Learning

I. **Simon on Learning**

II. **Genetic Algorithms and Genetic Programming**

III. **Models from Psychology: Reinforcement Learning, etc.**

I. Herbert Simon on Learning

“Any change in a system that allows it to perform better the second time on repetition of the same task or on another task drawn from the same population” — Simon (1983)

“Learning is any change in a system that produces a more or less permanent change in its capacity for adapting to its environment. ” — Simon (1996)

“Armchair speculation about expectations, rational or other, is not a satisfactory substitute for factual knowledge as to how human beings go about anticipating the future, what factors they take into account, and how these factors, rather than others, come within the range of their attention.” — Simon (1982)

Learning

Previous (e.g. Life, segregation) the agents are unchanging,

Now: change occurs in the agent's parameters (simplest) or in the form of the agent model (more complex).

Several sorts of models of learning:

1. **Artificial Neural Nets (ANNs): from machine learning, from biological simplifications of the brain's operation, (not covered here)**
2. **Evolutionary models, such as Genetic Algorithms (GAs) and Genetic Programming (GP), from natural evolution, and**
3. **Models from psychology experiments, such as Reinforcement Learning (RL).**

II. Modelling Learning in ACE Models

Two sorts of (deductive) learning have dominated ACE models: GA (genetic algorithms) & RL (reinforcement learning)

- **GA, with implicit learning as the population "learns" from generation to generation: either**
  - **a population of players (the single-population GA model), or**
  - **a population of routines, ideas, heuristics, with each player modeled as a population (the multi-population GA model).**
- **RL, where the probability of choosing an action that was effective last round increases.**
- **A third sort: Anticipatory (inductive), Belief-Based Learning — the future?**

Evolutionary Computation

Based on evolution with natural selection.

Fitter individuals have more offspring to pass their genes to; less fit individuals have fewer offspring.

Genes occur in chromosomes; each gene codes for one (or more) functions.

The *genotype* = the structure of the individual's chromosomes.

The *phenotype* = the expressed characteristics (behaviours) of the individual, coded by the genotype.

Many phenotypes emerge from the individual genotypes: difficult to predict.

Evolution …

1. **Populations evolve, not individuals.**
2. **Evolutionary change requires diversity at the genotype level in the population.**
   Clones are identical, and $\therefore$ so are their offspring.
3. **While a species changes and adapts to its environment, the environment itself might change, because of the species' actions (example?) or because of other species' actions.**
4. **Acquired skills die with the parent: only inherent characteristics are passed on.**
   But: For *Homo sapiens*, language means culture, which can be passed on through deliberate learning.

Wollemia nobilis

Holland's (1975) Genetic Algorithm mimics natural evolution.

1. A population of "individuals," each having a fitness, which is measured.
2. The fittest individuals are chosen to breed a new population of offspring, inheriting fitter traits and genotypes.
3. Return to 1. (Iterate)

In breeding, the GA uses the processes of:

- *Selection* of parents to breed new offspring,
- *Crossover* of parents' chromosomes to pass on a mixture of their two genotypes, and
- Random *mutation* of some genes.

Or: selection, exploitation or imitation (of fit phenotypes), and exploration (of the genotype space) which reduces the risk of local optima.

Four GA Design Choices:

1. *Measures of Fitness.*
   Depends on what's being modelled.
   e.g. utility, wealth, survival time, profit,
   The average fitness of successive populations is always monotonically increasing: local optima.

2. *Selection mechanisms.*
   In choosing parents, need to retain some diversity in genotypes: don't only choose the fittest individuals to mate.

   Tournament selection: choose pairs of individuals at random; take the fitter of the pair as a parent.
   Typically, retain 40% of new population from the old, to protect better traits (embodied in the genotypes).

3. ***Genetic Operators: crossover & mutation.***
Single-point crossover: take two chromosomes (one from each parent), cut at the same randomly chosen position, swap the cuts to create two offspring (use at least one). Possible to have more than one cut point, but not important.

Crossover preserves traits (from combinations of adjoining genes on the chromosome).

**Mutation: probability that a gene changes.
Creates novelty at the genome level: diversity.**

4. ***Population size.***
**Should considerably exceed the number of genes in each chromosome.
If the population is too small, then increased risk of convergence to a local optimum.
But 20 to 50 has been used successfully.**

Docking of a GA Model

In G&T 2nd edition, they dock Axelrod's (1987) model of the IPD:

Axelrod (1987) wrote in Pascal VS,

Marks (1988) wrote in C,

http://www.agsm.edu.au/~bobm/papers/niche.pdf

G&T (2005, pp. 239–247) write a GA in NetLogo !

http://cress.soc.surrey.ac.uk/s4ss/code/NetLogo/axelrod-ipd-ga.html

Szpiro (1997) used a GA to demonstrate the emergence of risk aversion.

GA Developments

Coding of the Genotype:
Nature uses the four nucleotide amino acids: G, A, T, C.
Holland's classical GA uses bits in a binary string.

But real numbers are possible: easier to deal with many real-valued problems.
Mutation: added a randomly chosen small number ($N(0, \sigma)$) to a small proportion of genes.

Or: use programs as genes → Genetic Programming (Koza 1992).
Need to ensure that crossover and mutation preserve a syntactically correct program, even if it performs very poorly.

***Classifier Systems* (Holland et al. 1986): Production Sets which can alter their rules by learning from feedback after it has acted.**

Best Individual or Whole Population?

GA largely used as an optimiser: asking what is the best value (of fitness) in the population?

But focussing on the value of the best individual throws away the population's emerging characteristics as a population.

It ignores the aggregate level of emerging phenomena.

Axelrod (1987) not only sought the best-performing strategy in the IPD,

but also asked questions at the aggregate level of the population, such as its stability against invasion by a different strategy: e.g. Tit for Tat's stability against Always Defect. (See Marks 1989.)

How many populations in the GA?

Vriend (2000): with a single population in the GA, we must distinguish *social learning* from *individual learning*.

Social learning occurs at the genotypic level: sexual reproduction means that parents can communicate (share information) with their offspring via crossover:

∴ over generations, fitter genes or traits can spread through a single population covertly, by inheritance of genetic material.

How many parents, grandparents, great grandparents, ... do you have?

"Individual" learning occurs at the phenotypic level as each individual interacts with others, is scored for its fitness, and is ∴ selected (or not) as a parent for the next generation, depending on its fitness ranking.

Multiple-population GA learning

With one population per player: Individual learning occurs only through arm's-length competition, and the selection of fitter individuals as future parents, not through inheritance of genetic material. (Illegal communication?)

When all members of a population are identical, then genetic inheritance is not a problem, since the aim is in general only to seek the fittest individual.

GAs with Co-evolution: Many Populations

When the environment in which the GA operates changes, and when such change is due to the behaviour of the species' competitors — co-evolution — then sharing of genetic material blurs the distinction between species.

Example: If the GA is being used to explore the behaviour of sellers in an oligopolistic market, genetic sharing can only model sub-rosa communication across brands.

This is illegal under most antitrust regimes, and therefore in general should not occur in the model, lest the results rely on it.

… so: How many populations?

The answer to the question: how many populations? is then: as many as there are distinct players, or distinct species coevolving.

Example: When each seller in an oligopoly has distinct costs, faces distinct demand, perhaps with a distinct actions set, then it should be modeled using a distinct population.

Perhaps because each GA has an internal population of individuals, a tendency to think of the GA as modeling heterogenous players.

But a single population assumes homogeneity.

Individuals in the GA

Each string in a GA population could be:

- **an individual brand (say), which I have argued above is unrealistic in general, or**
- **one possible decision, of a population that the agent could make — makes sense with a population per distinct player.**

So each new generation could be:

- **new individual decision makers (brands), or**
- **new ideas or heuristics belonging to long-lived players.**

Disputes about GAs in Economics:

Chattoe (1998) argues (correctly) that there has been confusion over the role of the GA:

- **an instrument to search a rugged solution space, or**
- **a model of firms' decision making and individual behaviour.**

Dawid (1999) argues that the GA is good at modelling the learning of populations of agents.

Curzon Price (1997): the GA provides a stream of hypothetical actions or strategies, which may or may not be used.

Duffy (2006) concludes that empirical evidence exists that GAs are reasonable "models of adaptive learning by populations of heterogeneous agents."

III. Explicit Agent-Based Learning

With populations in a GA, learning is implicit: it occurs at the population level, not at the individual level — it emerges.

Arthur (1991, 1993) was the first economist to model explicit agent learning, and to calibrate his models using data from human-subject experiments.

In his Reinforcement Learning (RL) model, how an agent chooses to act later is a function of the outcomes it experienced as a result of earlier choices — the Thorndike effect.

At first he calibrated individual learning, but with the artificial stock market (Arthur et al. 1997), he became interested in data at the aggregate level.

Arthur's RL Model, the earliest

His model: In round t, player i has a propensity $q_{ij}(t)$ to choose pure strategy j, and q_{ij} is updated:

$$q_{ij}(t+1) = q_{ij}(t) + (x - x_{\min}),$$

where x was the payoff for choosing strategy j previously, and

$x_{\min}$ is the lowest possible payoff.

$\therefore$ The propensity to choose a strategy j is *reinforced* if j has provided higher payoffs in the past, and vice versa.

$p_{ij}(t) = \frac{q_{ij}(t)}{\Sigma_{k=1}^{N} q_{ik}(t)}$ is the probability that agent i plays strategy j in period t, a function of all strategies' propensities.

Roth and Erev's generalisation

Roth & Erev (1995), Erev & Roth (1998) generalised Arthur's RL model to get a better fit with experimental data from multi-player games.

Initial propensities $q_{ij}(1)$ are equal across all strategies.

$\Sigma_j q_{ij}(1) = S_i(1) = S(1)$, an initial propensity parameter, equal across all players and strategies.

The rate of learning is proportional to $S(1)$.

Again, $p_{ij}(t) = \frac{q_{ij}(t)}{\Sigma_{k=1}^{N} q_{ik}(t)}$ is the probability that agent i plays strategy j in period t.

The Roth-Erev Model ...

Player i updates his propensity to play strategy j according to the rule:

$$q_{ij}(t+1) = (1-\phi)\, q_{ij}(t) + E_k(j, R(x)),$$

where $$E_k(j, R(x)) = \begin{cases} (1-\epsilon)\, R(x) \text{ if } j = k, \text{ or} \\ \frac{\epsilon}{N-1}\, R(x) \text{ otherwise,} \end{cases}$$

where $R(x) = x - x_{\min}$.

Three parameters:

- **initial-propensity parameter $S(1)$**
- **recency parameter ϕ: reduces the power of past experiences**
- **experimentation parameter ϵ**

When $\phi = \epsilon = 0$, Roth-Erev is Arthur.

Five Types of RL Models

Five types of RL models (Duffy 2006, Brenner 2006):

1. **the Arthur-Roth-Erev model above**
2. ***Q*-learning, which optimises long-term payoffs rather than immediate returns (Watkins & Dayan 1992)**
3. **multi-agent *Q* learning (Hu & Wellman 1998), and**
4. **Adaptive Play (Young 1998)**
5. **Another modification of RL: suppose that agents have certain "aspiration levels" in payoff terms that they are trying to achieve. This idea has a long history in economics dating back to Simon's (1955) notion of *satisficing*.**

   Could use x_{asp} instead of x_{min} above.

Anticipatory, Belief-Based Learning — Inductive

RL and GA-based learning models are deductive: respond to past actions and payoffs.
No attempt to anticipate and reason back, inductively.

***Belief-based learning*: agents form beliefs about other players' likely actions, and so respond to their beliefs. Inductive.**

Gjerstad & Dickhaut (1998): "heuristic belief learning": agents use heuristics to update their beliefs about others' actions (expressed as probabilities) — they found good convergence to competitive equilibrium and good fit with aggregate behaviour.

Timing of bids is crucial.

Selten's Directed Learning

Ex-post rationality determines adaptive behaviour.

Requires an ordering over the set of possible actions.

Players probabilistically move towards actions that would have been profitable had they been chosen earlier; and never move to lower their payoffs.

Hailu & Schilizzi (2004): use a mixed (i.e. probabilistic) strategy in a procurement (i.e. selling) tender:

- **if bid X won last auction, then slightly raise the bid, using $P(\frac{1}{2})X$ and $P(\frac{1}{2})\ X$+10% as next bid,**
- **if bid X was too high, then slightly lower the bid, using $P(\frac{1}{2})X$ and $P(\frac{1}{2})\ X-$10% as next bid, both bounded.**

Learning: How to optimise (LHTO), or how to predict (LHOP)?

LHTO: GA searches for actions or strategies that are best, lead to highest fitness (profits etc.)

LHTP: use the GA strings to encode how prices will change from period to period.

Used to calibrate GA output with human-subject experimental data.

For us humans, it seems that predicting prices is easier than predicting how to respond to changing prices.

Perhaps this suggests how markets help us solve difficult problems (see the *EJ* June 2005 feature discussed in Lecture 2).

Are Learning Strategies Better? (In Repeated Games)

Airiau et al. (2007) evaluate several learning and non-learning strategies in an evolutionary tournament where agents adopt successful strategies from the previous generation.

The testbed: all 57 distinct 2×2 games. Nine strategies:

Simple: ***Random R, Generalized Tit for Tat GTFT, Best Response BR, MaxiMin M.***
Sophisticated: ***Nash equilibrium N, Fictitious Play FP, Best Response to Fictitious Play BRFP, Bully, Saby.***

Static: ***R, N, M, Bully.***
Simple, purely reactive: ***GTFT, BR.***
Learning Strategies: ***FP, BRFP, Saby.***

Bully **chooses an action which maxes its payoff assuming the other will respond optimally to this action.**
Saby **plays its best response to an estimated probability distribution across the other's actions, given** ***Saby*****'s last move.**

Yes in Simple Round-Robins

1. In a round-robin with one player per strategy:

static strategies and *N* are dominated by the learning strategies: static strategies and *N* are not always efficient (Pareto-optimal outcomes)

***N* could significantly outperform only *R*.**

***FP* loses to *BRFP*, but beats *R*.**

Yes in Evolutionary Tournaments

2. **In an evolutionary tournament, where agents adopt the best strategies in the previous generation (i.e. selection, without the genetic operations of crossover and mutation):**

   the learning strategies (including *FP* and *BRFP*) outperform strategies such as *N*, and survive the evolutionary process.

   ***N* cannot survive when learning strategies are present in the first generation.**

   Learning strategies may not overwhelm all others: *Bully* can survive with them.

References

[1] Stéphane Airiau, Sabyasachi Saha and Sandip Sen (2007), Evolutionary Tournament-Based Comparison of Learning and Non-Learning Algorithms for Iterated Games, *Journal of Artificial Societies and Social Simulation* vol. 10, no. 3, http://jasss.soc.surrey.ac.uk/10/3/7.html

[2] Arthur, W.B. (1991), Designing economic agents that act like human agents: a behavioral approach to bounded rationality, *American Economic Review* 81: 353–359.

[3] Arthur, W.B. (1993), On designing economic agents that behave like human agents, *Journal of Evolutionary Economics* 3: 1–22.

[4] Arthur, W. B., J. Holland, B. LeBaron, R. Palmer and P. Tayler (1997), Asset pricing under endogenous expectations in an artificial stock market, in: W. B. Arthur, S. Durlauf and D. Lane, eds., *The Economy as an Evolving Complex System II* (Addison-Wesley, Reading, Mass.) 15–44.

[5] Axelrod R., The evolution of strategies in the iterated Prisoner's Dilemma, in L. Davis (ed.), *Genetic Algorithms and Simulated Annealing*, London, Pittman, 1987.

[6] Brenner T., 2006, Agent Learning Representation: Advice on Modelling Economic Learning, Ch 3 in the *Handbook of Computational Economics, Volume 2: Agent-Based Modeling*, edited by Leigh Tesfatsion and Kenneth L. Judd, (Amsterdam: Elsevier Science).

[7] Chattoe, E. (1998), Just how (un)realistic are evolutionary algorithms as representations of social processes? *Journal of Artificial Societies and Social Simulation* 1(3).
http://www.soc.surrey.ac.uk/JASSS/1/3/2.html

[8] Curzon Price, T. (1997), Using co-evolutionary programming to simulate strategic behavior in markets, *Journal of Evolutionary Economics* 7(3): 219–254.

[9] Dawid, H. (1999), *Adaptive Learning By Genetic Algorithms: Analytical Results and Applications to Economic Models* (Springer, Berlin), 2nd ed.

[10] Duffy J., 2006, Agent-Based Models and Human Subject Experiments, Ch 4 in the *Handbook of Computational Economics, Volume 2: Agent-Based Modeling*, ed. by L. Tesfatsion and K.L. Judd, (Amsterdam: Elsevier Science).

[11] Erev, I. & A.E. Roth (1998), Predicting how people play games: reinforcement learning in experimental games with unique mixed strategy equilibria, *American Economic Review* 88(4): 848–881.

[12] Gjerstad S. & J. Dickhaut (1998), Price formation in double auctions, *Games and Economic Behavior* 22: 1–29.

[13] A. Hailu & S. Schilizzi (2004), Are Auctions More Efficient Than Fixed Price Schemes When Bidders Learn? *Australian Journal of Management*, 29(2): 147–168.

[14] Holland J.H. (1975), *Adaptation in Natural and Artifical Systems,* Univ. of Michigan Press.

[15] Holland J.H. et al., *Induction: Processes of Inference, Learning and Discovery,* Bradford, Cambridge, Mass., 1986.

[16] Koza, J.R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT Press, Cambridge.

[17] Marks, R.E. (1988) Niche strategies: the Prisoner's Dilemma computer tournaments revisited, `http://www.agsm.edu.au/~bobm/papers/niche.pdf`

[18] Marks, R.E. (1989) Breeding optimal strategies: optimal behavior for oligopolists, *Proc. of the Third International Conference on Genetic Algorithms,* George Mason University, June 4–7, 1989, ed. by J.D. Schaffer, San Mateo, Calif.: Morgan Kaufmann Publishers, pp.198–207.

[19] Marks R.E., 2006, Market Design Using Agent-Based Models, Ch 12 in the *Handbook of Computational Economics, Volume 2: Agent-Based Modeling*, edited by Leigh Tesfatsion and Kenneth L. Judd, (Amsterdam: Elsevier Science).

[20] Roth, A.E. and I. Erev (1995), Learning in extensive form games: experimental data and simple dynamic models in the intermediate term, *Games and Economic Behavior* 8: 848–881.

[21] N.N. Schraudolf & J.J. Grefenstette, A User's Guide to GAucsd 1.4, Technical Report CS92-249, CSE Department, UC San Diego, 1992.

[22] Selten, R. (1998), Features of experimentally observed bounded rationality, *European Economic Review* 42: 413–436.

[23] Simon, H.A. (1955), A Behavioral Model of Rational Choice, *Quarterly Journal of Economics*, 69, 99–118.

[24] H. A. Simon. *Models of bounded rationality,* MIT Press, 1982.

[25] H. A. Simon (1983) Why should machines learn?, in *Machine Learning: An Artificial Intelligence Approach,* R. S. Michalski, J. G. Carbonell and T. M. Mitchell, editors 1983, Springer-Verlag. Berlin. pp. 25–37.

[26] H. A. Simon , *The Sciences of the Artificial,* The MIT Press, 2nd ed., 1996.

[27] Szpiro G., The emergence of risk aversion, *Complexity*, 2(4), 31–39, 1997.

[28] N. J. Vriend, An illustration of the essential difference between individual and social learning, and its consequences for computational analyses, *Journal of Economic Dynamics & Control,* 24: 1–19, 2000.