Learning

Macy, Michael W.; Benard, Steve; Flache, Andreas

Published in:
Simulating Social Complexity

DOI:
10.1007/978-3-319-66948-9_20

Publication date:
2017

Chapter 20
Learning

Michael W. Macy, Steve Benard, and Andreas Flache

Abstract Learning and evolution are adaptive or "backward-looking" models of social and biological systems. Learning changes the probability distribution of traits within an individual through direct and vicarious reinforcement, while evolution changes the probability distribution of traits within a population through reproduction and selection. Compared to forward-looking models of rational calculation that identify equilibrium outcomes, adaptive models pose fewer cognitive requirements and reveal both equilibrium and out-of-equilibrium dynamics. However, they are also less general than analytical models and require relatively stable environments. In this chapter, we review the conceptual and practical foundations of several approaches to models of learning that offer powerful tools for modeling social processes. These include the Bush-Mosteller stochastic learning model, the Roth-Erev matching model, feed-forward and attractor neural networks, and belief learning. Evolutionary approaches include replicator dynamics and genetic algorithms. A unifying theme is showing how complex patterns can arise from relatively simple adaptive rules.

Why Read This Chapter?
To understand the properties of various individual or collective learning algorithms and be able to implement them within an agent (where evolution is considered as a particular kind of collective learning).

M.W. Macy (✉)
Department of Information Science, Cornell University, Ithaca, NY 14853-7601, USA
e-mail: mwmacy@cornell.edu

S. Benard
Department of Sociology, Indiana University, Bloomington, IN, USA

A. Flache
Department of Sociology, ICS, University of Groningen, Groningen, The Netherlands

B. Edmonds, R. Meyer (eds.), *Simulating Social Complexity*,
Understanding Complex Systems, https://doi.org/10.1007/978-3-319-66948-9_20

20.1 Introduction

Evolution and learning are basic explanatory mechanisms for consequentialist theories of adaptive self-organization in complex systems.[1] These theories are consequentialist in that behavioral traits are selected by their outcomes. Positive outcomes increase the probability that the associated trait will be repeated (in learning theory) or reproduced (in evolutionary theory), while negative outcomes reduce it. Explanatory outcomes might be rewards and punishments (in learning theory), survival and reproduction (in evolutionary models), systemic requisites (in functionalism), equilibrium payoffs (in game theory), or the interests of a dominant class (in conflict theory).

An obvious problem in consequentialist models is that the explanatory logic runs in the opposite direction from the temporal ordering of events. Behavioral traits are the explanandum and their outcomes the explanans. This explanatory strategy collapses into teleology unless mechanisms can be identified that bridge the temporal gap. While expected utility theory and game theory posit a forward-looking and analytic causal mechanism, learning and evolution provide a backward-looking and experiential link. In everyday life, decisions are often highly routine, with little conscious deliberation. These routines can take the form of social norms, protocols, habits, traditions, and rituals. Learning and evolution explain how these routines emerge, proliferate, and change in the course of consequential social interaction, based on *experience* instead of calculation. In these models, *repetition*, not prediction, brings the future to bear on the present, by recycling the lessons of the past. Through repeated exposure to a recurrent problem, the consequences of alternative courses of action can be iteratively explored, by the individual actor (learning) or by a population (evolution).

Backward-looking rationality is based on rules rather than choices (Vanberg 1994). A choice is an instrumental, case-specific comparison of alternative courses of action, while rules are behavioral routines that provide standard solutions to recurrent problems. Rules can take the form of strategies, norms, customs, habits, morals, conventions, traditions, rituals, or heuristics. Rule-based decision-making is backward-looking in that the link between outcomes and the actions that produce them runs backward in time. The outcomes that explain the actions are not those the action will produce in the future; they are the outcomes that were previously experienced when the rule was followed in the past.

Learning alters the probability distribution of behavioral traits within a given individual, through processes of direct and vicarious reinforcement. Evolution alters the frequency distribution of traits within a population, through processes of reproduction and selection. Whether selection operates at the individual or population level, the units of adaptation need not be limited to human actors but may include larger entities such as firms or organizations that adapt their behavior

[1]Much of the material in this chapter has been previously published in Macy (1996, 1997, 1998, 2004) Macy and Flache (2002), and Flache and Macy (2002).

in response to environmental feedback. Nor is evolutionary adaptation limited to genetic propagation. In cultural evolution, norms, customs, conventions, and rituals propagate via role modeling, occupational training, social influence, imitation, and persuasion. For example, a firm's problem-solving strategies improve over time through exposure to recurrent choices, under the relentless selection pressure of market competition. Suboptimal routines are removed from the repertoires of actors by learning and imitation, and any residuals are removed from the population by bankruptcy and takeover. The outcomes may not be optimal, but we are often left with well-crafted routines that make their bearers look much more calculating than they really are (or need to be), like a veteran outfielder who catches a fly ball as if she had computed its trajectory.

20.2 Evolution

Selection pressures influence the probability that particular traits will be replicated, in the course of competition for scarce resources (ecological selection) or competition for a mate (sexual selection). Although evolution is often equated with ecological selection, sexual selection is at least as important. By building on partial solutions rather than discarding them, genetic crossover in sexual reproduction can exponentially increase the rate at which a species can explore an adaptive landscape, compared to reliance on trial and error. Paradoxically, sexual selection can sometimes inhibit ecological adaptation, especially among males. Gender differences in parental investment cause females to be choosier about mates and thus sexual selection to be more pronounced in males. An example is the peacock's large and cumbersome tail, which attracts the attention of peahens (who are relatively drab) as well as predators. Sexually selected traits tend to become exaggerated as males trap one another in an arms race to see who can have the largest antlers or to be bravest in battle.

Selection pressures can operate at multiple levels in a nested hierarchy, from groups of individuals with similar traits down to individual carriers of those traits, down to the traits themselves. Evolution through group selection was advanced by Wynne-Edwards (1962, 1986) as a solution to one of evolution's persistent puzzles—the viability of altruism in the face of egoistic ecological counterpressures. Pro-social in-group behavior confers a collective advantage over rival groups of rugged individualists. However, the theory was later dismissed by Williams in *Adaptation and Natural Selection* (Williams 1966), which showed that between-group variation gets swamped by within-group variation as group size increases. Moreover, group selection relies entirely on differential rates of extinction, with no plausible mechanism for the whole-cloth replication of successful groups.

Sexual selection suggests a more plausible explanation for the persistence of altruistic behaviors that reduce the chances of ecological selection. Contrary to Herbert Spencer's infamous view of evolution as "survival of the fittest," generosity can flourish even when these traits are ecologically disadvantageous, by attracting

females who have evolved a preference for "romantic" males who are ready to sacrifice for their partner. Traits that reduce the ecological fitness of an individual carrier can also flourish if the trait increases the selection chances of other individuals with that trait. Hamilton (1964) introduced this gene-centric theory of kin altruism, later popularized by Dawkins' in the *Selfish Gene* (Dawkins 1976).

Allison (1992) extended the theory to benevolence based on cultural relatedness, such as geographical proximity or a shared cultural marker. This may explain why gene-culture coevolution seems to favor a tendency to associate with those who are similar, to differentiate from "outsiders," and to defend the in-group against social trespass with the emotional ferocity of parents defending their offspring.

This model also shows how evolutionary principles initially developed to explain biological adaptation can be extended to explain social and cultural change. Prominent examples include the evolution of languages, religions, laws, organizations, and institutions. This approach has a long and checkered history. Social Darwinism is a discredited nineteenth-century theory that used biological principles as analogs for social processes such as market competition and colonial domination. Many sociologists still reject all theories of social or cultural evolution, along with biological explanations of human behavior, which they associate with racist and elitist theories of "survival of the fittest." Others, like the sociobiologist E. O. Wilson (1988, p. 167), believe "genes hold culture on a leash," leaving little room for cultural evolution to modify the products of natural selection. Similarly, evolutionary psychologists like Cosmides and Tooby search for the historical origins of human behavior as the product of ancestral natural selection rather than ongoing social or cultural evolution.

In contrast, a growing number of sociologists and economists are exploring the possibility that human behaviors and institutions may be heavily influenced by processes of social and cultural selection that are independent of biological imperatives. These include DiMaggio and Powell (the new institutional sociology), Nelson and Winter (evolutionary economics), and Hannan and Freeman (organizational ecology).

One particularly compelling application is the explanation of cultural diversity. In biological evolution, speciation occurs when geographic separation allows populations to evolve in different directions to the point that individuals from each group can no longer mate. Speciation implies that all life has evolved from a very small number of common ancestors, perhaps only one. The theory has been applied to the evolution of myriad Indo-European languages that are mutually incomprehensible despite having a common ancestor. In sociocultural models, speciation operates through homophily (attraction to those who are similar), xenophobia (aversion to those who are different), and influence (the tendency to become more similar to those to whom we are attracted and to differentiate from those we despise).

Critics counter that sociocultural evolutionists have failed to identify any underlying replicative device equivalent to the gene. Dawkins has proposed the "meme" as the unit of cultural evolution, but there is as yet no evidence that these exist. Yet Charles Darwin developed the theory of natural selection without knowing that

phenotypes are coded genetically in DNA. Perhaps the secrets of cultural evolution are waiting to be unlocked by impending breakthroughs in cognitive psychology.

The boundary between learning and evolution becomes muddied by a hybrid mechanism, often characterized as "cultural evolution." In cultural evolution, norms, customs, conventions, and rituals propagate via role modeling, occupational training, social influence, and imitation. Cultural evolution resembles learning in that the rules are soft wired and can therefore be changed without replacing the carrier. Cultural evolution also resembles biological evolution in that rules used by successful carriers are more likely to propagate to other members of the population. However, because cultural rules are soft wired, the rules can propagate without replacing the carriers. For example, norms can jump from one organism to another by imitation (Dawkins 1976; Durham 1992; Boyd and Richerson 1985; Lopreato 1990). A successful norm is one that can cause its carrier to act in ways that increase the chances that the norm will be adopted by others. Cultural evolution can also be driven by social learning (Bandura 1977) in which individuals respond to the effects of vicarious reinforcement. Social learning and role modeling can provide an efficient shortcut past the hard lessons of direct experience.

Imitation of successful role models is the principal rationale for modeling cultural evolution as an analog of natural selection (Boyd and Richerson 1985; Dawkins 1976). However, social influence differs decisively from sociobiological adaptation. Softwired rules can spread without replacement of their carriers, which means that reproductive fitness loses its privileged position as the criteria for replication. While "imitation of the fittest" is a reasonable specification of cultural selection pressures, it is clearly not the only possibility. Replication of hardwired rules may be a misleading model for cultural evolution, and researchers need to be cautious in using Darwinian analogs as templates for modeling the diffusion of cultural rules. In cultural models of "imitation of the fittest," actors must not only know which actor is most successful; they must also know the underlying strategy that is responsible for that success. Yet successful actors may not be willing to share this information. For very simple strategies, it may be sufficient to observe successful behaviors. However, conditional strategies based on "if-then" rules cannot always be deduced from systematic observation. Researchers should therefore exercise caution in using biological models based on Darwinian principles to model cultural evolution, which is a hybrid of the ideal types of evolution and learning.

20.3 Learning

The most elementary principle of learning is simple reinforcement. Thorndike (1898) first formulated the theory of reinforcement as the "law of effect," based on the principle that "pleasure stamps in, pain stamps out." If a behavioral response has a favorable outcome, the neural pathways that triggered the behavior are strengthened, which "loads the dice in favor of those of its performances which make for the most permanent interests of the brain's owner" (James 1981, p. 143). This

connectionist theory anticipates the error back propagation used in contemporary neural networks (Rumelhart and McClelland 1988). These models show how highly complex behavioral responses can be acquired through repeated exposure to a problem.

Reinforcement theory relaxes three key behavioral assumptions in models of forward-looking rationality:

1. Propinquity replaces causality as the link between choices and payoffs.
2. Reward and punishment replace utility as the motivation for choice.
3. Melioration replaces optimization as the basis for the distribution of choices over time.

We consider each of these in turn.

1. *Propinquity, not causality.* Compared to forward-looking calculation, the law of effect imposes a lighter cognitive load on decision makers by assuming experiential induction rather than logical deduction. Players explore the likely consequences of alternative choices and develop preferences for those associated with better outcomes, even though the association may be coincident, "superstitious," or causally spurious. The outcomes that matter are those that have already occurred, not those that an analytical actor might predict in the future. Anticipated outcomes are but the consciously projected distillations of prior exposure to a recurring problem. Research using fMRI supports the view that purposive assessment of means and ends can take place *after* decisions are made, suggesting that "rational choice" may be not so much a theory of decision but a theory of how decisions are rationalized to self and others.

   Reinforcement learning applies to both intended and unintended consequences of action. Because repetition, not foresight, links payoffs back to the choices that produce them, learning models need not assume that the payoffs are the intended consequences of action. Thus, the models can be applied to expressive behaviors that lack a deliberate or instrumental motive. Frank's (1988) evolutionary model of trust and commitment formalizes the backward-looking rationality of emotions like vengeance and sympathy. An angry or frightened actor may not be capable of deliberate and sober optimization of self-interest, yet the response to the stimulus has consequences for the individual, and these in turn can modify the probability that the associated behavior will be repeated.
2. *Reward and punishment, not utility.* Learning theory differs from expected utility theory in positing two distinct cognitive mechanisms that guide decisions toward better outcomes, *approach* (driven by reward) and *avoidance* (driven by punishment). The distinction means that aspiration levels are very important for learning theory. The effect of an outcome depends decisively on whether it is coded as gain or loss, satisfactory or unsatisfactory, pleasant or aversive.
3. *Melioration, not optimization*. Melioration refers to suboptimal gradient climbing when confronted with what Herrnstein and Drazin (1991) call "distributed choice" across recurrent decisions. A good example of distributed choice is the decision whether to cooperate in an iterated prisoner's dilemma game.

Suppose each side is satisfied when the partner cooperates and dissatisfied when the partner defects. Melioration implies a tendency to repeat choices with satisfactory outcomes even if other choices have higher utility, a behavioral tendency March and Simon (1958) call "satisficing." In contrast, unsatisfactory outcomes induce searching for alternative outcomes, including a tendency to revisit alternative choices whose outcomes are even worse, a pattern we call "dissatisficing." While satisficing is suboptimal when judged by conventional game-theoretic criteria, it may be more effective in leading actors out of a suboptimal equilibrium than if they were to use more sophisticated decision rules, such as "testing the waters" to see if they could occasionally get away with cheating. Gradient search is highly path dependent and not very good at backing out of evolutionary cul-de-sacs. Course correction can sometimes steer adaptive individuals to globally optimal solutions, making simple gradient climbers look much smarter than they need to be. Often, however, adaptive actors get stuck in local optima. Both reinforcement and reproduction are biased toward *better* strategies, but they carry no guarantee of finding the highest peak on the adaptive landscape, however relentless the search. Thus, learning theory can be usefully applied to the equilibrium selection problem in game theory. In repeated games (such as an ongoing prisoner's dilemma), there is often an indefinitely large number of analytic equilibria. However, not all these equilibria are learnable, either by individuals (via reinforcement) or by populations (via evolution). Learning theory has also been used to identify a fundamental solution concept for these games—stochastic collusion—based on a random walk from a self-limiting noncooperative equilibrium into a self-reinforcing cooperative equilibrium (Macy and Flache 2002).

20.4 Modeling Evolution

Replicator dynamics are the most widely used model of evolutionary selection (Taylor and Jonker 1978). In these models, the frequency of a strategy changes from generation to generation as a monotonic function of its "payoff advantage," defined in terms of the difference between the average payoff of that strategy and the average payoff in the population as a whole. The more successful a strategy is on average, the more frequent it tends to be in the next generation.

Replicator dynamics typically assume that in every generation, every population member encounters every other member exactly once, and replication is based on the outcome of this interaction relative to the payoff earned by all other members of the population. However, in natural settings, actors are unlikely to interact with or have information about the relative success of every member of a large population. The mechanism can also be implemented based on local interaction (limited to network "neighbors") and local replication (neighbors compete only with one another for offspring).

The outcomes of replicator dynamics depend on the initial distribution of strategies, since the performance of any given strategy will depend on its effectiveness in interaction with other strategies. For example, aggressive strategies perform much better in a population that is accommodating than one that is equally aggressive. It is also not possible for replicator dynamics to invent new strategies that were not present at the outset.

These limitations are minimized by using genetic algorithms. The genetic algorithm was proposed by Holland (1975) as a problem-solving device, modeled after the recursive system in natural ecologies. The algorithm provides a simple but elegant way to write a computer program that can improve through experience. The program consists of a string of symbols that carry machine instructions. The symbols are often binary digits called "bits" with values of 0 and 1. The string is analogous to a chromosome containing multiple genes. The analog of the gene is a bit or combination of bits that comprises a specific instruction. The values of the bits and bit combinations are analogous to the alleles of the gene. A one-bit gene has two alleles (0 and 1), a two-bit gene has four alleles (00, 01, 10, and 11), and so on. The number of bits in a gene depends on the instruction. An instruction to go left or right requires only a single bit. However, an instruction to go left, right, up, or down requires two bits. When the gene's instructions are followed, there is some performance evaluation that measures the program's reproductive fitness relative to other programs in a computational ecology. Relative fitness determines the probability that each strategy will propagate. Propagation occurs when two mated programs recombine through processes like "crossover" and "inversion." In crossover, the mated programs (or strings) are randomly split, and the "left" half of one string is combined with the "right" half of the other, and vice versa, creating two new strings. If two different protocols are each effective, but in different ways, crossover allows them to create an entirely new strategy that may combine the best abilities of each parent, making it superior to either. If so, then the new rule may go on to eventually displace both parent rules in the population of strategies. In addition, the new strings contain random copying errors. These mutations continually refresh the heterogeneity of the population, in the face of selection pressures that tend to reduce it.

To illustrate, consider the eight-bit string **10011010** mated with *11000101*. (The typefaces might represent gender, although the algorithm does not require sexual reproduction.) Each bit could be a specific gene, such as whether to trust a partner under eight different conditions (Macy and Skvoretz 1998). In mating, the two parent strings are randomly broken, say after the third gene. The two offspring would then be **100***00101* and *110***11010**. However, a chance copying error on the last gene might make the second child a mutant, with *110***11011**. At the end of each generation, each individual's probability of mating is a monotonic (often linear) function of relative performance during that generation, based on stochastic sampling (Goldberg 1989):

$$P_{ij} = \frac{F_i}{\sum_{n=1}^{N} F_n} \text{ for } j = 1 \text{ to } N, \; j \neq i \tag{20.1}$$

where P_{ij} is the probability that j is mated with i, F_i is i's "fitness" (or cumulative payoff over all previous rounds in that generation), and N is the size of the population. If the best strategy had only a small performance edge over the worst, it had only a small edge in the race to reproduce. With stochastic sampling, each individual, even the least fit, selects a mate from the fitness-weighted pool of eligibles. In each pairing, the two parents combined their chromosomes to create a single offspring that replaces the less fit parent. The two chromosomes are combined through crossover.

20.5 Learning Models

The need for a cognitive alternative to evolutionary models is reflected in a growing number of formal learning-theoretic models of behavior (Macy 1991; Roth and Erev 1995; Fudenberg and Levine 1998; Young 1998; Cohen et al. 2001). In general form, learning models consist of a probabilistic decision rule and a learning algorithm in which outcomes are evaluated relative to an aspiration level, and the corresponding decision rules are updated accordingly.

All stochastic learning models share two important principles, the law of effect and probabilistic decision-making (Macy 1989, 1991; Börgers & Sarin 1997; Roth and Erev 1995; Erev and Roth 1998, Erev et al. 1999; for more references cf. Erev et al. 1999). The law of effect implies that the propensity of an action increases if it is associated with a positively evaluated outcome, and it declines if the outcome is negatively evaluated. Probabilistic choice means that actors hold a propensity qX for every action X. The probability pX to choose action X then increases in the magnitude of the propensity for X relative to the propensities for the other actions.

Whether an outcome is evaluated as positive or negative depends on the evaluation function. An outcome is positive if it exceeds the actor's aspirations. There are basically three substantively different approaches for modeling the aspiration level, fixed interior aspiration, fixed zero aspiration, and moving average aspiration. Fixed interior aspiration assumes that some payoffs are below the aspiration level and are evaluated negatively, while other payoffs are above the aspiration level and are evaluated positively (e.g., Macy 1989, 1991; Fudenberg and Levine 1998). The fixed zero aspiration approach also fixes the aspiration level, but it does so at the minimum possible payoff (Roth and Erev 1995; Börgers and Sarin 1997; Erev and Roth 1998). In other words, in the fixed zero aspiration model, every payoff is deemed "good enough" to increase or at least not reduce the corresponding propensity, but higher payoffs increase propensities more than lower payoffs do. Finally, moving average aspiration models assume that the aspiration level approaches the average of the payoffs experienced recently, so that players get used to whatever outcome they may experience often enough (Macy and Flache 2002; Börgers and Sarin 1997; Erev and Rapoport 1998; Erev et al. 1999). Clearly, these assumptions have profound effects on model dynamics.

20.5.1 Bush-Mosteller Stochastic Learning Model

One of the simplest models of reinforcement learning is the Bush-Mosteller model (Bush and Mosteller 1950). The Bush-Mosteller stochastic learning algorithm updates probabilities following an action a as follows:

$$p_{a,t+1} = \begin{cases} p_{a,t} + (1 - p_{a,t}) \;\; l\, s_{a,t}, & s_{a,t} \geq 0 \\ \\ p_{a,t} + p_{a,t} \;\; l \;\; s_{a,t}, & s_{a,t} < 0 \end{cases}, \quad a \in \{C, D\} \tag{20.2}$$

In Eq. (20.2), $p_{a,t}$ is the probability of action a at time t, and $s_{a,t}$ is a positive or negative stimulus ($0 \leq s_{a,t} \leq 1$). The change in the probability for the action not taken, b, obtains from the constraint that probabilities always sum to one, i.e., $p_{b,t+1} = 1 - p_{a,t+1}$. The parameter l is a constant ($0 < l < 1$) that scales the learning rate. With $l \approx 0$, learning is very slow, and with $l \approx 1$, the model approximates a "win-stay, lose-shift" strategy (Catania 1992).

For any value of l, Eq. (20.2) implies a decreasing effect of reward as the associated propensity approaches unity, but an increasing effect of punishment. Similarly, as the propensity approaches zero, there is a decreasing effect of punishment and a growing effect of reward. This constrains probabilities to approach asymptotically their natural limits.

20.5.2 The Roth-Erev Matching Model

Roth and Erev (Roth and Erev 1995; Erev and Roth 1998; Erev et al. 1999) have proposed a learning-theoretic alternative to the Bush-Mosteller formulation. Their model draws on the "matching law" which holds that adaptive actors will choose between alternatives in a ratio that matches the ratio of reward. Like the Bush-Mosteller model, the Roth-Erev payoff matching model implements the three basic principles that distinguish learning from utility theory—experiential induction (vs. logical deduction), reward and punishment (vs. utility), and melioration (vs. optimization). The similarity in substantive assumptions makes it tempting to assume that the two models are mathematically equivalent or, if not, that they nevertheless give equivalent solutions.

On closer inspection, however, we find important differences, identified by Flache and Macy (2002). Each specification implements reinforcement learning in different ways and with different results. Roth and Erev (1995, Erev & Roth 1998) propose a baseline model of reinforcement learning with a fixed zero reference point. The law of effect is implemented such that the propensity for action X is simply the sum of all payoffs a player ever experienced when playing X. The probability to choose action X at time t is then the propensity for X divided by the sum of all action propensities at time t. The sum of the propensities increases over

time, such that payoffs have decreasing effects on choice probabilities. However, this also undermines the law of effect. Suppose, after some time, a new action is carried out and yields a higher payoff than every other action experienced before. The probability of repetition of this action will nevertheless be negligible, because its recent payoff is small in comparison with the accumulated payoffs stored in the propensities for the other actions. As a consequence, the baseline model of Roth and Erev (1995) fails to identify particular results, because it has the tendency to lock the learning dynamics into any outcome that occurs sufficiently often early on. Roth and Erev amend this problem by introducing a "forgetting parameter" that keeps propensities low relative to recent payoffs. With this, they increase the sensitivity of the model to recent reinforcement. Roth and Erev used a variant of this baseline model to estimate globally applicable parameters from data collected across a variety of human subject experiments. They concluded that "low rationality" models of reinforcement learning may often provide a more accurate prediction than forward-looking models. Like the Bush-Mosteller, the Roth-Erev model is stochastic, but the probabilities are not equivalent to propensities. The propensity q for action a at time T is the sum of all stimuli s_a a player has ever received when playing a:

$$q_{a,T} = \sum_{t=1}^{T} s_{a,t}, \quad a \in \{C, D\}\,. \tag{20.3}$$

Roth and Erev then use a "probabilistic choice rule" to translate propensities into probabilities. The probability p_a of action a at time $t+1$ is the propensity for a divided by the sum of the propensities at time t:

$$p_{a,t+1} = \frac{q_{a,t}}{q_{a,t} + q_{b,t}}, \quad (a, b) \in \{C, D\}\,, a \neq b \tag{20.4}$$

where a and b represent binary choices. Following action a, the associated propensity q_a increases if the payoff is positive relative to aspirations (by increasing the numerator in Eq. (20.4)) and decreases if negative. The propensity for b remains constant, but the probability of b declines (by increasing the denominator in the equivalent expression for $p_{b,t+1}$).

20.5.3 Artificial Neural Networks

Bush-Mosteller and Roth-Erev are very simple learning models that allow an actor to identify strategies that generally have more satisfactory outcomes. However, the actor cannot learn the conditions in which a strategy is more or less effective. Artificial neural nets add perception to reinforcement, so that actors can learn conditional strategies.

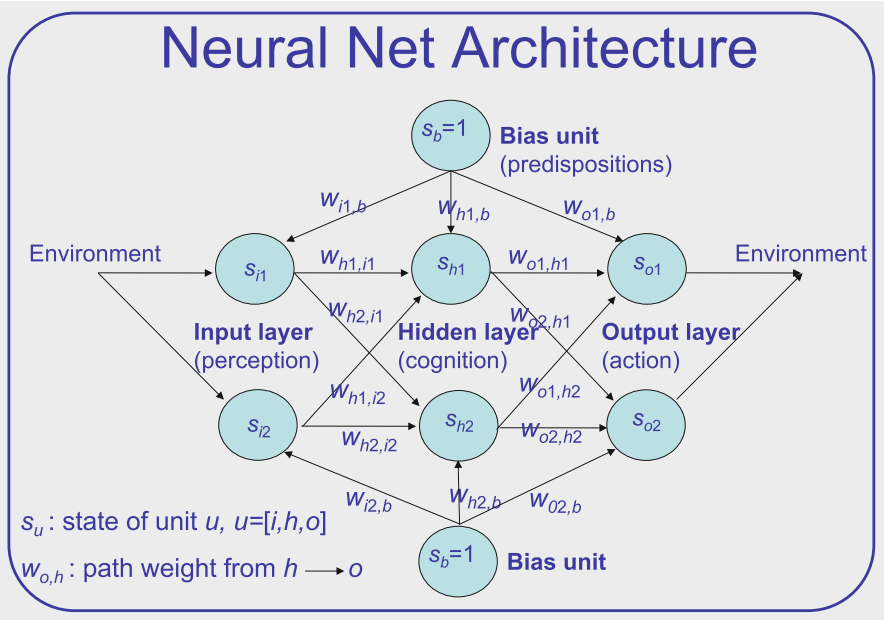

Fig. 20.1 Simple example of a feed-forward network with one hidden layer

An artificial neural network is a simple type of self-programmable learning device based on parallel distributed processing (Rumelhart and McClelland 1988). Like genetic algorithms, neural nets have a biological analog, in this case, the nerve systems of living organisms. In elementary form, the device consists of a web of neuron-like nodes (or neurodes) that fire when triggered by impulses of sufficient strength and in turn stimulate other nodes when fired. The magnitude of an impulse depends on the strength of the connection (or "synapses") between the two neurodes. The network learns by modifying these path coefficients, usually in response to environmental feedback about its performance.

There are two broad classes of neural networks that are most relevant to social scientists, feed-forward networks, and attractor networks. Feed-forward networks consist of four types of nodes, usually arranged in layers, as illustrated in Fig. 20.1. The most straightforward are input (sensory) and output (response) nodes. Input nodes are triggered by stimuli from the environment. In Fig. 20.1, there are two input nodes. I1 has been activated by the environment (+1), while I2 has not (−1). Output nodes, in turn, trigger action by the organism on the environment. In Fig. 20.1, output node O1 has been triggered, as indicated by the output value of +1.

The other two types of nodes are less intuitive. Intermediate (or "hidden") nodes link sensory and response nodes so as to increase the combinations of multiple stimuli that can be differentiated. Figure 20.1 shows a network with a single layer containing two hidden nodes, H1 and H2. The number of hidden layers and the

number of hidden nodes in each layer vary with the complexity of the stimulus patterns the network must learn to recognize and the complexity of the responses the network must learn to perform. Unlike sensory and response nodes, hidden nodes have no direct contact with the environment, hence their name.

Bias nodes are a type of hidden node that has no inputs. Instead, a bias node continuously fires, creating a predisposition toward excitation or inhibition in the nodes it stimulates, depending on the size and sign of the weights on the pathway from the bias node to the nodes it influences. If the path weight is positive, the bias is toward excitation, and if the path is negative, the bias is toward inhibition. The weighted paths from the bias node to other hidden and output nodes correspond to the activation thresholds for these nodes.

A feed-forward network involves two processes—action (firing of the output nodes) and learning (error correction). The action phase consists of the forward propagation of influence (either excitation or inhibition) from the input nodes to the hidden nodes to the output nodes. The influence of a node depends on the state of the node and the weight of the neural pathway to a node further forward in the network. The learning phase consists of the backward propagation of error from the output nodes to the hidden nodes to the input nodes, followed by the adjustment of the weights so as to reduce the error in the output nodes.

The action phase begins with the input nodes. Each node has a "state" which can be binary (e.g., 0 or 1 to indicate whether the node "fires") or continuous (e.g., 9 to indicate that the node did not fire as strongly as one whose state is 1.0). The states of the input nodes are controlled entirely by the environment and correspond to a pattern that the network perceives. The input nodes can influence the output nodes directly as well as via hidden nodes that are connected to the input nodes.

To illustrate, consider a neural network in which the input nodes are numbered from $i = 1$ to I. The ith input is selected, and the input register to all nodes j influenced by i is then updated by multiplying the state of i times the weight w_{ij} on the ij path. This updating is repeated for each input node. The pathways that link the nodes are weighted with values that determine the strength of the signals moving along the path. A low absolute value means that an input has little influence on the output. A large positive weight makes the input operate as an excitor. When it fires, the input excites an otherwise inhibited output. A large negative path weight makes the input operate as an inhibitor. When it fires, the input inhibits an otherwise excited output.

Next, the nodes whose input registers have been fully updated (e.g., the nodes in the first layer of hidden nodes) must update their states. The state of a node is updated by aggregating the values in the node's input register, including the input from its bias node (which always fires, e.g., $B_i = 1$).

Updating states is based on the activation function. Three activation functions are commonly used. Hard-limit functions fire the node iff the aggregate input exceeds zero. Sigmoid stochastic functions fire the node with a probability given by the aggregate input. Sigmoid deterministic functions fire the node with a

magnitude given by the aggregate input.[2] For example, if node k is influenced by two input nodes, i and j and a bias node b, then k sums the influence $ik = i^*w_{ij} + k^*w_{ik} + b^*w_{ib}$. Positive weights cause i, j, and b to activate k and negative weights inhibit. If j is hard limited, then if $ik > 0$, $k = 1$, else $k = 0$. If the activation function is stochastic, k is activated with a probability $p = 1/(1 + e(-ik)$. If the sigmoid function is deterministic, k is activated with magnitude p.

Once the states have been activated for all nodes whose input registers have been fully updated (e.g., the first layer of hidden nodes and/or one or more output nodes that are directly connected to an input node), these nodes in turn update the input registers of nodes they influence going forward (e.g., the second layer of hidden nodes and/or some or all of the output nodes). Once this updating is complete, all nodes whose input registers have been fully updated aggregate across their inputs and update their states, and so on until the states of the output nodes have been updated. This completes the action phase.

The network learns by modifying the path weights linking the neurodes. Learning only occurs when the response to a given sensory input pattern is unsatisfactory. The paths are then adjusted so as to reduce the probability of repeating the mistake the next time this pattern is encountered. In many applications, neural nets are trained to recognize certain patterns or combinations of inputs. For example, suppose we want to train a neural net to predict stock prices from a set of market indicators. We first train the net to correctly predict known prices. The net begins with random path coefficients. These generate a prediction. The error is then used to adjust the weights in an iterative process that improves the predictions. These weights are analogous to those in a linear regression, and like a regression, the weights can then be applied to new data to predict the unknown.

20.5.3.1 Back Propagation of Error

A feed-forward neural network is trained by adjusting the weights on the paths between the nodes. These weights correspond to the influence that a node will have in causing nodes further forward in the network to fire. When those nodes fire incorrectly, the weights responsible for the error must be adjusted accordingly. Just as the influence process propagates forward, from the input nodes to the hidden layer to the output nodes, the attribution of responsibility for errors propagates backward, from the output nodes to the hidden layers. The back propagation of error begins with the output nodes and works back through the network to the input nodes, in the opposite direction from the influence process that "feeds forward" from input nodes to hidden nodes to output nodes. First, the output error is calculated for each output node. This is simply the difference between the expected state of the ith output node ($\widehat{S}_i$), and the state that was observed (S_i). For an output node, error refers to

[2] A multilayer neural net requires a nonlinear activation function (such as a sigmoid). If the functions are linear, the multilayer net reduces to a single-layer I-O network.

the difference between the expected output for a given pattern of inputs and the observed output:

$$e_o = s_o\,(1 - s_o)\,(\widehat{s}_o - s_o) \tag{20.5}$$

where the term $s_o\ (1 - s_o)$ limits the error to the unit interval. If the initial weights were randomly assigned, it is unlikely that the output will be correct. For example, if we observe an output value of 0.37 and we expected 1, then the error is −0.53.

Once the error has been updated for each output node, these errors are used to update the error for the nodes that influenced the output nodes. These nodes are usually located in the last layer of hidden nodes, but they can be anywhere and can even include input nodes that are wired directly to an output.[3] Then the errors of the nodes in the last hidden layer are used to update error back further still to the hidden nodes that influenced the last layer of hidden nodes, and so on, back to the first layer of hidden nodes, until the errors for all hidden nodes have been updated. Input nodes cannot have error, since they simply represent an exogenous pattern that the network is asked to learn. Back propagation means that the error observed in an output node o $(\widehat{s}_o - s_o)$ is allocated not only to o but to all the hidden nodes that influenced o, based on the strength of their influence. The total error of a hidden node h is then simply the summation over all n_h allocated errors from the n nodes i that h influenced, including output nodes as well as other hidden nodes:

$$e_h = s_h\,(1 - s_h)\sum_{i=1}^{n} w_{hi}e_i \tag{20.6}$$

Once the errors for all hidden, bias, and output nodes have been back propagated, the weight on the path from i to j is updated:

$$w'_{ij} = w_{ij} + \lambda s_i e_j \tag{20.7}$$

where λ is a fractional learning rate. The learning algorithm means that the influence of i on j increases if j's error was positive (i.e., the expected output exceeded the observed) and decreases if j's influence was negative.

Note that the Bush-Mosteller model is equivalent to a neural net with only a single bias unit and an output, but with no sensory inputs or hidden units. Such a device is capable of learning *how often* to act, but not *when* to act, that is, it is incapable of learning conditional strategies. In contrast, a feed-forward network can learn to differentiate environmental cues and respond using more sophisticated protocols for contingent strategies.

[3] However, if an input node is wired to hidden nodes as well as output nodes, the error for this node cannot be updated until the errors for all hidden nodes that it influenced have been updated.

20.5.3.2 Attractor Neural Network

Feed-forward networks are the most widely used but not the only type of artificial neural network. An alternative design is the attractor neural network (Churchland and Sejnowski 1994; Quinlan 1991), originally developed and investigated by Hopfield (1982; Hopfield and Tank 1985). In a recent article, Nowak and Vallacher (1998) note the potential of these computational networks for modeling group dynamics. This approach promises to provide a fertile expansion to social network analysis, which has often assumed that social ties are binary and static. A neural network provides a way to dynamically model a social network in which learning occurs at both the individual and structural levels, as relations evolve in response to the behaviors they constrain.

Unlike feed-forward neural networks (Rumelhart and McClelland 1988), which are organized into hierarchies of input, hidden, and output nodes, attractor (or "feed-lateral") networks are internally undifferentiated. Nodes differ only in their states and in their relational alignments, but they are functionally identical. Without input units to receive directed feedback from the environment, these models are "unsupervised" and thus have no centralized mechanism to coordinate learning of efficient solutions. In the absence of formal training, each node operates using a set of behavioral rules or functions that compute changes of state ("decisions") in light of available information. Zeggelink (1994) calls these "object-oriented models," where each agent receives input from other agents and may transform these inputs into a change of state, which in turn serves as input for other agents.

An important agent-level rule that characterizes attractor networks is that individual nodes seek to minimize "energy" (also "stress" or "dissonance") across all relations with other nodes. As with feed-forward networks, this adaptation occurs in two discrete stages. In the action phase, nodes change their states to maximize similarity with nodes to which they are strongly connected. In the learning phase, they update their weights to strengthen ties to similar nodes. Thus, beginning with some (perhaps random) configuration, the network proceeds to search over an optimization landscape as nodes repeatedly cycle through these changes of weights and states.

In addition to variation in path strength, neural networks typically have paths that inhibit as well as excite. That is, nodes may be connected with negative as well as positive weights. In a social network application, agents connected by negative ties might correspond to "negative referents" (Schwartz and Ames 1977), who provoke differentiation rather than imitation.

Ultimately, these systems are able to locate stable configurations (called "attractors"), for which any change of state or weight would result in a net increase in stress for the affected nodes. Hopfield (1982) compares these equilibria to memories and shows that these systems of undifferentiated nodes can learn to implement higher-order cognitive functions. However, although the system may converge at a stable equilibrium that allows all nodes to be locally satisfied (i.e., a "local optimum"), this does not guarantee that the converged pattern will minimize overall dissonance (a "global optimum").

This class of models generally uses complete networks, with each node characterized by one or more binary or continuous states and linked to other nodes through endogenous weights. Like other neural networks, attractor networks learn stable configurations by iteratively adjusting the weights between individual nodes, without any global coordination. In this case, the weights change over time through a Hebbian learning rule: the weight w_{ij} is a function of the correlation of states for nodes i and j over time. Specifically, Hebbian learning implies the following rules:

- To the extent that nodes i and j adopt the same states at the same time, the weight of their common tie will increase until it approaches some upper limit (e.g., 1.0).
- To the extent that nodes i and j simultaneously adopt different states, the weight of their common tie will decrease until it approaches some lower limit (e.g., 0.0).

Although Hebbian learning was developed to study memory in cognitive systems, it corresponds to the homophily principle in social psychology (Homans 1951) and social network theory (McPherson and Smith-Lovin 1987), which holds that agents tend to be attracted to those whom they more closely resemble. This hypothesis is also consistent with structural balance theory (Cartwright and Harary 1956; Heider 1958) and has been widely supported in studies of interpersonal attraction and interaction, where it has been called "the law of attraction" (Byrne 1971; Byrne and Griffitt 1966).

An important property of attractor networks is that individual nodes seek to minimize "energy" (or dissonance) across all relations with other nodes—a process that parallels but differs from the pursuit of balanced relations in structural balance theory. These networks also posit self-reinforcing dynamics of attraction and influence as well as repulsion and differentiation.

Following Nowak and Vallacher (1998), Macy et al. (2003) apply the Hopfield model of dynamic attraction to the study of polarization in social networks. In this application, observed similarity/difference between states determines the strength and valence of the tie to a given referent. This attraction and repulsion may be described anecdotally in terms of liking, respect, or credibility and their opposites. In their application of the Hopfield model, each node has $N - 1$ undirected ties to other nodes. These ties include weights, which determine the strength and valence of influence between agents. Formally, social pressure on agent i to adopt a binary state s (where $s = \pm 1$) is the sum of the states of all other agents j, where influence from each agent is conditioned by the weight (w_{ij}) of the dyadic tie between i and j ($-1.0 < w_{ij} < 1.0$):

$$P_{is} = \frac{\sum_{j=1}^{N} w_{ij} s_j}{N - 1}, j \neq i \tag{20.8}$$

Thus, social pressure ($-1 < P_{is} < 1$) to adopt s becomes increasingly positive as i's "friends" adopt s ($s = 1$) and i's "enemies" reject s ($s = -1$). The pressure

can also become negative in the opposite circumstances. The model extends to multiple states in a straightforward way, where Eq. (20.8) independently determines the pressure on agent i for each binary state s. Strong positive or negative social pressure does not guarantee that an agent will accommodate, however. It is effective only if i is willing and able to respond to peer influence. If i is closed-minded or if a given trait is not under i's control (e.g., ethnicity or gender), then no change to s will occur. The probability π that agent i will change state s is a cumulative logistic function of social pressure:

$$\pi_{is} = \frac{1}{1 + e^{-10P_{is}}} \tag{20.9}$$

Agent i adopts s if $\pi > C + \chi$, where C is the inflection point of the sigmoid, χ is a random number drawn from a uniform distribution in the interval $[C - \varepsilon, C + \varepsilon]$, and ε is an exogenous error parameter ($0 \leq \varepsilon \leq 1$). At one extreme, $\varepsilon = 0$ produces highly deterministic behavior, such that any social pressure above the trigger value always leads to conformity and pressures below the trigger value entail differentiation. Following Harsanyi (1973), $\varepsilon > 0$ allows for a "smoothed best reply" in which pressure levels near the trigger point leave the agent relatively indifferent and thus likely to explore behaviors on either side of the threshold. In the Hopfield model, the path weight w_{ij} changes as a function of similarity in the states of node i and j. Weights begin with uniformly distributed random values, subject to the constraints that weights are symmetric ($w_{ij} = w_{ji}$). Across a vector of K distinct states s_{ik} (or the position of agent i on issue k), agent i compares its own states to the observed states of another agent j and adjusts the weight upward or downward corresponding to their aggregated level of agreement or disagreement. Based on the correspondence of states for agents i and j, their weight will change at each discrete time point t in proportion to a parameter λ, which defines the rate of structural learning ($0 < \lambda < 1$):

$$w_{ij,t+1} = w_{ijt}\,(1 - \lambda) + \frac{\lambda}{K} \sum_{k=1}^{K} s_{jkt} s_{ikt}, j \neq i \tag{20.10}$$

As correspondence of states can be positive (agreement) or negative (disagreement), ties can grow positive or negative over time, with weights between any two agents always symmetric.

Note one significant departure from structural balance theory. Although the agents in this model are clearly designed to maintain balance in their behaviors with both positive and negative referents, this assumption is not "wired in" to the relations themselves. That is, two agents i and j feel no direct need for consistency in their relations with a third agent h. Indeed, i has no knowledge of the jh relationship and thus no ability to adjust the ij relation so as to balance the triad.

Given an initially random configuration of states and weights, these agents will search for a profile that minimizes dissonance across their relations. Structural bal-

ance theory predicts that system-level stability can only occur when the group either has become uniform or has polarized into two (Cartwright and Harary 1956) or more (Davis 1967) internally cohesive and mutually antipathetic cliques. However, there is no guarantee in this model that they will achieve a globally optimal state in structural balance.

20.5.4 Belief Learning

Actors learn not only what is useful for obtaining rewards and avoiding punishments; they also update their beliefs about what is true and what is false. There are two main models of belief learning in the literature, Bayesian belief learning and fictitious play[4] (cf. Offerman 1997; Fudenberg and Levine 1998). These models differ in their assumptions about how players learn from observations. Both models assume that players believe that something is true with some fixed unknown probability p. In Bayesian learning, players then use Bayes' learning rule to rationally update over time their beliefs about p. In a nutshell, Bayes' learning rule implies that actors' assessment of the true value of p converges in the long run on the relative frequency of events that they observe. However, in the short and medium term, Bayesian learners remain suspicious in the sense that they take into account that observed events are an imperfect indication of p (Offerman 1997).

Fudenberg and Levine note that fictitious play is a special case of Bayesian learning. Fictitious play is a form of Bayesian learning that always puts full weight on the belief that the true value of p corresponds to the relative frequency observed in past events. Fudenberg and Levine (1998) note that it is an implausible property of fictitious play that a slight change in beliefs may radically alter behavior. The reason is that the best reply function always is a step function. As a remedy, Fudenberg and Levine introduce a smooth reply curve. The reply curve assigns a probability distribution that corresponds to the relative frequency of events. With strict best reply, the reply curve is a step function. Instead, a smooth reply curve assigns some probability to the action that is not strict best reply. This probability decreases in the difference between the expected payoffs. Specifically, when expected payoffs are equal, actors choose with equal probability, whereas their choice probabilities converge on pure strategies when the difference in expected payoffs approaches the maximum value.

The strict best reply function corresponds to the rule, "play X if the expected payoff for X is better than the expected payoff for Y, given your belief p. Otherwise play Y." Smooth best reply is then introduced with the modification to play the strict best reply strategy only with a probability of $1 - \eta$, whereas the alternative is played

[4]The Cournot rule may be considered as a third degenerate model of belief learning. According to the Cournot rule, players assume that the behavior of the opponent in the previous round will always occur again in the present round.

with probability η. The probability η, in turn, decreases in the absolute difference between expected payoffs $|\mathrm{u}X(p) - \mathrm{u}Y(p)|$, where $\eta = 0.5$ if players are indifferent.

Belief learning generally converges with the predictions of evolutionary selection models. The approaches are also similar in the predicted effects of initial conditions on end results. Broadly, the initial distribution of strategies in independent populations in evolutionary selection corresponds to the initial beliefs players' hold about their opponent. For example, when two pure strategy Nash equilibria are feasible, then the one tends to be selected toward which the initial strategy distribution in evolutionary selection or initial beliefs in belief learning are biased.

20.6 Conclusion

Evolution and learning are examples of backward-looking consequentialist models, in which outcomes of agents' past actions influence their future choices, either through selection (in the case of evolution) or reinforcement (in the case of learning). Backward-looking models make weaker assumptions about agents' cognitive capacities than forward-looking models and thus may be appropriate for settings in which agents lack the ability, resources, information, and motivation to engage in intensive cognitive processing, as in most everyday instances of collective action. Backward-looking models may also be useful for understanding behavior driven by affect, rather than calculation. Forward-looking models may be more appropriate in applications such as investment decisions, international diplomacy, or military strategy, where the stakes are high enough to warrant collection of all relevant information and the actors are highly skilled strategists. However, even where the cognitive assumptions of the models are plausible, forward-looking models are generally limited to the identification of static equilibria but not necessarily whether and how agents will reach those equilibria.

When implemented computationally, backward-looking models can show how likely agents are to reach particular equilibria, as well as the paths by which those equilibria may be reached. However, computational models are also less general than analytical models. Furthermore, backward-looking models will be of little help if change in the environment outpaces the rate of adaptation. These limitations underscore the importance of robustness testing over a range of parameter values.

Evolutionary models are most appropriate for theoretical questions in which adaptation takes place at the population level, through processes of selection. Biological evolution is the most obvious analog, but social and cultural evolution are likely to be more important for social scientists. However, as we note above, researchers must be cautious about drawing analogies from the biological to social/cultural dynamics.

Learning models based on reinforcement and Bayesian updating are useful in applications that do not require conditional strategies based on pattern recognition. When agents must learn more complex conditional strategies, feed-forward neural

networks may be employed. Furthermore, attractor neural networks are useful for modeling structural influence, such as conformity pressures from peers.

Models of evolution and learning are powerful tools for modeling social processes. Both show how complex patterns can arise when agents rely on relatively simple, experience-driven decision rules. This chapter seeks to provide researchers with an overview of promising research in this area and the tools necessary to further develop this research.

Further Reading

We refer readers interested in particular learning models and their application in agent-based simulation to Macy and Flache (2002), which gives a brief introduction into principles of reinforcement learning and discusses by means of simulation models how reinforcement learning affects behavior in social dilemma situations, whereas Macy (1996) compares two different approaches of modeling learning behavior by means of computer simulations. Fudenberg and Levine (1998) give a very good overview on how various learning rules relate to game-theoretic rationality and equilibrium concepts.

For some wider background reading, we recommend Macy (2004), which introduces the basic principles of learning theory applied to social behavior; Holland et al. (1986), which presents a framework in terms of rule-based mental models for understanding inductive reasoning and learning; and Sun (2008), which is a handbook of computational cognitive modeling.

References

Allison, P. (1992). The cultural evolution of beneficent norms. *Social Forces, 71*, 279–301.
Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
Börgers, T., & Sarin, R. (1997). Learning through reinforcement and replicator dynamics. *Journal of Economic Theory, 77*, 1–14.
Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.
Bush, R. R., & Mosteller, F. (1950). *Stochastic models for learning*. New York: Wiley.
Byrne, D. E. (1971). *The attraction paradigm*. New York: Academic.
Byrne, D. E., & Griffitt, D. (1966). A development investigation of the law of attraction. *Journal of Personality and Social Psychology, 4*, 699–702.
Cartwright, D., & Harary, F. (1956). Structural balance: A generalization of Heider's theory. *Psychological Review, 63*, 277–293.
Catania, A. C. (1992). *Learning* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
Churchland, P. S., & Sejnowski, T. J. (1994). *The computational brain*. Cambridge, MA: MIT Press.
Cohen, M. D., Riolo, R., & Axelrod, R. (2001). The role of social structure in the maintenance of cooperative regimes. *Rationality and Society, 13*(1), 5–32.
Davis, J. A. (1967). Clustering and structural balance in graphs. *Human Relations, 20*, 181–187.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
Durham, W. H. (1992). *Coevolution: genes, culture and human diversity*. Stanford, CA: Stanford University Press.
Erev, I., & Rapoport, A. (1998). Coordination, "magic", and reinforcement learning in a market entry game. *Games and Economic Behavior, 23*, 146–175.
Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review, 88*(4), 848–879.
Erev, I., Bereby-Meyer, Y., & Roth, A. E. (1999). The effect of adding a constant to all payoffs: Experimental investigation, and implications for reinforcement learning models. *Journal of Economic Behavior and Organizations, 39*(1), 111–128.
Flache, A., & Macy, M. W. (2002). Stochastic collusion and the power law of learning: A general reinforcement learning model of cooperation. *Journal of Conflict Resolution, 46*(5), 629–653.
Frank, R. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.
Fudenberg, D., & Levine, D. (1998). *The theory of learning in games*. Boston: MIT Press.
Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning*. New York: Addison-Wesley.
Hamilton, W. (1964). The genetic evolution of social behaviour. *Journal of Theoretical Biology, 17*, 1–54.
Harsanyi, J. (1973). Games with randomly disturbed payoffs. *International Journal of Game Theory, 2*, 1–23.
Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
Herrnstein, R. J., & Drazin, P. (1991). Meliorization: A theory of distributed choice. *Journal of Economic Perspectives, 5*(3), 137–156.
Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.
Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.
Homans, G. C. (1951). *The human group*. New York: Harcourt Brace.
Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, 79*, 2554–2558.
Hopfield, J. J., & Tank, D. W. (1985). "Neural" computation of decisions in optimization problems. *Biological Cybernetics, 52*, 141–152.
James, W. (1981). *Principles of psychology*. Cambridge, MA: Harvard University Press.
Lopreato, J. (1990). From social evolutionism to biocultural evolutionism. *Sociological Forum, 5*, 187–212.
Macy, M. (1989). Walking out of social traps: A stochastic learning model for the prisoner's dilemma. *Rationality and Society, 2*, 197–219.
Macy, M. (1991). Learning to cooperate: Stochastic and tacit collusion in social exchange. *American Journal of Sociology, 97*, 808–843.
Macy, M. (1996). Natural selection and social learning in prisoner's dilemma: Co-adaptation with genetic algorithms and artificial neural networks. *Sociological Methods and Research, 25*, 103–137.
Macy, M. (1997). Identity, interest, and emergent rationality. *Rationality and Society, 9*, 427–448.
Macy, M. (1998). Social order in an artificial world. *Journal of Artificial Societies and Social Simulation, 1*(1). http://jasss.soc.surrey.ac.uk/1/1/4.html
Macy, M. (2004). Learning theory. In G. Ritzer (Ed.), *Encyclopedia of social theory*. Thousand Oaks, CA: Sage.
Macy, M., & Flache, A. (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences, 99*, 7229–7236.
Macy, M., & Skvoretz, J. (1998). The evolution of trust and cooperation between strangers: A computational model. *American Sociological Review, 63*, 638–660.

Macy, M., Kitts, J., Flache, A., & Benard, S. (2003). Polarization in dynamic networks: A Hopfield model of emergent structure. In R. Breiger & K. Carley (Eds.), *Dynamic social network modeling and analysis: Workshop summary and papers* (pp. 162–173). Washington, DC: National Academy Press.

March, J. G., & Simon, H. A. (1958). *Organizations*. New York: Wiley.

McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: status distance and the composition of face to face groups. *American Sociological Review, 52*, 370–379.

Nowak, A., & Vallacher, R. R. (1998). Toward computational social psychology: Cellular automata and neural network models of interpersonal dynamics. In S. J. Read & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 277–311). Mahwah, NJ: Lawrence Erlbaum.

Offerman, T. (1997). *Beliefs and decision rules in public good games*. Dordrecht: Kluwer.

Quinlan, P. T. (1991). *Connectionism and psychology: A psychological perspective on new connectionist research*. Chicago, IL: University of Chicago Press.

Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in intermediate term. *Games and Economic Behavior, 8*, 164–212.

Rumelhart, D. E., & McClelland, J. L. (1988). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Schwartz, S. H., & Ames, R. E. (1977). Positive and negative referent others as sources of influence: A case of helping. *Sociometry, 40*, 12–21.

Sun, R. (Ed.). (2008). *The Cambridge handbook of computational psychology*. Cambridge: Cambridge University Press.

Taylor, P. D., & Jonker, L. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences, 40*, 145–156.

Thorndike, E. L. (1898). *Animal intelligence: An experimental study of the associative processes in animals*, *Psychological Review, Monograph Supplements, No. 8*. New York: Macmillan.

Vanberg, V. (1994). *Rules and choice in economics*. London: Routledge.

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.

Wilson, E. O. (1988). *On human nature*. Cambridge, MA: Harvard University Press.

Wynne-Edwards, V. C. (1962). *Animal dispersion in relation to social behaviour*. Edinburgh: Oliver & Boyd.

Wynne-Edwards, V. C. (1986). *Evolution through group selection*. Oxford: Blackwell Scientific.

Young, H. P. (1998). *Individual strategy and social structure. An evolutionary theory of institutions*. Princeton, NJ: Princeton University Press.

Zeggelink, E. (1994). Dynamics of structure: An individual oriented approach. *Social Networks, 16*, 295–333.