Avril Kenney

The Neuroscience of Decisions: Implications for Morality in Legal Policy

Introduction

Most people, when asked to explain the process by which they reached a decision they made, will be able to provide a reasonable-sounding answer. They will include information such as what knowledge they had, how they felt, what goals they were pursuing, and what consequences they considered. Usually the explanation will adequately account for the decision, based on the rules of logic that human reasoning is assumed to follow. In cases where a decision seems to have been made arbitrarily, for example among several equally desirable alternatives, people will say that they chose one for no particular reason, of their own free will, and in general such an explanation is completely acceptable.

Recently, however, researchers have shown incontrovertibly that people's decision-making process is far from rational. In making choices, we often consider factors that should not be relevant, ignore some that should be, and weight them in ways very different from what any rational method would come up with. How can this empirical observation be reconciled with the logic-based explanations that we provide for our own actions? The answer is simply that our explanations are often wrong.

It might seem odd for people to be incapable of correctly recounting how they reached even the most straightforward of decisions. But given how little we know about the inner workings of our minds, it should seem even more odd if we were actually able to observe our mental processes as they occurred. With all other aspects of the world, we learn from experience to have expectations about what circumstances or events will have what results. We can never directly observe causation, we can only infer it from patterns. If we see that an object falls every

time someone drops it, we learn that it falls *because* it was dropped, and if we see that a person eats food every time he feels hungry, we learn that he eats *because* he is hungry. These matchings do not represent absolute truth, they are only models to help us make predictions. In some situations, especially those where our irrational behavior differs from what our rational models predict, our explanations will differ significantly from the real causes.

Not surprisingly, this idea – that people are often incorrect about how they themselves reach their decisions – makes many people uncomfortable. It raises questions about free will and responsibility, an issue particularly relevant to legal policy. After all, if you cannot even explain why you did something, how can you be held accountable for a decision that your brain made without you? Should people with different brain anatomies be given different legal punishments? The problem with this question is that there is a difference between our moral sense of justice and our legal system of punishment.

Irrational Decisions and their Neural Correlates

Humans possess the powerful ability to use rule-based logic to determine new facts based on known information. It would seem wasteful ever not to use such an ability since we have it, so it is generally assumed that we reach all of our conclusions by these logical methods. The representations in our minds, and consequently in the legal system, expect people to take into account the facts available and follow them to the logical conclusions. Using these models, it is a simple matter to determine where a person's actions deviate from what is proper, and to treat him accordingly. However, research has shown that even "normal" human behavior does not fit predictions based purely on rationality. Several types of decision-making have been studied, and some have been localized to particular areas of the brain.

In general the exact consequence of a decision cannot be predicted, but the possible outcomes and their relative likelihoods can be approximated. A perfectly rational agent would

make whichever choice maximizes the "expected value" of the result – that is, maximizes the average value of all possible outcomes weighted by their probabilities. In many studies of such behavior, probabilities are provided explicitly, and money is used as the item to be valued, since it should rationally be valued on an absolute and linear scale. It has been found that real people, if they are using this method at all, tend to greatly misjudge both the values and the probabilities.

Ariely, Loewenstein, and Prelec (2006) performed a study in which subjects had to write down the last two digits of their social security numbers, decide whether they would pay that amount in dollars for a specified item, and then state the maximum amount they would pay for the item. Despite the fact that social security numbers have no relation to the question, people with higher numbers were on average willing to pay much larger amounts. This shows that our value judgments are measured relative to some baseline, which could be completely arbitrary and irrelevant.

Similar effects have been found when people have to value time in terms of money. When given the option of receiving $10 today or $11 tomorrow, many people would choose the slightly smaller amount at the earlier date, but if the dates were 365 days from now versus 366 days from now, the same people would likely choose the larger amount of money. Both cases are the same in that the person has to decide whether waiting one day is worth one extra dollar, but in the second version the one day is being valued relative to a baseline of a year, making it seem a much less significant interval to wait. An fMRI study by McClure et al. (2004) found that when people were faced with such decisions, immediate rewards tended to activate areas in the limbic system, while delayed rewards activated more cortical regions.

People also have what economists call "risk aversion" and "loss aversion," characteristics which cause them to avoid any chance of very bad outcomes and to prefer high-probability satisfactory outcomes over low-probability very good ones, even when these choices go against

the expected value calculations. Rangel, Camerer, and Montague (2008) report that signals in the striatum, insula, and orbitofrontal cortex can correspond to risk and expected value, while activation in the striatum and other areas is correlated with nonlinear functions of probabilities, which would account for the uneven weighting of different outcomes.

Rationalization and the Illusion of Introspection

Clearly, under many circumstances people with fully functional brains make decisions that are not rational, yet in many cases, when the irrationality is pointed out, they are surprised or confused by it. If we could directly observe the underlying causes of our actions, explanations obtained through introspection would always be able to account for observed behaviors, and would certainly not contradict them. In reality, though, people's explanations for their choices are often completely different from the actual causes, and sometimes they have no explanations at all.

In the example of the trolley problem, many people are unable to provide any reason for why they chose their answers. Even when they adamantly believe in their choice they cannot give an explanation, since their post-hoc reasoning can only use logical inferences, and the decision was not made logically. Hauser et al. (2007) performed a study in which subjects were asked to judge the permissibility of killing a person in variants of the trolley problem, and then to provide justifications for their decisions. Of the subjects who gave differing answers to the two versions (excluding those who added extra assumptions not stated in the question), only 30 percent were able to give adequate justification for their judgments. Furthermore, subjects who had exposure to readings on moral philosophy were significantly more likely to sufficiently justify their positions than those who had not, even though the two groups had the same likelihood of judging the acts permissible. This strongly implies that people cannot use introspection to find explanations for their decisions, and that the reasons they do give are from

external rather than internal information.

Additional compelling evidence comes from Nisbett and Wilson (1977), who describe a comprehensive set of studies showing that people often give incorrect accounts of what factors influence their decisions. As just one of many examples, in one experiment subjects were shown four pairs of stockings and had to decide which was the best quality and why they thought so. The pair farthest to the right was about four times as likely to be chosen as the one on the left (the stockings were all objectively of identical quality), yet none of the subjects cited position as having played a role in their decision-making process. In fact, almost all of them denied it even when specifically asked. Although in this scenario no individual can be proven to have selected the stocking based on position, some of them obviously did so, and if they had been at all capable of accessing the true causes of their decisions, they would have at least acknowledged the position factor when it was suggested.

A more recent study by Johansson et al. (2005) shows that people give qualitatively equivalent explanations for their choices regardless of whether they really made those choices. Subjects were shown two photos of faces and asked which was more attractive; they were then given the photo they had selected and asked to explain why they had picked it over the other. But in some of the trials, the experimenter switched the photos, giving the subject the one he had not actually chosen. People usually did not detect the switch, and there was no significant difference between their explanations for the decisions they had made and the decisions they had not made (measured by ratings of emotionality, specificity, and certainty). This finding implies that people have no direct evidence of where their decisions come from and are employing the same methods on their own mental processes as they would on any other observed events.

Further support for this claim is provided by brain-imaging studies in which the detection of a decision in the brain is compared with a subject's conscious perception of the decision. Soon

et al. (2008) conducted an fMRI study in which subjects watched letters flash on a screen and decided when to press one of two buttons. Each time a subject became consciously aware of the decision to press a button, he noted which letter was on the screen, so that the exact instant of the decision's entering his consciousness could be recorded. Two cortical brain regions – one in the frontopolar cortex and one in the parietal cortex – were found to reliably predict which of the buttons would be pressed, up to 10 seconds before the subject was conscious of the choice. Given such a delay, it is apparent that people do not directly see the procedures by which their decisions are made, since the information is present in the brain before they are aware of it.

How exactly people do reason about their decision-making is still uncertain. It is known that thinking about thoughts, as opposed to actions, characteristics, or physical states, selectively recruits certain parts of the brain – particularly the right temporo-parietal junction (right TPJ), but also the left TPJ and medial prefrontal cortex. Importantly, these areas are distinct from those involved in planning and performing actions, as well as from the mirror neuron system. Although the studies have primarily involved thinking about other people, a few experiments reported by Saxe (in press) suggest that these brain regions might correspond to thinking about mental states in general, whether one's own or someone else's. As of yet there have not been imaging studies requiring subjects to reflect on their own decision-making, so it remains to be seen whether this hypothesis holds. If confirmed, it would be consistent with the idea that people model their own mental activity in the same way as that of other people and not by direct observation.

The Relationship between Morality and Legal Policy

Scientific understanding of decision-making as recounted above is crucially relevant to legal policy in how laws are created as well as how they are applied. Given that even the most sensible people can make irrational decisions, such as by skewing value judgments or using an inappropriate anchoring point, we should think carefully about what criteria we use for making

laws. It is also important to moderate in individual court cases what information is provided and how it is presented to the judges and juries. When the actions of defendants are evaluated, irrationality cannot be a reason to call someone a criminal, rational explanations cannot excuse unacceptable behavior, and the lack of rational explanations cannot exempt someone from responsibility.

Our sense of morality provides us with certain intuitions about what is right or wrong, what is just, what people deserve. In the legal system, everything is based on rules, so it is important that policies be consistent and justifiable. Any ruling made by a court must be based on principles that can be explicitly stated and broadly applied. Laws can be as specific as necessary to cover all variants of a situation, but then they must be applied strictly to those situations to which they are relevant, with no space for exceptions, so that the results are "fair." If people could correctly provide the reasons for their moral intuitions, then laws could simply be made based on these reasons, and legal decisions would never conflict with moral decisions (at least to the extent that everyone's moral system is the same).

But in actuality, people's feelings about what is right can differ significantly from the law, a controversial problem because it is then uncertain whether laws should be changed to reflect our morality or whether morality should be disregarded in favor of the laws we have established. It could be argued, on the one hand, that moral judgments are clearly irrational and so should not be used for deciding how society should function, or on the other hand, that the purpose of law is to represent how people think and so it should as greatly as possible align with what people consider morally correct. Although the second view has some appealing qualities, the first is the one that should be followed, because laws are made to keep order in society, so they should be based on what their real consequences will be and not on what will make people feel good about them.

Morality, like all of our decision-making mechanisms, is a heuristic that often produces the appropriate result, but there is no reason to use it when better models are available. In the trolley problem, for example, normal moral people differentiate between pulling a lever and physically pushing someone, so they might give different punishments to someone depending on which of these actions he took. From a legal standpoint, the purpose of punishing someone is not to give him what he deserves, but to provide the greatest overall benefit for society. Therefore, the legal punishment for an action should be whatever is most likely to compensate for losses and prevent future crimes.

In determining what the consequences should be for someone's behavior, there is a tendency to differentiate between situations based on how much the person is considered "responsible" for his actions. This has a valid evolutionary basis, since historically it has made sense that someone should not be punished for crimes that he was forced to commit by external influences, he should only be punished for actions that he made the choice to perform. A problem arises when biological bases are discovered for psychological phenomena – if someone has a brain structure that causes him to behave in a certain way, it can be claimed that his actions are forced on him just as if from an external source. This has led many people and some legal systems to treat people differently based on such biological evidence.

The emerging scientific findings presented here challenge this intuition by pointing out that in a fundamental sense we are not ever capable of making any choice other than the one we ultimately do make. The notion of "free will," that we are free to choose among alternatives and always could decide differently if we so desired, is a convenient concept to fill in the unpredictable part of our mental model of human behavior, but it can be misleading when people try to combine it with what actually occurs biologically. All decisions are directly caused by the interaction between our environment and our brains, regardless of what we consciously perceive

the causes to be or whether we are “normal” or “abnormal.” Responsibility in the moral sense rests on the assumption that people have free will, but such an assumption does not hold in a scientific or legal context, in which responsibility is just a distractor in the attempt to determine appropriate consequences.

Conclusion

People generally conceive of themselves as rational agents, with the ability to consciously make choices and see where those choices come from. There is now robust data indicating that our decisions are not based on rationality and our explanations of them are not directly related to their actual causation. This knowledge should be a central consideration in the practice of law, which relies on models of human behavior that must be as accurate as possible. When forming legal policies, people should keep in mind that moral judgments are subject to the many errors in natural decision-making, so laws should be not be made only according to morality. And when assigning punishments to those who have committed crimes, people should focus not on how much choice there was in whether to commit the crime, but on what consequences the punishments will have for the individual and for society.

References

Ariely, D., Loewenstein, G., Prelec, D. (2006). "Tom Sawyer and the construction of value." *Journal of Economic Behavior & Organization* 60(1), 1-10.

Eastman, N., Campbell, C. (2006). "Neuroscience and legal determination of criminal responsibility." *Nature Reviews Neuroscience* 7(4), 311-318.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D. (2001). "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293, 2105-2108.

Hauser, M., Cushman, F., Young, L., Jin, R.K., Mikhail, J. (2007). "A Dissociation Between Moral Judgments and Justifications." *Mind & Language* 22(1), 1-21.

Johansson, P., Hall, L., Sikström, S., Olsson, A. (2005). "Failure to Detect Mismatches Between Intention and Outcome in a Simple Decision Task." *Science* 310, 116-119.

McClure, S.M., Laibson, D.I., Loewenstein, G., Cohen, J.D. (2004). "Separate Neural Systems Value Immediate and Delayed Monetary Rewards." *Science* 306, 503-507.

Nisbett, R.E., Wilson, T.D. (1977). "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84(3), 231-259.

Rangel, A., Camerer, C., Montague, P.R. (2008). "A framework for studying the neurobiology of value-based decision making." *Neuroscience* 9, 1-12.

Rosen, J. (2007). "The Brain on the Stand." *The New York Times*, 11 March 2007.

Saxe, R. (in press). "The happiness of the fish: Evidence for a common theory of one's own and others' actions." In K. Markman, B. Klein, J. Suhr (eds.), *The Handbook of Imagination and Mental Simulation*.

Soon, C.S., Brass, M., Heinze, H., Haynes, J. (2008). "Unconscious determinants of free decisions in the human brain." *Nature Neuroscience* 11, 543-545.