Review of *The Undoing Project: A Friendship that Changed Our Minds,* by Michael Lewis

Gary Marcus, Uber AI Labs and Dept. of Psychology, New York University
Ernest Davis, Dept. of Computer Science, New York University,

The hardest decisions in life are often the ones in which the facts themselves are uncertain; cognitive psychologists call this reasoning and decision-making under uncertainty. The two psychologists who have told us the most about how humans reason under uncertainty are, without a doubt Amos Tversky (1937-1996) and Daniel Kahneman (b, 1934). Kahneman won the Nobel Prize, in 2002, for the influence his ideas have had on economics, and Tversky would surely have joined in that prize had he lived until it was awarded. The two are also the subject of a new book, *The Undoing Project* by the extraordinary journalist Michael Lewis, best known for his books *Liar's Poker* and *The Big Short*, on the financial world, and *Moneyball.* on how an increased understanding of statistics influenced professional baseball, The latter was turned into a movie with Brad Pitt, and, as Lewis notes in an introduction, it also led to the current book.

The core of the book, discussed below, is the science that Tversky and Kahneman created together (Disclosure: Kahneman is a recent friend of Gary Marcus, the first author of this review.) Threaded around that often complex and technical science (which we shall discuss in a moment) is a crystal clear and entertaining account of the two men, and, almost as a separate story, the tale of their remarkable collaboration, peppered with shorter sketches of some of the many notable figures in science, medicine and public policy whose lives and works were deeply impacted by Kahneman and Tversky's discoveries: the psychologist Amnon Rappoport, the medical scientist Don Redelmeier, the economist Richard Thaler, the legal scholar Cass Sunstein, and others (Kahneman's recent best-selling book, *Thinking Fast and Slow* has a much more extensive account of his own scientific work, both with Tversky and since Tversky's death.)

Both men, Tversky and Kahneman, would be impressive figures even if they hadn't made major contributions to science. As Lewis notes, both men served ably and bravely in Israeli Army, in the wars in 1956, 1967, and 1973. Tversky won an award for bravery for, against orders, running out to rescue a fellow-soldier who had fainted while carrying a live torpedo during a training exercise. Kahneman, having survived the Holocaust in hiding in France, served in the psychology department of the Israel Defense Force and used his expertise to achieve many improvements, large and small, in the effectiveness of the military and the well-being of the soldiers, designing a screening procedure for soldiers that is still in use today.

Kahneman and Tversky started working together in 1969 and rapidly developed an extraordinarily intense and productive working relation. They worked together for hours every day, in non-stop talk punctuated with frequent laughter. They would sit together at a single typewriter to write their papers. Neither man remembered or cared which of them

had come up with which idea; the ideas emerged from their interaction. They were equal co-authors on all their papers and alternated the order of their names: their first paper (by coin flip) was by "Tversky and Kahneman", their second by "Kahneman and Tversky" and so it continued. (As it happens, we have followed their pattern in our writings, including in this review.)

The teamwork was perfect for a decade; then things started to go sour. Though all their work was published on equal terms, Tversky was showered with accolades — a position at Stanford, a MacArthur fellowship, honors and invitations from all over — while Kahneman was, in this period, comparatively passed over and ignored. The authors of this review can attest that no one meets or hears Danny Kahneman without being impressed; but by all accounts Amos Tversky (whom neither of us ever met) was simply amazing in person, a whirlwind of intellectual energy. Tversky's greater success was, inevitably, galling to Kahneman; Tversky himself was seriously angered when the MacArthur was awarded to him with no mention of Kahneman. But the reaction of the outside world in itself would have been bearable; what was not bearable was Kahneman's impression that Tversky himself increasingly thought of Kahneman as the junior partner. Collaborative work, though still absolutely engaging, was now terribly painful and fraught. In 1996, Kahneman broke off the friendship. Three days later, Tversky telephoned him with the news that he had been diagnosed with terminal cancer and had at most six months to live Lewis' account of this remarkable partnership is fascinating, sympathetic, sensitive, humorous, and moving.

The work of Tversky and Kahneman is best known for a series of brilliantly designed, breathtakingly straightforward experiments that demonstrate that, faced with very simple decisions or problems involving chance, people make decisions and give answers that make no logical sense.

- Experimental subjects are asked to estimate the number of seven-letter words of the form `_ _ _ _ ing' that would appear on average in 2000 words of a novel. Their average estimate is 13.4 words. The same subjects are later asked to estimate the number of seven letter words of the form `_ _ _ _ _ n _' that would appear in 2000 words of a novel. The average estimate is 4.7 words. Of course, any word that fits the first form necessarily fits the second form, so there cannot be fewer words of the second form.

- Subjects read the following account (this was in 1980, when the tennis player Björn Borg was at the height of his career.)."Suppose that Björn Borg reaches the Wimbledon final. Please rank order the following outcomes from most likely to least likely::

    A. Borg will win the match.
    B. Borg will lose the first set.
    C. Borg will lose the first set but win the match
    D. Borg will win the first set but lose the match

72% of subjects answered that C was more likely than B. But it is completely impossible that C is more likely than B; in any circumstance where C is true, B is also true.

- Subjects were told that a person named Dick had been chosen randomly from a pool of 100 persons, of whom 70 were lawyers and 30 were engineers. The subjects then, correctly, predicted that there was a 70% probability that Dick was a lawyer. But then the subjects were given the following description of Dick:

  > Dick is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

  This bland information obviously sheds no light at all on the question of whether Dick is a lawyer or an engineer, and therefore does not affect the probabilities: There is still a 70% chance that Dick is a lawyer, since he was chosen from a pool of 70 lawyers and 30 engineers. However, once subjects were given the description, they almost all answered that there was now a 50% chance that he was a lawyer and a 50% chance that he was an engineer. They fixated on the useless specifics of the individual description, and ignored the underlying frequencies.

- When people are asked to choose between alternatives, their choice depends on the way the outcomes are presented, not just on the content of the outcomes. An example is the "Asian disease" problems. One group of subjects is presented with the following story:

  > *Story 1*: Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative program to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs is as follows:
  >
  > If Program A is adopted, 200 people will be saved.
  >
  > If Program B is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved.
  >
  > Which program would you favor?

The overwhelming majority of subjects prefer Program A.

The second group got a story with the same set up, but with the outcomes described as follows:

> *Story 2:*

> If Program A is adopted, 400 people will die.
>
> If Program B is adopted, there is a 1/3 probability that no one will die and a 2/3 probability that all 600 people will be saved.
>
> Which program would you favor?

In this case, the overwhelming majority of subjects prefer Program B.

The point is that the two stories are *identical:* the consequence of program A is that 400 will die and 200 will live; the consequence of B is that there is a 2/3 chance that all 600 will die and a 1/3 chance that all 600 will live. The two stories differ only in that story 1 phrases the outcome in terms of the number saved and story 2 phrases the outcome in terms of the number who die. And yet readers of two stories react in diametrically opposite ways.

Kahneman and Tversky also came up with a number of experiments of the same flavor which, if not exactly logical errors, were at least very strange when viewed from a rational standpoint. One of their studies dealt with the degree of regret that people feel. Subjects were told the following:

> You have participated in a lottery and have bought a single expensive ticket in the hope of winning the single large prize that is offered. The ticket was drawn blindly from a large urn, and its number is 107358. The results of the lottery are now announced, and it turns out that the winning number is 618379.

Subjects are asked to rate their unhappiness on a scale of 1 to 10; they answer that they would be moderately unhappy. However, if you change the winning number in the story to be 207358 and show the new story to subjects, they report that they would be very unhappy. This is not exactly a logical error, since there is no logical theory that prescribes how unhappy you should be under various circumstances. But what sense does it make, since the subjects are in fact equally badly off in both cases, and the numbers are completely random?

Tversky and Kahneman identified these errors in laboratory experiments with hypothetical experiments, but the same mistakes are made in real situations with important consequences. Doctors seeing a symptom will conjecture the cause that is most characteristic of the symptom, without taking into account the inherent likelihood of the cause, making the same mistake as in the experiment with Dick the lawyer. Students registering for a conference are more likely to register early if it is stated that there is a penalty for late registration than if it is stated that there is a discount for early registration, making the same mistake as in the experiment with the Asian disease.

To explain these and similar results, Kahneman and Tversky theorized that people are solving these kinds of problems using rough-and-ready techniques rather than careful

thinking. The *availability heuristic* leads people to estimate the likelihood of events in terms of the ease with which they can of them, as in the seven letter words ending with "ing". The *representativeness heuristic* leads people to judge likelihood in terms of the degree to which they were similar to stereotypes; a narrative of Borg losing the first set but winning the match corresponds to the stereotype of a Borg match; the bald statement that Borg lost the first set does not.

From the first, the work of Tversky and Kahneman met with strong reactions, both positive and negative. To many scientists in psychology, economics, medicine, management science, and other fields, the work felt like a revelation; they could now understand why their idealized models of human reason were unreal and why people make the mistakes they do; and, understanding that, they could address themselves to the questions of what can be done to improve human judgment. Entire fields of inquiry, such as behavioral economics, owe their start to the work of Kahneman and Tversky. , Medicine, too, has been greatly influenced.

But other scientists found the work repellent and rejected it, sometimes furiously, almost as if they felt it as an affront to human dignity. (Lewis quotes the psychologist Eldar Shafir: "Give people a visual illusion and they say, 'It's only my eyes.' Give them a linguistic illusion. They're fooled but they say, 'No big deal.' Then you give them one of Amos and Danny's examples and they say, 'Now you're insulting me.'") To some, questioning human reason in this way seemed not only incorrect, but even immoral. The assumption that people were more or less rational was central to established views of psychology, decision theory, and economics; to throw these out because subjects could be tricked by a few wise-guy experiments seemed absurd and retrograde.

Consequently, there grew up a small cottage industry in refuting Tversky and Kahneman.. First, it was found that under some circumstances, the effects they described could be nullified or reduced; for instance, if you give people the Asian disease problem in a language they learned as adults rather than their native language, then they are much more consistent in their answers across the two versions. Results of this kind are certainly valuable, but in our eyes they suggest only that the significance of the original experiments needs to be sharply demarcated, not that the experiments should be rejected.

Second, some critics argued that, viewed in the proper context, the answers that the subjects gave are actually correct. For instance, in the Asian flu story, the subjects were interpreting "200 will live" as "at least 200 will live" and "400 will die" as "at least 400 will die". In the Borg experiment the subjects interpreted the option "Borg lost the first set," as meaning something like "Borg will play poorly." However this approach, which strikes us as a form of post hoc rationalization, is problematic. First there is generally no way to determine how subjects are interpreting the sentence without changing the experiment. Second, if you always want the subjects' answers to be right, then you sometimes have to resort to quite far-fetched explanations of what they are thinking. Finally, this line of argument runs opposite to the first category of objection; if you think that subjects are right in answering differently for story 2 than story 1 when they read the

story in English, then why congratulate them on answering consistently when they read the story in French?

More generally, the impulse to preserve the theory of human rationality by reinterpreting seemingly incorrect answers as correct ones leads to a research methodology in which any possible experimental outcome supports the claim that humans are rational. The critics were able to come up with an explanation of why it might be reasonable for subjects to answer that Borg was more likely to lose the first set and win the match than to lose the first set.. But suppose the experiment had come out the other way; suppose that the subjects had all agreed that Borg was more likely to lose the first set than to lose the first set and win the match. Would these same critics then have been scratching their heads at the irrationality of the answer and the non-natural interpretation of the question? Not at all; everyone would be basking in the reassurance that subjects got this answer right. At that point, the claim that people are rational ceases to be a theory to be studied by empirical experimentation, and becomes a matter of dogma rather than science.

We would have thought this was obvious, but for all of the great work that Kahneman and Tversky did, the rationalist fallacy that they by rights should have buried still lingers. In recent decades, a large and influential school of cognitive psychology has arisen that takes as its central tenet that human cognitive processes are rational, or even optimal; or at least as good as they could be, given certain fundamental limitations on cognitive power. Often this claim, in our eyes a non-starter, yet remarkably popular, is tied into a Panglossian view of evolution: Humans have evolved with perfect cognitive powers, because cognitive flaws are clearly maladaptive. The hope seems to be that evolution would have fully weeded out all flaws over time. (As one of us (Marcus) pointed out in a book *Kluge*, we should expect no such thing: evolution often alights on "local maxima", solutions that are better than other nearby possibilities, without being the best possible solution). The neo-Rationalists theorists often make a fetish of Bayes' law, a basic probabilistic law that determines the likelihood of a hypothesis after relevant observations have been made; as such they are therefore often known as Bayesians.

We find this puzzling, a step backwards in our understanding of of psychology. To take one example, subjects' answers in the experiment with Dick the lawyer, described above, are direct violations of Bayes' law. The Bayesians' only hope is to explain away or ignore results like those of Tversky and Kahneman, and a myriad other results demonstrating the limits of the human mind and the proneness of the human mind to bias, confusion, and all manner of error; and they must dismiss as an illusion the conventional wisdom, so obvious as to be banal, that people do lots of unnecessary foolish things and make lots of unnecessary mistakes.

Anyone who has studied cognitive science, and especially anyone who has compared the power of the human mind to the limitations of artificial intelligence technology, must stand in awe of the capacities of human cognition, even in its most ordinary manifestations: a human going about his daily routine, an infant exploring its new world. We currently understand scientifically only a tiny fraction of the workings of the mind and how the brain effectuates them. Without question, the major task facing cognitive

science is to explain how the mind works as well as it does; explaining why it does not work perfectly is very much secondary. But there is nothing to be gained by pretending that it does work perfectly and ignoring all the contrary evidence. On the contrary: if we want to study what the mind is actually doing, it is critical to rid ourselves of the preconception that the mind conforms to an idealized mathematical standard. Therein lies the enormous importance of Kahneman and Tversky's work.