OR/MS Today - *April 2010*

Forum

Neuroscience and operations research: a two-way street

By Stuart Dreyfus

In 1986, my brother Hubert, a professor of philosophy, and I wrote the book "Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer" [Dreyfus and Dreyfus 1986] in which we argued that the brain produces skillful coping behavior in familiar types of situations by using involved intuition rather than by detached thinking. By "thinking" we meant the kind of reasoning, symbol-manipulating, rule-following, theory-based procedures, etc. that we are consciously aware of as we face novel problems. Our primary goal was to argue that the belief held by most researchers designing expert systems at that time — that experts use reasoning and rules — was misguided. Our argument was phenomenological, meaning based on careful observation of both novice and expert naturalistic behavior.

In chapter 6 on Managerial Art and Management Science, we applauded the construction of O.R. models of structured domains such as inventory control or queueing phenomena and of novel situations, but we questioned the advisability of developing models of unstructured situations such as business managerial or public policy issues that are based on the interrogation of experts about what they considered the important facts describing the situation (state variables), the rules by which they would change over time given decisions (dynamics), and a measure of quality of the resulting sequence of events (criterion).

We also questioned the use in familiar types of situations of decision analysis requiring that experts furnish probabilities of events and utilities of skeletally described outcomes. We could, however, offer no convincing refutation of the belief prevalent in artificial intelligence research and implicitly held in operations research that, while experienced experts in familiar types of situations make intuitive decisions rapidly and effortlessly, they must be doing so by unconscious thinking, presumably based on shortcuts and rule compilations acquired during their experience.

With trepidation we offered the conjecture that the intuitive brain may store a large repertoire of remembered situations that had been successfully handled in the past, and may somehow access one similar to the current situation and then use that information to produce its decisions. By 1988, when a paperback version of our book was published, we had learned enough about neural networks to renounce our separately remembered situation (i) view in favor of synaptic-based pattern discrimination and association, but we in no way anticipated the neuroscientific events described below. While this explanation of intuition survives today [Klein 2003], modern behavioral neuroscience is finding otherwise [Dreyfus 2004], and operations research has played a fundamental role in this conclusion.

Behavioral neuroscientists

Behavioral neuroscientists are little concerned with the biophysical and biochemical details of neural nerve cell behavior, the traditional domain of neuroscientific research, but rather with how neuronal activity produces thought and action. This research is speculative but marshals considerable brain knowledge supporting its hypotheses. A major tool of the behavioral neuroscientific field is brain imaging, generally using either the modern functional magnetic resonance imaging (fMRI) or the older electroencephalographic (EEG) recordings. Information is also gleaned from electrode readings from the brain of performing lab animals. Furthermore, much is learned from the impact of damage to brain areas. Research generally involves the computational simulation of neural networks employing architectures consistent with what is known about connectivities among various brain areas and about neural activity observed in those areas under various conditions. Theories are buttressed by the ability of

computational brain simulations to explain known experimental psychological data.
We now turn to ways behavioral neuroscience and O.R. do or can inform each other. An influential group of behavioral neuroscientists has hypothesized that the satisfaction of Bellman's dynamic-programming equation expressing mathematically what he called the principle of optimality is accomplished by the brain by means of a successive approximation procedure. They call their use of this principle involving learning in the course of acting by the name temporal difference reinforcement learning (TDRL) [Dayan and Niv 2008], taking their lead from research in machine learning where the idea was first developed [Sutton and Barto 1998].
These behavioral neuroscientists refer to Bellman's optimal value function as the critic, because it assigns to each occurring current state the expected future reward accruable during a sequential decision process and is used during learning to criticize a chosen action. The brain area responsible for learning a good action or decision given the current state — Bellman's optimal policy function — is called the actor area. The goal of an organism is taken to be the maximization, or at least the successive improvement, of the expected future reward from carrying out a sequential behavior. Reward, it is generally agreed by neuroscientists [Montague, Hyman and Cohen 2004], is represented in the brain by the level of production of the neuromodulatory chemical compound dopamine and depends upon events and the current goal of the organism.
This group of behavioral neuroscientists proposes that the brain, while learning a skill through many repeated trial-and-error sequences in a domain, starts with little or no knowledge of correct actions or the expected future reward associated with encountered states, but simultaneously improves its estimate of each. Since the learning organism, be it human or lower animal, does not initially know a model of the environment (its dynamical relations etc.), many of these neuroscientists propose that, rather than learning a model based on observed experiences in the real world the way a physicist or engineer might, what needs to be learned by the brain is merely the correct critic value of various states and the appropriate action. This is called model-free learning.
The learning principle of this actor-critic TDRL approach is that, given an estimate of the expected future reward of the current state of the organism and given an action chosen based on the current action-generating brain area, suppose that the environment then moves the organism to a new state for which the brain will already have an estimated expected future reward. Then, if (1) the immediate reward, if any, given the current transition plus the estimated future reward from the organism's new state exceeds (2) the estimate of the future reward of the original state, then the estimated reward of the original state is incrementally increased by modification of the synapses producing the critic's estimate. Furthermore, the synapses of the action area change to become more likely to choose that action again if, on a later realization of a process involving the same skill, the organism is again in the same state. The discrepancy between (1) above and (2) above, if one exists, is called the temporal difference error and drives the learning.
Should the chosen decision produce a negative temporal difference error the expected future reward is reduced and the chosen action is penalized in future realizations. This process is repeated at the states encountered during future realizations of the same sequential skill. In order to allow the exploration required to produce an improved action, the chosen action is always a perturbation of what the actor area suggests. Of course, this procedure will lead, at best, to successive improvement of performance with no assurance of global optimality, but it is unlikely that an organism's learned behavior in the real world is always optimal. More often it achieves a local maximum unimprovable by small variations in actions (ii).

Perspective Power

These behavioral neuroscientists have furthermore conjectured, based on solid evidence, that the brain is capable of gating the incoming stimuli from sense organs (or presumably in some cases from an observed constellation of relevant information) so that the importance of different stimuli can vary depending upon circumstances. A change in gating is generally termed by neuroscientists a changed goal or rule [Miller and Cohen 2001], [Rougier et al. 2005]. In my research with my brother [Dreyfus and Dreyfus 1986] we, instead, used the term "perspective" to describe the way stimuli were salienced. Attaining the ability to skillfully cope requires that a certain brain area learn when a particular goal or perspective is appropriate for attaining maximal future reward and under what circumstances to change its goal or perspective (iii).
We have seen, then, that model-free actor-critic TDRL based on operations research's Bellman equation together with learned stimuli-gating to realize perspective seems to explain what my brother and I called intuitive expertise. This theory leaves no need for conjecturing, as we did, the use of a library of remembered separate successful experiences. It explains both the speed and effortlessness of naturalistic skillful coping accomplished entirely by synaptic modifications to achieve learning and by spreading neural activations to produce behavior.

The idea of learning when circumstances dictate a change in goal during a sequential process is beginning to infiltrate the machine-learning literature [Botvinick, Niv and Barto 2008], and operations research models of novel sequential processes could probably benefit from this same elaboration.

But there is a far more significant byproduct of the recent brain research described above that speaks to applied operations researchers. Behavioral neuroscientists clearly distinguish between the network of brain areas, largely in the celebrated cerebral cortex with its convoluted grey matter, that collaborate to produce executive functions such as planning, reasoning, theorizing and rule following that are described as detached conscious thinking and a separate, mainly subcortical system, that, among other things, implements the actor-critic method of learning and produces involved, intuitive, experience-based, skillful coping behavior. Behavioral observations show that damage to the executive system, caused by stroke or lesion, that eliminates all of what we have called conscious thinking, fails to affect the execution of previously learned coping skills. For a striking example see [Sacks 2007].

Relating this to the modeling profession of applied operations research, the above implies that novices or experts facing unfamiliar situations, who typically use learned rules or reasoning implemented in cerebral cortex to figure out what to do, may indeed be able to reliably report to a modeler the basis of their coping behavior. Skilled performers, however, when facing familiar types of situations, who don't use thought but rather act intuitively based on synaptic modifications mainly in their subcortical brain resulting from the implementation of actor-critic learning, cannot be expected to reliably so report.

Enamored, as humans tend to be, with their conscious thinking brain, most subjects, if interrogated about an act or choice, will probably make up some sophisticated version of the procedure they had used as novices, mistakenly thinking that it must be what they were subconsciously doing. This, however, when seen from the perspective of model-free TDRL, should not be taken as the true basis of their intuitive naturalistic skillful coping. This raises serious questions about models of unstructured situations based on information elicited from experienced experts (iv). Perhaps O.R. modeling should be trusted only in novel situations, and this includes situations where problems are familiar but the problem environment is novel such as during an international financial crisis where no intuitive experienced experts exist, or in structured situations where the states, dynamics and criterion or other model parameters are directly observable without consulting experts.

It may well turn out that, while behavioral neuroscience has greatly profited from a fundamental result of operations research, its discoveries offer our field something of equal importance in return. To those applied operations researchers who, with hubris, attempt to go beyond the modeling of novel or structured operational situations and instead ask intuitive model-free experts in types of situations concerning unstructured management and policy problems with which they are familiar to articulate models that then become the basis of their computational recommendations, neuroscience may be offering compelling reasons for humility.

Stuart Dreyfus *(dreyfus@ieor.berkeley.edu) is professor emeritus at the IEOR Dept., U.C. Berkeley.* *Support for this research was provided by Statoil ASA of Norway through Project Academy StatoilHydro. The author would appreciate receiving comments or questions concerning this article. For extended notes accompanying this article, see the online edition of* OR/MS Today.

References

1. Botvinick, M.M., Niv, Y., and Barto, A.C., 2007, "Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective," www.princeton.edu/~yael/NIPSWorkshop/BotvinickNivBarto.pdf.
2. Dayan, P. and Niv, Y., 2008, "Reinforcement Learning: The Good, The Bad and The Ugly," *Current Opinion in Neurobiology,* Vol. 18, pp. 185-196.
3. Dreyfus, H. and Dreyfus, S., 1986, "Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer," The Free Press, New York.
4. Dreyfus, S. 2004, "Totally Model-Free Learned Skillful Coping," *Bulletin of Science,* Technology & Society, Vol. 24, No. 3, pp. 182-187
5. Klein, G., 2003, "Intuition at Work," Doubleday, New York.
6. Miller, E.K. and Cohen, J.D., 2001, "An Integrative Theory of Prefrontal Cortex Function," Annual Review

of Neuroscience, Vol. 24, pp. 167-202.
7. Montague, P.R., Hyman, S.E. and Cohen, J.D., 2004, "Computational roles for dopamine in behavioural control," Nature, Vol. 431, pp. 760-767.
8. Nonaka, I. and Takeuchi, H., 1995, The Knowledge-Creating Company, Oxford University Press, New York and Oxford.
9. Rougier, N.P., Noelle, D.C., Braver, T.S., Cohen, J.D., and O'Reilly, R.C., 2005, "Prefrontal cortex and flexible cognitive control: Rules without symbols," Proceedings of the National Academy of Sciences, Vol. 201, No. 20, pp. 7,338-7,343.
10. Sacks, O., 2007, "Musicophilia," Knopf, New York, pp. 187-213
11. Sutton, R.S. and Barto, A.C., 1998, "Reinforcement Learning: An Introduction," MIT Press, Cambridge, Mass.

Notes

i. Warning: In year 2000, without consulting us, a paperback edition of our 1986 book was published. The new preface and changes we had inserted in the 1988 paperback have consequently become history.

ii. Shortly after the actor-critic model-free method of intuitive skill acquisition requiring no learning of the real-world's dynamics or other elements of a world model was developed in the machine-learning literature, a companion approach called Q-learning was introduced. It was elegantly proven, under suitable conditions, to lead to optimal solutions of Markovian decision processes. When the seminal book [Sutton and Barto 1998] on temporal difference reinforcement learning was produced by its developers, only three pages concerned the less elegant actor-critic idea. Most behavioral neuroscientists, however, seem quite reasonably to believe that, except when the organism must choose one of a very small number of discrete actions in each state, a situation rare in the real world of coping but common in psychological experiments and experiments requiring animal task learning, the computations required in Q-learning are infeasible. This certainly is the case where actions are chosen from a continuum of possibilities, such as the bodily movements of an animal chasing its prey or an athlete catching a ball. The activity level (firing rate) of a neuron can be any of a continuum of values, so a neural network is ideally suited to the production of a continuous function rather than a discrete-valued one. Consequently, the actor-critic version of TDRL is used in most behavioral neuroscientific simulations.

iii: In imaging experiments concerned with goals and their change, the neuroscientific literature generally studies volitional goal changes where a subject is told a set of possible goals and what action each requires and, if during a session the experimenter changes the goal, the subject is informed that the usual action is now incorrect. (The Wisconsin card sort task is often used.) Figuring out a different goal when told an action is incorrect produces neural activity in an area of prefrontal cortex, and when this area is damaged, the ability appropriately to change goals is compromised.

I doubt, however, if these studies of the conscious decision to change goals is informative concerning what brain area is involved when a coping organism, after considerable naturalistic learning experiences, automatically changes perspective when the stimuli demand it. For example, a dog, after many learning experiences, when chasing a squirrel will not continue to run toward it, but will veer to intercept it if the squirrel heads for a tree. I believe that the same can be said of the experienced baseball outfielder running to catch a fly ball, who certainly used his thinking brain when initially learning his skill. Once skilled, however, if the visual stimuli produced by watching the ball's overhead flight demands it, he will

automatically change perspective as he suddenly turns to field the ball after it strikes the outfield wall. It is easy to believe that he suddenly thinks, "I'm not going to be able to catch it and should therefore change my goal to fielding it," but I doubt that this is likely or necessary, given that the dog in the first example doesn't have or need this thinking capacity. This leads me to suspect that the neuroscientists studying largely subcortically produced actor-critic TDRL behavior are wrong in their picture of real-world effortless and automatic skillful coping when they include a prefrontal cortex goal-setting area rather than an as yet unidentified, and likely subcortical, area. They are right, of course, for the kind of situations experiments currently demand.

iv: What management consultants call experience-based tacit knowledge is presumably really synapse-based know how produced by the sum of one's relevant experiences in the way actor-critic neuroscientists speculate. It then does not take the form of a tacit mental model as certain management consultants are wont to believe [Nonaka and Takeuchi 1995], and there is no way of converting the tacit into the explicit. What passes for such a conversion can be no better than either the recitation of a few remembered experiences or else the production of a mental model of the sort that, according to our five-stage model of skill acquisition, competent performers use.

- *Table of Contents*
- *OR/MS Today* Home Page

Lionheart Publishing, Inc.
506 Roswell Rd., Suite 220, Marietta, GA 30060 USA
Phone: 770-431-0867 | Fax: 770-432-6969
E-mail: lpi@lionhrtpub.com
URL: http://www.lionhrtpub.com