

Decision Making and Reward, Computational Perspectives

This entry reviews historical developments in the study of the role of incentives in decision making, emphasizing recent computational approaches that model decision-making's physical basis.

Making a decision means selecting an action from a discrete set of alternatives. The role of reward in decision-making has been a focus of interest in psychology at least since the early 1900s, when Edward Thorndike proposed the Law of Effect to describe how rewards shape animal behavior. Efforts to develop mathematical descriptions of perception and behavior after World War II, for example, led to signal detection theory – which assumes reward-maximizing behavior and uses it to characterize basic perceptual abilities – and to theories of economic preference such as R. Duncan Luce's choice axiom, in which the probability of one choice over another is independent of the set of third options available. Behaviorism reached the zenith of its influence around this time as well, using B. F. Skinner's automated experimental techniques in an attempt to characterize behavior in terms of its consequences for reinforcement. The result was a valuable body of data and a set of robust behavioral regularities (such as Richard Herrnstein's 'matching law') that continue to constrain theories today. From a contemporary perspective, though, most theorizing during this period — especially behaviorist theorizing — was non-computational: that is, it did not involve simulating or mathematically modeling a causal process of reinforcement-guided behavior at any level of physical description, either with a machine or with pen-and-paper calculations. This

changed when the post-war computer revolution encouraged researchers to regard cognition as a physical, computational process determined by the state of an organism's brain.

Reinforcement Learning

Reinforcement learning (RL) has been a powerful force in machine learning, psychology and neuroscience since the 1980s. It blends a computational approach to decision making with the behaviorist/classical-economic assumption that agents act so as to maximize, or at least improve, earnings. RL theory developed from the theoretical foundations established in control theory by Richard Bellman in the 1950s and '60s, but specifically exploited the recursive structure of equations for predicting future reward as a function of an action policy (specifying which action to take in every state of the environment) applied to a discrete state-space representation of the world. (A discrete state-space is a representation consisting of a list of all the unique states in which an agent could find itself; in contrast, a *continuous* set, such as the set of real numbers used in calculus, is uncountable and could never be exhaustively written down, even in an infinitely long list.) Using the discrete state-space approach, Richard Sutton and Andrew Barto showed that unsupervised, online learning by trial and error was an effective method for creating artificial agents without programming in all possible relevant knowledge — indeed, without the programmer even having this knowledge. Despite this success, RL approaches to decision making are often hindered by their frequent reliance on a *compound-serial* representation of time: an extremely memory-intensive representation in which a binary state variable is assigned to every relevant subinterval of a time period

and linked in a chain (a binary state variable is a memory slot for a 1 or a 0 indicating whether the world is in the corresponding environmental state).

Decision Making in Continuous Time

The study of decision making in continuous time, however, has remained a pre-eminent concern in psychology, since response time (RT) data are continuous. Psychological decision-making models account for RT data using a variety of real-time mechanisms. Many of these also assume finely discretized time (time divided into a large number of small subintervals), but unlike compound-serial models, they do not require a unique representational state variable in memory for every discrete subinterval. For example, in the case of a two-alternative, stimulus-discrimination task (e.g., determining whether a light is bright or dim), an agent might initiate two response processes (requiring one real-valued state variable apiece). Each process samples the stimulus repeatedly, thereby accumulating information and racing the other process to a *threshold* for producing its corresponding response. Accumulation-to-threshold can be conceived as the tallying of votes in favor of one hypothesis about stimulus identity over another (e.g., bright vs. dim).

When the two tallies compete with each other directly — i.e., when a new vote for one hypothesis takes away a vote for the alternative — the difference between tallies is a number that traces out a random walk as it changes over time, much like a stock market price over many days of trading. The resulting model can implement the sequential probability ratio test (SPRT) of statistical decision theory, which chooses between two

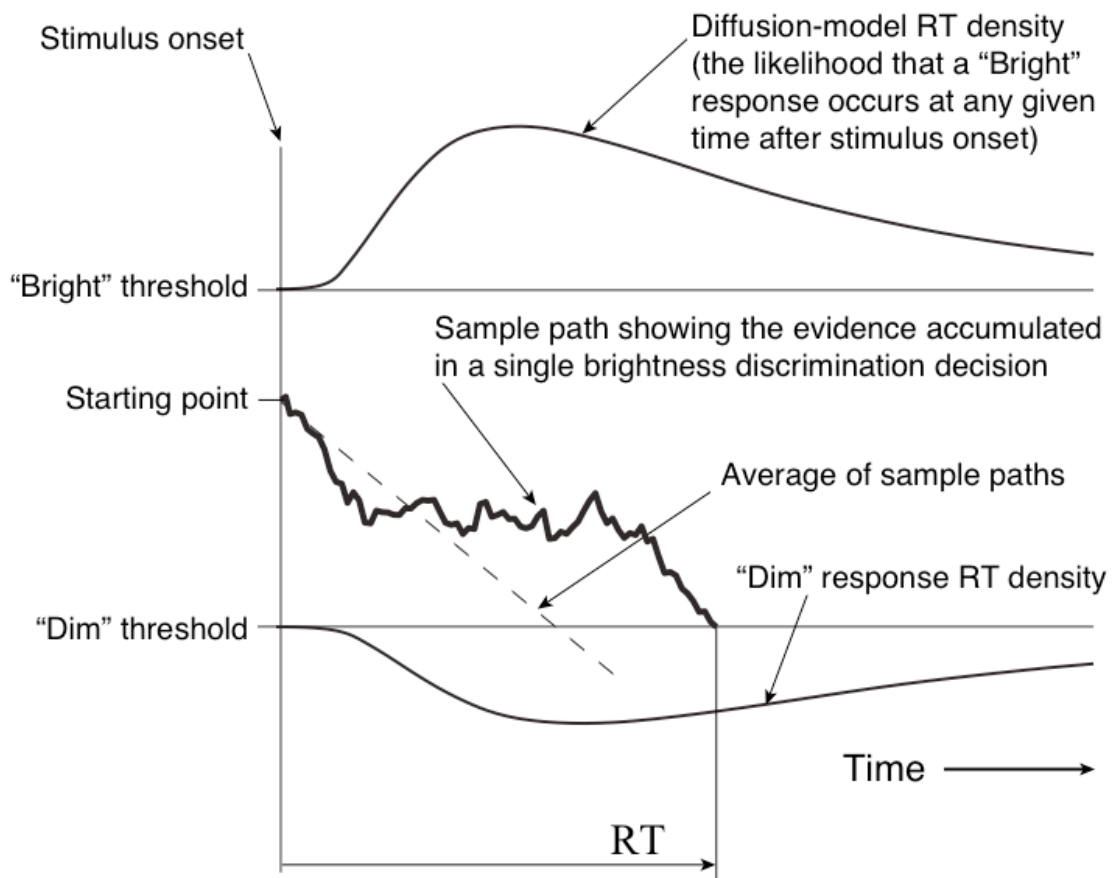


Figure. 1: The diffusion model of decision making. Here, one example of a computer-simulated evidence accumulation process (called a “sample path”) is shown in bold. It begins from a starting point representing prior beliefs about whether the stimulus will be bright or dim. When the stimulus begins, the process is driven downward at a constant rate (on average) by a dim stimulus and upward by a bright stimulus. Noise perturbs the process as it drifts downward. On this trial, a response is made some fixed amount of time after the process crosses the “dim” threshold. The time at which it crosses either threshold defines the model’s “decision time”; the sum of decision time and an additional, small, fixed duration defines the model’s “response time” (RT); and a large number of RTs produces a distribution, illustrated as a separate “RT density” function for each decision threshold.

hypotheses by repeatedly sampling a data source. This is appealing since the SPRT minimizes the average number of stimulus-samples needed to achieve any given level of accuracy. It therefore lends itself well to maximizing the expected rate of rewards that could be earned by making a series of correct decisions. It also explains naturally why

people and animals typically produce a *speed-accuracy tradeoff*: when response thresholds are low in the SPRT (i.e., when thresholds for decisions in favor of each hypothesis are more easily achieved), less time needs to be spent collecting information in order to make a response, but the chance of an error increases.

Integrating Different Approaches

Recently, work on integrating RL with continuous-time models of decision making (such as Roger Ratcliff's *diffusion model*, Fig. 1) has produced novel explanations of basic phenomena in RT data that have resisted any widely accepted explanation. For example, an explanation of faster RTs and increased response probability for more preferred responses comes from models in which the threshold of a random walk process is adapted to maximize the rate of rewards. When this happens, agents can rapidly adapt to a near-optimal speed-accuracy tradeoff in response to changes in the pace of a task. Threshold-adaptation approaches are complemented by others in which the random walk itself is biased to head in one direction by reward history. Both approaches have lower-level counterparts that simulate neural processing in the brain using spiking, integrate-and-fire models.

These encouraging results are still very limited compared to the scope of discrete-time RL. Nevertheless, the nascent integration of discrete and continuous-time approaches may lead to better artificial agent performance and a more thorough understanding of the neural circuits underlying biological agent performance, because it draws on methods of adaptation suitable for the widest range of physical systems – both systems for which a

discrete state space is the best description, and those represented optimally with a continuous state space.

Patrick Simen

See also Decision Making; Neural Underpinnings; Neuroeconomics; Reinforcement Learning; Psychological Perspectives

Further Readings

Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review*, *113*, 700–765.

Busemeyer, J. & Townsend, J. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Simen, P., Cohen, J. D. & Holmes, P. (2006). Rapid decision threshold modulation by reward rate in a neural network. *Neural Networks*, *19*, 1013–1026.

Soltani, A. & Wang, X.-J. (2006). A biophysically based neural model of matching law behavior: melioration by stochastic synapses. *Journal of Neuroscience*, *26*, 3731–3744.

Sutton, R. & Barto, A. (1998). *Reinforcement Learning*. Cambridge, MA: MIT Press.