A Realistic Model of Rationality

Stephen Slade
Information Systems Department
Stern School of Business, New York University
44 West 4th Street
New York, NY 10012-1126
(212) 998-0818 / (212) 995-4228 (fax)
sslade@stern.nyu.edu

Abstract

The economic model of rationality has a mathematical elegance and appeal. Unfortunately, standard decision theory is not generally feasible as a computational, psychological, or social model of rationality.

We propose an alternative model based on realistic assumptions of computational power, psychological ability, and social interaction. The traditional task for demonstrating rationality is decision making. We describe the VOTE program which simulates human decision making based on our model. We discuss the application of the decision making model to the related social, multi-agent phenomena of advice, persuasion, and negotiation.

Introduction

There are many possible ways to view rationality.

Economic decision theory proposes a prescriptive, mathematical model. [Raiffa, 1968] Given knowledge of the options' probabilities and payoffs, decision theory can identify the choice with the highest expected value. According to economics, that is the rational choice.

However, outside of a casino, the decision making agent rarely knows the probabilities or payoffs for a given choice. The agent may not even know what all the options are. In certain circumstances when all the information is available, the agent still may not have the computational ability to compute the optimal answer, as in the game of chess. [Newell *et al.*, 1958] Evaluating all possible outcomes is not computationally feasible.

Simon [Simon, 1982a, Simon, 1982b] discussed these problems and proposed *bounded rationality* which incorporates information processing constraints in an effort to reflect the limitations of human cognition. Simon recognized that an agent may lack processing capability or information, and could not *optimize* a choice, but rather must *satisfice*. Simon and Newell describe decision making as a search problem. [Newell and Simon, 1972]

In this paper, we present an alternative model of rationality in the tradition of bounded rationality intended to make realistic assumptions about the decision maker. [Slade, 1994] In making a decision, the rational agent does not optimize or satisfice, but rather *justifies* her choice.

We make the following assumptions.

- *An agent has many goals with varying preferences.* We view goals quite broadly. A goal may be something specific like SATISFY THIRST or ANSWER THE PHONE, or something general like SAVE MONEY or BE HONEST. Some goals are more important than other goals.

- *An agent executes plans to achieve goals.* A plan is some sequence of actions with the intended consequence of satisfying specific goals. Generating and executing plans requires resources.

- *An agent has limited resources.* A resource is anything that may satisfy an enabling condition for a plan. Time and money are typical resources. Other resources include skills, credentials, and authority, as well as cognitive resources, such as knowledge and memory.

- *Different agents have different goals and resources.* Decision making is subjective. If it were not, markets would not exist. No buyer could find a seller.

- *An agent allocates resources to achieve her preferred goals.* If resources were unlimited, then an agent could achieve all her goals. Goals are in conflict when they vie for the same resource. A rational agent expends her resources as a reflection of her goal preferences.

Since knowledge is considered a resource, an agent is not irrational if she fails to achieve a goal from lack of knowledge.

If an agent prefers car A to car B, but cannot afford car A, then it may be rational for the agent to purchase the less expensive car B instead. Suppose she could have

negotiated a better price for car A, but did not. Her behavior is not irrational if she was unaware of the negotiation option. In this case, lacking knowledge can be considered equivalent to lacking money. If she had more money, she could have purchased car A. It is not irrational that she did not have more money.

- *Emotions are a reflection of goal states.* There is a cognitive dimension to emotions which serves to communicate goal information to other agents.

It may seem odd to discuss emotions in the context of a model of rationality. Emotional behavior is usually considered the antithesis of rational behavior. However, using Roseman's model [Roseman, 1982], it is possible to incorporate emotions as a means of reflecting an agent's state of goal pursuit.

Achieving a goal leads to a positive emotion, such as happiness or pride, whereas failing to achieve a goal results in a negative emotion, such as hate or frustration. Furthermore, the strength of the emotion reflects the importance of the related goal. An agent is passionate about what is most important.

In our view, it would be irrational for an agent to display an inappropriate emotion when winning a gold medal or losing a child. Emotions contribute to the social dimension of rationality and decision making.

- *An agent has relationships, positive and negative, with other agents, with varying strengths.* An agent is not alone in the world. An agent may have family, friends, and colleagues. In business, an agent may have employees, investors, customers, and competitors. Each of these other agents has her own set of goals, preferences, and resources.

- *Through a relationship, an agent adopts the goals of the other agent with a preference related to the strength of the relationship.* For a positive relationship, an agent may allocate resources on behalf of the other agent. For a negative relationship, an agent may engage in counterplanning. [Carbonell, 1979]

Adopted goals are processed in a uniform manner with intrinsic goals. This process of goal adoption provides a principled model of social interaction. Agents may engage in cooperative or competitive behavior based on goals adopted through interpersonal relationships.

- *Decisions require justification.* Given that most choices will not be optimal, and are based on subjective goals and beliefs, decisions need to be justified or explained. The explanation indicates how the agent arrived at the decision, so that other agents may better understand and evaluate the choice. An explanation reflects the decision strategy.

When a judge renders a verdict, or a member of Congress votes on a bill, or a corporate executive closes a plant, the decision needs an explanation. There are many possible ways to arrive at a choice. The explanation tells us how the decision was made, or, at least, how the agent wishes us to interpret her decision.

VOTE

The VOTE decision making model is based on the explicit representation of goals, choices, relationships, and strategies, and the use of natural language to produce explanations. [Slade, 1994]

The VOTE program simulates the roll call voting of members of the United States House of Representatives. Given a member of Congress and a specific bill, VOTE tries to determine how that member would vote and then produces a natural language explanation of the resulting decision in English or French.

Below is an example of the VOTE program simulating Congressman Morris Udall voting on a bill banning flag burning.

```
> (vote 'udall 'hr-2978)

* Member:        Morris K. Udall
* Bill:          Flag Desecration
* Bill banning the desecration of the flag.
— OMITTING INTERMEDIATE OUTPUT —

* English rationale:
```

Morris K. Udall votes against bill HR-2978, the flag desecration bill. After weighing the implications, he believes that provisions of this bill are not constitutional. He completely supports the United States Constitution and the Bill of Rights. Udall readily endorses the right of freedom of speech. Even so, Udall realizes that members of the Democratic party oppose the right of burning the American flag in protest.

```
* French rationale:
```

Morris K. Udall s'oppose au projet de loi HR-2978, la loi de la profanation du drapeau. Après une considération approfondie, il croit que les dispositions de ce projet de loi ne sont pas constitutionelles. Il est un champion de la Constitution américaine et de la déclaration des Droits. Udall désire vivement appuyer le droit de libre expression. Cependant, Udall comprend que les membres du parti Démocratique s'opposent au fait de brûler le drapeau américain lors d'une manifestation.

The natural language explanation above is not canned text, but is generated automatically by VOTE. Similarly, the French text is not a translation of the English text,

but is generated from the underlying knowledge representations.

The VOTE program relies on a set of interrelated databases, including issues (over 200 currently in the database), constituency groups (150), bills (42), members (67), and decision strategies (16). We note that multiple decision strategies are required since the explanation of the decision depends on the strategy employed. It is not enough to use one simple strategy of summing the weights of the conflicting issues and relationships.

The purpose of VOTE is not to predict individual voting decisions, but rather to demonstrate the computational feasibility of a particular model of interpersonal relationships and decision making. Having said that, we observe empirically that VOTE's accuracy rate on thousands of predictions exceeds 75%.

VOTE embodies the realistic decision making assumptions stated above.

- *Many goals.* Each member in VOTE has dozens of goals, such as opposing gun control or abortion. Many of these goals may actually be in conflict. For example, a member may have reasons both to favor and to oppose the death penalty. In a logical model, P and $not P$ results in contradiction and mayhem. In a psychological model, conflict reflects normal cognitive dissonance.

- *Limited resources.* A member has only one vote to cast. Even without internal conflict, a member may face a conflict with a particular bill, which may call for a choice between balancing the budget and increased defense spending.

- *Subjective decisions.* On the same bill, different members vote differently. Members who oppose a bill may do so for different reasons and generate different explanations. The members share the same knowledge of issues, bills, and constituency groups. However, members have different goals, relationships, and voting records.

- *Allocate resources to achieve preferences.* The main resource here is power to vote for or against a bill. Given the frequent conflicts, the member must try to make decisions that will be consistent with her most important goals.

- *Adopt goals through relationships.* The member has relationships with various constituency groups, such as labor, business, women, minorities, or environmentalists. Each group has an issue agenda which is adopted by the member at a level of importance reflecting the strength of the relationship. In the example, Udall has a positive relationship with the Democratic party, which has the goal of banning flag burning. This adopted goal creates a conflict.

- *Explain decisions.* It is not enough to arrive at a decision. An agent must also justify her choice. VOTE uses explicit strategies in arriving at a decision. These strategies are higher level schemata that provide an organizing rationale for the decision. In the example, Udall's decision strategy is to oppose the bill as being unconstitutional.

We assert that VOTE's decisions are rational, even though they may not be optimal. Given the lack of complete knowledge in this domain, it would be irrational to assume the feasibility of achieving an optimal decision. What is rational is for a member to consider her preferences and those of her constituents, the consequences of the legislation, and the decision's explanation.

From the political science literature, Kingdon [Kingdon, 1973] notes that voting strategies often hinge on the role of explanation. Members of Congress report that for a given vote they either need to have a good explanation or avoid the vote that would require an explanation. Given that a member is elected by the voters of her district, her ideology and beliefs are likely to reflect those of her constituents. Thus, generally a member's votes will not require explanations. Furthermore, once she has established a voting record, she can avoid explanation by being consistent in her future votes. That is, if a member votes on bill X the way in which she has always voted on similar bills in the past, then she should not have to explain that vote.

Kingdon quotes a representative who opposed a measure providing for the direct election of the president, but nonetheless voted for it:

> 'Very frankly, if I had a chance to sit down with all my constituents for 15 minutes and talk to them, I'd have voted against the whole thing. But I don't have that chance. They wanted to change. If I voted against it, it would appear to them that I was against change, and I wouldn't have a chance to explain myself.' [Kingdon, 1973]

Kingdon notes that the importance or intensity of an issue can also affect the justification of a vote.

> The effect of this need to explain oneself is somewhat related to the weighing of intensities . . . If the congressman feels intensely about the matter, he will take the trouble to explain his position. If he does not feel so strongly, it is likely that he will avoid the situation in which he is obliged to explain, by voting with his constituents. Because there are many occasions on which a segment of his constituency has strong preferences and the congressman's preferences are not so strong, this tendency to avoid the uncomfortable confrontation probably contributes

a good deal to effective representation of such interests. [Kingdon, 1973]

In certain cases, a vote may seem irrational. In Congress, a member may cast a vote that appears to violate the preferences of the member and her constituents.

For example, black members of Congress will occasionally vote against civil rights legislation. This action appears bewildering in the absence of an explanation. The black members can claim that they were registering a protest vote, and wanted to encourage the passage of stronger legislation. Typically, the protest votes do not result in stopping passage of the bill. Thus, the members can have their cake and eat it too. The VOTE decision strategy NOT GOOD ENOUGH incorporates this explanation.

The need for explanations is a reflection of the fact that agents cannot make optimal decisions. There are many possible decision strategies. A rational decision maker provides an explanation to illuminate her decision.

Multi-agent Interaction

Decision making is usually viewed from the perspective of a single agent. The VOTE model suggests a principled way to enlarge decision making as a social process through the adoption of goals from interpersonal relationships.

There are other multi-agent phenomena that are explicitly social that may be examined from our model of decision making.

- *Advice.* An agent may provide objective information to another agent who is having trouble making a decision. In the VOTE model, useful, dispassionate advice may include additional preferences, choices, and consequences that were not previously considered. The information might also comprise new decision strategies or explanations. The Congressional Research Office can provide a member with statistical data or other background information related to a vote.

- *Persuasion.* Persuasion is like advice, except that persuader is not merely providing additional information or knowledge, but is actually arguing for a particular outcome from which she may benefit. Given that persuasion aims to achieve a goal of the persuader, it helps if there is a positive relationship between the decision maker and the persuader. A lobbyist may make a campaign contribution to establish a positive relationship with a member of Congress, and thereby make it more likely that the member will adopt the lobbyist's goals.

- *Negotiation.* Negotiation is a variation of persuasion. Here each side stands to benefit from the other's loss. Successful negotiation depends on finding some common ground between the parties. Two opposing members of Congress may hammer out a compromise on a given bill, or offer to trade votes on other bills.

In each case, advice, persuasion, and negotiation, the parties need to understand each other's preferences and beliefs. The basic VOTE model of decision making provides the foundation for these other social interactions.

Realistic Irrationality

In presenting our model of rationality, we have avoided the prescriptive view of good and bad decisions. If an agent has an internal set of goals, and makes decisions consistent with those preferences, we consider those choices to be rational. It would appear that almost any decision can be viewed as rational by this account.

This is a problem. How can you have rationality if there is no irrationality? Is there good without evil? Is there hot without cold? Can there be Democrats without Republicans?

We know that people make bad decisions. The author has met such people. We know that individual decision makers often seek out the advice of others when facing a difficult choice. It is possible to improve decisions. Our model can accomodate these data.

One cause of irrationality is due to the subjective frame of reference. In VOTE, a member of Congress may have a conflict in goals adopted from two constituency groups. The group on the losing side may view the decision as irrational from their frame of reference.

In addition to goals adopted through relationships, there are societal norms: standard sets of preferences and beliefs to which an agent may subscribe when making a decision. For example, economics provides a decision maker with an agenda derived from the laws of supply and demand, which are usually summarized with the dictum *maximize profits*. Similarly, most religions provide codes of ethical behavior such as the *Ten Commandments*, suggesting that agents refrain from theft and murder.

Is it irrational for an executive to steal from or even kill her competitor? Society has decided that religion wins out over economics.

In many cases, society has stipulated normative behavior. It is considered irrational if not illegal to violate these norms. There is an implicit rule that agents adhere to the norms of society. Subjective decision making has its limits. A decision which may be rational for a single agent becomes irrational in a social context.

An agent faced with a hard choice then has several reasons to get outside advice, based on the fundamental assumption of limited knowledge. There are a number of questions which may arise.

- What other options or choices may be available?

- What additional goals or preferences are relevant?

- What are the expected consequences for the available options?

- What explanations are appropriate for each option?

- What other agents may be affected by the decision?

A decision made in the absence of such information may be considered irrational. However, there is no axiomatic set of knowledge describing society's rational expectations.

We assume that an agent knows that by shooting a gun, you can kill someone. We do not necessarily assume that by investing in derivatives, you can lose a billion dollars.

Different groups and situations have different norms. *Common sense* is the normative common denominator. Different norms exist for specific areas, such as economics, law, medicine, sports, computer science, and artificial intelligence. What is rational for the lawyer may be irrational for the physician.

The irrationality of a decision depends on the societal norms. Earlier, we argued that knowledge, like money, was a resource, and that lacking knowledge did not necessarily make a decision irrational. We now qualify that statement by suggesting that society assumes a certain level of consensual knowledge that is a resource common to agents in similar circumstances.

Conclusion

We have discussed a particular model of decision making demonstrated by the VOTE program. According to this model, a rational decision maker should should perform the following actions.

- allocate scarce resources to achieve preferred goals.

- use emotions to communicate goal states.

- adopt goals through interpersonal relationships.

- justify decisions.

Knowledge about goal preferences and plan consequences fall outside this definition of rationality. Nevertheless, viewing knowledge as a resource, an agent should acquire such knowledge to increase her goals that can be achieved, just as she might want to acquire more money to achieve more goals. However, it is not irrational to lack knowledge, just as it is not irrational to lack money.

We suggest that our decision making model is realistic, and could be extended for multi-agent interactions such as advice, persuasion, and negotiation.

Software

The current version of VOTE, in Common LISP, is available as an Internet resource at the URL `ftp://is.stern.nyu.edu/pub/vote/`.

References

[Carbonell, 1979] J. Carbonell. *Subjective Understanding: Computer Models of Belief Systems*. PhD thesis, Yale University, 1979. Technical Report 150.

[Kingdon, 1973] J.W. Kingdon. *Congressmen's Voting Decisions*. Harper and Row, New York, 1973.

[Newell and Simon, 1972] A. Newell and H. Simon. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ, 1972.

[Newell *et al.*, 1958] A. Newell, J.C. Shaw, and H.A. Simon. Chess-playing programs and the problem of complexity. *IBM Journal of Research and Development*, 2:320–335, October 1958.

[Raiffa, 1968] H. Raiffa. *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Addison-Wesley, Reading, MA, 1968.

[Roseman, 1982] I. Roseman. *Cognitive Aspects of Discrete Emotions*. PhD thesis, Yale University, 1982.

[Simon, 1982a] H.A. Simon. *Models of Bounded Rationality (Volume 1): Economic Analysis and Public Policy*. MIT Press, Cambridge, MA, 1982.

[Simon, 1982b] H.A. Simon. *Models of Bounded Rationality (Volume 2): Behavioral Economics and Business Organization*, volume 2. MIT Press, Cambridge, MA, 1982.

[Slade, 1994] S. B. Slade. *Goal-based Decision Making: An Interpersonal Model*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1994.