people continue to follow that fixed rule, we have a new equilibrium with short midday wait times. However, with new, shorter wait times, people who had not been going to the emergency room for sprained ankles or chest colds may now decide to go. That equilibrium relies on people choosing suboptimal actions, such as not going to an emergency room even though they would not have to wait. If people learn, we cannot rely on past data to predict outcomes under a policy change. This insight, known as the *Lucas critique*, is a variant of *Campbell’s law*, which states that people respond to any measure or standard in ways that render it less effective.²³

The Lucas Critique

Changes in a policy or the environment likely produce behavioral responses by those affected. Models estimated with data on past human behaviors will therefore not be accurate. Models must take into account the fact that people respond to policy and environmental changes.

As should be clear at this point, there exists no best solution for how to model people. How rational we make them or how adaptive we make their rules depends on the circumstances. We should exercise our best judgment in each situation. Given the uncertainties, we should err on the side of more models rather than fewer.

Even if we are predisposed to dismiss rational-choice models as unrealistic, we must recognize their tractability, their capacity to reveal the directional forces of incentives, and their value as a benchmark. Simple rule-based behaviors, such as zero intelligence, are also unrealistic. Though wrong, they can be of use. They are easy to analyze and can reveal how much intelligence matters in a given setting.

Human behavior occurs within the extremes of zero intelligence and full rationality, so it makes sense to construct models in which individuals adapt using rules. Those rules should take into account the fact that people vary in their cognitive attachment and capabilities within a domain. We should therefore expect behavioral diversity. We might also expect some consistency within groups. This too can be included in models.²⁴

In sum, given the complexities involved in modeling humans, we have abundant reasons to apply multiple diverse models. We may not be able to predict exactly what people will do, but we may be able to identify the set of possibilities. If we can, we have benefited from constructing models because we know what could happen.

We conclude with a plea for humility and empathy. In constructing models of people, a modeler must be humble. Given the challenges of diversity, social influence, cognitive errors, purpose, and adaptation, our models will inevitably be wrong, which is why we take a many-model approach. Austere models of behavior fit some situations well and allow us to focus on other aspects of the environment. Richer behavioral models will be more appropriate when we have better data. We must maintain modest expectations. People are diverse, purposive, adaptive, biased, and socially influenced, and we possess a degree of agency. How can we not expect any single model of human behavior to be wrong? It must be. Our aim is to construct many models that as an ensemble will be useful.