The DNA of Decision Science

Sam L. Savage

This work was supported in part by the Air Force Office of Scientific Research, Bolling AFB (Wash. DC)

The decision sciences are evolving rapidly. For a curriculum of study in this area to survive, it must evolve in parallel. By considering a trait common to many evolving systems, I will explore a potential path of future evolution for the decision sciences and their teaching.

I. Some Issues Pertaining to Evolving Systems

A. Aggregation/Sublimation

Evolving systems are often comprised of building blocks which were once at the evolutionary forefront of the system themselves.

In biology, for example, ammonia and methane were at one time, the “King of Beasts”. Later, they formed the basis of amino acids which in turn became the components of DNA.

This may be viewed as an evolutionary process of *aggregation* of simple building blocks into more complex building blocks. In recent discussions with B. Curtis Eaves, he has suggested the term *sublimation* to describe what happens to the components which comprise the aggregate. I suspect there may be a more formal statement of this principle of aggregation in the emerging theory of Complexity. For example, see Waldrop (1992).

Memes

The evolutionary building blocks of ideas are discussed by Richard Dawkins in *The Selfish Gene*, (1976). He defines a cultural analog of the gene as the "meme". Memes can mutate, evolve or become extinct as they are passed from generation to generation. Evolutionary Aggregation is often discernible in the memes associated with technologies.

For example, the ancient memes of mechanical engineering; the wheel, lever and pump, formed the basis for the heat engines of the early 1800’s. For nearly half a century, these “engines” were only operated by “engineers,” in large industrial applications such as the pumping of flooded mine shafts. Eventually, as sublimated along with the transmission and pneumatic tire into the aggregate of the automobile, the engine was finally of benefit to the individual.

The memes of electronics; resistors, capacitors, coils and vacuum tubes, were the building blocks of early computers in the 1940’s. The transistor led to the aggregation of far more electronic components into the computers of the 1960’s. Until the early 1980’s, computers were only operated by engineers. Integrated circuitry has raised the level of aggregation to millions of components per square inch, and led to computers for individual use.

For technological systems, the aggregation principle may be restated as:

Today's systems are often tomorrow's subsystems.

B. Ergonomics

Much of modern industrial design is rightly focused on the *ergonomics*, or ease of use of technologies. The cases of the automobile and personal computer show that:

Complex systems are often easier to use than their simpler ancestors

Today's automobiles are far more complex than those of Henry Ford's, yet far easier to drive. The starting crank attached directly to the engine, which could occasionally break your wrist, has been replaced by the ignition key.

The lap-top computer on which I am typing this, is more complex than John Von Neumann could have dreamt of stuffing into a dirigible hanger. Yet it is much easier to use than his first simple machine..

Arthur C. Clarke's famous comment on this issue is:

Any Sufficiently Advanced Technology is Indistinguishable from Magic

C. Standardization - The Network Externality

Another important factor in technological evolution is the establishment of standards such as the distance between railroad tracks, or the number of threads per unit length on bolts. These are known in economics as *network externalities*, and can spell the difference between the proliferation or extinction of technologies.

Some technological standards evolve by design, others by chance. But whatever their origin, once established they must not be ignored.

II. The Decision Sciences

A. Aggregation, Ergonomics and Standardization

The earliest memes of Decision Science were the laws of arithmetic required for decisions involving the sharing and bartering of commodities. Technologies built on arithmetic, such as algebra and probability, form the building blocks of modern Decision Science.

Because most decision makers in industry and government do not view their decisions in algebraic and probabilistic terms, there has traditionally been a serious ergonomic problem with these techniques. I refer to this as an "Algebraic Curtain" separating the Decision Maker from Decision Science.

Furthermore, few standards have been established for the wide dissemination and application of the methods of Decision Science. This has limited their proliferation.

B. The Developmental Necessities of Applications - DNA

The meteoric evolution of the microcomputer has resulted in a new level of aggregation for the Decision Sciences. The algebraic and probabilistic building blocks of the formative years are being sublimated into computerized models containing what I refer to as the Developmental Necessities of Applications, or DNA. Unlike the algebraic representations of the past, this digital DNA is "alive". It contains just enough representative data so that a user can experimentally infer its structure. It is ergonomic in that a user with little mathematical understanding, may expand the DNA, input data, and get results as output. Furthermore standard formats are rapidly being established in which models may be rapidly and widely disseminated. In short, it is a “seed” of knowledge from which applications may grow.

Transformation vs. Formulation

Developing an application from DNA differs dramatically from the traditional Decision Science approach. Instead of formulating a model from scratch, the decision maker merely transform existing ones, which may be combined (recombinant DNA) to form larger applications.

A couple of analogies demonstrate that the right building blocks can have a multiple order of magnitude effect on development times.

A biological analogy: consider the daunting task of constructing a flamingo directly from amino acids as opposed to the trivial one of incubating a fertilized flamingo egg.

An electronic analogy: Imagine how quickly Von Neumann would have developed a computer had he been able to visit one of today’s electronic supply houses and fill his shopping cart with 100mhz motherboards, gigabyte disk drives, and megabyte memory chips.

An Example Decision Science DNA: A Spreadsheet Linear Program Model

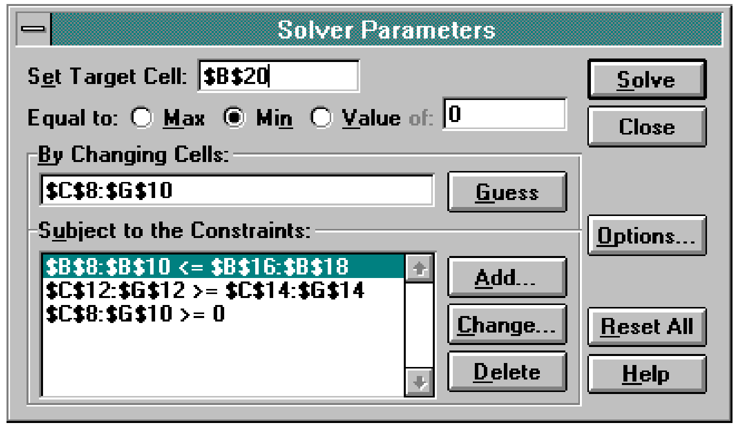
For several years, Microsoft has included limited mathematical optimization capability with its Excel spreadsheet. Roughly 1 million of these packages are sold annually. The package includes small worksheet examples for several applications. Below is a view of the dense transportation linear programming model shipped with Excel.

	A	B	C	D	E	F	G	H
1	**Transportation Problem.**							
2	Minimize the costs of shipping goods from production plants to warehouses near metropolitan demand							
3	centers, while not exceeding the supply available from each plant and meeting the demand from each							
4	metropolitan area.							
6			*Number to ship from plant x to warehouse y (at intersection):*					
7	*Plants:*	*Total*	*San Fran*	*Denver*	*Chicago*	*Dallas*	*New York*	
8	S. Carolina	5	1	1	1	1	1	
9	Tennessee	5	1	1	1	1	1	
10	Arizona	5	1	1	1	1	1	
12	Totals:		3	3	3	3	3	
14	*Demands by Whse -->*		180	80	200	160	220	
15	*Plants:*	*Supply*	*Shipping costs from plant x to warehouse y (at intersection):*					
16	S. Carolina	310	10	8	6	5	4	
17	Tennessee	260	6	5	4	3	6	
18	Arizona	280	3	4	5	5	9	
20	*Shipping:*	**$83**	$19	$17	$15	$13	$19	

Formulas relate the various parts of the model.

		Number to ship from plant x to warehouse y (at intersection):				
Plants:	*Total*	*San Fran*	*Denver*	*Chicago*	*Dallas*	*New York*
S. Carolina	=SUM(C8:G8)	0	0	0	80	220
Tennessee	=SUM(C9:G9)	0	0	180	80	0
Arizona	=SUM(C10:G10)	180	80	20	0	0

A separate window contains specifications to Excel's built-in solver to minimize the shipping cost in cell B20, by changing the values of the amounts shipped in cells C8 through G10, while satisfying both demand and capacity constraints.



Instantiation

Of course, no one has the **exact** transportation problem described above. However, this worksheet may be *instantiated* to create a simple application through transformations such as the addition or deletion of rows or columns, and entry of actual data. A click of the solve button then invokes the mathematical optimization routines. This addition of information, and transformation of the digital DNA to create an actual application is analogous to the fertilization and incubation of biological DNA.

Replication

A measure of success in evolving systems is the degree to which they are *replicated*. Of course this sort of success does not guarantee utility. Just as the DNA of the giant panda is a relative failure compared to that of the cockroach, many potentially useful applications of decision science have failed to catch on. On the other hand, some computer viruses represent successful forms of digital DNA with even lower utility than that of the cockroach.

Model Classes and Transformations

Geoffrion (1992) in his work on structured modeling, describes *classes* of mathematical models. For example, one can refer abstractly to the class of dense transportation linear programming models such as the one above, without ever mentioning a particular instance of such a problem. There is great expressive power in the ability to describe model classes as opposed to model instances.

Decision Science DNA serves as a representative of its model class. For DNA to be an effective modeling device, there must exist simple transformations which map it to any instance of the class. The types of transformations available in a modeling environment significantly influence the scope of this DNA mapping.

Scaling and Hyper-scaling

Two of the most important transformations in this regard are *scaling* and *hyper-scaling*. Scaling changes the cardinality within any dimension of the DNA, as in adding or deleting warehouses or plants in the above transportation model. Hyper-scaling changes the number of dimensions themselves, as in creating a multi-period transportation model by adding a time dimension.

Spreadsheets, Algebraic Modeling Languages, and OLAP

Several readily available modeling environments exist in which DNA may be created and transformed.

Spreadsheets

Electronic spreadsheets (such as Microsoft Excel and Lotus 1-2-3) are appealing, ergonomic modeling environments because of their interactivity, and ability to quickly graph results. More importantly, with approximately 20 million total users, they have become the modeling vernacular among decision makers. This greatly increases the likelihood that spreadsheet DNA will be replicated.

Because of this, spreadsheets provide a single environment in which to teach and prototype almost any sort of Decision Science applications. However, for industrial applications there can be drawbacks. Spreadsheet models are difficult to document, and they scale only moderately, often requiring editing or copying of formulas in the process. They are fundamentally 2 dimensional but usually allow a limited third dimension. They do not hyper-scale well.

Algebraic modeling languages

Modeling languages such as AMPL, GAMS, and LINGO have been designed primarily with mathematical optimization in mind, and are not suited to most other Decision Science applications. They are highly documentable and scalable, and reasonably hyper-scalable. Unfortunately, they are non-interactive in the spreadsheet sense. Furthermore they are in effect programming languages with an algebraic perspective that is un-ergonomic for direct use by most decision makers. As a result, they have a user base only 10 thousand or so.

For industrial size optmization problems they are today’s best solution. For teaching, however, they present a formidable learning curve which precludes their use in courses not devoted to optimization.

An expression of the transportation model DNA in LINGO appears below.

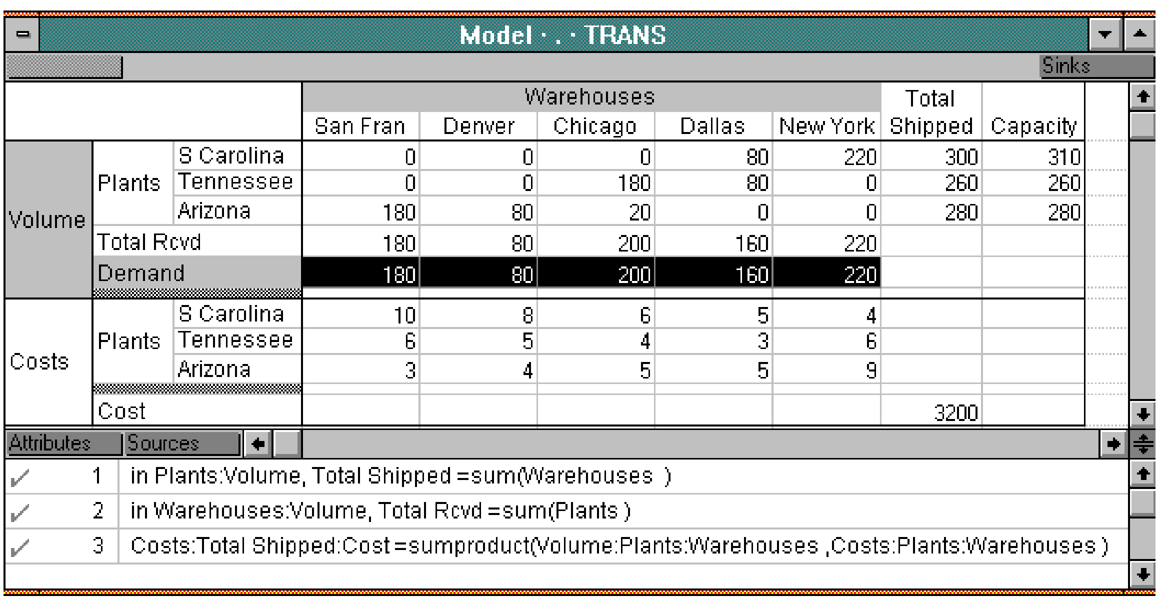
```
MODEL:
  1]
  2]SETS:
```

```
WAREHOUSE /SAN_FRAN, DENVER, CHICAGO, DALLAS, NEW_YORK /:DEMAND, RECEIVED;
PLANTS / S_CAROLINA, TENNESSEE, ARIZONA / : CAPACITY, SUPPLIED;
ROUTES(PLANTS, WAREHOUSE)  : VOLUME, COST;
ENDSETS

@FOR(PLANTS(J) : SUPPLIED(J) < CAPACITY(J));
@FOR(WAREHOUSE(I) : RECEIVED(I) > DEMAND(I));
MIN = @SUM( ROUTES(I,J): VOLUME(I,J) * COST(I, J));
@FOR(WAREHOUSE(I) : RECEIVED(I) = @SUM( PLANTS(J): VOLUME(J, I)));
@FOR(PLANTS(J) : SUPPLIED(J) = @SUM( WAREHOUSE(I): VOLUME(J, I)));
DATA:
CAPACITY = 310, 260, 280;
COST = 10, 8, 6, 5, 4, 6, 5, 4, 3, 6, 3, 4, 5, 5, 9;
DEMAND = 180, 80, 200, 160, 220;
ENDDATA

END
```

On Line Analytical Processing

New classes of data manipulating software known as On Line Analytical Processors (OLAP) or multi-dimensional modelers (MDMs) have recently emerged, see PC Magazine (1993), InfoWorld (1994), and Codd et. al.(1993). These are interactive, documentable, and scale and hyper-scale smoothly. They possess some of the best features of spreadsheets, data bases and modeling languages. Because they are new, it is difficult to predict their long term future, but roughly 10^5 have been marketed as of 1994. They range in price from a few hundred dollars for small stand alone systems (Lotus Improv) to tens of thousands dollars for client server systems with access to gigabytes of data scattered across several data bases (Ess base). It appears that it is the client server environment in which these multi-dimensional products are currently having the most success.

The transportation model DNA as expressed in IMPROV is shown below.

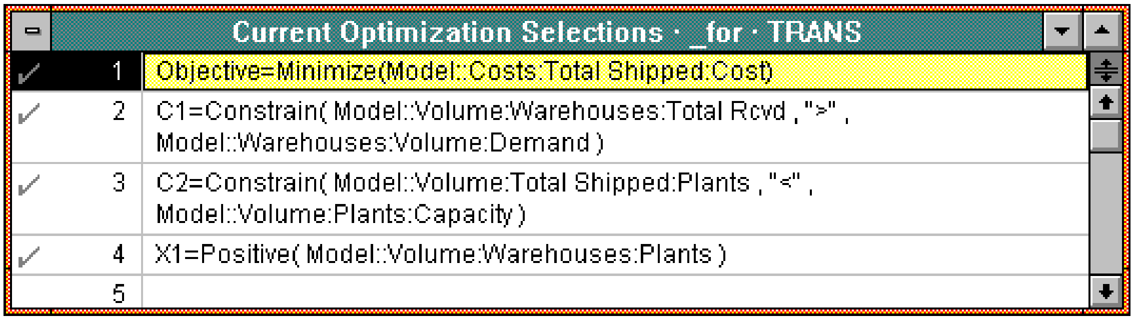


Model · . · TRANS

			Warehouses					Total	
			San Fran	Denver	Chicago	Dallas	New York	Total Shipped	Capacity
Volume	Plants	S Carolina	0	0	0	80	220	300	310
		Tennessee	0	0	180	80	0	260	260
		Arizona	180	80	20	0	0	280	280
	Total Rcvd		180	80	200	160	220		
	Demand		180	80	200	160	220		
Costs	Plants	S Carolina	10	8	6	5	4		
		Tennessee	6	5	4	3	6		
		Arizona	3	4	5	5	9		
	Cost							3200	

1	in Plants:Volume, Total Shipped =sum(Warehouses)
2	in Warehouses:Volume, Total Rcvd =sum(Plants)
3	Costs:Total Shipped:Cost =sumproduct(Volume:Plants:Warehouses ,Costs:Plants:Warehouses)

Although Improv does not provide optimization capability on its own, the author has developed a system for Primal Solutions Inc., under a grant from the Air Force Office of Scientific Research, to aggregate Improv and an algebraic modeling language into a single system. The optimization specifications are also stored in Improv.



Upon invoking the optimization command, the model along with the optimization specification are translated into an algebraic modeling language in which it may be solved directly on a PC, or in client server mode on a remote workstation. The LINGO model shown earlier was created using this system.

I believe that the OLAP and MDM environments will be central to the future of the Deceision Sciences, in which data manipulation and decision making become increasingly integrated. However, because these products are still in flux, it may be too soon to spend much time on them in Decision Science courses.

III. DNA and the Decision Science Curriculum

A. A Matter of Timing

How should the curriculum of the Research University respond to new stages of technological evolution? Neither too quickly nor too slowly.

Not too quickly, because in its early stages, it is difficult to distinguish a “flash in the pan” from a new technological standard. For example, one should keep an eye on OLAP and MDM environments, but not introduce them into the curriculum until they have stabilized.

Not too slowly, because as technologies become obsolete, or sublimated within other technologies, a curriculum becomes irrelevant. Thus, an understanding of the simplex algorithm is of no more use to someone taking a general course on the methods of Decision Science, than an understanding of the Otto cycle of internal combustion is to someone taking driver’s education.

B. We’re Overdue

There is ample evidence that curricula in the decision sciences are due for a significant change. In no particular order:

- Enrollment is off, especially in Management Science courses in business schools.

- Decision Science texts and curricula have not changed fundamentally since the microcomputer revolution of 15 years ago.
- By now it is clear that the microcomputer is not a “flash in pan”.
- Two professional societies in the United States that spearhead the decision sciences, The Operations Research Society of America (ORSA) and The Institute of Management Science (TIMS), have recently sublimated themselves into a single Institute of Operations Research and the Management Sciences (INFORMS).

C. Some Directions for Change

1. Focus Less on algorithms and More on mathematical pitfalls

As algorithms become sublimated within other technologies, such as the linear and nonlinear solvers in millions of copies of Microsoft Excel, it becomes more important to teach concepts related to using these tools rather than developing new tools. Thus efficient formulations of integer models, and convexity, convergence and stability issues should be stressed. Algorithmists need not despair, however. In the long run, nothing should increase the demand for robust algorithms faster than the current wide dissemination of optimization.

2. Focus Less on model formulation and More on model classes and data

The direct teaching of mathematical modeling is notoriously difficult and ineffective. It is more effective to introduce the DNA of important model classes and show how it may be transformed to meet individual needs. For example, in keeping with my overall theme, once the DNA of the product mix LP and dense transportation LP are presented to students, it is elementary to aggregate these to form more complex models involving both production and shipping. Furthermore, the data collection for Decision Science applications, which used to require specialized programming in its own right, can now be integrated with the applications themselves through dynamic data exchange (DDE) and client server technology.

3. Focus Less on Statistics and More on Monte Carlo Simulation

Many people manage to make it through statistics courses without ever understanding the concept of a probability distribution. Introducing random number generators into interactive mathematical models, and performing Monte Carlo Simulation, can provide an intuitive link to such concepts as the Central Limit Theorem and functions of a random variable.

4. Use Industry Standard Software Where Possible

Students are far more likely to benefit from using software that they will see again in the workplace even if it is less suited to the task at hand than a specialized product. Do not underestimate network externalities.

5. Provide DNA on disks

There is no point in taking my approach unless students leave class with a disk containing DNA. See Gardner (1992), Plane (1994) and Savage (1993, 1994). Once the students have comprehended a number of these seeds of knowledge, a natural homework assignment is to ask them to create new ones.

6. Teach less to more people

I have found that demonstrating DNA, and providing it on disk allows a more rapid coverage of topics. Of course all topics covered will not be absorbed immediately, but the student can easily reexamine them experimentally at a later date.

The bad news from the Decision Science faculty perspective is that the total course hours required for a given topic have been reduced. The good news is that there is now the potential for a much larger audience of students, and the best news of all is that those who go on to become decision makers are more likely than ever to actually use the Decision Science they have been taught.

IV. References

Codd, E.F., Codd, S.B., and Salley, C.T. 1993 *Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate*, E.F. Codd & Associates.

Dawkins, Richard. 1976. *The Selfish Gene*, Oxford University Press, Oxford, p.206

Gardner, Everette S. 1992. *The Spreadsheet Operations Manager*, McGraw-Hill, New York.

Geoffrion, Arthur M. 1992. The SML Language for Structured Modeling: Levels 1 and 2. *Operations Research*, **40**, pp. 38-57

InfoWorld, January 24, 1994, *Multidimensional Models Boost Viewing Options*, p.69

PC Magazine, September 28, 1993, *New Dimensions in Spreadsheets*, p.183

Plane, Donald R. 1994, *Management Science A Spreadsheet Approach*, The Scientific Press, Danvers MA

Savage, Sam L. 1993. *Fundamental Analytic Spreadsheet Tools for Quantitative Management*, McGraw-Hill, New York

Savage, Sam L. 1994. *Fundamental Analytic Spreadsheet Tools for Production and Operations Management*, McGraw-Hill, New York

Waldrop, Mitchell M. 1992. *Complexity*, Touchstone, New York