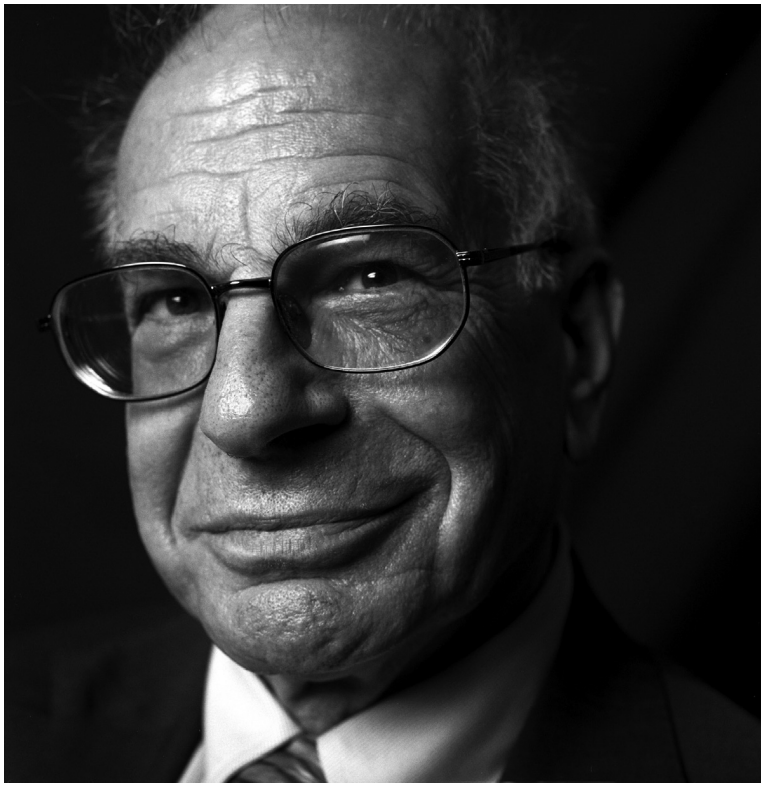

3
DANIEL KAHNEMAN

CAN WE TRUST OUR INTUITIONS?

Daniel Kahneman is the world's pre-eminent investigator of the ways in which the limits of our cognitive abilities shape our judgements. Since the late 1960s, his work, much of which was carried out in close collaboration with the late Amos Tversky (1937–1996), has focused on our intuitive judgements. One of Kahneman and Tversky's favoured methods of studying such judgements involves asking subjects relatively simple questions about cases. An example is the famous 'Linda Case', in which subjects are given the following description of the protagonist:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.

Subjects were then given a list of eight possible outcomes describing Linda's present employment and activities. Besides a number of miscellaneous possibilities (e.g. elementary school teacher, psychiatric social worker), this list included the descriptions 'Linda is a bank teller' and 'Linda is a bank teller active in the feminist movement'. Subjects were then asked to rank these descriptions by the probability that they were

true. A large majority responded that Linda was less likely to be a bank teller than a bank teller active in the feminist movement. This is an obvious mistake, since it cannot be more likely for Linda to possess attributes X and Y than for her to possess attribute X. The explanation Kahneman offered for the erroneous majority judgement was that respondents were implicitly using a heuristic — a mental short-cut — to arrive at their judgements. This particular heuristic involved replacing the attribute that was the target of the question (the relative probability of the description's truth) with an attribute that comes more easily to mind (the relative resemblance of the description to the introductory statement about Linda). In other words, respondents were using the degree to which the description of her current activities resembled Linda as a quick way of judging the likelihood that the description was true.

The Linda Case illustrates several aspects of intuitive judgements: they typically spring to mind quickly and automatically, without much effort, and are difficult to control or modify, even in the face of conflicting evidence. (As the biologist Stephen Jay Gould wrote about this example: 'I know [the right answer], yet a little homunculus in my head continues to jump up and down, shouting at me, "But she can't just be a bank teller, read the description!" ') Moreover, while the heuristics that give rise to such judgements may generally be useful — they economize on our mind's scarce computing time and ability — they can also lead us astray.

Kahneman and Tversky also famously drew attention to another way in which our intuitive judgements may fail to conform to rational principles: irrelevant variations in the description of alternatives can evoke a change in judgement, because each description elicits a different intuitive mental representation of the alternatives. An example is their Asian Disease Case:

> Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programmes to combat the disease have been proposed. Assume that the exact scientific

estimates of the consequences of the programmes are as follows:

> If programme A is adopted, 200 people will be saved.
>
> If programme B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.

Which of the two programmes would you favour?

In this version of the problem, a substantial majority of respondents favours programme A. Other respondents, however, received the same cover story followed by two differently described options:

> If programme A* is adopted, 400 people will die.
>
> If programme B* is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

Given these options, a clear majority favours programme B*. Of course, A and A* are identical, as are B and B*. Nonetheless, subjects are significantly more likely to choose the option in which 200 people will certainly be saved and 400 will certainly die over the risky option when the description draws attention to lives saved rather than to lives lost.

Kahneman and Tversky explained this pattern as a result of people's intuitive tendency to represent outcomes as involving 'gains' or 'losses' relative to an imagined baseline and apply different decision-making rules depending on whether outcomes are represented as gains or as losses. The first description of the problem draws attention to lives saved, eliciting a 'gain' representation relative to a baseline in which everyone dies. The second description draws attention to lives lost, eliciting a 'loss' representation relative to a baseline in which everyone survives. Now, when considering *gains*, people are generally *risk-averse*: they favour a certain gain of a given number of lives over

a gamble in which the sum of probability-weighted lives gained is just as great. This makes the certain gain in A of 200 lives seem attractive relative to the gamble in B, which involves a one-third probability of saving 600 lives and a two-thirds probability of saving no lives at all. (The sum of probability-weighted lives saved in B is therefore $(1/3 \times 600) + (2/3 \times 0) = 200$, which is equal to the number of lives saved for certain in A.) When considering *losses*, however, people are generally *risk-seeking*: to avoid a certain loss, they will take a gamble in which the sum of probability-weighted lives lost is just as great as the number of lives lost in the certain option. This makes a risky option like B*, which involves a one-third chance of no loss of life and a two-thirds chance of a loss of 600 lives, relatively attractive when compared to A*, which involves a certain loss of 400 lives. Kahneman and Tversky described the resulting difference in people's preferences as a 'framing effect': different ways of presenting the same decision problem elicit different responses, even though rationality requires the same pattern of response.

Kahneman and Tversky's work on a variety of heuristics and framing effects, and their innovative theory of how people choose in risky situations, spawned a huge research programme in psychology and economics, and earned Kahneman the Nobel Prize in Economics in 2002. Since the mid-1980s, Kahneman and others have also investigated the use of heuristics and the existence of framing effects in moral judgements. This work is relevant to moral theory because one common procedure of moral enquiry is to employ the method of reflective equilibrium, which involves working back and forth among our intuitive judgements about particular cases and the principles we believe to govern them, revising any of these elements wherever necessary in order to achieve an acceptable coherence among them. We reach reflective equilibrium when our intuitive case judgements and moral principles are consistent with each other and some of these judgements and principles provide support for or provide a best explanation of others. For those who employ this method, it is obviously important to find out when these case judgements are liable to error, and if they are, what, if anything, we can do to correct them.

I meet Kahneman to discuss these issues in London in September 2006. I am immediately struck by his energy and the sense of excitement with which he approaches his work. We've arranged to meet following an afternoon seminar at a central London think tank. The seminar runs late, and most participants push out of the crowded, airless room looking exhausted. Kahneman, however, emerges with a spring in his step. During our walk through Covent Garden to my office at the London School of Economics, he enthusiastically recalls his meeting earlier that day with a Ph.D. student, Benedetto De Martino, at the nearby Wellcome Department of Imaging Neuroscience. De Martino had studied the areas of people's brains that become active while they are making choices. He had found that subjects who are strongly susceptible to the aforementioned gain/loss framing effects have higher activity in a part of the brain associated with emotional processing (the amygdala), whereas subjects who are less susceptible to such effects display higher activity in parts of the brain associated with analytical processing (areas in the prefrontal cortex). The study therefore offered some support for the view that the intuitive processes that Kahneman and Tversky had uncovered operate in a distinctive part of the brain and determine judgement unless overridden by more deliberative reasoning. As Kahneman is explaining the suggestions he made to De Martino for further research, we cross the wide, busy Kingsway. In the middle of the pedestrian crossing, he stops short, oblivious to the approaching traffic, and exclaims 'Damn! I made a mistaken suggestion. I need to e-mail this guy right away!' A few minutes later, after he has fired off a quick message to De Martino, I start the interview by asking how he became interested in moral questions.

DANIEL KAHNEMAN: In my teens, I was interested in philosophical questions — God's existence, the reasons not to do what people think is wrong. But I discovered very early on that I was more interested in what made people believe in things than in the correctness of the beliefs themselves. For example, I was more curious about what made people believe in God than whether He really existed, and was

more occupied with the origins of people's moral intuitions than with moral philosophy.

ALEX VOORHOEVE: *'Intuition' is a vague term. How would you define it?*

DANIEL KAHNEMAN: I haven't always been consistent in my use of the term. The first time that I used it in print, I think, was in a paper with Amos Tversky on the 'Law of Small Numbers'.[1] In that paper, we wrote something to the effect that 'the Law of Large Numbers is not part of the repertoire of human intuition'. What we meant there by 'intuition' was a descriptive generalization made by an objective observer about the rule that appears to generate the judgements of an individual about specific cases. And this can be an accurate description by the outside observer, even if the individual would reject this rule if he were to consider it.

I have also used the term to describe an intuitive *system*, a way of generating thoughts that are called intuitive thoughts. The operations of this system are typically automatic, quick, effortless, associative, and often emotionally charged. Usually, they are not open to introspection, and difficult to control or change. This intuitive system generates involuntary impressions that come to mind spontaneously, like percepts. Intuitive judgements directly reflect these impressions with little modification from the reasoning system, which functions very differently from the intuitive system: the reasoning system's operations are step by step,

[1] The Law of *Large* Numbers holds that given a sample of independent and identically distributed random variables with a finite expected value, the average of these observations will eventually approach and stay close to the expected value. 'Expected value' here does not mean what it does in ordinary English; it is the sum of the possible outcomes weighted by the probability that they will occur. For example, if one rolls a fair die, the expected value of the outcome is $(1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$. The Law of Large Numbers predicts that the average stabilizes around the expected value of 3.5 as the number of times one rolls the die becomes large. The Law of *Small* Numbers is a psychological tendency to treat small samples as if they were large samples.

slow, effortful, and more likely to be consciously controlled and flexible.

ALEX VOORHOEVE: *There appears to be a difference between what you call intuitive judgements, understood as a product of the intuitive system, and the type of case judgements that philosophers use to build and test their moral theories, which they also sometimes refer to as 'intuitions'. Philosophers typically take these to be considered case judgements, that is, judgements that arise after reflection on the case and on our reasons for judging it as we do, whereas your category of intuitive judgements does not appear to involve such extensive reflection. Perhaps the relation between the two types of judgements is as follows. Philosophers start with a judgement about a case that comes to mind quickly and automatically and that may be emotionally laden—what you would call an intuitive judgement. They then examine whether they can find reasons to regard it as unreliable or biased. If they find no such reason, it is presumptively treated as a valid case judgement, and they try to find principles that explain the judgement, in part by reflecting further on what caused them to arrive at the judgement in question. An example of someone who uses this approach is Frances Kamm, who, in conversation, describes her method as follows:*

> I don't really have a considered judgement about a case until I have a visual experience of it. I have to deeply imagine myself in a certain situation with an open mind. It is almost as if you are looking at something with no preconceptions. You have to attend to it, and then things will pop out at you. First you may get the intuitive judgement of what you really should do in the circumstance you are imagining. Then you wonder, 'Why am I reaching this conclusion?' And your inner eye focuses on one factor as driving this judgement.

What do you think of this method of arriving at considered case judgements?

DANIEL KAHNEMAN: Well, there is a very interesting contrast between the way a psychologist would go at it and the way a philosopher would go at it, and it has a lot to do with the idea that the philosopher can know *why* she is reaching this conclusion. A basic assumption in psychological analysis is that you first have an intuition in a given situation. Then, when you ask yourself why you have it, you tell yourself a story. But the story you come up with does not necessarily identify the cause of your intuitions, because you typically do not have access to what causes your intuitions. Moreover, in her work, Kamm seems to assume that the cause of her intuitive judgement is the recognition of something that counts as a reason. But this need not be true. For example, the psychologists Dale Miller and Cathy McFarland asked subjects to determine the appropriate level of compensation for a man who was shot in the arm during the robbery of a grocery store. Some respondents were told that the robbery happened in the victim's regular store; others were told that the victim was shot in a store that he visited for the first time that day, because his usual store happened to be closed. The second version is more poignant, because it is easier to imagine the counterfactual undoing of an unusual event than of a regular occurrence. This difference in poignancy led to a difference of $100,000 in the median compensation that respondents awarded, so it was a clear cause of the size of the award. But further research showed that subjects do not regard poignancy as a good reason to award higher compensation in such cases. So I would very sharply separate the different phases in Kamm's description of her method: the first phase, where she knows what she would do, and the second phase, in which she thinks deeply as to why she would do it and comes up with a reason. It is at least possible that the reason she comes up with was not the cause of her judgement.

You know, many psychologists believe that consciousness is a story that we tell ourselves about ourselves. And in many cases, the story does not correspond to reality. I mean, it is very easy to create cases where you know the story isn't true. Let me give you an example, just

to push the point a little. There are experiments with post-hypnotic suggestion, where you tell somebody, 'I'll clap my hands and then you'll get up and open the window.' The person wakes up, you clap your hands, and he gets up and opens the window. If you ask him, 'Why did you open the window?' he'll say something like, 'The room felt very warm.'

ALEX VOORHOEVE: *But aren't these cases of hypnosis simply pathological?*

DANIEL KAHNEMAN: No, they are the best examples! The beauty of these examples is that you know why the subjects are doing it. They are doing it because they were given the instruction, and then somebody clapped his hands. But they have a completely different experience of why they are doing it. What's more, people are never at a loss for a reason why they did it. They can be made to do absolutely absurd things with post-hypnotic suggestion, and yet these things seem to make sense to them when they do them. The conclusion I draw from this is that the mental operation of making sense of our intuitive judgements is a very different cognitive activity from having these intuitions. This takes us to the core of my disagreement with Frances Kamm. To me, her confidence is very much like the confidence of the hypnotic subject who claims he knows why he opened the window.

ALEX VOORHOEVE: *Still, the case of the hypnotic seems to show only that we cannot always trust our introspective judgements about why we act or judge as we do; it doesn't show that we can never trust these judgements. And the method employed by Kamm and others involves more than simple introspective judgement about what is driving our judgements in isolated cases. Philosophers employing this method go on to consider whether the reason they think they might be responding to in a particular case determines their judgement in other cases also.*

DANIEL KAHNEMAN: But this raises another methodological problem. There is a distinction between what I have called the 'within-subject'

and the 'between-subject' method of considering cases. The within-subject method involves noting the intuitions a given subject has when she is considering multiple contrasting cases. The between-subject method involves noting the intuitions a subject has in a single case without engaging in explicit comparison with other cases. The two methods can elicit very different intuitions about the cases considered. Now, because its aim is to find rules that are applicable across different cases, moral philosophy is, by its nature, restricted to the within-subject method. That is, a moral philosopher is always conscious of two or more cases and has intuitions about the differences in the two cases and about whether these differences are relevant to how she should respond. And I think this is a major limitation of moral philosophy, because, in their everyday life, people are confronted with problems one at a time, so their relevant intuitions are about cases that occur one at a time. And the moral philosopher's stance prevents her from identifying the moral intuitions that are relevant to individuals who live their lives in this way. For this purpose, the between-subject approach, which involves asking one group of people about one case and another group of people about another case, is superior.

ALEX VOORHOEVE: *That's very interesting . . . I suppose I see how people's tendency to respond differently to the same case depending on whether it is framed as involving gains or losses, as in the Asian Disease Case, would only clearly emerge in a between-subject experiment. Can you give some other examples of cases in which the two methods lead to different judgements?*

DANIEL KAHNEMAN: Let me give you two examples. First, the intuitive judgements in the grocery store cases I mentioned before, which revealed the effect of poignancy on financial compensation for the victim, could only have been elicited by the between-subject method. For if one considers the two cases side-by-side, whether the victim was at his regular grocery store or not would have appeared irrelevant, and so would have made no difference to the compensation

awarded. Second, in research I carried out with Cass Sunstein, David Schkade, and Ilana Ritov, we found that the punishments people thought appropriate when considering in isolation cases of business fraud differed significantly from punishments they thought appropriate when they compared cases of business fraud with wrongdoing involving bodily injury. Our hypothesis was that outrage at an act was a significant contributor to the size of the punishment people judged appropriate. When they considered a case of business fraud in isolation, people implicitly compared the case to other cases of business fraud, so their level of outrage was determined by how egregious the conduct was as compared to other cases of business fraud. A particularly egregious case of business fraud would therefore be very heavily punished. However, in comparing cases of business fraud with cases of a different category altogether, like bodily injury, the relative importance of the category to which the wrongdoing belongs became relevant as well. Since bodily injury was regarded as generally worse than fraud, this could make the same case of fraud seem less worthy of severe punishment.

ALEX VOORHOEVE: *But these cases seem to* vindicate *rather than undermine the within-subject approach used by philosophers. For in these cases, the within-subject approach of having a single person consider multiple cases side by side seems to prevent mistakes that would occur if we considered each case separately.*

DANIEL KAHNEMAN: Well, to some extent I agree, and in that paper with Sunstein, Schkade, and Ritov, we indeed try to argue that the within-subject comparison has an advantage because it gives you a better shot at approaching a consistent view of the world than judging single cases in isolation would, even though it doesn't always work. But when we are thinking about policies and about applying moral rules to people, then it's important to remember that these people will experience cases as participants in the between-subject experiment. What is morally compelling to us as we consider the contrast between the two cases may not seem morally relevant to them at all, and so the

rules we may wish to impose on them may not seem relevant to them. There is, I think, a genuine dilemma here between the demands of coherence and the need not to impose on people principles that violate their judgements. A solution that we arrived at in that paper is to partially accommodate both demands. So our recommendation was to measure outrage at the wrongful behaviour, considered in isolation, and use outrage as an input in determining the severity of punishment. We would not recommend allowing outrage alone to determine punishment, as it appears to do in some jury trials. Our thinking was that while a public policy that merely reflects outrage is grotesque, a policy that is insensitive to outrage is not going to be acceptable to people. So public policy should be sensitive to outrage, but not dominated by it. That is the best we could do. It is not a unique solution, but it acknowledges that moral intuition about a specific case can neither be trusted nor altogether ignored.

ALEX VOORHOEVE: *This proposed solution seems to treat people's emotions and judgements as simply given, when they are not. As you said, moral emotions like outrage are judgement-sensitive; how outraged we are depends on how awful we consider the wrongdoing to be. So if people's sense of outrage is based on a mistaken judgement of the turpitude of the crime, then that judgement needs to be corrected rather than accommodated.*

DANIEL KAHNEMAN: I think that is an excellent point, and I think that intuitions are indeed malleable to some extent and in some cases can change through education. Here's a nice example. We did a survey once in Canada where we found that people are truly offended by the fact that car accident insurance rates are determined by where they live, so that if they live in an area where there are many accidents, they pay a high insurance rate. People initially thought that was deeply immoral, because they felt that insurance rates should be determined by the driver's behaviour alone. But this is an intuition you could train people to recognize as mistaken. For they recognized that the insurance rate it is proper for people to pay is proportional to the probability of their being in an accident, even if they don't cause

it, and they saw that the area they lived in affected the probability that they would be in an accident. But not all intuitions are malleable in this way.

ALEX VOORHOEVE: *Can you give an example of a case in which you think our intuitions are not responsive to reflection?*

DANIEL KAHNEMAN: Take moral philosophers' favourite pair of so-called 'trolley cases'.[2] The case of the fat man on the bridge is an example where I think it is clear that people's intuitive judgements follow the rule that using direct physical violence against an innocent person is unacceptable, and the more directly physical the violence is, the more unacceptable it is. Now, on reflection, this looks like a poor rule—it doesn't seem to pick out a morally relevant factor. But applied to individual cases it is going to be powerful every time. You (or at least I) could blame an individual who did not divert the trolley, but I cannot imagine blaming anyone for not throwing the fat man in front of the trolley.

ALEX VOORHOEVE: *I agree that in the form in which the case of the fat man is traditionally stated, the harmful direct physical contact may play some role in generating people's intuitive judgement that pushing him is wrong. But my guess would be that the judgement that it is wrong to use him to stop the trolley would persist if the case was described as one in which one had to press a lever which opened a trapdoor and dropped him into the path of the trolley. So our judgement in the two contrasting trolley*

[2] In the Side-track Case, a runaway trolley is headed towards and will kill five people unless a lever is pulled which will divert it onto a dead-end side-track where it will kill one person. Most people intuitively believe it is permissible—and some believe it is obligatory—to divert the trolley in this case. In the contrasting Bridge Case, the trolley will kill the five unless you push a large bystander, whose weight is sufficient to stop the trolley, off a bridge and into its path, thereby killing him. Most people intuitively believe it is impermissible to push the large man in this case. This pair of judgements stands in need of explanation, since both diverting the trolley and pushing the bystander involve taking an action that will kill one and save five.

cases appears to be better explained by the following general principle: while it may be permissible to pursue the greater good in ways that have lesser harm as a side-effect (as one would do if one diverted the trolley onto a side-track where it kills the one), it is impermissible to use someone harmfully as a means to a greater good (as one would do if one used the fat man to stop the trolley). Something like this principle would seem to make more sense than the 'Don't use direct physical violence against innocents' rule you cited, because it seems to incorporate a significant moral idea, which is, roughly, that innocent people are not to be used harmfully as a means to others' ends without their consent.

DANIEL KAHNEMAN: I am sceptical about the potential of this idea to explain and justify this intuitive judgement. The aversion to pushing the fat man is linked to intuitions about *causes*—you feel you have a more direct causal role in the man's death if you push him, or push a lever that moves him toward the train, than if you move a lever that moves the train toward him. When you call the killing of the single person in the latter case an unintended side-effect, you draw on an intuition about causality that I do not find morally compelling. So I find it hard to believe that the two cases differ in morally relevant ways. However, since the fat man scenario evokes an extraordinarily powerful intuition, you should not have a rule that ignores it. That is, if anyone had a system that would condone pushing the bystander to save the five, then that system would not be viable, that system would not be acceptable. Because pushing him is just deeply repugnant. When I say this, I am not stating a piece of moral knowledge; I am simply making a sociological and psychological prediction.

ALEX VOORHOEVE: *Well, what would 'work' in practice—in the sense of finding a rule that people would accept and act on—is a pragmatic question. But I am interested in whether it* really *is morally permissible to use the fat man to save the five, whether or not I can get others to agree to it. You seem to think there is no justification for your intuitive judgement that it is impermissible to use him harmfully as a means in this case. You also seem to think that rationally, the case where we use him*

to stop the trolley is just like the case where we divert the trolley onto the side-track, because both involve taking an action that will kill one in order to save five. If you also think that you should divert the trolley because this brings about the greater good, then I think you should simply revise your judgement that it is impermissible to push the fat man. Should you ever be confronted with a case like that, then you should tell yourself, 'I should overcome my irrational repugnance, and push him.'

DANIEL KAHNEMAN: I find that unconvincing. My intuitions about abstract theories like consequentialism that tell you always to pursue the greater good are just weaker than my intuitive judgements about cases. To me, consequentialism is just a story, and a pretty good story at that, even though it has some holes in it, but it doesn't have the power of our intuitions about particularly compelling cases.

ALEX VOORHOEVE: *You believe, however, that we can sometimes be motivated to abandon our initial intuitions if we come to believe they are generated by morally irrelevant factors, as in the grocery store cases, or if we realize that they result from insufficient consideration of relevant information, as in the business fraud versus bodily harm cases and the accident insurance cases. So you believe we have the ability to revise our intuitive case judgements in the process of searching for reflective equilibrium, at least in certain cases. Why don't you think that can happen in this case?*

DANIEL KAHNEMAN: I don't dispute that reflective equilibrium is something we must strive to achieve. But we start the search for reflective equilibrium from a position in which our basic intuitions appear inconsistent. That is to say, we start out with strong, basic intuitions, which are accompanied by powerful emotional reactions, and which strike us as self-evidently correct, so that we think that we can generalize fairly naturally from them. However, we find that when we generalize from our intuitions we hit a point where other intuitions, naturally generalized, will lead us to contrary conclusions. This is true in many areas, including judgements of probability and

non-moral decision-making. So what we do then is construct systems that ultimately suppress some of these contradictory intuitions and build mostly on others. My expectation is that there is no uniquely reasonable way to do this, and that rhetoric, reigning cultural habits, and so on are going to play very important roles in determining how we do this. Part of the rhetorical success of a system stems from its ability to sustain the conviction that comes from some of the basic intuitions and to generalize from those intuitions whilst drawing people's attention away from conflicting intuitions. For example, the motivational force of the consequentialist story depends heavily on the rhetoric you use to back it up and on the ability to anchor yourself in some intuitions and draw on the emotional and motivational powers of those intuitions, while keeping your mind away from other intuitions. And some consequentialists, like Peter Singer, are very good at this. But it is clear to me that you can start with one set of intuitions or with another set, and that depending on where you start and on the rhetoric you employ, you are going to end up in a different place. In some sense, this makes the enterprise unending, because there is no unique solution. So, given that we have powerful but profoundly inconsistent intuitions, and given that there appears to be no uniquely compelling way to resolve the inconsistencies, I am sceptical about the hope of achieving a unique and fully satisfactory reflective equilibrium.

ALEX VOORHOEVE: *Are you saying that if we are clear-headed enough to see that there is no uniquely compelling and coherent system of moral thought, we will not be sufficiently confident about the rightness of our more general principles to override strong intuitions that are inconsistent with those general principles? Is that why you could not see yourself as ready to revise your intuitive judgement that pushing the fat man is impermissible in the light of a general principle that it is always better to act in a way that kills one and saves five?*

DANIEL KAHNEMAN: Obviously, I am quite aware that my position is not internally coherent. But I have a reason not to try to achieve

coherence: I believe coherence is impossible. Some moral quandaries evoke intuitions that have the form of a sensible rule—the case where you turn the trolley onto a side-track is an instance. Other cases evoke intuitions that do not have the form of a sensible rule but that are easily abandoned when they conflict with sensible rules or with other intuitions. (The grocery store cases are of this kind.) Cases of a third kind, however, evoke powerful intuitions that do not suggest a sensible rule and that nevertheless do not yield to conflicting rules or intuitions. The fat man scenario is a prime example. I do not believe a sensible rule is generating the intuition that one shouldn't push the fat man, nor do I believe one could be invented to justify it; nonetheless, I find pushing the fat man deeply repugnant. Because of this third class of cases, our basic intuitions are likely to contain contradictions that cannot be resolved. I believe that the search for coherence is admirable, and that it should be diligently pursued. But I also believe it is important to remember that it will inevitably fail.

References and further reading

For an overview of Daniel Kahneman's work, see his Nobel Prize Lecture, 'A Perspective on Judgement and Choice: Mapping Bounded Rationality', *American Psychologist* 58 (2003): 697–720. His autobiography is published online at <http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahnemanautobio.html>. The quotation by Stephen Jay Gould is from *Bully for Brontosaurus: Reflections in Natural History* (New York: Norton, 1991): 469. The brain imaging research mentioned is reported in Benedetto De Martino, Dharshan Kumaran, Ben Seymour, and Ray Dolan, 'Frames, Biases and Rational Decision-making in the Human Brain', *Science* 313 (2006): 684–7.

The method of reflective equilibrium in moral theory was first proposed in John Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 2nd edn 1999), sect. 4. See also Norman Daniels, 'Reflective Equilibrium', in *The Stanford Encyclopaedia of Philosophy Online* (2003).

The research on poignancy as a determinant of compensation is reported in Dale Miller and Cathy McFarland, 'Counterfactual Thinking and Victim Compensation: A Test of Norm Theory', *Personality and Social Psychology Bulletin* 12 (1986): 513–19. For a discussion of the ways in which we construct erroneous *post hoc* justifications for our intuitive judgements, see Jonathan Haidt, 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgement', *Psychological Review* 108 (2001): 814–34. The study on intuitions about punishment to which Kahneman refers is Cass Sunstein, Daniel Kahneman, David Schkade, and Ilana Ritov, 'Predictably Incoherent Judgements', *Stanford Law Review* 54 (2001/2): 1153–215. For other discussions of moral heuristics see Daniel Kahneman, Jack Knetsch, and Richard Thaler, 'Fairness as a Constraint on Profit-seeking: Entitlements in the Market', *American Economic Review* 76 (1986): 728–41, Kahneman's Tanner Lecture on Human Values, 'The Cognitive Psychology of Consequences and Moral Intuition' (unpubl. ms, 1994), Tamara Horowitz, 'Philosophical Intuitions and Psychological Theory', *Ethics* 108 (1998): 367–85, Frances Kamm, 'Moral Intuitions, Cognitive Psychology, and the Harming/Not-Aiding Distinction', in *Intricate Ethics* (Oxford: Oxford University Press, 2006): 422–49, and Cass Sunstein, 'Moral Heuristics', *Behavioural and Brain Sciences* 28 (2005): 531–73. The ways that people respond to trolley cases are reported in Marc Hauser, *Moral Minds* (New York: HarperCollins, 2006): ch. 3.