How Judgments Happen

There is no limit to the number of questions you can answer, whether they are questions someone else asks or questions you ask yourself. Nor is there a limit to the number of attributes you can evaluate. You are capable of counting the number of capital letters on this page, comparing the height of the windows of your house to the one across the street, and assessing the political prospects of your senator on a scale from excellent to disastrous. The questions are addressed to System 2, which will direct attention and search memory to find the answers. System 2 receives questions or generates them: in either case it directs attention and searches memory to find the answers. System 1 operates differently. It continuously monitors what is going on outside and inside the mind, and continuously generates assessments of various aspects of the situation without specific intention and with little or no effort. These *basic assessments* play an important role in intuitive judgment, because they are easily substituted for more difficult questions—this is the essential idea of the heuristics and biases approach. Two other features of System 1 also support the substitution of one judgment for another. One is the ability to translate values across dimensions, which you do in answering a question that most people find easy: "If Sam were as tall as he is intelligent, how tall would he be?" Finally, there is the mental shotgun. An intention of System 2 to answer a specific question or evaluate a particular attribute of the situation automatically triggers other computations, including basic assessments.

Basic Assessments

System 1 has been shaped by evolution to provide a continuous assessment of the main problems that an organism must solve to survive: How are things going? Is there a threat or a major opportunity? Is everything normal? Should I approach or avoid? The questions are perhaps less urgent for a human in a city environment than for a gazelle on the savannah, aalenc and e: How , but we have inherited the neural mechanisms that evolved to provide ongoing assessments of threat level, and they have not been turned off. Situations are constantly evaluated as good or bad, requiring escape or permitting approach. Good mood and cognitive ease are the human equivalents of assessments of safety and familiarity.

For a specific example of a basic assessment, consider the ability to discriminate friend from foe at a glance. This contributes to one's chances of survival in a dangerous world, and such a specialized capability has indeed evolved. Alex Todorov, my colleague at Princeton, has explored the biological roots of the rapid judgments of how safe it is to interact with a stranger. He showed that we are endowed with an ability to evaluate, in a single glance at a stranger's face, two potentially crucial facts about that person: how dominant (and therefore potentially threatening) he is, and how trustworthy he is, whether his intentions are more likely to be friendly or hostile. The shape of the face provides the cues for assessing dominance: a "strong" square chin is one such cue. Facial expression (smile or frown) provides the cues for assessing the stranger's intentions. The combination of a square chin with a turned-down mouth may spell trouble. The accuracy of face reading is far from perfect: round chins are not a reliable indicator of meekness, and smiles can (to some extent) be faked. Still, even an imperfect ability to assess strangers confers a survival advantage.

This ancient mechanism is put to a novel use in the modern world: it has some influence on how people vote. Todorov showed his students pictures of men's faces, sometimes for as little as one-tenth of a second, and asked them to rate the faces on various attributes, including likability and competence. Observers agreed quite well on those ratings. The faces that Todorov showed were not a random set: they were the campaign portraits of politicians competing for elective office. Todorov then compared the results of the electoral races to the ratings of competence that Princeton students had made, based on brief exposure to photographs and without any political context. In about 70% of the races for senator, congressman, and governor, the election winner was the candidate whose face had earned a higher rating of competence. This striking result was quickly confirmed in national elections in Finland, in zoning board elections in England, and in various electoral contests in Australia, Germany, and Mexico. Surprisingly (at least to me), ratings of competence were far more predictive of voting outcomes in Todorov's study than ratings of likability.

Todorov has found that people judge competence by combining the two dimensions of strength and trustworthiness. The faces that exude competence combine a strong chin with a slight confident-appearing smile. There is no evidence that these facial features actually predict how well politicians will perform in office. But studies of the brain's response to winning and losing candidates show that we are biologically predisposed to reject candidates who lack the attributes we value—in this research, losers evoked stronger indications of (negative) emotional response. This is an example of what I will call a *judgment heuristic* in the following chapters. Voters are attempting to form an impression of how good a candidate will be in office, and they fall back on a simpler assessment that is made quickly and automatically and is available when System 2 must make its decision.

Political scientists followed up on Todorov's initial research by identifying a category of voters for whom the automatic preferences of System 1 are particularly likely to play a large role. They found what they were looking for among politicalr m="5%">Todoly uninformed voters who watch a great deal of television. As expected, the effect of facial competence on voting is about three times larger for information-poor and TV-prone voters than for others who are better informed and watch less television. Evidently, the relative importance of System 1 in determining voting choices is not the same for all people. We will encounter other examples of such individual differences.

System 1 understands language, of course, and understanding depends on the basic assessments that are routinely carried out as part of the perception of events and the comprehension of messages. These assessments include computations of similarity and representativeness, attributions of causality, and evaluations of the availability of associations and exemplars. They are performed even in the absence of a specific task set, although the results are used to meet task demands as they arise.

The list of basic assessments is long, but not every possible attribute is assessed. For an example, look briefly at figure 7.

A glance provides an immediate impression of many features of the display. You know that the two towers are equally tall and that they are more similar to each other than the tower on the left is to the array of blocks in the middle. However, you do not immediately know that the number of blocks in the left-hand tower is the same as the number of blocks arrayed on the floor, and you have no impression of the height of the tower that you could build from them. To confirm that the numbers are the same, you would need to count the two sets of blocks and compare the results, an activity that only System 2 can carry out.

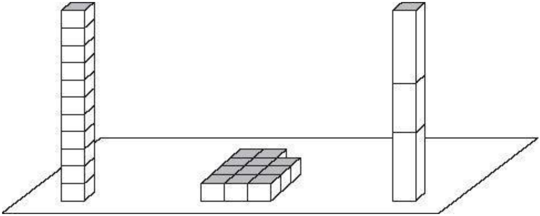

Figure 7

Sets and Prototypes

For another example, consider the question: What is the average length of the lines in figure 8?

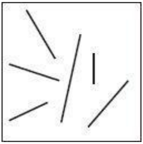

Figure 8

This question is easy and System 1 answers it without prompting. Experiments have shown that a fraction of a second is sufficient for people to register the average length of an array of lines with considerable precision. Furthermore, the accuracy of these judgments is not impaired when the observer is cognitively busy with a memory task. They do not necessarily know how to describe the average in inches or centimeters, but they will be very accurate in adjusting the length of another line to match the average. System 2 is not needed to form an impression of the norm of length for an array. System 1 does it, automatically and effortlessly, just as it registers the color of the lines and the fact that they are not parallel. We also can form an immediate impression of the number of objects in an array—precisely if there are four or fewer objects, crudely if there are more.

Now to another question: What is the total length of the lines in figure 8? This is a different experience, because System 1 has no suggestions to offer. The only way you can answer this question is by activating System 2, which will laboriously estimate the average, estimate or count the lines, and multiply average length by the number of lines.
estimaight="0%">

The failure of System 1 to compute the total length of a set of lines at a glance may look obvious to you; you never thought you could do it. It is in fact an instance of an important limitation of that system. Because System 1 represents categories by a prototype or a set of typical exemplars, it deals well with averages but poorly with sums. The size of the category, the number of instances it contains, tends to be ignored in judgments of what I will call *sum-like variables*.

Participants in one of the numerous experiments that were prompted by the litigation following the disastrous *Exxon Valdez* oil spill were asked their willingness to pay for nets to cover oil ponds in which migratory birds often drown. Different groups of participants stated their willingness to pay to save 2,000, 20,000, or 200,000 birds. If saving birds is an economic good it should be a sum-like variable: saving 200,000 birds should be worth much more than saving 2,000 birds. In fact, the average contributions of the three groups were $80, $78, and $88 respectively. The number of birds made very little difference. What the participants reacted to, in all three groups, was a prototype—the awful image of a helpless bird drowning, its feathers soaked in thick oil. The almost complete neglect of quantity in such emotional contexts has been confirmed many times.

Intensity Matching

Questions about your happiness, the president's popularity, the proper punishment of financial evildoers, and the future prospects of a politician share an important characteristic: they all refer to an underlying dimension of intensity or amount, which permits the use of the word *more*: more happy, more popular, more severe, or more powerful (for a politician). For example, a candidate's political future can range from the low of "She will be defeated in the primary" to a high of "She will someday be president of the United States."

Here we encounter a new aptitude of System 1. An underlying scale of intensity allows *matching* across diverse dimensions. If crimes were colors, murder would be a deeper shade of red than theft. If crimes were expressed as music, mass murder would be played fortissimo while accumulating unpaid parking tickets would be a faint pianissimo. And of course you have similar feelings about the intensity of punishments. In classic experiments, people adjusted the loudness of a sound to the severity of crimes; other people adjusted loudness to the severity of legal punishments. If you heard two notes, one for the crime and one for the punishment, you would feel a sense of injustice if one tone was much louder than the other.

Consider an example that we will encounter again later:

> Julie read fluently when she was four years old.

Now match Julie's reading prowess as a child to the following intensity scales:

How tall is a man who is as tall as Julie was precocious?

What do you think of 6 feet? Obviously too little. What about 7 feet? Probably too much. You are looking for a height that is as remarkable as the achievement of reading at age four. Fairly remarkable, but not extraordinary. Reading at fifteen months would be extraordinary, perhaps like a man who is 7'8".

What level of income in your profession matches Julie's reading achievement?
Which crime is as severe as Julie was precocious?
Which graduating GPA in an Ivy League college matches Julie's reading?

Not very hard, was it? Furthermore, you can be assured that your matches will be quite close to those of other people in your cultural milieu. We will see that when people are asked to predict Julie's GPA from the information about the age at which she learned to read, they answer by translating from one scale to another and pick the matching GPA. And we will also see why this mode of prediction by matching is statistically wrong—although it is perfectly natural to System 1, and for most people except statisticians it is also acceptable to System 2.

The Mental Shotgun

System 1 carries out many computations at any one time. Some of these are routine assessments that go on continuously. Whenever your eyes are open, your brain computes a three-dimensional representation of what is in your field of vision, complete with the shape of objects, their position in space, and their identity. No intention is needed to trigger this operation or the continuous monitoring for violated expectations. In contrast to these routine assessments, other computations are undertaken only when needed: you do not maintain a continuous evaluation of how happy or wealthy you are, and even if you are a political addict you do not continuously assess the president's prospects. The occasional judgments are voluntary. They occur only when you intend them to do so.

You do not automatically count the number of syllables of every word you read, but you can do it if you so choose. However, the control over intended computations is far from precise: we often compute much more than we want or need. I call this excess computation the *mental shotgun*. It is impossible to aim at a single point with a shotgun because it shoots pellets that scatter, and it seems almost equally difficult for System 1 not to do more than System 2 charges it to do. Two experiments that I read long ago suggested this image.

Participants in one experiment listened to pairs of words, with the instruction to press a key as quickly as possible whenever they detected that the words rhymed. The words rhyme in both these pairs:

VOTE—NOTE
VOTE—GOAT

The difference is obvious to you because you see the two pairs. VOTE and GOAT rhyme, but they are spelled differently. The participants only heard the words, but they were also influenced by the spelling. They were distinctly slower to recognize the words as rhyming if their spelling was discrepant. Although the instructions required only a comparison of sounds, the participants also compared their spelling, and the mismatch on the irrelevant dimension slowed them down. An intention to answer one question evoked another, which was not only superfluous but actually detrimental to the main task.

In another study, people listened to a series of sentences, with the instruction to press one key as quickly as post="lly desible to indicate if the sentence was literally true, and another key if the sentence was not literally true. What are the correct responses for the following sentences?

Some roads are snakes.
Some jobs are snakes.
Some jobs are jails.

All three sentences are literally false. However, you probably noticed that the second sentence is more obviously false than the other two—the reaction times collected in the experiment confirmed a substantial difference. The reason for the difference is that the two difficult sentences can be metaphorically true. Here again, the intention to perform one computation evoked another. And here again, the correct answer prevailed in the conflict, but the conflict with the irrelevant answer disrupted performance. In the next chapter we will see that the combination of a mental shotgun with intensity matching explains why we have intuitive judgments about many things that we know little about.

Speaking of Judgment

“Evaluating people as attractive or not is a basic assessment. You do that automatically whether or not you want to, and it influences you.”

“There are circuits in the brain that evaluate dominance from the shape of the face. He looks the part for a leadership role.”

“The punishment won’t feel just unless its intensity matches the crime. Just like you can match the loudness of a sound to the brightness of a light.”

“This was a clear instance of a mental shotgun. He was asked whether he thought the company was financially sound, but he couldn’t forget that he likes their product.”