

Going to Extremes

How Like Minds Unite and Divide

Cass R. Sunstein

OXFORD
UNIVERSITY PRESS

2009

judgments were far more negative than the average of individual judgments.²⁵ In many cases, group members decided that the behavior was really very unfair, even though individuals believed that the behavior was only mildly unfair. Interestingly, the groups' conclusions were typically more extreme than were people's individual judgments after deliberation. But such judgments were nonetheless more negative, and thus more extreme, than predeliberation individual judgments.

These findings are remarkably similar to those involving juror outrage, where, as we have seen, groups are more outraged than their median member. We now have a strong clue about the sources of protest movements, a topic that I explore in due course. For the moment, let us try to explain group polarization.

CHAPTER 2

Extremism

Why and When

In this chapter, my major goal is to answer two questions: Why do like-minded people go to extremes? And when do they do so? As we shall see, the answers to those questions bear on an exceedingly wide range of social puzzles, including the immense power of authorities, the nature of "evil," the idea of groupthink, and social cascades, by which large groups of people move in new directions in terms of their investments, their political choices, and even their religious convictions.

The most important reason for group polarization, and a key to extremism in all its forms, involves the exchange of new information. Group polarization often occurs because people are telling one another what they know, and what they know is skewed in a predictable direction. When they listen to each other, they move.

NEW INFORMATION

Suppose that you are in a group of people whose members tend to think that Israel is the real aggressor in the Mideast conflict, that eating beef is unhealthy, or that same-sex unions are a good idea. In such a group, you will hear many arguments to that effect. Because of the initial distribution of views, you will hear relatively fewer opposing views. It is highly likely that you will have heard some, but not all, of the arguments that emerge from the discussion. After you have heard all of what is said, you will probably shift further in the direction of thinking that Israel is the real aggressor, opposing eating beef, and favoring civil unions. And even if you do not shift—even if you are impervious to what others think—most group members will probably be affected.

When groups move, they do so in large part because of the impact of information.¹ Happily, people tend to respond to the arguments made by other people—and the pool of arguments, in a group with a predisposition in a particular direction, will inevitably be skewed in the direction of the original predisposition.

Certainly this can happen in a group whose members tend to support aggressive government regulation to combat climate change. Group members will hear a number of arguments in favor of aggressive government regulation and fewer arguments the other way. If people are listening, they will have a stronger conviction, in the same direction from which they began, as a result of deliberation. If people are worried about climate change, the arguments they offer will incline them toward greater worry. If people start with the belief that climate change is a hoax and a myth, their discussions will amplify and intensify that belief. And indeed, a form of “environmental tribalism” is an important

part of modern political life. Some groups are indifferent to environmental problems that greatly concern and even terrify others. The key reason is the information to which group members are exposed. If you hear that genetically modified food poses serious risks, and if that view is widespread in your community, you might end up frightened. If you hear nothing about the risks associated with genetically modified food, except perhaps that some zealots are frightened, you will probably ridicule their fear. And when groups move in dangerous directions—toward killing and destruction—it is usually because the flow of information supports that movement.

CORROBORATION

Those who lack confidence and who are unsure what they should think tend to moderate their views.² Suppose that you are asked what you think about some question on which you lack information. You are likely to avoid extremes. It is for this reason that cautious people, not knowing what to do, tend to choose some midpoint between the extremes.³ But if other people seem to share their views, people become more confident that they are correct. As a result, they will probably move in a more extreme direction.

In a wide variety of experimental contexts, people's opinions have been shown to become more extreme simply because their initial views have been corroborated and because they have been more confident after learning of the shared views of others.⁴ Suppose that other people share your view that the United States is not to be trusted, that the attacks of 9/11 were staged, or that Iran poses a serious threat to the rest of the world. If so, your own view will be more deeply felt after you hear what they have to say. Note

that there is an obvious connection between this explanation and the finding that Republican appointees on a panel of three Republican appointees are likely to be more extreme than Republican appointees on a panel with only two such judges. The existence of unanimous confirmation, from two others, will strengthen confidence—and hence strengthen extremity.⁵

What is especially noteworthy is that this process—of increased confidence and increased extremism—is often occurring simultaneously for all participants. Suppose that a group of four people is inclined to distrust the intentions of the United States with respect to foreign aid. Seeing her tentative view confirmed by three others, each member is likely to feel vindicated, to hold her view more confidently, and to move in a more extreme direction. At the same time, the very same internal movements are also occurring in *other* people (from corroboration to more confidence, and from more confidence to more extremism). But those movements will not be highly visible to each participant. It will simply appear as if others “really” hold their views without hesitation. As a result, our little group might conclude, after a day’s discussion, that the intentions of the United States, with respect to foreign aid, cannot be trusted at all.

We have a clue here about the great importance of social networks, on the Internet and in ordinary life, in creating movements of various sorts. Social networks can operate as polarization machines because they help to confirm and thus amplify people’s antecedent views. Those who are inclined to support a cause or a candidate may become quite excited if support is widespread on their social network. In 2008, President Barack Obama greatly benefited from this process, in a way that created extreme enthusiasm for his candidacy. Some of this was planned; his campaign self-consciously promoted social networks that spread favorable information.

But some of this was spontaneous. Obama supporters, especially young people, worked hard on their own to take advantage of existing networks and create new ones that would turn curiosity and tentative support into intense enthusiasm and active involvement.

A very different example is provided by Islamic terrorism, which is also fueled by spontaneous social networks, in which like-minded people discuss grievances with potentially violent results.⁶ Terrorism specialist Marc Sageman explains that at certain stages, “the interactivity among a ‘bunch of guys’ acted as an echo chamber, which progressively radicalized them collectively to the point where they were ready to collectively join a terrorist organization. Now the same process is taking place online.”⁷ The major force here is not Web sites, which people read passively; it consists of Listserve’s, blogs, and discussion forums, “which are crucial in the process of radicalization.”⁸ As we shall see in more detail, Islamic terrorism is a product, in significant part, of group polarization.

These are examples from the political domain, but there are plenty of other illustrations. Why are some cars popular in some areas, but not at all popular in others? Why are some foods enjoyed, or thought to be especially healthy, in some places, whereas the same foods are disliked, or thought to be unhealthy, in other places? Joseph Heinrich and his coauthors note that “[m]any Germans believe that drinking water after eating cherries is deadly; they also believe that putting ice in soft drinks is unhealthy. The English, however, rather enjoy a cold drink of water after some cherries; and Americans love icy refreshments.”⁹ Why is the same music liked, or hated, among groups of teenagers? Here, too, corroboration greatly matters.

A less innocuous example: In some nations, strong majorities believe that Arab terrorists were not responsible

for the attacks of September 11, 2001. According to the Pew Research Institute, 93 percent of Americans believe that Arab terrorists destroyed the World Trade Center, whereas only 11 percent of Kuwaitis believe that Arab terrorists destroyed the World Trade Center.¹⁰ With respect to daily life, a great deal of what we believe, like, and dislike is influenced by the exchange of information and by corroboration.

REPUTATION

A few years ago, I was discussing group polarization with a philosopher who works on the topic of animal rights and animal welfare. He is strongly committed to reducing the suffering of animals, and he told me the following story: "On Friday of a three-day conference, we are perfectly sensible, by my lights. But by Sunday, we stop thinking straight! We become much too extreme. By Sunday, people start saying that no experiment on animals ever produced useful knowledge for human beings. By Sunday, people start saying that it is never acceptable to eat meat, even if animals lived a very long and very happy life, and died of natural causes. Some of us have, in a way, lost our minds." The philosopher told me that this change in view—a form of polarization—was not adequately explained by the exchange of new information or by increased confidence.

What he had in mind was a third explanation, involving social comparison. That explanation begins with the claim that people want to be perceived favorably by other group members, and also to perceive themselves favorably. Sometimes our views are, to a greater or lesser extent, a function of how we want to present ourselves. Of course, some people are more concerned than others with their self-presentation. But once we hear what others believe, some of

us will adjust our positions at least slightly in the direction of the dominant position, to hold onto our preserved self-presentation. We might contain our opposition; we might voice somewhat more enthusiasm for the majority view than we really feel.

Some people might want to show, for example, that they are not timid or cautious, especially in an entrepreneurial group that disparages these characteristics and favors boldness and risk-taking. In business, people often want to seem to be risk takers. In such a group, people will frame their position so that they do not appear timid or cautious by comparison with other group members. And when they hear what other people think, they might find that they occupy a somewhat different position, in relation to the group, from what they hoped. They will shift accordingly.¹¹ This might be because they want others to see them in a certain way. Or it might be because they want to see themselves a certain way, and a shift is necessary so that they can see themselves in the most attractive light.

Suppose, for example, that group members believe that they are somewhat more opposed to capital punishment than are most people. Such people might shift a bit after finding themselves in a group of people who are strongly opposed to capital punishment, simply to maintain their preferred self-presentation. Does the example seem unrealistic? Consider the otherwise inexplicably extreme behavior of many Republicans and many Democrats in the debate over the Bush-Gore presidential vote in Florida in 2000. Reasonable people could differ at the time. Each side had something to say. But many members of both parties, talking and listening mostly to one another, suggested that the other party was trying to "steal the election." This is one example of what happens in nearly all presidential elections. In 2008, for example, many supporters of Senator John McCain ended

up in unfounded and indefensible positions, urging and apparently thinking that President Obama “palled around with terrorists” and might even be disloyal to the country.

The phenomenon occurs in many contexts. People might wish not to seem too enthusiastic about, or too restrained in their enthusiasm for, affirmative action, feminism, or an increase in national defense; hence their views shift when they see what other group members think. The result is to press the group’s position toward one or another extreme, and also to induce shifts in individual members.

To understand the importance of social comparison, consider the important finding that low-status members of groups become ever more reluctant, over the course of discussion, to repeat privately held information,¹² that is, information that they hold but that others do not. Those in the group who are inexperienced, or are thought to be low on the hierarchy, are particularly loath to emphasize their privately held information as discussion proceeds. Suppose that the leaders of a religious organization are suspected of wrongdoing. How many people, low on the totem pole, will hold them to account?

The empirical findings suggest that group members, and especially lower status ones, are nervous about emphasizing information that most group members lack. Indeed, lower status members will often drop uniquely held information very rapidly—partly because of the difficulty of establishing its credibility and relevance, and partly because they risk the group’s disapproval if they press a line of argument that others reject. In many deliberating groups, people who emphasize uniquely held information take an obvious social risk, and they know it. Note in this regard that group members typically *underestimate* the performance of low-status members and typically *overestimate* the performance of high-status members, in a way that gives high-status

members a degree of deference that is not warranted by reality.¹³

In the same vein, those who discuss shared information obtain rewards in the form of an enhanced sense of competence in the eyes of others—and in their own eyes as well.¹⁴ Important but perhaps obvious: If someone tells you something you already know, you are likely to like that person a little bit better as a result. Important and less obvious: If someone tells you something you already know, you are likely to like *yourself* a bit better as a result! In face-to-face discussions and in purely written tasks, people give higher ratings (in terms of knowledge, competence, and credibility) both to themselves and to others after receiving information that they knew already. The general problem is that deliberating groups often move to unjustified extremes because they fail to elicit information that could steer them in the right directions.

A political example: In the presidency of George W. Bush, many failures occurred because of an unfortunate culture that encouraged, rather than combated, group polarization.¹⁵ In the words of Scott McClellan,

Bush’s way of managing the problems in Iraq was proving inequivalent to the task . . . [H]e was insulated from the reality of events on the ground and consequently began falling into the trap of believing his own spin. He failed to spend enough time seeking independent input from a broad range of outside experts, those beyond the White House bubble who had firsthand experience on the ground in Iraq, and—perhaps most important—those with differing points of view, including those who disagreed with his position.¹⁶

By contrast, Lincoln’s presidency has been described as a healthy Team of Rivals,¹⁷ in which Lincoln self-consciously chose diverse people who could challenge his inclinations

and test one another's arguments in the interest of producing the most sensible judgments. Unfortunately and even tragically, the Bush administration turned into a Team of Unrivals, in which internal diversity and dissent were squelched as disloyal. With respect to the Iraq war, tax policy, regulation, and spending, group polarization operated in full force, and the administration's leaders took no steps to combat it. Reputational pressures, of a particularly acute form, ensured extremism, confidence, and uniformity.

Genocide offers an especially grim example of this phenomenon. How can apparently ordinary people turn into killers? Information plays a major role. When people are informed that killing is right or even necessary, they might be willing to kill. In the words of a participant in the genocide in Rwanda: "When you have been prepared the right way by the radios and the official advice, you obey more easily, even if the order is to kill your neighbors."¹⁸ But as another killer suggested, reputational pressures produce killing as well: "If you proved too green with the machete, you could find yourself deprived of rewards, to nudge you in the right direction. If you got laughed at one day, you did not take long to shape up. If you went home empty-handed, you might even be scolded by your wife or your children."¹⁹

TWO FUNCTIONS OF POLARIZATION

We should distinguish between two different accounts of group polarization. One account suggests that polarization reveals hidden beliefs and desires. A very different account insists that polarization creates new beliefs and desires.

On the first account, people often have a suppressed but deep-seated set of concerns. These concerns do not ordinarily materialize in social life; they usually remain unspoken.

The concerns are not unthought, but they are, in extreme cases, unthinkable, in the sense that they really cannot be voiced in public without creating serious risks of social disapproval or even ostracism.²⁰ Now imagine that group members speak with one another, and those suppressed concerns come to the surface. As people exchange tales and reactions, the unthinkable comes out into the open. One result is more extremism, as people feel outrage about practices that used to produce self-silencing.²¹ Consider the context of disability, where this is a plausible account. Among disabled people, the objections to the status quo are there, but they are sometimes buried, and discussion brings them out.

Compare the issue of sex equality. The whole idea of consciousness-raising is designed to signal the existence of repressed angers and objections; once people speak with one another, consciousness is raised in the sense that those angers and objections come to the surface. What was once suppressed, perhaps on the ground that powerful people would object, is now voiced; people articulate their concerns as a result of group discussions. What was once unthinkable is now in the public domain. On this view, deliberation can create a kind of self-discovery, in which the authentic inner voice becomes articulate.²² Here deliberation *reveals* something that unquestionably existed before.

The area of sexual harassment is a particularly revealing example. Women did not exactly like being harassed, but before the practice was unlawful, or even had a name, their anger was muted. Once women spoke to one another in an open way, and in the midst of the emerging women's movement, a silenced group was ready to speak out. Attacking sexual harassment was once, in a sense, unthinkable. Even the phrase did not exist. In many places, defending sexual harassment is now unthinkable (even if

significant numbers of men may not believe that it is quite so bad).

On the competing account, group polarization can occur even if there is no initial sense of grievance; little or perhaps nothing was suppressed. Instead social influences, involving the efforts of *polarization entrepreneurs*, give rise to intensifying objections and growing protest. For many group members, the views that end up being extreme are entirely *generated* by group interactions. People may not have a deep-seated belief that climate change is occurring or that some apparent opponent is bad or corrupt or badly motivated. But as they speak with one another, their inclination to accept that belief is intensified. Here deliberation creates, for some or many, a series of objections that had previously been absent.

We can imagine this phenomenon in the political domain, as people develop an initial concern with some practice or person, and that initial concern intensifies as a result of internal discussions. For some disabled people and some women, this competing account undoubtedly captures reality. In the important domain of ethnic identification, we will encounter some important examples. Religious beliefs and practices often arise and intensify in exactly this way.

From the point of view of those who are subject to it, group polarization is often entirely rational. You are in a group of people, discussing climate change or same-sex marriage. You hear a set of arguments. Your initial inclinations are confirmed. You like the other group members, and you want them to like you. In these circumstances, increased extremism, on your part, may be a perfectly rational reaction to what you learn and to what you care about. This sunny picture of polarization—from the standpoint of those who fall prey to it—undoubtedly captures

much of reality.²³ When people shift in groups, it is often for perfectly sensible reasons. The point applies broadly and in some settings where sense isn't exactly the currency of the realm; it suggests that political extremists and conspiracy theorists may well be responding rationally to what they hear and learn.²⁴

But there are two major wrinkles. The first is that to the extent that people are motivated by a concern for their reputation, they might not be moving because of information and good arguments. If people who believe in animal rights are shifting not because of what they hear, but because of how they want to seem, their shift might make little sense on the merits. The second and subtler wrinkle is that much of the time, people do not seem to have anything like an adequate sense of the partiality and skew of the groups in which they find themselves. If you are in a group of people who lean to the left or to the right, you should adjust your reactions to what they say, simply because of the inclinations of those in the group. If people in your own company are especially optimistic about a certain course of action and dismissive about the plans of a competitor, you might want to take into account the likely biases that surround you. If those who surround the president or the governor seem to think that a certain economic plan is terrific, the president or the governor ought to consider the possibility that the group's members start out in favor of the plan, and are hardly a representative sample of expert opinion.

I suspect, in fact, that group polarization often occurs because of people's failure to adjust their reactions to the skewed compositions of the groups in which they find themselves.²⁵ We act as if those groups reflect an impartial sum of information, even when there is a systematic bias. This tendency can get us into a lot of trouble in many areas, warping our judgments not only about politics but also

about health, money, and religion. Indeed, financial crises often stem from processes of this kind, as groups with their own biases create speculative bubbles in (say) real estate and Internet stocks—and then produce panics.

Recall the importance of distinguishing between two different kinds of polarization: planned and spontaneous. As we have seen, some people act as polarization entrepreneurs: They attempt to create communities of like-minded people, and they are aware that these communities will not only harden positions but also move them to a more extreme point. But sometimes polarization arises spontaneously, through entirely voluntary choices, without the slightest kind of planning. Consider, for example, people's reading patterns, which suggest an unmistakable form of self-sorting into liberal and conservative networks.²⁶ Or consider the blogosphere itself, which shows a similar kind of spontaneous sorting and polarization.²⁷ Or consider simple geographical choices; like-minded people, in essential agreement on political issues, may end up living in the same area simply because that is what they want to do.²⁸ We shall encounter many examples of both planned and spontaneous polarization.

"RHETORICAL ADVANTAGE" AND SKEWED DEBATES

A Mysterious Finding

In the context of punitive damage awards by juries, a particular finding deserves special attention. Recall that jurors were asked to record their dollar judgments in advance of deliberation and then to deliberate together to produce dollar verdicts. The principal effect was to make nearly *all* awards go up, in the sense that the jury's dollar

award typically exceeded the median award of individual jurors.²⁹ There is a further point. The effect of deliberation in increasing dollar awards was most pronounced in the case of high awards. For example, the median *individual* judgment, in the case involving the defective yacht, was \$450,000, whereas the median *jury* judgment, in that same case, was \$1 million. But awards shifted upward for low awards as well.

Here is the mystery: Why did all awards go up? Why didn't the low ones, at least, go down? A tempting explanation, consistent with group polarization, is that any positive median award suggests a pre-deliberation tendency to punish—and as usual, deliberation aggravates that tendency by increasing awards. But even if this explanation is correct, it does not seem nearly specific enough. The striking fact is that those arguing for higher awards seem to have an automatic *rhetorical advantage* over those arguing for lower awards. The intriguing possibility is that in many domains, one point of view has such a rhetorical advantage over other points of view, with predictable results for both thought and behavior.

Daniel Kahneman, David Schkade, and I conducted a subsequent study that supported our speculation, at least for punitive damage awards. We asked a large group of University of Chicago law students to participate in an odd little experiment. We told them that they were to assume that they were sitting on a jury that was deadlocked on the question of appropriate punishment, with some people supporting a greater award and others supporting a lower award. We asked half of the students to devise arguments that would support a higher award *assuming that they knew nothing about the particular case*. Believe it or not, the law students produced a number of such arguments. For example, they stressed the need to deter this particular wrongdoer, the

need to deter other potential wrongdoers, and the importance of ensuring that an injured party received more money.

We asked the other half of the students to devise arguments that would support a lower award—again assuming that they knew nothing about the particular case. Here, too, the law students produced a number of such arguments. They stressed the risk that a large award would stop companies from engaging in beneficial activity, the danger that a big award might go mostly to lawyers, and the fact that the injured person should not get a windfall benefit. Then we asked both groups whether it was easier to argue for a higher award or a lower one.

The answer was clear: Most people find it easier, just in the abstract, to defend higher punitive awards against corporations than to defend lower awards.³⁰ Those defending the higher awards have an automatic rhetorical advantage. Even when people know absolutely nothing about the facts of individual cases, they are able to generate appealing arguments in favor of higher awards. It is much harder to produce plausible-sounding arguments in favor of lower awards. Those seeking higher penalties have a built-in advantage.

Doctors, Altruists, and Others

Rhetorical advantages have been found in seemingly distant areas. Suppose that a group of doctors is deciding what steps to take to resuscitate apparently doomed patients. Are individual doctors less likely, or more likely, to support heroic efforts than teams of doctors?

The evidence suggests that as individuals, doctors are less likely to support heroic efforts than teams. The apparent reason is that in cases of conflict, those who favor such efforts have a rhetorical advantage over those who do

not.³¹ Doctors do not want to seem, to one another, to be willing to give up on a patient and condemn him to death, even when the chance of success is low. For many doctors operating in groups, giving up suggests an indifference to the sanctity of human life, a lack of a strong commitment, perhaps even a lack of confidence in one's own competence. Hence teams of doctors are willing to do more to save people than are individual doctors. In a sense, medical teams turn out to be more extreme. (Patients and family members, take note.)

Individuals behave very differently from teams in the Dictator Game, an experiment used by social scientists to study selfishness and altruism.³² In this game, a subject is told that she can allocate a sum of money, say \$10, between herself and some stranger. What will the subject do? The standard economic prediction is that most subjects will keep all or almost all of the money for themselves; why should we share money with complete strangers? But the standard prediction turns out to be wrong. Most people choose to keep somewhere between \$6 and \$8 and to share the rest.³³

My question here, however, is not individual behavior but how behavior in the Dictator Game is affected if people are placed in teams—if people decide in groups rather than as individuals. Are groups more altruistic than individuals? The answer is that team members choose still more equal divisions.³⁴ Once placed in groups, people show a significant shift toward greater generosity.

Why is this? A good answer points to a rhetorical advantage, one that disfavors selfishness even within a group that stands to benefit from it. If you are on a team of people deciding how selfish to be, you might well be less selfish than you would be on your own—just because you do not want to appear to be particularly selfish. Imagine, for example, that you are deliberating with a group of people about how

much money to give to charity. Chances are good that the group will end up being more charitable than the median individual, simply because people do not want to appear to be greedy. People's concern for their reputation plays a large role. People's self-conception also matters: Who wants to feel like a greedy person?

Of course, the outcomes here would change if the team in the Dictator Game had some reason to be hostile to those who would benefit from their generosity. We can easily imagine a variation of the Dictator Game in which, for example, people of a relatively poor religious group are deciding how much to allocate to another religious group that is thought to be both hostile and far wealthier. In this variation, the rhetorical advantage would favor greater selfishness.

Rhetorical Advantage Why? Rhetorical Advantage When?

All this leaves some important questions unanswered: What produces a rhetorical advantage? When will we see one? How can we know in what direction the advantage will go?

The simplest answer points to the particular norms that prevail within the group, and norms, of course, vary across time and place. Among most Americans, current norms make it easy to argue for high penalties against corporations for serious misconduct. But we can easily imagine subcommunities within America (corporate headquarters?) in which the rhetorical advantage runs exactly the other way. In such groups, the level of punishment might be expected to decrease, not to increase, as a result of social interactions. And of course, social norms and reputational influences are closely entangled. Given existing norms, most juries know

that they are likely to seem odd if they want to impose little punishment for really bad corporate misconduct.

In any case, it is easy to envisage many other contexts in which one or another side has an automatic rhetorical advantage. Consider debates over penalties for drug dealers and over changing tax rates. In contemporary American political debates, those favoring higher penalties and lower taxes have a strong upper hand. If one group is arguing for maintaining the current tax rates and another for increasing them, the second will have a real uphill battle. And if some people are arguing for lower penalties for criminal offenses, they had better have some unusually strong arguments. Or imagine discussion within a firm about whether to run a risk, or within a family about whether to take some precautions against a threat that family members face from, say, crime in the neighborhood, a bad economy, or a car that isn't particularly safe. Such a firm might well end up taking the risk, just because those who favor taking the risk have a rhetorical advantage, and for the same reason, such a family might be inclined to take precautions.

Of course, there are limits on the effects of rhetorical advantages. No reasonable person wants taxes to disappear or to impose life sentences for minor drug crimes. But when a rhetorical advantage is involved, group deliberation will produce significant changes in individual judgments. Undoubtedly legislative behavior—involving national security, tax policy, and criminal punishment—is affected by rhetorical advantages. Many movements within judicial panels can be explained in similar terms. True, the governing norms vary from one nation to another. In the United Kingdom and Germany, for example, it is much easier to argue for tax increases than in the United States, especially on polluting behavior; no strong rhetorical advantage is enjoyed by those opposing taxes. In some parts of the world, those

resisting restrictions on abortion have a rhetorical advantage; in other parts of the world, they are at a severe disadvantage.

Are rhetorical advantages unhelpful or damaging? In the abstract, this question cannot be answered sensibly. Shifts, including extreme movements, must be evaluated on their merits. Perhaps the higher punitive awards that follow deliberation are simply better. So, too, perhaps, are the movements by doctors toward taking more heroic measures, and by groups deciding to divide funds more equally. The only point is that such advantages exist and that they help to explain social movements, including extreme ones. It would be a surprising stroke of luck if such movements were always benign. When groups become violent, for example, it is often because a rhetorical advantage favors those who press toward more severe responses to real or imagined grievances.

MORE EXTREMISM, LESS EXTREMISM

Group polarization is not a social constant. It can be increased or decreased, and even eliminated, by certain features of group members or their situation.

Extremists Move Most

Recall that in the study of protests, people who started out at a more extreme point showed the greatest shift as a result of group discussion. The point is quite general: Extremists are especially prone to polarization. When people start out at an extreme point and are placed in a group of like-minded people, they are likely to go especially far in the direction toward which they started.³⁵ There is a lesson here about the sources of terrorism and political violence in general. And because there is a link between confidence and extremism,

the confidence of particular members also plays an important role; confident people are more prone to polarization.³⁶

Recall that people moderate their opinions if they are unsure whether they are right. And other things being equal, confident people have an advantage in social deliberations. It follows that if group members tend toward extremism, and if the group is dominated by confident people, it is exceedingly likely to shift. In a brilliant essay, Russell Hardin writes that extremists suffer from a *crippled epistemology*.³⁷ He argues that extremists are often far from irrational. The problem is that they know very little, and what they know supports their extremism. No one doubts that some extremists know a great deal; sometimes extremism is defensible or even right. (The American revolutionaries were extremists; so were Martin Luther King Jr. and Nelson Mandela.) But when groups make unjustified extreme movements—in the direction, for example, of terrorism or genocide—a crippled epistemology is often the reason. Those who start out in an extreme position will be all the more subject to the influences discussed here.

The general point—that extremists are especially prone to significant further shifts—is not limited to the most obvious extremists. The point certainly applies in the business world. Members of a corporate board, inclined to take unusual risks, fall in the same category; the Enron disaster occurred in part as a result of group polarization. The same processes occur within members of a student organization committed, say, to gay rights or to reducing a university's investments in Sudan. So, too, for a government that is determined to avoid, or to make, war. I have suggested that the deliberations of the American government under George W. Bush, culminating in the Iraq war, are a clear example.³⁸ Tragically, the relatively extreme movement toward war was fueled by antecedent extremism and by an

absence of dissenting voices, produced by intense pressure on those who would reject the party line.

Solidarity and Affective Ties Increase Polarization

If members of the group think that they have a shared identity and a high degree of solidarity, there will be heightened polarization.³⁹ One reason is that if people feel united by some factor (family, politics, or religious convictions), dissent will be dampened. If individual members tend to perceive the others as friendly, likable, and similar to them, the size and likelihood of the shift will increase.⁴⁰ The existence of such ties reduces the number of diverse arguments and also intensifies social influences on choice. A clear implication is that mistakes are likely to increase when group members are united mostly through bonds of affection and not through concentration on a particular task; alternative views are least likely to find expression.

By contrast, people are less likely to shift if the direction advocated is being pushed by unfriendly group members or by members who are in some sense "different." A sense of "group belongingness" affects the extent of polarization. In the same vein, physical spacing tends to reduce polarization; a sense of common fate and intragroup similarity tend to increase it, as does the introduction of a rival outgroup.

An interesting experiment investigated the effects of group identification on polarization.⁴¹ Some people were given instructions in which their group membership was made salient (the "group immersion" condition), whereas others were given no such instructions (the "individual" condition). For example, those in the group immersion condition were told that their group consisted solely of first-year psychology students and that they were being tested as group members rather than as individuals. The

relevant issues involved affirmative action, government subsidies for the theater, privatization of nationalized industries, and phasing out of nuclear power plants. The results were stunning. Polarization generally occurred, but it was significantly greater when group identity was emphasized. This experiment shows that polarization is highly likely to occur, and to be most extreme, when group membership is made salient.

Compare a related experiment designed to see how group polarization might be dampened.⁴² The experiment involved the creation of four-person groups. The experimenters began with tests to establish that all of the groups included equal numbers of persons on two sides of political issues—whether smoking should be banned in public places, whether sex discrimination is a thing of the past, and whether censorship of material for adults infringes on human liberties. People's judgments were registered on a scale running from +4 (strong agreement) to 0 (neutral) to -4 (strong disagreement). In half of the cases (the "uncategorized condition"), people were not made aware that the group consisted of equally divided subgroups. In the other half (the "categorized condition"), people were told that they would find a sharp division in their group, which had equally divided subgroups. They were also informed who was in which group and told that they should sit around the table so that one subgroup was on one side facing the other subgroup.

In the uncategorized condition, discussion generally led to a dramatic reduction in the gap between the two sides. The result was a convergence of opinion toward the middle of the two opposing positions. But things were very different in the categorized condition. Here the shift toward the median was much less pronounced, and frequently there was barely any shift at all. In short, *calling attention to group membership made people far less likely to shift*

in directions urged by people from different groups. This little experiment offers a large lesson: If people are told that they are defined by their membership in a certain group—Catholics, Jews, Irish, Russians, Democrats, conservatives—they will be less likely to listen carefully to those who are defined in different terms.

Exit

Over time, group polarization can be fortified by "exit," as moderate members leave the group because they dislike the direction in which things are heading. In a leading study of Islamic terrorism, Marc Sageman emphasizes the importance of this fact. As group members move toward the possibility of violence, there is a situation of voluntary sorting and self-selection in which "only the true believers remain." Those believers regard themselves as "best friends and a substitute for family."⁴³ These are the most dangerous conditions of all: The groups include extremists, unified by bonds of affection and solidarity, and prone to discussions only among themselves.

The more general point is that when people are prone to exit, the group is likely to become more extreme. The group will end up smaller; its members will be both more like-minded and more willing to take extreme measures. In a kind of vicious circle, that very fact will mean that internal discussions will produce still more extremism. The shifts of student groups in the United States in the 1960s—from relatively moderate forms of left-wing thought to real radicalism and even violence—can be explained partly in these terms. And indeed, this account fits some of the dynamics of the White House under President George W. Bush, as moderate and dissenting officials left the government, leading to the Team of Univrals that I have mentioned.

It follows that in an important sense, a group is more likely to show extreme movement if it makes it easy for people to leave. If only loyalists stay, the group's median member will be more extreme, and deliberation will produce increasingly extreme movements. Making exit difficult prevents the group from shrinking. But it also ensures that the group will include people who favor relative moderation and tend to discipline its movement toward extremes.

There is a clear connection between these points and Albert Hirschman's important analysis of "exit" and "voice" as responses to disagreement with groups and organizations.⁴⁴ Hirschman shows that when exit is freely available, people might simply leave and not use their voices to ensure improved performance. He offers the example of competition between public schools and private schools. If public schools deteriorate, people might exit in favor of private schools. This result will impose some pressure toward improving the public schools, but it will also cause the more significant "loss to the public schools of those member-customers who would be most motivated and determined to put up a fight against the deterioration if they did not have the alternative of the private schools."⁴⁵

What is true for schools is also true for groups that are inclined to go to extremes. An easy exit option will reduce the number of dissenting voices and thus produce greater radicalism. At the same time, the difficulty of exit, combined with strong social pressures, might also reduce dissent, especially because members are likely to be highly dependent on the good will of group members.

Informed Members and Facts

When one or more people in a group are confident that they know the right answer to a factual question, the group might

well shift in the direction of accuracy.⁴⁶ For such problems, sometimes described as “eureka problems,” groups do well; they do not polarize. It is for this reason that groups tend to perform impressively on crossword puzzles. On puzzles, members hardly go to extremes. They accept the correct answer once it is announced. If there is immediate recognition of the correct answer, then groups will arrive at it. With eureka problems, for which the answer, once revealed, is clear to all, deliberation appears to produce accuracy rather than extremism.

Suppose, for example, that the question is how many people were on the earth in 1940, or the number of home runs hit by Barry Bonds, or the distance between Paris and London. Suppose, too, that one or a few people know the right answer. If so, there is a good chance that the group will not polarize, but instead converge on that answer. When this is so, the reason is simple: The person who is confident that she knows the answer will speak with assurance and authority, and she is likely to be convincing for that very reason. If one member of a group is certain that Barry Bonds hit 766 home runs, and if other members are uncertain, then the group might well end up agreeing that he hit 766 home runs.

Of course, it is not inevitable that the result will be agreement on the truth. Social pressures can lead people to blunder even on the simplest factual issues. An impressive study demonstrates that majority pressures can be powerful even for factual questions on which some people know the right answer.⁴⁷ The study involved 1,200 people, forming groups of four, five, six members. Individuals were asked true-false questions involving art, poetry, public opinion, geography, economics, and politics. They were then asked to assemble into groups, which discussed the questions and produced answers. The majority played a large role in

determining the group's answers. The truth played a role, too, but a lesser one. If a majority of individuals in the group gave the right answer, the group decision moved toward the majority in 79 percent of the cases. If a majority of individuals in the group gave the wrong answer, the group decision nonetheless moved toward the majority in 56 percent of the cases.

Hence the truth did have an influence—79 percent is higher than 56 percent—but the majority's judgment was the dominant influence. And because the majority was influential even when wrong, the average group decision was right only slightly more often than the average individual decision (66 percent vs. 62 percent).

This study demonstrates that groups might err even when some of their members know the truth. In some cases, however, group members who are ignorant will be tentative, and members who are informed will speak confidently. This is enough to promote convergence on truth rather than polarization.

Equally Opposed Subgroups

Return to our study of political beliefs in Boulder and Colorado Springs. What would have happened if we had mixed people from the two places? A tempting response would be that the answer lies in the predeliberation median. If the group's median member favored same-sex unions, perhaps most people would shift in that direction, even if people from Boulder were mixed with those from Colorado Springs.

This might well have happened, but we cannot be sure. The reason is that polarization may not be found when the relevant group consists of individuals drawn equally from two extremes.⁴⁸ Suppose that people who initially favor caution are put together with people who initially

favor risk-taking. If so, the group judgment may well move toward the middle. Consider a study⁴⁹ of six-member groups specifically designed to contain two subgroups (of three persons each) initially committed to opposed extremes; the effect of discussion was to produce movement toward the center. One reason is the existence of relevant information in both directions.

Not surprisingly, this study of equally opposed subgroups found the greatest "depolarization" with obscure matters of fact that carried no emotional resonance—for example, the population of the United States in 1900. It found the least depolarization with highly visible public questions—for example, whether capital punishment is justified. In cases of that kind, people simply stuck with what they thought before. Matters of personal taste depolarized a moderate amount—for example, preference for basketball or football, or for colors to paint a room. It follows that long-debated issues are not likely to depolarize. With respect to such issues, people are simply less likely to shift at all, in part because the arguments are familiar to everyone, and nothing new will emerge from discussion.

We can now offer four conclusions about what might happen within mixed groups.

1. For many issues and many groups, the median point of view, in advance of deliberation, is the best predictor of the direction of the shift; this was indeed what we observed in our study of punitive damage awards by juries.
2. When groups contain equally opposed subgroups, do not hold rigidly to their positions, and listen to one another, members will shift toward the middle; they will depolarize. The effect of mixing will be to produce moderation.

3. When people are dealing with "eureka problems," for which the right answer, once announced, is clear to most or all, mixed groups will find the right answer.

4. Sometimes people will stay exactly where they are. Those with entrenched views on capital punishment, the conflict in the Middle East, or abortion may not be much moved to hear what their adversaries have to say.

These capsule summaries help to explain when one or another of these outcomes is most likely. Standard polarization will occur if there is a well-defined predeliberation tendency in one direction and if people have sufficient open-mindedness that they are likely to listen to one another. Depolarization will occur if group members are split fairly evenly and if people are willing to listen. People will converge on truth if they know it when someone announces it. No movement will occur if people know what they think and think that those who disagree are knaves or fools.

In this regard, return to our studies of judicial behavior. On almost every issue, we observe the pattern I have described, in which Democratic and Republican appointees differ and in which that difference is significantly heightened on all-Democratic and all-Republican panels. But as we have seen, that pattern is not always observed. On three issues, the two sets of appointees do differ, but they do not polarize. Their voting patterns remain the same regardless of whether they are sitting with zero, one, or two people from their own party. In advance, what would you have guessed that the three issues were?

Recall the answer: abortion, capital punishment, and national security. In those domains, Democratic and Republican appointees are simply unable to influence one another. There is a large lesson here for domains in which people's beliefs, preferences, and values are so fixed that social

influences are powerless to affect them. And indeed, there is one court of appeals (of twelve) in which Republican and Democratic appointees are generally uninfluenced by one another and in which both sets of appointees do not show more extreme voting patterns on unified panels. I am speaking of the U.S. Court of Appeals for the Sixth Circuit, on which—according to informal lore—Democratic and Republican appointees really don't like each other. Our statistical analysis tends to support the informal lore.

Biased Assimilation

Another set of empirical findings bear directly on the nature and limits of polarization. Suppose that you produced a group of people, half of whom favor capital punishment, and half of whom reject it. Suppose that you gave to the entire group a set of balanced, substantive readings, offering arguments in both directions. What result would you predict? Many people think that we would observe more moderation and hence depolarization. Having seen sensible arguments on the other side, both groups might move to uncertainty, and in that sense to the center.

Surprisingly, this is not what is usually observed.⁵⁰ After reading balanced materials offering arguments both ways, opponents of capital punishment are strengthened in their opposition; they become more extreme. Advocates of capital punishment also harden. At least on some issues, people show "biased assimilation."⁵¹ Reading a set of arguments, they discount uncongenial points as silly or stupid and find congenial ones to be smart and pertinent. Hence they are strengthened in their original convictions.

The finding of biased assimilation has important implications for many issues in politics and elsewhere. People often ignore powerful contrary evidence. Some radical movements

prosper even when their members are surrounded by information that seems flatly inconsistent with their beliefs. That information can be, and is, discounted as mere propaganda; indeed, its very existence is taken to support people's radical beliefs. Closer to home, our affections, our fears, our judgments, and our preferences often stay fixed, and we retain confidence in them, even when we know enough to shift. Extremists are strongly committed to their beliefs, and when they see evidence that cuts the other way, or even evidence that seems balanced, they can become still more committed, not less so.

So while we have hoped that mixed groups, confronted with balanced information, would polarize less, the opposite is sometimes true. Suppose that the group contains five people who greatly fear climate change and five people who believe that the risks are small. After talking together, and after hearing balanced information, all ten might actually have a stronger commitment to what they thought before they started to talk—and the two groups would be further apart, not less so. I will return shortly to the circumstances in which this unhappy outcome will occur.

Here is an especially disturbing finding. When people's false beliefs are corrected, they might become even firmer in their commitment to those beliefs!⁵² Suppose, for example, that supporters of the Iraq war were told, by an apparently credible news source and at an early stage, that Iraq did not, in fact, have weapons of mass destruction. Remarkably, such corrections often do not reduce misperceptions, and sometimes they actually increase and strengthen them.

Return here to the problem of terrorism and note the suggestion that intense group dynamics, spawning what Marc Sageman calls a process of "in-group love," ensure that "the group acts as an interactive 'echo chamber,' encouraging escalation of grievances and beliefs in conspiracy

to the point of hatred."⁵³ Group members come to rely exclusively on one another to validate new information, and everything that they believe is a product of interactions within their enclaves. Thus "they discard information refuting their beliefs as propaganda from the West."⁵⁴ Here is a clear case of biased assimilation, in a way that promotes group polarization.

How can these findings be explained? And where and when do biased assimilation and attitude polarization occur?

Motivated Assimilations

The simplest point is that people appear to process information in a way that is distorted by their emotions and their motivations. Consider the well-established finding that after purchasing a product, people tend to seek out information confirming that their purchase was a sensible one. People are seeking to be reassured that they made the right decision. They wish to reduce cognitive dissonance, which makes people credit and seek out congruent information, and discredit and avoid incongruent information. More generally, people process information in a way that fits with their desires. They credit arguments that fit with what they already think, and they discredit arguments that point the other way.

Prior Convictions and Biases

Suppose that society consists of two groups of people, the Sensibles and the Haters, and that members of both groups have strong prior convictions. Suppose that the Sensibles have a strong antecedent commitment to a certain view—say, that the Holocaust actually happened, that Al Qaeda was responsible for the attacks of 9/11, that the president is not a Communist spy. Suppose that the Sensibles read balanced materials on these three questions.

The materials that support their antecedent view will not only seem convincing; they will also offer a range of details that will fortify the prior beliefs of most Sensibles. By contrast, the materials that contradict those beliefs will seem implausible, incoherent, ill-motivated, possibly a bit mad. The result is that people's antecedent convictions will be strengthened. Of course the opposite pattern will be observed for the Haters, who begin with the belief that the Holocaust did not happen, that the United States was itself responsible for the attacks on 9/11, that the president is a Communist spy. Biased assimilation can therefore be predicted from the mere existence of strong antecedent convictions and the effects of those convictions on (rational) judgments about new information.

When Biased Assimilation and When Not

This simple account helps to explain why biased assimilation will occur little, or perhaps not at all, if groups begin with a weak prior commitment. Suppose that the Sensibles are weakly committed to the propositions above and that the Haters disagree with them, but without much conviction. If both groups are exposed to balanced materials, they might tend to coalesce—at least if they do not have significantly asymmetrical trust.

Biased assimilation should be easy to understand in this light. It is in large part a product of strong prior convictions and also of divergences in trust. The Sensibles will trust some people and distrust others, and the Haters will show the opposite pattern. When they read materials from both sides, it is not exactly stunning that they end up learning from, and discounting, different sides. If, by contrast, people begin with weak prior convictions and do not suffer from asymmetrical trust, they will converge. We can also see in this light why people are often moved from their prior

convictions, not by their usual antagonists and opponents, but by people with whom they typically identify.

Self-Defeating Corrections

Turn now to the case of correction. Suppose that people believe that the Holocaust did not happen and that Al Qaeda was not responsible for the attacks of 9/11. After reading materials that purport to be corrections, many people will be unlikely to change their views. On the contrary, the purported correction may be, in a sense, self-defeating. Perhaps the correction serves mostly to anger people; if so, it might strengthen their commitment to what they believed before. Perhaps the correction focuses people's attention on the issue and the debate in question, and in that sense leads them to commit themselves, more strongly than before, to what they vaguely believed. It is well established that when people are given information suggesting that they have no reason to fear what previously seemed to be a small risk, their fear often increases. This mysterious finding might be explained by the fact that the information focuses people's attention on that risk, and when attention is focused on a risk, fear increases. So too, perhaps, with corrections of false reports of wrongdoing: By focusing people's attention on those reports, they increase the sense that wrongdoing has occurred.

On purely cognitive grounds, it does seem harder to explain situations in which corrections actually strengthen (false) beliefs. But on certain assumptions, the very existence of the correction may attest to its falsehood. An attempted refutation by an untrustworthy source can be taken as additional evidence in favor of those beliefs. For example, the attempt might not have been made if the beliefs were not true. Why correct an error, unless there is not something to it?

Many corrections will of course not be self-defeating. If people do not have strong motivations for accepting a falsehood, if their prior knowledge is weak, and if they have a degree of trust in those who are providing the correction, then false beliefs will dissipate. Outcomes will thus be different among different social groups. Some groups will be strongly motivated, for example, to accept a terrible rumor about a politician or an institution, whereas other groups will be strongly motivated to reject it.

The Deliberative Opinion Poll

In some influential work, James Fishkin has pioneered the idea of a "deliberative opinion poll," in which small groups, consisting of highly diverse individuals, are asked to come together and deliberate about various issues.⁵⁵ Fishkin has conducted deliberative opinion polls on numerous questions and in several nations, including the United States, England, and Australia. Fishkin finds some noteworthy shifts in individual views, in a way that suggests that deliberation is having a significant effect, but he does not find a systematic tendency toward group polarization. In his studies, individuals shift both toward and away from the median of pre-deliberation views. In England, for example, deliberation led to reduced interest in using imprisonment as a tool for combating crime.⁵⁶ Similar shifts were shown in the direction of greater enthusiasm for procedural rights of defendants and increased willingness to explore alternatives to prison.⁵⁷

On some issues, the effect of deliberation was to create an increase in the intensity with which people held their pre-existing convictions.⁵⁸ But in deliberative opinion polls, this was hardly a uniform pattern. On some questions, deliberation increased the percentage of people holding a minority

position (with, for example, a jump from 36 percent to 57 percent of people favoring policies making divorce "harder to get").⁵⁹ These changes are very different from what we observed in Colorado, and they are not what would be predicted by group polarization.

How can we explain Fishkin's findings? At least three factors distinguish the deliberative opinion poll from standard tests of group polarization. First, Fishkin's groups were overseen by a moderator, concerned to ensure a level of openness and likely to alter some of the dynamics discussed here. Second, and probably more important, Fishkin's studies presented people with a set of written materials that tried to be balanced and that contained detailed arguments supporting sides. At least if people did not start with strong convictions, the likely result would be to move people in different directions from those that would be expected by simple group discussion, unaffected by external materials inevitably containing a degree of authority. Indeed, it would be easy to produce a set of such materials that would predictably shift people's views in the direction favored by the experimenter. And even without a self-conscious attempt at manipulation, or a general effort to be neutral and fair, the materials will undoubtedly affect the direction that deliberation will take group members.

Third, Fishkin's participants did not deliberate to a group decision, and the absence of such a decision probably weakened the influences that produce extremism. When people have committed themselves to a group judgment, it is likely that their individual responses, even if subsequent and anonymous, will be affected by the commitment. To be sure, group polarization has been found after mere exposure to the views of other group members, but it is typically smaller than after discussion

and group judgment.⁶⁰ These three factors undoubtedly contribute to Fishkin's results.

GROUPS OVER TIME: "POLARIZATION GAMES"

Most studies of group polarization involve one-shot experiments. Consider, for example, the Colorado experiment, in which people were brought together, asked to talk, and then told to go home. Let us notice an intriguing implication of the experiments, an implication with special importance for people who meet with each other not once, but on a regular basis.

Suppose that participants engage in repeated discussions. Suppose that they meet each month, express views, and take votes. If so, there should be repeated shifts toward, and past, specific extreme points. Suppose that a group of citizens is thinking about genetic engineering of food, climate change, or the war on terror. The consequence of their discussions, over time, should be to lead in quite extreme directions. In these repeated *polarization games*, deliberation over time might well produce a situation in which people eventually come to hold positions more extreme than those of any individual member before the series of deliberations began.

In fact, the idea of repeated polarization games seems far more realistic than the processes studied in one-shot experiments. Groups typically meet many times, not just once. There appear to be few studies of such repeated polarization games. But it is not difficult to think of real-world groups in which the consequence of deliberation, over time, appears to be to shift both groups and individuals to positions that, early on, they could not possibly have accepted. Shifts of this kind clearly occurred with student groups in the 1960s.⁶¹ They also seem to have occurred with Islamic terrorists in the aftermath of the attacks of 9/11.⁶²

On the other hand, it is just not true that members of political organizations typically operate this way, even though they meet on a continuing basis. In the United States, Democrats do not usually move more and more to the left, and Republicans do not usually move more and more to the right. Why is this? One reason is that people are sensible and know what they think, or don't think, and their sense limits their movements. Another reason is the existence of *external constraints on extreme movements*. If Democrats shift far to the left, they will find themselves with fewer voters, and that fact imposes real discipline on the effects of internal deliberations. Political organizations are interested in attracting members and in achieving their goals, and this interest has significant limiting effects on potential movements. More generally, the direction and extent of extreme movements will often depend on the existence of external constraints. Market-type pressures, of the kind faced by political parties, often impose significant limits.

So, too, in the domain of business: Suppose that a group of people who lead a company go in an extreme direction. Suppose that the result is to produce inferior products. The company will be punished if consumers do not like those products. Life offers a number of reality checks, and these checks can limit shifts in our beliefs and our actions.

PEOPLE ARE DIFFERENT: OF THRESHOLDS AND TIPPING POINTS

Different people have different "thresholds" for moving as a result of new information or social pressure.⁶³ Such thresholds are important for understanding the dynamics of extremism.

Suppose that you believe that climate change is a serious problem and that the world should enter into an agreement

to impose stringent limits on greenhouse gas emissions. It is possible that you hold this belief without much conviction, in the sense that if certain people told you that you were wrong, you might shift. Suppose several friends tell you that the best way to handle the problem of climate change is through modest limits on emissions that increase over time, alongside funds to help poor countries adapt to warmer climates. Perhaps their statements are enough to persuade you. If they are not, it may be because your threshold for changing your mind is very high, and you will not adopt a different view unless you are given detailed arguments from real authorities. The basic point is that some people will readily shift their views on hearing a different position, whereas others will shift with more difficulty, and still others will shift only when presented with truly overwhelming reasons to do so.

These points help to explain why different people will move in different degrees in a group setting: why some people will not move at all, and why some groups are more prone to major movement than others. Two things matter: the direction of people's original convictions and their thresholds for changing them. Recall that among federal judges, there is no polarization on the issues of abortion, national security, and capital punishment, apparently because the threshold for changing views is exceedingly high. When group members begin with firm convictions, they require a great deal of information or social pressure (or both) to change their views. If social influences are strong enough, such people will likely move, but the extent of their movement is limited because of relatively high thresholds for accepting certain beliefs or engaging in certain behavior.

Tipping points can be immensely important to extreme movements. Suppose, for example, that a group of people is

deciding whether to undertake some action—say, to engage in violent protest. If only 10 percent of the group favors violent action and if majority rule is used, no violence will occur. But suppose that there are interdependencies among group members, so that what one person will do depends on what other people do. Suppose that people have diverse thresholds, and that most group members will opt to engage in violence if enough other members favor that course. If those who are clearly committed to violence make their views known at any early stage, others with relatively low thresholds will join them. If those with high thresholds resist and are sufficiently numerous, the first group will be outvoted. But suppose, instead, that there is a sequence in which the violence-prone state their views first, followed by those with low thresholds, and then followed by those with mildly higher thresholds. We could easily imagine a kind of cascade in the direction of violence.

The general point is that once a sufficient number of people converge on violence, a tipping point will occur, in the sense that those with higher thresholds will “tip,” and eventually most group members will become willing to support violence. To know whether violence will occur, a great deal depends on who speaks or acts first, and also on the distribution of privately held views. It also follows that small and seemingly random variables can play a large role in moving large groups of people toward extremism.⁶⁴ Radical movements are sometimes impossible to predict, even though they seem inevitable in hindsight. The difficulty of prediction stems from the fact that observers do not have access to people’s private thoughts and have no idea what kinds of thresholds would lead people to move in radical directions. The fall of communism had a great deal to do with processes of this kind.⁶⁵ When large changes occur that seemed unforeseeable, it is often because of diverse

thresholds within the population. Once people start to shift, dramatic movements suddenly become possible.

AUTHORITY AND OBEDIENCE

Now let us turn to some of the most famous and most alarming findings in modern social science.⁶⁶ The experiments, conducted by the psychologist Stanley Milgram, involved influence not by the judgments of peers, but by the will of an experimenter. For better or for worse, these experiments almost certainly could not be performed today because of restrictions on the use of human subjects. But they are of independent interest, because they have large implications for social influences on judgments of both morality and facts. Indeed, it is not possible to understand extremism without understanding obedience, and it is not easy to understand obedience without understanding Milgram’s work.

The experiments asked people to administer electric shocks to a person sitting in an adjacent room. Milgram’s subjects were told, falsely, that the purpose of the experiments was to test people’s memories and to see whether punishment might help people remember better. Unbeknownst to the subject, the victim of the electric shocks was a confederate, and there were no real shocks. The apparent shocks were delivered by a simulated shock generator with thirty clearly delineated voltage levels, ranging from 15 to 450 volts, accompanied by verbal descriptions ranging from “Slight Shock” to “Danger: Severe Shock.” As the experiment unfolded, people were asked to administer increasingly severe shocks for incorrect answers to memory questions—with the shocks going to and past the “Danger: Severe Shock” level, which began at 400 volts.

In Milgram's original experiments, the subjects included forty men between the ages of twenty and fifty. They came from a range of occupations, including engineers, high school teachers, and postal clerks. They were paid a small amount for their participation—and also told that they could keep the money no matter how the experiment went. The “memory test” involved remembering word pairs; every mistake, by the confederate/victim, was to be met by an electric shock and a movement to one higher level on the shock generator. To ensure that everything seemed authentic, the subject was, at the beginning of the experiment, given an actual sample shock at the lowest level. But the subject was also assured that the shocks would not cause long-term harm, with the experimenter declaring, in response to a prearranged question from the confederate, “Although the shocks can be extremely painful, they cause no permanent tissue damage.”⁶⁷

In the original experiments, the victim did not make any protest until the 300-volt shock, when he loudly kicked the wall of the room where he was bound to the electric chair. After that point, the victim did not answer further questions and was heard from only after the 315-volt shock, when he pounded on the wall again—and was not heard from thereafter, even with increases in shocks to and past the 400-volt level. If the subject indicated an unwillingness to continue, the experimenter offered prods of increasing firmness, from “Please go on” to “You have no other choice; you *must* go on.”⁶⁸ But the experimenter had no power to impose sanctions on subjects.

What do you think that people would do, when placed in this experiment? Most people predict that in such studies more than 95 percent of subjects would refuse to proceed to the end of the series of shocks. When people are asked to predict what people would do, the expected break-off point

is “Very Strong Shock,” 195 volts.⁶⁹ But in Milgram's original experiment, *every one of the forty subjects went beyond 300 volts*. The mean maximum shock level was 405 volts. A strong majority—twenty-six of forty, or 65 percent—went to the full 450-volt shock, two steps beyond “Danger: Severe Shock.”⁷⁰

Later variations on the original experiments produced even more remarkable results. In those experiments, the victim expressed a growing level of pain and distress as the voltage increased.⁷¹ Small grunts were heard from 75 volts to 105 volts, and at 120 volts, the subject shouted, to the experimenter, that the shocks were starting to become painful. At 150 volts, the victim cries out, “Experimenter, get me out of here! I won't be in the experiment anymore! I refuse to go on!”⁷² At 180 volts, the victim says, “I can't stand the pain.” At 270 volts, he responds with an agonized scream. At 300 volts, he shouts that he will no longer answer the questions. At 315 volts, he screams violently. At 330 volts and after, he is not heard.

In this version of the experiment, there was no significant change in Milgram's results: Twenty-five of forty participants went to the maximum level, and the mean maximum level was above 360 volts. In a somewhat gruesome variation, the victim says, before the experiment begins, that he has a heart condition, and his pleas to discontinue the experiment include repeated references to the fact that his heart is “bothering” him as the shocks continue.⁷³ This, too, did not lead subjects to behave differently. Notably, women do not behave differently from men in these experiments; they show the same basic patterns of responses.

Milgram himself explains his results as involving obedience to authority, in a way that explains certain forms of extremism, including the behavior of Germans under Nazi rule. Indeed, Milgram conducted his experiments partly to

understand how the Holocaust could have happened.⁷⁴ Milgram concluded that ordinary people will follow orders, even if the result is to produce great suffering in innocent others. Undoubtedly, simple obedience is part of the picture. But I want to urge an explanation that connects closely with group polarization.⁷⁵

The explanation involves the information conveyed by the instructions of an apparently legitimate authority. People who are invited to an academic setting, to participate in an experiment run by an apparently experienced scientist, might well defer to the experimenter's instructions, thinking that the experimenter is likely to know what should be done, all things considered, and that the experimenter is not likely to inflict serious harm for no good reason. In short, people are following a kind of heuristic or mental shortcut: "If an experimenter had an established institution asks me to do something, it is probably the right thing to do, or at least not a terrible thing to do." If the experimenter asks people to proceed, most of them might believe, reasonably, that the harm apparently done to the victims is not serious and that the experiment actually has significant benefits for society. On this account, the experimenter has special expertise. And on this account, many of the subjects put their moral qualms to one side, not because of blind obedience, but because of a judgment that their qualms are likely to have been ill-founded. That judgment must have been based, in turn, on a belief that the experimenter is not likely to ask subjects to proceed if the experiment is really objectionable.

On this view, Milgram's subjects were responding to an especially loud informational signal—the sort of signal sent by a real specialist in the field. And note that in fact, those who obeyed the authority, in Milgram's experiment, turned out to be right: No suffering was inflicted. The serious

problem here, and what Milgram revealed, is that the heuristic—in favor of obedience of apparently trusted authorities—does not always work well. In real-world cases, it leads to terrible moral errors.

A subsequent study, exploring the grounds for obedience, offers support for this reading of Milgram's experiments.⁷⁶ In that study, a large number of people watched the tapes of those experiments and were asked to rank possible explanations for compliance with the experimenter's request. Deference to expertise was the highest-rank option. This is not definitive, of course, but an illuminating variation on the basic experiment, conducted by Milgram himself, provides further support.⁷⁷ In this variation, the subject was placed among *three* people asked to administer the shocks. Two of those people, actually Milgram's confederates, refused to go past a certain level (150 volts for one and 210 volts for the other). In such cases, the overwhelming majority of subjects—92.5 percent—defied the experimenter. This was by far the most effective of Milgram's many variations on his basic study, all designed to reduce the level of obedience.

It is clear that in Milgram's experiments, the influence came from the experimenter's own position—that the shocks should continue and that no permanent damage would be done. But when the subject's peers defied Milgram's experimenter, the experimenter's position was effectively negated by the information conveyed by the refusals of peers. Hence subjects could rely on their own moral judgments, or perhaps follow the moral signals indicated by the peers' refusals. Milgram himself established, in yet another variation, something nice about human nature. Without any advice from the experimenter and without any external influences at all, the subject's moral judgment was clear: *Do not administer shocks above a very low level.*⁷⁸

The general lessons are straightforward. Group polarization occurs because of the informational and reputational signals given by others. When an authority tells people to do something, both of those signals can be very loud. If an authority tells you to do something apparently harmful or cruel, you might do exactly that, either because you think that it is the right thing to do or because you do not want to risk your reputation. In one experiment, for example, twenty of twenty-one nurses were willing to follow a doctor's orders to give a 20-milliliter dose of a drug called "androgen"—even though the label clearly stated that 5 milliliters was the usual dose and warned that 10 milliliters was the maximum.⁷⁹ Similar deference to authority can be found outside social science experiments. Almost half of surveyed nurses responded that they could remember a time when they had actually "carried out a physician's order that you felt could have had harmful consequences to the patient."⁸⁰

In these cases, the nurses seemed to be following a sensible heuristic, to the following effect: "Follow doctors' orders, because doctors know what is in the best interest of patients." Under plausible assumptions, this heuristic also works pretty well. Medical care would probably be worse, not better, if nurses were regularly in the business of second-guessing the decisions of doctors. The problem, as in Milgram's experiments, is that the heuristic can produce significant errors. Doctors do blunder, and sometimes nurses would do better to make an inquiry.

Consider here the fact that no fewer than sixty-eight fast-food restaurants have been subject to successful strip-search scams, in which a male caller, masquerading as a police officer named Scott, informs an assistant store manager that an employee at the restaurant has committed theft.⁸¹ Having

learned a great deal about the local conditions, "Officer Scott" asks the manager for the name of an attractive female employee who, Scott says, has been engaged in theft and is likely to have contraband on her. Officer Scott is then allowed to talk to the employee, and he tells her that she has two choices. She can come to police headquarters to be strip-searched or instead be strip-searched at that very moment by a fellow employee. Believing herself to be innocent, the employee agrees. Officer Scott then instructs that fellow employee to search the young woman's most private places, with the store's video cameras looking on. This is a clear example of how a sensible heuristic, in favor of obedience to authority, can go badly wrong. People should usually obey police officers—but not when they ask women to submit to a strip-search for no legitimate purpose.

The case of Officer Scott is a scam, of course, but it suggests that extreme movements often occur simply because someone in a position of authority has initiated them. Real atrocities, including torture and even genocide, can be explained in part by reference to mechanisms of this sort.⁸² Consider these words from a participant in the genocide in Rwanda: "When you receive a new order, you hesitate but you obey, or else you're taking a risk. When you have been prepared the right way by the radios and the official advice, you obey more easily, even if the order is to kill your neighbors. The mission of a good organizer is to stifle your hesitations when he gives you instructions . . . You obey freely."⁸³ And after a time, what was required may become in the nature of habit. As another put it, "At first killing was obligatory; afterward we got used to it. We became naturally cruel. We no longer needed encouragement or fines to kill, or even orders or advice. Discipline was relaxed because it wasn't necessary anymore."⁸⁴

There is an important point here about the nature of ordinary moral inhibitions and the importance of strengthening the moral intuitions that underlie them. As Tzvetan Todorov writes, "What the crimes of the Nazis teach us is that those who enforce the law are more dangerous than those who break it. If only the guards had given themselves over to their instincts! Unfortunately, they followed the rules."⁸⁵ In his account, the predominant type of guard in the concentration camps was "a conformist, willing to serve whoever wielded power and more concerned with his own welfare than with the triumph of doctrine."⁸⁶ When the system is working well, prison guards usually should obey their superiors, but when the system is not working well, they should be prepared to disobey.

SITUATIONISM, PRISON ABUSE, AND THE STANFORD PRISON EXPERIMENT

These points suggest two different answers to a perennial question: Why do human beings commit despicable acts? One answer points to individual dispositions; a different answer, suggested by Milgram's work, emphasizes situational pressures. In 2005, Secretary of State Condoleezza Rice stressed the importance of individual dispositions in describing terrorists as "simply evil people who want to kill." So-called situationists reject this view. They believe that horrible acts can be committed by perfectly normal people. The most extreme situationists insist that in the right circumstances, most of us, and perhaps almost all of us, might be led to commit atrocities.

The situationist view receives strong support from Milgram's experiment, from studies of group polarization, and also from Philip Zimbardo's influential study of situational influences, known as the Stanford Prison Experiment.⁸⁷

Because Zimbardo's experiment bears on extreme behavior in multiple domains, it will be useful to spend some time with it here.

The experiment started with an ad in a local newspaper, asking for volunteers for a study of prison life, lasting two weeks and paying \$15 a day (about \$75 in current dollars). Seventy of those who answered the ad were called to Stanford for interviews and a series of psychological tests. All seventy were American college students; most had completed summer school courses at Stanford or Berkeley. Twenty-four of them were selected on the ground that they were the healthiest and most normal. Half were randomly assigned to be prison guards; the other half were randomly assigned to be prisoners. All of them indicated that they would prefer to be prisoners, in part because they could not imagine being a prison guard after college, but they could imagine being in jail, and they thought they might learn from the experience. All of them agreed to participate through informed consent forms. They were also informed that if they were assigned the role of prisoners, they would suffer deprivations of their civil rights and have only minimally adequate diet and medical care. Those assigned to be prisoners were also told to wait at home on a particular Sunday, when they would be contacted to begin the experiment.

On that day, they were surprised to find themselves "arrested" by actual Stanford police officers (enlisted by Zimbardo), who handcuffed them, searched them, advised them of their rights, and booked them at police headquarters. Brought to a mock prison in the basement of the Stanford psychology department, they were stripped, deloused, and made to wear smocks, without underwear, and with numbers sewn on front and back. They were also forced to wear ankle chains and nylon stocking caps (not

having been asked to shave their heads). They walked in uncomfortable rubber thongs. Having worked with one of Zimbardo's graduate students, the guards read the prisoners a series of rules: "prisoners will be allowed 5 minutes in the lavatory," "prisoners must address each other by number only," "prisoners must never refer to their condition as an 'experiment' or a 'simulation,'" and others. Somewhat ominously, prisoners were told that the last rule was the most important: "Failure to obey any of the above rules may result in punishment."

The first day of the experiment was awkward for guards and prisoners alike, and not terribly eventful. Some of the guards did seem to relish their role, asking prisoners to do push-ups as "punishment" for laughing at some of the guards' comments. Whenever a prisoner showed an irrelevant attitude, he was likely to be asked to do more push-ups. Some guards engaged in acts of arbitrary cruelty—say, by leaning on prisoners and pushing them back with billy clubs. Things got much worse on Monday. On that day, the prisoners staged a rebellion, ripping off their numbers, refusing to obey commands, and mocking the guards.

Zimbardo asked the guards to take steps to control the situation. They did exactly that. Their responses consisted of forcing the prisoners to do jumping jacks and push-ups; stripping them naked in their cells; depriving them of meals, pillows, blankets, and beds; and placing them in solitary confinement. Some of the prisoners were baffled by the sheer aggressiveness of the response, with one screaming wildly, "No, no, no! This is an *experiment*! Leave me alone! Shit, let go of me, fucker! You're not going to take our fucking beds!" The rebellion was effectively crushed.

As the behavior of the guards became increasingly aggressive and humiliating, one of the prisoners, named

Doug, broke down and asked to be released. Zimbardo, having adopted the role of "prison superintendent," met with him privately. Zimbardo told Doug that he would forfeit his payment if he quit early, asked him to serve as an informer in return for "special privileges," and generally convinced him to continue. Returning to the prison, Doug falsely announced to the other prisoners that they could not leave. Shortly thereafter, his own stress reactions appeared to become hysterical, even pathological, as he threatened violence against both the guards and himself, and he was indeed released. On each of the next three days, another prisoner showed acute stress reactions and had to be released. The remaining prisoners became subdued and "zombie-like."

What of the guards? The picture was one of growing cruelty, aggression, and dehumanization. Sometimes without provocation, the guards stripped the prisoners naked, hooded them, chained them, denied them food or bedding privileges, put them into solitary confinement, and made them clean toilet bowls with their bare hands. There was sexual humiliation as well. On Thursday, one of the most aggressive guards, nicknamed John Wayne, called out to several of the prisoners, "See that hole in the ground? Now do twenty-five push-ups, *fucking* that hole! You hear me!" The prisoners dutifully obeyed. He continued, "Now, you two, you're male camels. Stand behind the female camels and *hump* them." Submitting to the order, the prisoners simulated sodomy.

The experiment ended prematurely after Zimbardo enlisted the help of Christina Maslach, a recent Stanford PhD in psychology who was starting her career as an assistant professor at Berkeley. In Maslach's own words, "I looked at the line of hooded, shuffling, chained prisoners, with guards shouting orders at them . . . I was overwhelmed by

a chilling, sickening feeling." Refusing to engage Zimbardo's claim that this was "amazing stuff," Maslach ended up in a heated argument with him (notwithstanding the fact that they were romantically involved at the time). She describes the "fight" as "too long and too traumatic," but eventually Zimbardo acknowledged that the experiment had had an adverse effect on him, as well as on the student subjects. He decided to halt the experiment on Friday.

Zimbardo himself draws some large lessons from his experiment. He insists that individual dispositions are far less important than we tend to think and that situational pressures can lead decent people to commit terrible acts. Recall that the prisoners and the guards were randomly assigned to their roles. "The line between Good and Evil, once thought to be impermeable, proved instead to be quite permeable."⁸⁸ Those assigned to be prisoners behaved as prisoners and were in a sense broken by the role. Those assigned to be as guards behaved badly, even viciously, notwithstanding their general normality. Zimbardo writes, "At the start of this experiment, there were no differences between the two groups; less than a week later, there were no similarities between them."⁸⁹ Notably, the prisoners were skeptical of the claim of random assignment and insisted, after the conclusion of the experiment, that the guards were taller than they were. (They were wrong; the two groups had the same average height.)

In pointing to the apparent normality of those involved in Nazi war crimes, Zimbardo gives a social science twist to Hannah Arendt's claims about the "banality of evil." And in explaining what makes atrocities possible, Zimbardo places a large emphasis on deindividuation—a process by which both perpetrators and victims become essentially anonymous and are thereby transformed into a type or a role. The very decision to wear a uniform can have significant behavioral

effects; warriors who change their appearance in preparation for war are more likely to brutalize their enemies. During the process of deindividuation, people enter a state of arousal in which they do not face the ordinary social sanctions and in which their own moral doubts are silenced. In this account, deindividuation ensures the triumph of "the Dionysian trait of uninhibited release and lust" over the "Apollonian central trait" of "constraint and the inhibition of desire."⁹⁰

These general points, and the Stanford Prison Experiment in particular, seem to help to explain the horrific behavior of American soldiers at Abu Ghraib. Recall the well-publicized incidents, some of them photographed, in which soldiers humiliated prisoners by leading them around by dog leashes, forcing them to simulate fellatio, and making them masturbate in front of a cigarette-smoking female soldier (herself giving a high-five salute of approval). American personnel also threatened male detainees with rape, beat them with broom handles and chairs, punched and kicked them, and forced them to wear women's underwear. Perhaps such abuses were a predictable consequence of situational forces, not (as prominent military leaders have urged) of the dispositions of rogue soldiers or a few bad apples.

In the Stanford Prison Experiment, the most interesting puzzle is the behavior of the guards. How could ordinary college students show such a high level of aggression and cruelty? It is true that unlike in the Milgram experiments, no authority was issuing specific orders. But Zimbardo specifically instructed guards to assume a particular role, in which they "have total power" with the task of producing "the required psychological state in the prisoners for as long as the study lasted." Zimbardo, a professor at Stanford, told college students to make the students "feel as though they were in prison." These instructions, alongside the very role of the

guard, conveyed certain information about what should be done. Those who find themselves operating as prison guards know that they should behave in certain ways. This is no less true in an experimental setting than elsewhere. Indeed, the experimental setting might have aggravated the behavior of some of the guards, who knew that certain safeguards were in place and that their specific task was to induce "the required psychological state."

We might draw some large lessons from this conclusion. Perhaps those who engage in extreme behavior are led to do so by their role and their context; perhaps all of us, under certain circumstances, could commit atrocities. Chillingly, Milgram himself said, "If a system of death camps were set up in the United States of the sort that we had seen in Nazi Germany, one would be able to find sufficient personnel for those camps in any medium-sized American town."⁹¹

At Abu Ghraib in Iraq, otherwise ordinary members of the military, both male and female, understood various forms of torture and humiliation as "standard operating procedure."⁹² Sabrina Harman, a soldier who famously appeared in photographs in which prisoners were sexually humiliated, observed, "That's the only way to get through each day, to start blocking things out. Just forget what happened. You go to bed, and then you have the next day to worry about. It's another day closer to home. Then that day's over, and you just block that one out."⁹³ Tim Dugan, another soldier, said that the soldiers were told, "We got a chance to break this unlawful insurgency, and the people in an unlawful insurgency have no protection under the Geneva Conventions.' . . . If the fuckin' secretary of defense designates the motherfucker an unlawful insurgency, I mean, what the fuck am I supposed to say? It's an unlawful insurgency, wouldn't you think? He's the second-highest motherfucker in the country during the war."⁹⁴

Alison Des Forges, an investigator of the Rwandan genocide with Human Rights Watch, concluded:

This behavior lies just under the surface of any of us. The simplified accounts of genocide allow distance between us and the perpetrators of genocide. They are so evil we couldn't ever see ourselves doing the same thing. But if you consider the terrible pressure under which people were operating, then you automatically reassert their humanity—and that becomes alarming. You are forced to look at the situation and say, "What would I have done?" Sometimes the answer is not encouraging.⁹⁵

Des Forges is undoubtedly right, and the behavior of American soldiers at Abu Ghraib supports her point. Of special note, for purposes of understanding that behavior, is the fact that the soldiers did not learn the prisoners' names; on the contrary, they gave them nicknames, turning them into "cartoon characters, which make them comfortably unreal."⁹⁶ In the words of one of the soldiers: "I had one guy whose breath just stank. I called him Yuck Mouth. We had a guy—probably the tallest Iraqi I've ever seen—and his nose kind of looked like Big Bird off Sesame Street. I called him Big Bird. I had Trap Jaw, because he had real sharp teeth, looked like he could chew a brick. I had one that I called Gomer Pyle."⁹⁷ This kind of deindividuation of the victims of abuse is characteristic of what happens when people are asked to play certain social roles.

But for purely situational accounts of human behavior, there is an evident problem. The Stanford Prison Experiment uncovered significant differences among both prisoners and guards. Some of the prisoners could not handle the situation and essentially screamed, "Let me out of here!"—in part, perhaps, as a strategic effort to escape a terrible situation. Some of the guards did their jobs, but without cruelty, and they did various favors for the

prisoners. These identifiably “good guards” were altogether different from others, whose behavior was sadistic. Dispositions did matter. There is a real difference between the actual perpetrators and those who simply stood by. The same is true of American soldiers at Abu Ghraib, with a few enthusiastic about acts of abuse, and a few others seeming to revel in them. Sometimes one of the soldiers “would see something happening with a prisoner, and say, ‘Hey, this is wrong,’ or, ‘Operationally, we can’t do this.’ But when they said nothing,” the worst of the soldiers would feel free to act.⁹⁸ And recall, too, that Christina Maslach, the assistant professor involved in a romantic relationship with Zimbardo, expressed outrage and asked for the experiment to end, notwithstanding the obvious pressure simply to go along, and perhaps to marvel. As Zimbardo himself emphasizes, many human beings are able to resist situational pressures and to engage in forms of heroism. Even when group polarization is under way, some people, some of the time, will hold fast to their convictions and stay where they are, especially if group members go in destructive or violent directions.

Here is one way to think about the Stanford Prison Experiment and its real-world analogues, which might help us to sort out the relationship between dispositions and social contexts. In experimental settings and in the real world, most people will be reluctant to harm others. Most of them have strong moral commitments, and it will not be so easy for the situation to lead them to put those commitments to one side. Often their reluctance can be overcome with appropriate incentives and the right information. If people can be assured that any harm is small or nonexistent, or necessary to produce some greater good, they might well put their moral qualms to one side. (Recall Milgram’s experiments.) If people can be assured that any harm is

deserved, or part of legitimate punishment, then they might well be willing to inflict harm. (Prison guards do not refuse to put recalcitrant prisoners in solitary confinement.) But—and this is the key point—different people have radically different thresholds that must be met before they will be willing to harm others. Even at Abu Ghraib, there were significant differences in the attitudes and the behavior of American soldiers who lived in a situation that encouraged cruelty and apparent sadism. Some soldiers even turned out to be heroes, alerting the authorities to what was happening. Studies of genocide show disparities as well, even when killing is pervasive. As one killer recalled, “We became more and more cruel, more and more calm, more and more bloody. But we did not see that we were becoming more and more killers. The more we cut, the more cutting became child’s play to us. For a few, it turned into a treat, if I may say so.”⁹⁹

Some people—life’s “bad guards”—have a real capacity for sadism and cruelty; that capacity is built into their dispositions. If such people are instructed to act sadistically, or merely authorized to do so, they will. Other people have somewhat higher thresholds. They will require strong situational assurance that harming others is justified or acceptable, all things considered. Still other people—life’s heroes or those who refuse to act in accordance with role and culture that lead most people in terrible directions—have exceedingly high thresholds, or perhaps their moral convictions operate as an absolute barrier. The resistance of the heroes seems to be a product of a deeply engrained moral sense, whose roots are not well understood, but which undoubtedly comes, for many people, from background factors that enable or even require people to say a firm: “No!” A continuum of thresholds exists from the sadists to the heroes, or from the devils to the saints.

If all of this is right, we can understand why different prison experiments, and different prisons, might have different outcomes. A great deal depends on the initial mix of dispositions. A group of low-threshold guards will behave very differently from a group of high-threshold guards, in part because of their antecedent inclinations, and in part because of social interactions among them. Because of group polarization, a set of low-threshold guards might well become very cruel indeed, whereas a set of high-threshold guards will probably behave pretty well. With mixed groups, we could easily imagine a range of outcomes, ranging from extreme cruelty to comparative generosity. If the low-threshold guards act first and influence their high-threshold colleagues, cruelty is likely; if the high-threshold guards act first and influence the low-threshold types, the outcome will be much better. If heroes are present, and if they are clear and confident, they might be able to ensure a good outcome. Hierarchical relationships at many organizations—including schools, workplaces, and religious organizations—can be understood in roughly analogous terms. Teachers, employers, and religious leaders can take on some of the characteristics of aggressive prison guards, or not, and individual thresholds and social interactions make all the difference.

A great deal depends as well on the specific incentives and on existing information. Most low-threshold types will not show cruelty unless they are given at least some incentive to do so. Those with relatively high thresholds might be willing to show considerable aggression if their incentives are strong enough. Of course, beliefs can have a significant impact. Suppose that people are informed that aggression is justified or necessary in the circumstances. Perhaps they learn, or are told, that the victims of their aggression are wrongdoers who deserve whatever they get. Or perhaps

they learn that they are a part of a group of people (ethnic, religious, national) who have been systematically humiliated by others and who are entirely justified in responding to past humiliation. Or perhaps they learn that certain individuals or certain groups are bad by disposition, or perhaps even subhuman, and must be treated accordingly. Dispositions are partly a product of beliefs, contributing to low or high thresholds, and once belief-driven dispositions are in place, social situations can add fresh information, often overcoming the relevant threshold.

What emerges is a clear challenge to the most ambitious claims for situationism and a more complicated understanding of the relationship between individual dispositions and social situations. That understanding fits the Stanford Prison Experiment, and it helps to explain why different social contexts, and different social roles, can produce such radically diverse results.

The prison experiment shows that the very assumption of a particular social role automatically conveys a great deal of information about appropriate behavior. But social roles are not fixed. Prison guards need not feel free to brutalize prisoners. Perhaps the largest lesson is that a constant sense of moral responsibility should be taken to be a part of, rather than inconsistent with, a wide range of social roles.

A NOTE ON THE INTERNET—AND THE ARCHITECTURE OF SERENDIPITY

Many people have expressed concern about the social influences that are exerted via the mass media and the Internet.¹⁰⁰ Perhaps some of these influences produce unjustified extremism. A general problem is one of fragmentation, or “cyberbalkanization.” The Internet is making it possible for people to design a kind of Daily Me—their personal

communications packages, which include only the topics and opinions that they like and exclude troublesome issues and disfavored voices. With greater specialization, people are increasingly able to avoid general interest newspapers and magazines and to make choices that reflect their own predispositions. Many people appear to be hearing more and louder versions of their own views, thus reducing the benefits that come from exposure to competing views and unnoticed problems. Long before the Internet, it was possible to discuss the "racial stratification of the public sphere" by reference to divergences between white and African American newspapers.¹⁰¹ The Internet creates a much more dramatic "stratification," as groups of multiple kinds can sort themselves into like-minded types.

We should be clear about the nature of the problem. In any free society, you can read and see what you like, and you are allowed to exclude the rest. But with daily newspapers and evening news shows, we often live with a kind of *architecture of serendipity*—that is, a situation in which we will have a number of serendipitous encounters with topics and points of view. These encounters can have a large impact; sometimes they can even change our lives.

We might think, for example, that we have no interest in some problem in Turkey or India, but a story on these nations might spark our interest and divert our attention, possibly prompting action. We might think that we have a certain view on climate change or on labor unions, but a story might suggest that our thoughts are badly wrong and that we should consider a different perspective. If the architecture of serendipity is transformed into an architecture of control, people may well restrict themselves to topics and views that they find congenial.

In a way, of course, this is freedom in action. But an understanding of group polarization explains why a

fragmented communications market may create serious problems. If people on the Internet are deliberating mostly with like-minded others, their views will not merely be reinforced; they will instead be shifted to more extreme points. Indeed, the Internet would seem to be replicating the Colorado experiment, and doing so every hour of every day. With the Internet, it is exceedingly easy for each of us to find like-minded types. Views that would ordinarily dissolve, simply because of an absence of social support, can be found in large numbers on the Internet, even if they are understood to be exotic, indefensible, or bizarre in most communities. As Marc Sageman writes, "Let's assume that a very few people in the world share the same strange belief, say, that the moon is made of green cheese. Through a process of self-selection, they find each other on the same forum. . . . Soon, they will assume that everyone shares this conviction because only the true believers air their views and the rest stay silent."¹⁰² Recall that group polarization sometimes occurs because people do not take sufficient account of the fact that the views of group members are biased, or worse, and do not really represent the convictions of most people in the community. The problem is especially severe on the Internet, where it is so easy to find support for judgments that are held by only a (bizarre, confused, or hateful) few.

This point is strengthened by the fact, noted previously, that polarization is all the greater and all the more likely when people are attached by bonds by affection, commonality, or solidarity. Many Internet discussion groups are unified by a sense of shared identity. Hence a "plausible hypothesis is that the Internet-like setting is most likely to create a strong tendency toward group polarization when the members of the group feel some sense of group identity."¹⁰³ Here as elsewhere, this cannot be said

to be bad by itself. Perhaps the increased extremism is good. But it is certainly troublesome if diverse social groups are led, through predictable mechanisms, toward increasingly opposing and ever more extreme views. Mutual misunderstandings, even anger and contempt, are nearly inevitable.

In the modern era, terrorism is the most dramatic example. In Sageman's words, the "structure of the Internet has become the structure of global Islamic terrorism. It has evolved organically through the search and exploration of new safe methods of interaction by thousands of terrorist sympathizers given the fact that their physical habitat had become very hostile post-9/11."¹⁰⁴ Until 2004, face-to-face interactions played the key role in producing terrorist networks. More recently, the Internet has assumed great importance. Sageman emphasizes that the traditional hierarchy of terrorist groups is undermined by the Internet, which leads to a form of spontaneous self-organization. Chat rooms and dedicated forums help to inspire many young Muslims to join the Islamic terrorist movement. "The new forums have the same influence that these radical mosques played in the previous generation of terrorists."¹⁰⁵ Conspiracy theories, fueling outrage, are spread in rapid fashion, as "individuals seek and select the rooms most compatible with their views and abandon the ones they disagree with. In a sense, the followers vote with their mice and select the views they like."¹⁰⁶ In the context of terrorism, a kind of "leaderless jihad" is a result. "Thanks to the Internet, global Islamic terrorism may fade away, but will never completely die."¹⁰⁷

This is an extreme example. But in countless domains, the Internet produces a process of spontaneous creation of groups of like-minded types, fueling group polarization. People who would otherwise be loners, or isolated

in their objections and concerns, congregate into social networks.

HOMOPHILY AND CURIOSITY

Of course, it is true that people are curious, and many of us actually like serendipity. We seek, and do not deplore, a situation in which we are exposed to new ideas and competing views. In our own way, we combat group polarization, simply because we resist information cocoons and groups that consist solely of like-minded types. In business and in government, successful leaders seek divergent views and fresh opinions, precisely because of their intuitive awareness of the risks of polarization. In the United States, Presidents Abraham Lincoln and Franklin Delano Roosevelt are the foremost examples; they made special efforts to ensure that they did not live in echo chambers. It turns out that humility and curiosity help to ensure better decisions, in large part because they increase the pool of information. The Internet can help to prevent polarization if people use it to find novel points of view. In many societies, group polarization is countered, every day, by people's desire to test their own judgments against those of dissimilar others.

Nonetheless, there does seem to be a strong human tendency to self-segregate along the kinds of lines that promote polarization. In sociology, a detailed empirical literature explores "homophily"—the process by which "similarity breeds connection."¹⁰⁸ People who are similar along relevant dimensions tend to seek out one another and to live in the same social networks. In small groups, people who are unified by such demographic characteristics as age, education, race, religion, and ethnicity show a distinct tendency to self-segregate. The same is true of those unified along the lines of aspirations, attitudes, and intelligence. For

present purposes, what is most important is "value homophily," which includes the "considerable tendency for adults to associate with those of their own political affiliation."¹⁰⁹ Within the United States, many people believe that most sensible people share their political convictions, if only because those with whom they associate tend to think as they do.

Why does homophily occur? As a matter of history, geography has played a large role. Family, work, and other organizations also create strong ties among like-minded types. But these structural sources are complemented by voluntary ties and personal choices. People of similar religious views often choose to associate with one another. The hostility between believers and nonbelievers is in part a product of polarization. Miller McPherson and colleagues find that in many domains, "attraction is affected by perceived similarity," and people "associate with similar others for ease of communication, shared cultural tastes, and other features that smooth the coordination of activity and communication."¹¹⁰ In the era of the Internet, a great deal of work remains to be done on the extent to which homophily is creating niches of like-minded types. But a lot of evidence supports the view that cultural tastes, including tastes for music, spread through a process involving homophily.¹¹¹ To a large degree, people's tastes are shaped through interaction with others who have similar inclinations.¹¹² What is true for cultural preferences is undoubtedly true for political judgments and risk attitudes as well.

Existing work on homophily has not been brought into contact with the phenomenon of group polarization. This is a serious gap. It is clear that if birds of a feather are flocking together, extreme movements are to be expected. What is important is that the extent of the flocking depends both on social architecture and on prevailing norms. If people

naturally encounter those who are unlike themselves, or if workplaces and media facilitate such encounters, homophily will be counteracted. And if social norms encourage people to cultivate the tendency toward curiosity, and even to delight in new topics and opinions, then groups will contain birds of many different feathers.

GROUPTHINK AND GROUP POLARIZATION

We are now in a position to assess groupthink, a widely discussed phenomenon in the 1970s and 1980s, one that bears directly on my concerns here. Developed by Irving Janis, the idea of groupthink is designed to capture processes of decision that predictably lead to social blunders, catastrophes, and even forms of extremism.¹¹³ Janis's term drew directly and self-consciously on George Orwell's 1984 and, in particular, on Orwell's term *doublethink*. Stated briefly, Janis's suggestion was that certain groups stifle dissent, value consensus over correctness, fail to examine alternatives and consequences, and as a result, end up producing fiascos. Janis's plea was for a process of decision that would be "vigilant" in the sense that it would ensure careful attention to alternative courses of action and to the risks associated with those alternatives.

To support his argument, Janis relied on a number of actual policy decisions. According to Janis, groupthink was largely responsible for President Kennedy's disastrous decision to authorize the Bay of Pigs invasion. When President Johnson and his advisers escalated the Vietnam War during 1964-67, it was because the relevant group stifled dissent, sought consensus, and did not think well about consequences. The idea of groupthink has been applied to the Watergate cover-up,¹¹⁴ Neville Chamberlain's policy of appeasing Nazi Germany,¹¹⁵ the Ford Motor Company's

decision to market the Edsel, NASA's launch of the *Challenger* space shuttle, Nazi Germany's invasion of the Soviet Union in 1941, and the decision by Chemie Grunenthal to market thalidomide, which caused serious birth defects in children.¹¹⁶ President George W. Bush's decision to launch the Iraq war can easily be understood in terms of groupthink.¹¹⁷ In Janis's view, groupthink leads to many problems of defective decision making, including incomplete survey of alternatives and objectives, failure to examine the risks of the preferred choice, poor information search, selective bias in processing information, and failure to assess alternatives.¹¹⁸

Janis argued that groupthink involves several "types" of symptoms.¹¹⁹ These include close-mindedness, involving a collective effort "to rationalize" so as to discount warnings or information that might lead to reconsideration, and stereotyped views of enemies, as either too evil to warrant efforts at negotiation or "too weak and stupid to counter" the group's risky choices. Organizations susceptible to groupthink impose pressures toward uniformity. Here Janis refers to self-censorship on the part of group members, who minimize the importance of their own doubts and counter-arguments. Self-censorship is connected with an illusion of unanimity. This illusion is fostered by direct pressure on any members who argue against the group's stereotypes, illusions, and commitments.

Janis added that groupthink has a set of identifiable causes. The first and most important is cohesiveness; a group that lacks that quality is not "likely to display symptoms of defective decision-making." But groupthink requires additional conditions. These include insulation of the policy-making group, which reduces the chance of receiving expert advice and critical evaluation from outside; lack of a tradition of impartial leadership, meaning that leaders will not

encourage open inquiry and critical evaluation; lack of procedures for promoting good decision making; and homogeneous social backgrounds and ideology on the part of members.

Janis contended that the remedy for groupthink involves vigilant processing of information.¹²⁰ Leaders should encourage critical evaluation by giving high priority to objections and doubts. To promote diversity of view, independent policy-planning and evaluation groups should work on the same problems, with different leaders. Group members should be assigned the role of devil's advocate, bringing a new perspective to bear. Outside experts and qualified people not directly involved in the issue at hand should be encouraged to challenge prevailing views. In support of these ideas, Janis found that groupthink was absent in many successful decisions, such as the Kennedy administration's peaceful resolution of the Cuban Missile Crisis and the Marshall Plan for rebuilding Europe after World War II.

How do Janis's claims bear on my argument here? I have emphasized that groups can go to extremes. I have also urged that social pressures, both informational and reputational, are heightened if group members have a high degree of solidarity and affection. In this light, many of Janis's examples can be seen as case studies in group polarization, as groups move to more extreme points in line with their original tendencies. Janis's emphasis on self-censorship, heightened by social pressures, fits well with my basic claims. As he shows, many examples of group polarization require an appreciation of the role of leaders, whose views count for far more than those of other group members. If a leader does not encourage dissent and is inclined to an identifiable conclusion, it is highly likely that the group as a whole will move toward that conclusion.

In my view, the idea of group polarization is far more helpful, in explaining both extremism and error, than the idea of groupthink. It should be clear that Janis does not suggest any simple hypothesis that might be tested. Empirical work on the groupthink phenomenon has suggested a mixed verdict,¹²¹ and there is a lively debate over Janis's claims.¹²² Much of the debate stems from uncertainty about the relationship between Janis's claimed symptoms and policy fiascos. Critics have urged that that "support for the posited groupings of groupthink characteristics derives from anecdote, casual observation, and intuitive appeal rather than rigorous research."¹²³

A careful study of successful and unsuccessful decision making in seven prominent American companies (including Chrysler, Coca-Cola, and CBS News) tried to test whether such companies exhibit groupthink and, if so, whether a lack of success is correlated with it.¹²⁴ In support of Janis's claims, the authors did find a strong relationship between a group's decision-making process and its likelihood of success. When information was processed well, companies were more likely to make good decisions. On the other hand, the successful groups showed some features of groupthink. In fact, those groups had strong leaders who attempted to persuade others that they were right. Such leaders produced mistakes only if they created "absolutist cults," defined as organizations centralizing power in a single person.¹²⁵ Such centralization, more than anything else, is associated with bad outcomes.

This study finds analogues in many others that have found some, but not complete, support for the groupthink model.¹²⁶ A systematic exploration of Janis's own examples concluded that groupthink characteristics were indeed correlated with failures.¹²⁷ In particular, the study found that defective decision making was strongly correlated with the

structural faults of groups, including insulation and homogeneity. But when group members are friends rather than strangers, have worked together in the past, or are asked to wear group labels, they have not shown more self-censorship than other groups, and it is not at all clear that such cohesive groups make worse decisions.¹²⁸ It may well be that if members trust one another and share norms of disclosure and dissent, there will be less self-censorship than in groups of strangers, for in such groups, people might fear that a dissident view will create serious friction.

But some of Janis's claims have fared well. Insulated groups have been found to consider fewer alternatives and make worse decisions than noninsulated groups.¹²⁹ Also in support of Janis's claims, groups with highly directive leaders have been found to suggest fewer alternatives, to use less information, to suppress dissent, and generally to show inferior decision-making processes.¹³⁰ Most studies also find that poor decision-making procedures, under Janis's criteria, produce less disagreement and worse decisions than do good procedures.¹³¹

How do all these findings bear on the analysis here? What is the relationship between groupthink and group polarization? The most obvious point is that group polarization offers a simple and clear prediction: As a statistical regularity, deliberating groups will end up in a more extreme point in line with their predeliberation tendencies. The idea of groupthink is far more complex and unruly, without any simple predictions. Working from real-world examples, Janis generalized a set of points about when groups are most likely to blunder. The generalizations are suggestive and helpful, but they do not offer a clear account of what characteristics of groups will lead to extremism, blunders, or catastrophes.

CASCADES

My emphasis thus far has been on group polarization. But extremism can also be fueled by a closely related phenomenon: *social cascades*. As cascades occur, beliefs and perspectives spread from some people to others, to the point where many people are relying, not on what they actually know, but on what (they think) other people think. This belief may well be erroneous, because people are relying not on their private information, but on the judgments of trusted others. When people conclude that the United States or Israel was responsible for the attacks of 9/11, or that doctors were responsible for the spread of AIDS among African Americans, or that a certain investment can't miss, cascades are typically responsible. Cascades play a large role in the stock market and in real estate. When certain stocks become suddenly popular, cascades are usually involved.¹³² Companies do their best to create cascades; the iPhone and the iPod are both terrific products, but they have definitely benefited from cascade effects. When people are suddenly fearful of a new risk, cascades are usually fueling their fear. Conspiracy theories in general tend to spread from one person to another through a cascade-like process. Social cascades come in two varieties: informational and reputational.

Informational Cascades

To see how informational cascades work, imagine a deliberating group that is deciding whether some person or group has engaged in unfair or even outrageous conduct, warranting some kind of punishment or reprisal.¹³³ Assume that the group members are announcing their views in sequence. From his own knowledge and experience, each member has some private information about what that person or

group has done. But each member also attends, reasonably enough, to the judgments of others.

Andrews is the first to speak. He suggests that bad conduct has indeed occurred. Barnes now knows Andrews's judgment; it is clear that she, too, should certainly conclude that there is unfairness if she agrees independently with Andrews. But if her independent judgment is otherwise, she would—if she trusts Andrews no more and no less than she trusts herself—be indifferent about what to think or do, and she might simply flip a coin. Now turn to a third person, Carlton. Suppose that both Andrews and Barnes have said that outrageous conduct has occurred, but that Carlton's own information, though not conclusive, suggests that they are wrong. In that event, Carlton might well ignore what he knows and follow Andrews and Barnes. It is likely, after all, that both Andrews and Barnes had reasons for their conclusion, and unless Carlton thinks that his own information is better than theirs, he should follow their lead. If he does, Carlton is in a cascade.

Now suppose that Carlton is speaking in response to what Andrews and Barnes did, not on the basis of his own information, and also that later people, in our little queue, know what Andrews, Barnes, and Carlton said. On reasonable assumptions, they will do exactly what Carlton did. That is, they will agree that outrageous conduct has occurred, regardless of their private information (which, we are supposing, is relevant but inconclusive). This will happen even if Andrews initially blundered. That initial blunder, in short, can start a process by which a number of people participate in creating serious mistakes.

If this is what is happening, there is a major social problem: People who are in the cascade do not disclose the information that they privately hold. In the example just given, the judgment of group members will not reflect the

overall knowledge, or the aggregate knowledge, of those within the group—even if the information held by individual members, if actually revealed and aggregated, would produce a better and quite different conclusion. The reason for the problem is that individuals are following the lead of those who came before. And if people are doing this, then they might end up in quite extreme directions. They might also converge on a judgment about climate change, or the right investments, or Iran, or China, or the intentions of the United States, that defies reality, and that produces dangerous action.

Does all this seem unrealistic? It should not; cascades often occur in the real world. The real estate boom of the early twenty-first century, culminating in the subprime crisis, was a product of a cascade.¹³⁴ When there are speculative bubbles, people are typically relying not on fundamentals but on their judgments about what other people are likely to think and do. Hence prices can continue to go up simply because people think that other people are investing—until a crash occurs. It was widely thought, and said, that real estate prices always go up, even though this is false. By historical standards since 1940, home prices jumped spectacularly only in one period: from 1997 to 2004. In that period, many people thought, and said, that it is in the nature of home prices to increase over time, and people's behavior tracked their belief. But the belief was demonstrably false. For many decades, home prices were relatively stable, until the unprecedented boom that began in 1997.

As Robert Shiller has shown, the best explanation of the real estate bubble greatly overlaps with the best explanation of the stock market bubble of the late 1990s: In both cases, people were greatly influenced by a process of social contagion that amounted to an informational cascade. This belief

produced wildly unrealistic projections, with palpable consequences for home purchases and mortgage choices. In 2005, Shiller and Karl Case conducted a survey among San Francisco home buyers. The median expected price increase, over the next decade, was 9 percent per year! In fact, one-third of those surveyed thought that the annual increase would be much higher than that. Their baseless optimism was based on two factors: salient price increases in the recent past and the apparent, and contagious, optimism of other people.

Of course the stock of public knowledge depends not merely on word-of-mouth and on visible sales, but on the media as well. In the late 1990s and early 2000s, it was widely reported that home prices were rapidly increasing (true) and that the prices would continue to increase over time (not true). If the apparent experts confirm "what everyone knows," then seemingly risky deals, of the sort that have led so many people to disaster, will seem hard to resist. The Internet bubble of the late 1990s was a result of similar forces, producing its own form of extremism. Notwithstanding the underlying evidence about values, people believed that continued growth was highly likely, because of what other people thought (combined with recent events); terrible investment choices resulted.

Do cascades occur for cultural products, such as art, music, movies, and literature? They certainly do, and they can produce unpredictable extreme movements. For a fascinating example, consider a study of music downloads. Matthew Salganik and his coauthors¹³⁵ created an artificial music market, with 14,341 participants who were visitors to a Web site popular with young people. The participants were given a list of previously unknown songs from unknown bands. They were asked to listen to a brief selection of any songs that interested them, to decide which songs

(if any) to download, and to assign a rating to the songs they chose. About half of the participants were asked to make their decisions independently, based on the names of the bands and the songs and their own judgment about the quality of the music. The other half could see how many times each song had been downloaded by other participants. These participants were also randomly assigned to one or another of eight possible "worlds," with each evolving on its own; those in any particular world could see only the downloads in their own world. The key question was whether people would be affected by the choices of others—and whether different music would become popular in the different "worlds."

Did cascades develop? Were there extreme movements? There is not the slightest doubt. In all eight worlds, individuals were far more likely to download songs that had been previously downloaded in significant numbers—and far less likely to download songs that had not been so popular. Most strikingly, the success of songs was quite unpredictable. The songs that did well or poorly in the control group, where people did not see other people's judgments, could perform very differently in the "social influence worlds." In those worlds, most songs could become very popular or very unpopular, with much depending on the choices of the first downloaders. The identical song could be a hit or a failure, simply because other people, at the start, were seen to choose to download it or not. As Salganik and his coauthors put it: "In general, the 'best' songs never do very badly, and the 'worst' songs never do extremely well," but (and this is the remarkable point) "almost any other result is possible."

As we have seen, similar findings have been made in the context of jury judgments about punitive damage awards. There is a great deal of unpredictability for identical cases, in

part because social influences among jurors can spur juries to make extremely high awards. As with jury judgments, so, too, with music (and movies and books and political views): Because people pay attention to one another, an early movement in a particular direction can operate as a spark that ignites a fire, leading to unexpected and dramatic outcomes. Many domains have what economists call "multiple equilibria"—a range of possible outcomes, all stable, and all possible with modest differences in starting points. People are often tempted to think, after the fact, that an outcome was entirely predictable and that the success of a musician, an actor, an author, or a politician was inevitable in light of his or her skills and characteristics. Social influences suggest that we should beware of that temptation. Small interventions and even coincidences, at a key stage, can produce large variations in the ultimate outcome.

For a less entertaining example, consider the existence of widely divergent group judgments about the origins and causes of AIDS—with some groups believing, falsely, that the first cases were observed in Africa as a result of sexual relations between human beings and monkeys, and with other groups believing, also falsely, that the virus was produced in government laboratories.¹³⁶ These and other views about AIDS are a product of social interactions and, in particular, of cascade effects. Deliberation often fails, and extreme views often spread, as a result. When groups come to believe some alleged fact about the egregious misconduct of some person or nation, an informational cascade is often at work.

Reputational Cascades

In a reputational cascade, people think that they know what is right, or what is likely to be right, but they nonetheless go

along with the crowd to maintain the good opinion of others. Suppose that Albert suggests that global warming will produce catastrophic harm in the near future and that Barbara concurs with Albert, not because she actually thinks that Albert is right, but because she does not wish to seem, to Albert, to be ignorant or indifferent to environmental protection. If Albert and Barbara say that global warming will produce catastrophic harm in the near future, Cynthia might not contradict them publicly and might even appear to share their judgment—not because she believes that judgment to be correct, but because she does not want to face their hostility or lose their good opinion.

It should be easy to see how this process might generate a cascade. Once Albert, Barbara, and Cynthia offer a united front on the issue, their friend David might be reluctant to contradict them, even if he thinks that they are wrong. The apparently shared view of Albert, Barbara, and Cynthia carries information; that view might be right. But even if David has reason to believe that they are wrong, he might not want to take them on publicly. The problem, of course, is that the group will not hear what David knows. Reputational cascades often help to account for the spread of extreme views. Especially when people live in some kind of enclave, they may silence themselves in the face of an emerging judgment or opinion, even if they believe it to be wrong.

In the actual world of group decisions, people are, of course, uncertain whether publicly expressed statements are a product of independent knowledge, participation in an informational cascade, or reputational pressure. Much of the time, listeners and observers overstate the extent to which the actions of others are based on independent information rather than social pressures. Deliberating groups often move to extreme points as a result.

Politics

There is every reason to think that cascade effects occur for issues of politics and morality—and that such effects can produce dramatic and extreme movements. Suppose that people are asking whether a politician would make a good nominee for high office. Informational cascades are highly likely; indeed, an informational cascade helped to account for the Democratic nominations of both John Kerry in 2004¹³⁷ and Barack Obama in 2008. When Democrats shifted from Howard Dean to John Kerry, or from Hillary Clinton to Barack Obama, it was not because each Democratic voter made an independent judgment on behalf of Kerry or Obama. It was in large part because of a widespread perception that other people were flocking to the eventual winner. With respect to Kerry, Duncan Watts's account is worth quoting at length, because it captures the general dynamic so well:

A few weeks before the Iowa caucuses, Kerry's campaign seemed dead, but then he unexpectedly won Iowa, then New Hampshire, and then primary after primary. How did this happen? ... When everyone is looking to someone else for an opinion—trying, for example, to pick the Democratic candidate they think everyone else will pick—it's possible that whatever information other people might have gets lost, and instead we get a cascade of imitation that, like a stampeding herd, can start for no apparent reason and subsequently go in any direction with equal likelihood. Stock market bubbles and cultural fads are the examples that most people associate with cascades ... but the same dynamics can show up even in the serious business of Democratic primaries. ... We think of ourselves as autonomous individuals, each driven by our own internal abilities and desires and therefore solely responsible for our own behavior, particularly when it comes to voting. No voter ever admits—even to herself—that she chose Kerry because he won New Hampshire.

A similar process greatly benefited Obama, who focused on the Iowa primary at a time when he was trailing badly in the national polls. After he won that primary, both informational and reputational cascades developed on his behalf, propelling him to the nomination. Information about his qualities spread rapidly among people who had theretofore known nothing about him. Those who admired Obama and made favorable statements about him received reputational benefits; social pressures worked to his advantage.

Social cascades can be found for many contested political questions, including the legitimacy of same-sex marriage, abortion, particular wars, and capital punishment. Perspectives on both environmental and economic issues are often a product of cascade effects. Few of us have thought long and hard about these questions. We often end up thinking what we think others think—at least if we think that those others think like we do. When “political correctness” moves people dramatically to the left or to the right, cascades are typically involved. These points raise an additional warning flag about any situation in which citizens sort themselves into communities of like-minded others. In such communities, cascades are almost inevitable, and they might well be based on poor thinking and confusion. The problem is that the same forces that produce factual errors operate in the moral and political domains as well.

CHAPTER 3

Movements

An understanding of group polarization and cascade effects has implications for all sorts of social beliefs and movements. Let us now consider some examples. For any of them, of course, a whole volume would be necessary to give a full sense of the underlying dynamics. My goal here is not to provide that full sense, but to say enough to suggest that the social influences explored here have played a crucial role.

OPPOSITIONAL MOVEMENTS AND GEOGRAPHIC ISOLATION

Under what circumstances will a group of people, with some degree of commonality, form a shared sense not only of identity but also of grievance, and ultimately seek to oppose existing social practices? Why do oppositional movements occur?

The real world is messy; it is not a controlled experiment. But for a clue, consider Sharon Groch's discussion of the