

Rationality

*Decision Behavior:
Theory and Evidence*
Todd Davies

A Framework for Rationality

1. Actions. A set (one or more) of moves executed in C.
2. Constraints. A set of characteristics that apply in (i.e. that characterize) C, relating inputs to outputs.
3. Evidence. Any data available as inputs to the constraints that apply in C.
4. Objectives. The desired outcomes of actions executed in C.



Consider...

- In 1981, John Hinckley, Jr., shot then-President Ronald Reagan in the chest to win the love of the actress Jodie Foster. Hinckley believed that if he shot Reagan, Foster would fall in love with him. He turned out to be wrong about that.
- Can Hinckley's actions be described as rational? Why or why not?
- How might Hinckley's actions be mapped onto the framework for rationality?

Now consider...

- Agent A sees a dial-faced clock on a wall, without a seconds hand in early April. The clock reads 6:30. Outside, the sun is setting. A concludes that the current time is about 6:30 pm.
- Is A's conclusion rational?

But suppose that...

- A returns the following day at noon and finds that the clock still reads 6:30.
Does this change the rationality of A's inference from the previous day?
- A had no evidence from outside the room - i.e. could not look outside.
Does this change the rationality of the clock inference?

A possible mapping...

- Evidence: Clock reads 6:30, It is early April, [Sun setting outside]
- Constraints: “6:30” may refer either to AM or PM, Sun sets around 6:30 pm in early April, Clock time relates to real time only if the clock is working properly
- Objective: Infer the correct time
- Action: Conclude that the time is 6:30 PM.

Levels of Rationality for a Single Agent's Objectives (corrected)

1. Subjective rationality. The actions of an agent A are subjectively rational in a context C to the extent that A 's actions, given the evidence available to and the constraints as understood by A , are consistent with the achievement of A 's objectives.
2. Intersubjective rationality. The actions of an agent A are intersubjectively rational in a context C to the extent that A 's actions, given the evidence available to and the constraints as understood by the collection A of agents in C , are consistent with the achievement of A 's objectives.
3. Objective rationality. The actions of an agent A are objectively rational in a context C to the extent that A 's actions, given all of the evidence and constraints applying in C , are consistent with the achievement of A 's objectives.

Goals for Evaluating Rationality

1. Normative analysis. How well do possible actions meet objectives in a context C at a given level of rationality?
2. Descriptive analysis. How well do actually chosen actions meet objectives in a context at some level of rationality?

What do normative and descriptive analyses have to say about...

- The Hinckley example?
- The clock example when external indications of the time of day (i.e. visibility of the sun) are available?
- The clock example when external indications of time of day are not available?

Are people generally rational?

- What does this question mean?
- How would one go about answering it?
- Does the answer matter? Why or why not?

Are people rational? Some possible responses...

Yes, usually

Yes, by definition

Yes, within the limits of time, information, etc. that they usually have

Yes, for the problems they evolved to cope with

Yes, in the aggregate

No, and we can do something about it

No, and there isn't much we can do to improve