

14. The simulation heuristic

Daniel Kahneman and Amos Tversky

Our original treatment of the availability heuristic (Tversky & Kahneman, 1973, 11) discussed two classes of mental operations that “bring things to mind”: the retrieval of instances and the construction of examples or scenarios. *Recall* and *construction* are quite different ways of bringing things to mind; they are used to answer different questions, and they follow different rules. Past research has dealt mainly with the retrieval of instances from memory, and the process of mental construction has been relatively neglected.

To advance the study of availability for construction, we now sketch a mental operation that we label the simulation heuristic. Our starting point is a common introspection: There appear to be many situations in which questions about events are answered by an operation that resembles the running of a simulation model. The simulation can be constrained and controlled in several ways: The starting conditions for a “run” can be left at their realistic default values or modified to assume some special contingency; the outcomes can be left unspecified, or else a target state may be set, with the task of finding a path to that state from the initial conditions. A simulation does not necessarily produce a single story, which starts at the beginning and ends with a definite outcome. Rather, we construe the output of simulation as an assessment of the ease with which the model could produce different outcomes, given its initial conditions and operating parameters. Thus, we suggest that mental simulation yields a measure of the propensity of one’s model of the situation to generate various outcomes, much as the propensities of a statistical model can be assessed by Monte Carlo techniques. The ease with which the

This chapter is drawn from the Katz-Newcomb Lecture in Social Psychology “On the Psychology of Possible Worlds,” Ann Arbor, Michigan, April, 1979. The work was supported by the Office of Naval Research under Contract N00014-79-C-0077 to Stanford University.

simulation of a system reaches a particular state is eventually used to judge the propensity of the (real) system to produce that state.

We shall argue that assessments of propensity and probability derived from mental simulations are used in several tasks of judgment and also that they play a significant role in several affective states. We first list some judgmental activities in which mental simulation appears to be involved. We then describe a study of the cognitive rules that govern the mental undoing of past events, and we briefly discuss the implications of these rules for emotions that arise when reality is compared with a favored alternative, which one had failed to reach but could easily imagine reaching. We conclude this brief sketch of the simulation heuristic by some remarks on scenarios, and on the biases that are likely to arise when this heuristic is used.

1. *Prediction.* Imagine the first meeting between two persons that you know well, who have never met before. How do you generate predictions such as "They will get on famously" or "They will grate on one another"?

2. *Assessing the probability of a specified event.* How do you assess the likelihood of American armed intervention to secure the oilfields of Saudi Arabia in the next decade? Note the difference between this task and the preceding one. The simulation in the present case has a specified target-state, and its object is to obtain some measure of the "ease" with which this target state can be produced, within the constraints of a realistic model of the international system.

3. *Assessing conditioned probabilities.* If civil war breaks out in Saudi Arabia, what are the likely consequences? Note that this simulation exercise differs from mere prediction, because it involves a specified initial state, which may diverge more or less from current reality. The assessment of remote contingencies, in particular, involves an interesting ambiguity: What changes should be made in one's current model *before* the "run" of the simulation? Should one make only the minimal changes that incorporate the specified contingency (e.g., civil war in Saudi Arabia), subject to elementary requirements of consistency? Or should one introduce all the changes that are made probable by the stipulation of the condition? In that case, for example, one's model of the political system would first be adjusted to make the civil war in Saudi Arabia as unsurprising as possible, and the simulation would employ the parameters of the revised model.

4. *Counterfactual assessments.* How close did Hitler's scientists come to developing the atom bomb in World War II? If they had developed it in February 1945, would the outcome of the war have been different? Counterfactual assessments are also used in many mundane settings, as when we judge that "she could have coped with the job situation if her child had not been ill."

5. *Assessments of causality.* To test whether event *A* caused event *B*, we may undo *A* in our mind, and observe whether *B* still occurs in the simulation. Simulation can also be used to test whether *A* markedly

increased the propensity of *B*, perhaps even made *B* inevitable. We suggest that a test of causality by simulation is involved in examples such as “You know very well that they would have quarreled even if she had not mentioned his mother.”

Studies of undoing

Our initial investigations of the simulation heuristic have focused on counterfactual judgments. In particular, we have been concerned with the process by which people judge that an event “was close to happening” or “nearly occurred.” The spatial metaphor is compelling and has been adopted in many philosophical investigations: It appears reasonable to speak of the distance between reality and some once-possible but unrealized world. The psychological significance of this assessment of distance between what happened and what could have happened is illustrated in the following example:

Mr. Crane and Mr. Tees were scheduled to leave the airport on different flights, at the same time. They traveled from town in the same limousine, were caught in a traffic jam, and arrived at the airport 30 minutes after the scheduled departure time of their flights.

Mr. Crane is told that his flight left on time.

Mr. Tees is told that his flight was delayed, and just left five minutes ago.

Who is more upset?

Mr. Crane Mr. Tees

It will come as no surprise that 96% of a sample of students who answered this question stated that Mr. Tees would be more upset. What is it that makes the stereotype so obvious? Note that the objective situation of the two gentlemen is precisely identical, as both have missed their planes. Furthermore, since both had expected to miss their planes, the difference between them cannot be attributed to disappointment. In every sense of the word, the difference between Tees and Crane is immaterial. The only reason for Mr. Tees to be more upset is that it was more “possible” for him to reach his flight. We suggest that the standard emotional script for this situation calls for both travelers to engage in a simulation exercise, in which they test how close they came to reaching their flight in time. The counterfactual construction functions as would an expectation. Although the story makes it clear that the expectations of Mr. Tees and Mr. Crane could not be different, Mr. Tees is now more disappointed because it is easier for him to imagine how he could have arrived 5 minutes earlier than it is for Mr. Crane to imagine how the 30 minutes delay could have been avoided.

There is an Alice-in-Wonderland quality to such examples, with their odd mixture of fantasy and reality. If Mr. Crane is capable of imagining unicorns – and we expect he is – why does he find it relatively difficult to

imagine himself avoiding a 30-minute delay, as we suggest he does? Evidently, there are constraints on the freedom of fantasy, and the psychological analysis of mental simulation consists primarily of an investigation of these constraints.

Our understanding of the rules of mental simulations is still rudimentary and we can present only early results and tentative speculations in a domain that appears exceptionally rich and promising. We have obtained preliminary observations on the rules that govern a special class of simulation activity – undoing the past. Our studies of undoing have focused on a situation in which this activity is especially common – the response of surviving relatives to a fatal accident. Here again, as in the case of Mr. Tees and Mr. Crane, we chose to study what we call the emotional script for a situation. For an example, consider the following story:

Mr. Jones was 47 years old, the father of three and a successful banking executive. His wife has been ill at home for several months.

On the day of the accident, Mr. Jones left his office at the regular time. He sometimes left early to take care of home chores at his wife's request, but this was not necessary on that day. Mr. Jones did not drive home by his regular route. The day was exceptionally clear and Mr. Jones told his friends at the office that he would drive along the shore to enjoy the view.

The accident occurred at a major intersection. The light turned amber as Mr. Jones approached. Witnesses noted that he braked hard to stop at the crossing, although he could easily have gone through. His family recognized this as a common occurrence in Mr. Jones' driving. As he began to cross after the light changed, a light truck charged into the intersection at top speed, and rammed Mr. Jones' car from the left. Mr. Jones was killed instantly.

It was later ascertained that the truck was driven by a teenage boy, who was under the influence of drugs.

As commonly happens in such situations, the Jones family and their friends often thought and often said, "If only . . .", during the days that followed the accident. How did they continue this thought? Please write one or more likely completions.

This version (labeled the "route" version) was given to 62 students at the University of British Columbia. Another group of 61 students received a "time" version, in which the second paragraph read as follows:

On the day of the accident, Mr. Jones left the office earlier than usual, to attend to some household chores at his wife's request. He drove home along his regular route. Mr. Jones occasionally chose to drive along the shore, to enjoy the view on exceptionally clear days, but that day was just average.

The analysis of the first completion of the "If only" stem is given in Table 1. Four categories of response were found: (i) Undoing of route; (ii) Undoing of time of departure from the office; (iii) Mr. Jones crossing at the amber light; (iv) Removing the drugged boy from the scene.

A particularly impressive aspect of the results shown in Table 1 is an

Table 1. Analysis of first completion of the "if only" stem

Response categories	Time version	Route version
(i) Route	8	33
(ii) Time	16	2
(iii) Crossing	19	14
(iv) Boy	18	13
(v) Other	1	3

event that fails to occur: Not a single subject mentioned that if Mr. Jones had come to the intersection two or three seconds earlier he would have gone through safely. The finding is typical: Events are not mentally undone by arbitrary alterations in the values of continuous variables. Evidently, subjects do not perform the undoing task by eliminating that necessary condition of the critical event that has the lowest prior probability – a procedure that would surely lead them to focus on the extraordinary coincidence of the two cars meeting at the intersection. Whatever it is that people do, then, is not perfectly correlated with prior probability.

The alterations that people introduce in stories can be classified as downhill, uphill, or horizontal changes. A downhill change is one that removes a surprising or unexpected aspect of the story, or otherwise increases its internal coherence. An uphill change is one that introduces unlikely occurrences. A horizontal change is one in which an arbitrary value of a variable is replaced by another arbitrary value, which is neither more nor less likely than the first. The experimental manipulation caused a change of route to be downhill in one version, uphill in the other, with a corresponding variation in the character of changes of the timing of Mr. Jones's fatal trip. The manipulation was clearly successful: Subjects were more likely to undo the accident by restoring a normal value of a variable than by introducing an exception. In general, uphill changes are relatively rare in the subjects' responses, and horizontal changes are non-existent.

The notion of downhill and uphill changes is borrowed from the experience of the cross-country skier, and it is intended to illustrate the special nature of the distance relation that can be defined for possible states of a system. The essential property of that relation is that it is not symmetric. For the cross-country skier, a brief downhill run from *A* to *B* is often paired with a long and laborious climb from *B* to *A*. In this metaphor, exceptional states or events are peaks, normal states or events are valleys. Thus, we propose that the psychological distance from an exception to the norm that it violates is smaller than the distance from the norm to the same exception. The preference for downhill changes is perhaps the major rule that mental simulations obey; it embodies the essential constraints that lend realism to counterfactual fantasies.

A notable aspect of the results shown in Table 1 is the relatively low

proportion of responses in which the accident is undone by eliminating the event that is naturally viewed as its cause: The insane behavior of the drugged boy at the intersection. This finding illustrates another property of mental simulation, which we label the focus rule: Stories are commonly altered by changing some property of the main object of concern and attention. In the present case, of course, the focus of attention was Mr. Jones, since the subjects had been instructed to empathize with his family. To test the focus rule, a new version of the accident story was constructed, in which the last paragraph was replaced by the following information:

It was later ascertained that the truck was driven by a teenage boy, named Tom Searler. Tom's father had just found him at home under the influence of drugs. This was a common occurrence, as Tom used drugs heavily. There had been a quarrel, during which Tom grabbed the keys that were lying on the living room table and drove off blindly. He was severely injured in the accident.

Subjects given this version of the story were asked to complete the stem "If only . . .," either on behalf of Mr. Jones's relatives or on behalf of Tom's relatives. Here again, we consider the first response made by the subjects. The majority of subjects who took the role of Tom's relatives (68%) modified the story by removing him from the scene of the accident – most often by not allowing the fatal keys on the table. In contrast, only a minority (28%) of the subjects identifying with Mr. Jones's relatives mentioned Tom in their responses.

We have described this study of undoing in some detail, in spite of its preliminary character, to illustrate the surprising tidiness of the rules that govern mental simulation and to demonstrate the existence of widely shared norms concerning the counterfactual fantasies that are appropriate in certain situations. We believe that the cognitive rules that govern the ease of mental undoing will be helpful in the study of a cluster of emotions that could be called counterfactual emotions, because of their dependence on a comparison of reality with what might or should have been: Frustration, regret, and some cases of indignation, grief, and envy are all examples. The common feature of these aversive emotional states is that one's hedonic adaptation level is higher than one's current reality, as if the unrealized possibilities were weighted into the adaptation level, by weights that correspond to the ease with which these possibilities are reached in mental simulation.

Remarks on scenarios

In the context of prediction and planning under uncertainty, the deliberate manipulation of mental models appears to be sufficiently important to deserve the label of a distinctive simulation heuristic. The clearest example of such activities is the explicit construction of scenarios as a procedure for the estimation of probabilities.

What makes a good scenario? In the terms already introduced, a good scenario is one that bridges the gap between the initial state and the target event by a series of intermediate events, with a general downhill trend and no significant uphill move along the way. Informal observations suggest that the plausibility of a scenario depends much more on the plausibility of its weakest link than on the number of links. A scenario is especially satisfying when the path that leads from the initial to the terminal state is not immediately apparent, so that the introduction of intermediate stages actually raises the subjective probability of the target event.

Any scenario is necessarily schematic and incomplete. It is therefore of interest to discover the rules that govern the selection of the events that are explicitly specified in the scenario. We hypothesize that the "joints" of a scenario are events that are low in redundancy and high in causal significance. A non-redundant event represents a local minimum in the predictability of the sequence, a point at which significant alternatives might arise. A causally significant event is one whose occurrence alters the values that are considered normal for other events in the chain that eventually leads to the target of the scenario.

The elaboration of a single plausible scenario that leads from realistic initial conditions to a specified end state is often used to support the judgment that the probability of the end state is high. On the other hand, we tend to conclude that an outcome is improbable if it can be reached only by invoking uphill assumptions of rare events and strange coincidences. Thus, an assessment of the "goodness" of scenarios can serve as a heuristic to judge the probability of events. In the context of planning, in particular, scenarios are often used to assess the probability that the plan will succeed and to evaluate the risk of various causes of failure.

We have suggested that the construction of scenarios is used as a heuristic to assess the probability of events by a mediating assessment of the propensity of some causal system to produce these events. Like any other heuristic, the simulation heuristic should be subject to characteristic errors and biases. Research is lacking in this area, but the following hypotheses appear promising: (i) The search for non-redundant and causally significant "joints" in scenario construction is expected to lead to a bias for scenarios (and end-states) in which dramatic events mark causal transitions. There will be a corresponding tendency to underestimate the likelihood of events that are produced by slow and incremental changes. (ii) The use of scenarios to assess probability is associated with a bias in favor of events for which one plausible scenario can be found, with a corresponding bias against events that can be produced in a multitude of unlikely ways. Such a bias could have especially pernicious consequences in a planning context, because it produces overly optimistic estimates of the probability that the plan will succeed. By its very nature, a plan consists of a chain of plausible links. At any point in the chain, it is

sensible to expect that events will unfold as planned. However, the cumulative probability of at least one fatal failure could be overwhelmingly high even when the probability of each individual cause of failure is negligible. Plans fail because of surprises, occasions on which the unexpected uphill change occurs. The simulation heuristic, which is biased in favor of downhill changes, is therefore associated with a risk of large and systematic errors.