

# A Psychological View of Moral Intuition

By Jonathan Baron

**A**T LEAST SINCE THE PUBLICATION OF JOHN RAWLS' *A Theory of Justice* in 1971, it has been standard practice in moral philosophy to develop theories by trying to explain and systematize our moral intuitions. Rawls made an analogy with linguistics. A few years before Rawls wrote, Noam Chomsky had advanced the field of linguistics through a similar move. Chomsky developed a mathematical theory of the structure of sentences by trying to account simply for his own intuitions about what was a sentence and what was not.

Rawls's view of theory construction through "reflective equilibrium" was subtle and elaborate. Much has been written about it. But a glance at current philosophy journals suggests that a simpler method has become common practice. Typically, the author presents a few carefully constructed cases and then tries to account for her own moral judgments about these cases.

This is especially true in the study of the set of issues concerned with causing harm through action *versus* omission, or causing harm *versus* letting it happen. For example, H. M. Malm describes a case in which two children, John and Mary, are in a crushing machine.<sup>1</sup> If you push a button, John will be killed but Mary will be unharmed. If you do not push, the reverse will happen. Malm thinks that you are wrong to push the button because it would be difficult to come up with a reason for doing so, while you do not need to give a reason for not pushing. Malm uses this example to derive substantive moral conclusions.

The question of where these intuitions come from and how they can be explained is indeed an interesting one. It may tell us something about human psychology. But, to someone who has been reading the literature on the psychology of heuristics and biases, it is unlikely to tell one much about the correct moral solution to these cases or others.<sup>2</sup>

It is easiest to see the problem if we get away from the moral domain and look at human intuitions about other cases where there is more agreement about the right

**JONATHAN BARON IS** Professor of Psychology at the University of Pennsylvania. He is interested in decision making, moral judgment, and their relation to public issues. He received a B.A. in psychology from Harvard in 1966 and a Ph.D. in psychology from the University of Michigan in 1970.

---

answer. Psychologists have discovered — under different names — a variety of cases in which intuitions seem to be systematically incorrect.

One set of cases goes by the name of “naïve theories.” These are most easily found in science. In one study, students graduating from Harvard were asked why it was hot in the summer and cold in the winter. Many students said that the earth was closer to the sun in the summer.<sup>3</sup> This is reasonable: if you move closer to a fire, you get warmer. If this were the explanation, though, it would be warmer in July than in January in the Southern Hemisphere too. People’s intuitions about science are often based on everyday experience, but this experience can mislead.

Many kinds of naïve theories have been discovered. Some people think that a thermostat is like the accelerator in a car, so, when they walk into a cool house, they turn it from 60 to 80 degrees so that the house warms up (to 70) more quickly. Again, not an unreasonable theory, just false. Of course, in science, we can show that the theory is false, though this may be more difficult in ethics. My point is that the same mechanisms that lead to incorrect intuitive theories in science may operate in ethics too.

A related phenomenon — probably the same phenomenon under a different name — is the use of heuristics in judgment and decision making. If you flip a coin six times, which is more likely, HHHTTT or HTHHTT? Many people say the latter, because it looks more similar to other random sequences. (The two sequences each have probability 1/64.) People tend to judge probability in terms of similarity, or “representativeness.” Again, use of similarity to judge probability usually works, but it fails here.

Psychologists have cataloged dozens of such effects in decision making. In the “sunk cost effect,” people consider the resources put into an option in deciding whether to abandon it, holding future consequences constant. For example, Ernie bought one TV dinner on sale for \$4, decided to invite Bert over for dinner, then bought another for \$8 (the sale being over). When both dinners were in the oven, Bert called to say he couldn’t come. Which dinner should Ernie eat if he can eat only one, or does it matter? About 40 percent of adults (and more children) say that he should eat the \$8 one, because he wastes less money that way. The rule against waste is a good one, but it does not apply here. The money is already spent.

The “status-quo effect” increases the value of whatever is perceived as being the status quo. In one study, half of the (Cornell) students in a room were told that they had a Cornell mug, and they were asked how much money they would accept for it. The other half were asked how much money they would need so that they would prefer getting that amount of money to getting a Cornell mug. The former group gave answers about twice as high as the latter. Note that both groups were choosing between the same two outcomes. The only difference was in what they were told about the status quo. In general, sticking to the status quo is a reasonable heuristic; change is costly and risky. But in this case, these differences were not there. The heuristic was used anyway.

For another example of the status-quo effect, consider the following two cases:

- A. The government plans to put a hazardous waste site 50 miles away and reduce your taxes, unless you (and others) are willing to give up the reduction, in which case the waste site will be 500 miles away;

- 
- B. The government plans to put the site 500 miles away unless you (and others) are willing to accept a tax reduction, in which case it will be 50 miles away.

In each case, what tax reduction would make you indifferent between the two locations? Many people would want more in B than in A. The move from 500 to 50 is a more serious loss than the move from 50 to 500 is a gain, despite the fact that the site is not built yet.<sup>4</sup>

Notice the analogy between these examples and Malm's example of John and Mary. Malm's intuition is easily explained as an extension of the same heuristic that yields the status-quo effect. Again, the point is not that Malm is necessarily wrong, but rather that this intuition can be understood without thinking that intuition is the royal road to moral truth.

A similar distinction occurs in cases of action *versus* omission. In questionnaire studies and in real life, most drivers given the choice of reducing auto-insurance costs by limiting their right to sue will not take this option if they must act to do so. But if they must act to prevent the change, most people will accept it. The latter was the case in Pennsylvania, the former in New Jersey. About twice as many drivers chose the reduction in New Jersey as in Pennsylvania.

In general, the harm caused by action is more regrettable than the harm caused by omission, for a variety of reasons: acts are more often intentional; omissions are usually more easily corrected; and so on. So people develop a general rule against causing harm through omissions. When an action has both costs and benefits, people shy away from it, even when the usual factors that distinguish acts and omissions are absent. This intuition may lie behind some of the opposition to trade agreements, even among those who admit that the benefits outweigh the harms.

Suppose that a flu epidemic is going to kill 10 out of 10,000 children. A vaccine will completely prevent the flu, but the vaccine may kill some number of children. You can tell nothing about which children will die from the flu or from the vaccine. All children have an equal chance in both cases. How much vaccine risk would you tolerate before you would decide not to vaccinate your child? (Or, if you were a policy maker, how much risk would you tolerate before you would no longer recommend the vaccine?) Many people will not tolerate very much risk. Some will tolerate none at all.

Notice that, if this intuition is put into effect, more children would die than if more children were vaccinated. This may have happened. In England and Japan, news stories about rare, isolated cases of brain damage or death possibly resulting from pertussis vaccine led many parents not to vaccinate their children. As a result, hundreds of children died or suffered brain damage from whooping cough, which the vaccine would have prevented. In the U.S. some parents still resist vaccination even though they admit that the risk of vaccination is lower than that of nonvaccination. The Sabin polio vaccine is another case.

If you were a rational child, you would probably want your risk to be minimized, and you would not care whether minimal risk was associated with someone's action or inaction. If we accept this account of what a child would rationally want and then fail to vaccinate the child, we are going against the child's interests, and not because someone else's interests are served instead, but rather because of our own moral beliefs. Why should we advocate moral principles—such as the relevance

---

of the act-omission distinction—that go against some interests and serve no interests? (We could, of course, describe the situation *ex-post* as a conflict between those children who would die from the vaccine and those who would die from the disease. But the *ex-post* and *ex-ante* descriptions are the same case described differently. If we make a different decision, we must explain why the description is morally relevant.)

I have argued that moral intuitions can be understood as naïve theories, or heuristics, so that they cannot be relied upon as the basis of normative moral theory.<sup>5</sup> Given that these heuristics exist and that at least some of them are clearly erroneous, how can we be sure that our moral intuitions are not in the same category? All of these heuristics can be understood as approximations that work most of the time (at best). If that is what they are, then they are not reliable guides to moral truth.

This view is consistent with the fact that people differ considerably in the heuristics they use and in the strength of their commitment to those heuristics.

This is true in most of the examples I have given. In another study Mark Spranca and I gave subjects a story about a tennis player named John who wanted to beat Ivan Lendl (then ranked first in the world) by inducing Lendl to eat cayenne pepper. John knew that the house salad dressing at the club dining room where they were eating contained cayenne. He planned to recommend it to Ivan. In one version, John was about to do so, and Ivan ordered it himself, so John said nothing. In the other version, John recommended it. Some subjects felt quite strongly that John's behavior was morally different in the two cases (although they acknowledged the equivalence of intention). Others felt equally strongly that the behavior was equivalent. One subject wrote angrily, in large letters, that this distinction was at the root of Nazism: those who did nothing to stop it were fully responsible for their omissions. Possibly, philosophers also differ in the strengths of their intuitions, and this accounts in part for their disagreements.

---

*“I have argued that moral intuitions can be understood as naïve theories, or heuristics, so that they cannot be relied upon as the basis of normative moral theory.”*

---

Cultures, too, differ in their intuitions. U.S. college students, for example, have strong intuitions about the importance of autonomy and noninterference, but these intuitions were almost totally absent in a sample of students in India. Reliance on intuition can result in a culture-bound morality.<sup>6</sup>

I shall deal briefly here with three objections to my argument. One is that some moral intuitions are held very strongly by everyone, so they must be correct. It isn't clear to me why this follows. However, it is also true that some decision-making biases are held very strongly even though they are demonstrably illogical.

Second, it may be that the best moral system is to follow our strong intuitions,

---

even if we cannot understand how these intuitions arise. I must grant this as a possibility. This is especially likely to be true when we can understand our intuitions as good approximations to ideal rules, and when any attempts to apply the ideal rules directly are more likely than not to lead to error. But this seems to me to be a practical matter, which I have called prescriptive rather than normative. If we want to understand morality, our intuitions are not enough, although as a guide to behavior they may be very good.

Third, if we disallow intuitions as primary data in moral philosophy, is anything left? The analogy with linguistics suggests a negative answer. I cannot answer this question here, although I and others have tried to answer it elsewhere. In brief, I think that alternative bases of moral philosophy can be found by asking ourselves about the ultimate aim of the enterprise.<sup>7</sup> ◻

---

### Endnotes

<sup>1</sup> "Killing, letting die, and simple conflicts." *Philosophy and Public Affairs* 18 (1989): 239-258.

<sup>2</sup> For a review of this literature, and for discussions of points for which no citations are given, see my book, *Thinking and Deciding* (2nd ed.), New York: Cambridge University Press, 1994.

<sup>3</sup> Matthew H. Schneps, "A private universe," Santa Monica, CA: Pyramid Film and Video.

<sup>4</sup> Jonathan Baron, "The effect of normative beliefs on anticipated emotions." *Journal of Personality and Social Psychology*, 1992, 63: 320-330.

<sup>5</sup> See my paper, "Nonconsequentialist decisions," *Behavioral and Brain Sciences*, 1994, 17: 1-42, and my book *Morality and rational choice* (Dordrecht: Kluwer, 1993), for additional discussion of the arguments made here.

<sup>6</sup> This is based on still-unpublished work conducted in collaboration with Joan Miller.

<sup>7</sup> Again, see my "Nonconsequentialist decisions" and *Morality and rational choice*, as well as Richard Hare's *Moral thinking: Its levels, method, and point* (Oxford: Clarendon Press, 1981).