

#republic

DIVIDED DEMOCRACY
IN THE AGE
OF SOCIAL MEDIA

CASS R.
SUNSTEIN

With a new afterword by the author

PRINCETON UNIVERSITY PRESS
Princeton and Oxford

information. An advantage of a system with general-interest intermediaries and public forums—with broad access by speakers to diverse publics—is that it ensures a kind of social spreading of information. At the same time, an individually filtered speech universe is likely to produce too few of what the philosopher Edna Ullmann-Margalit has called *solidarity goods*—goods whose value increases with the number of people who are consuming them.³⁰ A presidential debate is a classic example of a solidarity good.

The third and final difficulty has to do with the proper understanding of freedom and the relationship between consumers and citizens. If we believe in consumer sovereignty, and if we celebrate the power to filter, we are likely to think that freedom consists in the satisfaction of private preferences—in an absence of restrictions on individual choices. This is a widely held view about freedom. Indeed, it is a view that underlies much current thinking about free speech. It is mostly right. But it is also inadequate—a big part of the picture, true, but hardly the whole thing.

Of course free choice is important. But freedom, properly understood, consists not simply in the satisfaction of whatever preferences people have, but also in the chance to have preferences and beliefs formed under decent conditions—in the ability to have preferences formed after exposure to a sufficient amount of information as well as an appropriately wide and diverse range of options. There can be no assurance of freedom in a system committed to the Daily Me.

POLARIZATION

The Internet can bring people together rather than draw them apart. Countless people are using social media to build larger and more diverse communities. But there is narrowing as well, in the form of communities of niches. One of my own fields is behavioral science, and with the help of Twitter, those of us who are interested in that field can easily find each other. If you want to learn about the latest developments, Twitter is a great help. For example, behavioral scientists are interested in “loss aversion,” which means that people dislike losses more than they like corresponding gains. If you’re interested in new examples, exceptions, or elaborations of the behavioral finding, Twitter is terrific. And yes, there is a #lossaversion.

That’s great, but for academics in any particular field, there’s a risk that Twitter will contain echo chambers just for them. And of course people can find hashtags to signal topics or points of view of many different kinds. In 2015, #NeverTrump was popular for a time, in opposition to #MakeAmericaGreatAgain, and Clinton supporters favored #ImWithHer, which was opposed by #LockHerUp. Those who wanted genetically modified organism (GMO) labels on food looked for #JustLabelIt, and #BlackLivesMatter competed with #AllLivesMatter.

Many people are using the Internet in exactly the sense prophesied by those who celebrate the Daily Me, and in a way that invites the continuing emergence of highly specialized websites, discussion groups, and feeds of innumerable sorts. What problems would be created as a result?

FLAVORS, FILTERS, AND VOTES

It is obvious that if there is only one flavor of ice cream and only one kind of toaster, a wide range of people will make the same choice. (Some people will refuse ice cream, and some will rely on something other than toasters, but that is another matter.) It is also obvious that as choice is increased, different individuals and different groups will make increasingly different choices. It is obvious, finally, that as the costs of reaching people with particular interests, and of interacting with them, as well as interacting start to shrink, then we will see a massive increase in niche markets of multiple kinds. This has been the growing pattern over time with the proliferation of communications options.

On YouTube, like-minded people can, in a sense, congregate to discuss and focus on one or more of those possibilities—not least when the clip casts ridicule on a particular person or point of view. Consider the celebratory words of David Bohnet, founder of *geocities.com* (no longer in service, but at one time the third most visited website): “The Internet gives you the opportunity to meet other people who are interested in the same things you are, no matter how specialized, no matter how weird, no matter how big or how small.”²¹ This is undoubtedly true, but it is not only an occasion for celebration.

To see this point, it is necessary to think a bit about why people are likely to engage in filtering. The simplest reason is that people often know, or think they know, what they like and dislike. A friend of mine is keenly interested in Russia; he subscribes to a service that provides him with some two dozen stories about Russia each day. If you are bored by news stories involving Russia or the Middle East, or if you have no interest in Wall Street, you might turn your mind off when these are discussed, and if you can filter your communications universe accordingly, you might think that it’s even better. In addition, many people like hearing discussions that come from a perspective that they find sympathetic. If you are a

Republican, you might prefer a newspaper with a Republican slant, or at least without a Democratic one. And indeed, many Americans with conservative leanings prefer to get their news from avowedly conservative sources, such as Fox News or the *Wall Street Journal*, whereas many Americans with liberal leanings work hard to avoid those very sources.

Does that matter? An ingenious study by Gregory J. Martin and Ali Yurukoglu of Stanford University explores whether people’s voting behavior really is influenced by what they see on cable news.²² Their research starts with the fact that in different parts of the United States, the stations found on specific channels vary. It turns out that when it comes to the total number of viewers, channel location matters a lot. People are more likely to watch stations in the lower positions. For historical reasons, Fox News and MSNBC have sometimes received advantageous channel positions—but sometimes have not.

Across recent time periods and in various parts of the country, Martin and Yurukoglu examined the relationships among channel positions, people’s intended votes, county-level presidential vote shares, and individual viewership. With several large data sets, they tested the effects on voting of watching Fox News and MSNBC. One of their findings is that Fox and MSNBC have both grown more ideologically defined, and Republicans and Democrats alike are aware of that. In 2000 and 2004, a typical Democrat was no more likely than a typical Republican to watch MSNBC. By 2008, a typical Democrat was 20 percentage points more likely to watch MSNBC. In 2004, a Republican was only 11 points more likely than a Democrat to watch Fox. By 2008, the gap had widened to more than 30 points.

The authors also found that both Fox and MSNBC have real effects on people’s likely votes. For those who end up watching Fox because of channel position, just four additional minutes of weekly viewing increases the probability of intending to vote for the Republican presidential candidate by 0.9 percentage points. For those

who watch MSNBC, four such extra minutes decreases the likelihood of intending to vote for the Republican presidential candidate by about 0.7 percentage points.

With one hour of viewing per week, the effects are greater. In 2008, an hour of MSNBC decreased the likelihood of a Republican vote by about 3.6 percentage points. In the same year, watching Fox for an hour increased the probability of a Republican vote by 3.5 points. At the level of individual voters, that may not be such a huge deal. But considered across the United States, the effects are large. The researchers estimate that in 2004 and 2008, if there had been no Fox News on cable television, the Republican vote share (as measured by voters' expressed intentions) would have been 4 percentage points lower. And if MSNBC had had CNN's more moderate ideology, the Republican share of the 2008 presidential vote intention would have been about 3 percentage points higher. (In general, Fox has more success in converting viewers than MSNBC does; it also has a much larger audience.)

These are disturbingly big numbers. Fox News and MSNBC do not merely attract like-minded people. They also heighten divisions among voters, contributing to political polarization—and they affect people's ultimate votes. We're speaking here of television rather than websites or social media, but the phenomenon is quite general.

APPROPRIATELY SLANTED

Many people are most willing to trust, and most enjoy, “appropriately slanted” stories about the events of the day. Their particular choices are designed to ensure that they can trust what they read. Or maybe they want to insulate themselves from opinions that they find implausible, indefensible, or invidious. Everyone considers some points of view beyond the pale, and we filter those out if this is at all possible. For many years, I lived in Chicago, and I loved the Chicago Bears. (I still love them—just a bit less.) When they were on national television, I turned off the sound and listened to the

local announcers. I did this not only because the local announcers were better but also because they were biased in the Bears' favor, and when the Bears did badly, their hearts broke along with mine.

Of course the fact that after people buy a new car, they often love to read advertisements that speak enthusiastically about the same car that they have just obtained. Those advertisements tend to be comforting because they confirm the wisdom of the decision to purchase that particular car. If you are a member of a particular political party or have strong convictions, you might want support, reinforcement, and ammunition, not criticism.

We can make some distinctions here. Members of some groups want to wall themselves off from most or all others simply in order to maintain a degree of comfort and possibly a way of life. For the same reason, some religious groups self-segregate. Such groups tolerate pluralism and are interested largely in self-protection; they do not have large ambitions or seek to proselytize to others. Political regimes can act similarly. In 2016, the Chinese government took new steps to restrict the activities of foreign nonprofits in China. An evident goal was to protect its own interests. Its focus was on China and its environs, not on remote lands.

Other groups have a self-conscious social project or even a kind of “combat mission” to convert others, and their desire to self-segregate is intended to strengthen their members' convictions in order to promote long-term recruitment and conversion plans. Terrorists operate in just this way. That is one reason that they use social media. Political parties sometimes think in similar terms, and they frequently ignore the views of others, except when they hold those views up to ridicule. When links are provided to other sites, it is often to show how dangerous or contemptible competing views really are. Tweeters and bloggers routinely do exactly this.

OVERLOAD, GROUPISM, AND *PLURIBUS PLURES*

In the face of dramatic increases in communications options, there is an omnipresent risk of information overload—too many options,

too many topics, too many opinions, a cacophony of voices. Indeed the risk of overload and the need for filtering go hand in hand. In my view, Bruce Springsteen's music is timeless, but his hit from the 1990s, "57 Channels and Nothing On," is hilariously out of date in light of the number of current programming options (just 57?). Filtering, often in the form of narrowing, is inevitable in order to avoid overload and impose some order on an overwhelming number of sources of information. Your Twitter feed will be restricted to what you want to see, and your Facebook friends will be a small subset of humanity.

By itself, this is not a problem. But when options are so plentiful, many people will take the opportunity to listen only or mostly to those points of view that they find most agreeable. For many of us, of course, what matters is that we enjoy what we see or read, or learn from it, and it is not necessary that we are comforted by it. But there is a natural human tendency to make choices with respect to entertainment and news that do not disturb our preexisting view of the world.

I am not suggesting that the Internet is a lonely or antisocial domain; it is hardly that. In contrast to television, many of the current options are extraordinarily social, dramatically increasing people's capacity to form bonds with individuals and groups that would otherwise have been entirely inaccessible. Facebook, Twitter, Instagram, Medium, and Vine provide increasingly remarkable opportunities, not for isolation, but for the creation of new groups and connections. This is the foundation for the concern about the risk of fragmentation.

Consider in this regard a revealing little experiment from a number of years ago.³ Members of a nationally representative group of Americans were asked whether they would like to read news stories from one of four sources: Fox (known to be conservative), National Public Radio (NPR, known to be liberal), CNN (often thought to be liberal), and the British Broadcasting Network (whose politics are not widely known to Americans). The stories came in different news categories: US politics, the war in Iraq, "race in America," crime, travel, and sports. For the first four categories, Republicans

chose Fox by an overwhelming margin. In contrast, Democrats split their "votes" among NPR and CNN—and showed a general aversion to Fox. For travel and sports, the divide between Republicans and Democrats was much smaller. And independents showed no preference for any particular source.

That's not exactly amazing, but there was a more surprising finding: *people's level of interest in the same exact news stories was greatly affected by the network label.* For Republicans, the identical headline became far more interesting and the story became far more attractive if it carried the Fox label. In fact, the Republican "hit rate" for the same news stories was three times higher when it was labeled "Fox." (Interestingly, the hit rate was doubled when sports and travel stories were so labeled.) Democrats showed a real aversion to stories labeled "Fox," and the CNN and NPR labels created a modest increase in their interest.

The overall conclusion is that Fox attracts substantial Republican support, and that Democratic viewers and readers take pains to avoid Fox—while CNN and NPR have noticeable but weak brand loyalty among Democrats. There is every reason to suspect that the result would generalize to online behavior—that people with identifiable leanings are consulting sources, including websites, that match their predilections, and are avoiding sources that do not cater to those predilections.

All this is just the tip of the iceberg. To cite some wise words from long ago:

Because the Internet makes it easier to find like-minded individuals, it can facilitate and strengthen fringe communities that have a common ideology but are dispersed geographically. Thus, particle physicists, Star Trek fans, and members of militia groups have used the Internet to find each other, swap information and stoke each others' passions. In many cases, their heated dialogues might never have reached critical mass as long as geographical separation diluted them to a few parts per million.⁴

It is worth underlining the idea that people are working to “stoke each others’ passions,” because that notion will play a large role in the discussion to follow. Of course, many of those with committed views on one or another topic—gun control, abortion, or immigration—are speaking mostly with each other. Social media feeds and linking behavior follow a similar pattern.

All this is perfectly natural and even reasonable. Those who visit what they see as appropriately slanted sites are likely to want to visit similarly slanted sites, and people who create a site with one point of view are unlikely to want to promote their adversaries. (Recall that collaborative filtering works because people tend to like what people like them tend to like.) And many people who consult sites with a distinctive perspective hardly restrict themselves to like-minded sources of information. But what we now know about individual behavior supports the general view that many people are mostly hearing more and louder echoes of their own voices. To say the least, this is undesirable from the democratic standpoint.

I do not mean to deny the obvious fact that any system that allows for freedom of choice will create some balkanization of opinion. Long before the advent of the Internet, and in an era of a handful of television stations, people made self-conscious choices among newspapers and radio stations. In any era, many people want to be comforted rather than challenged. Magazines and newspapers, for example, often cater to people with definite interests in certain points of view. Since the early nineteenth century, African American newspapers have been widely read by African Americans, and these newspapers offer significantly different coverage of common issues than white-oriented newspapers and also make dramatically different choices about what issues are important.⁵ Whites rarely read such newspapers.

What is emerging nonetheless counts as a significant change. With a dramatic increase in options and a greater power to customize comes a corresponding increase in the range of actual choices, and those choices are likely, in many cases, to match demographic characteristics, preexisting political convictions, or both. Of course

this has many advantages; among other things, it greatly increases the aggregate amount of information, the entertainment value of choices, and the sheer fun of the options. But there are problems too. If diverse groups are seeing and hearing quite different points of view, or focusing on quite different topics, mutual understanding might be difficult, and it might be increasingly hard for people to solve problems that society faces together.

Consider a few examples. Many Americans fear that certain environmental problems—abandoned hazardous waste sites, genetic engineering of food, climate change—are extremely serious and require immediate government action. But others believe that the same problems are imaginative fictions, generated by zealous and self-serving politicians. Many Americans think that most welfare recipients are indolent and content to live off the work of others. On this view, “welfare reform,” to be worthy of the name, consists of reduced handouts—a step necessary to encourage people to fend for themselves. But many other Americans believe that welfare recipients generally face severe disadvantages and would be entirely willing to work if decent jobs were available. On this view, welfare reform, understood as reductions in benefits, is an act of official cruelty. Many people believe that the largest threat to American security remains terrorism, and that if terrorism is not a top priority, catastrophic attacks are likely to ensue. Many others believe that while terrorism presents serious risks, the threat has been overblown, and that other problems, including climate change, deserve at least equal attention. To say the least, it will be difficult for people armed with such opposing perspectives to reach anything like common ground or make progress on the underlying questions. People might believe opposing “facts.” Consider how these difficulties will increase if people do not know the competing view, consistently avoid speaking with one another, and are unaware how to address divergent concerns of fellow citizens.

Numerous websites are created and run by hate groups and extremist organizations. They appear to be achieving a measure of success, at least if we measure this by reference to “hits.” Some

such groups have had hundreds of thousands or even millions of visitors. What is also striking is that many extremist organizations and hate groups provide links to one another, and expressly attempt to encourage both recruitment and discussion among like-minded people.

We can sharpen our understanding here if we attend to the phenomenon of *group polarization*. This phenomenon raises serious questions about any system in which individuals and groups make diverse choices, and many people end up in echo chambers of their own design. On the Internet, polarization is a real phenomenon; we might even call it “cyberpolarization.” To understand how it works, we need to investigate a little social science.

AN EXPERIMENT IN COLORADO

The term “group polarization” refers to something simple: after deliberation, people are likely to move toward a more extreme point in the direction to which the group’s members were originally inclined. With respect to the Internet and social media, the implication is that groups of like-minded people, engaged in discussion with one another, will typically end up thinking the same thing that they thought before—but in a more extreme form.

For an initial glimpse of the problem, let us put the Internet to one side and consider a small experiment in democracy that was held in Colorado in 2005.⁶ About sixty US citizens were brought together and assembled into ten groups, each consisting of six people. Members of each group were asked to deliberate on three of the most controversial issues of the day: *Should states allow same-sex couples to enter into civil unions? Should employers engage in “affirmative action” by giving a preference to members of traditionally disadvantaged groups? Should the United States sign an international treaty to combat global warming?*

As the experiment was designed, the groups consisted of “liberal” and “conservative” members—the former from Boulder, the latter from Colorado Springs. It is widely known that Boulder

tends to be liberal and that Colorado Springs tends to be conservative. The groups were screened to ensure that their members conformed to these stereotypes. In the parlance of election years, there were five “blue state” groups and five “red state” ones—five groups whose members initially tended toward liberal positions on the three issues, and five whose members tended toward conservative positions on those issues. People were asked to state their opinions individually and anonymously both before and after fifteen minutes of group discussion, and also try to reach a public verdict before making their final anonymous statements as individuals. What was the effect of discussion?

The results were simple. In almost every group, members ended up with more extreme positions after they spoke with one another. Discussion made civil unions more popular among liberals; discussion made civil unions less popular among conservatives. Liberals favored an international treaty to control global warming before discussion; they favored it more strongly after discussion. Conservatives were neutral on that treaty before discussion; they strongly opposed it after discussion. Mildly favorable toward affirmative action before discussion, liberals became strongly favorable toward it after discussion. Firmly negative about affirmative action before discussion, conservatives became even more negative about it afterward.

Aside from increasing extremism, the experiment had an independent effect: it made both liberal and conservative groups significantly more homogeneous—and thus squelched diversity. Before members started to talk, many groups displayed a fair bit of internal disagreement. The disagreements were reduced as a result of a mere fifteen-minute discussion. Even in their anonymous statements, group members showed far more consensus after discussion than before. It follows that discussion helped to widen the rift between liberals and conservatives on all three issues. Before discussion, some liberal groups were, on some issues, fairly close to some conservative groups. The result of discussion was to divide them far more sharply.

GROUP POLARIZATION

The Colorado experiment is vivid evidence of group polarization. The basic phenomenon has been found in over a dozen nations.⁷ Consider a few examples:

- Members of a group of moderately profeminist women became more strongly profeminist after discussion.⁸
- After discussion, citizens of France, initially critical of the United States and its intentions with respect to economic aid, became more critical still.⁹
- After discussion, whites predisposed to show racial prejudice offered more negative responses to the question of whether white racism is responsible for conditions faced by African Americans in US cities.¹⁰
- After discussion, whites predisposed not to show racial prejudice offered more positive responses to the same question.¹¹
- Republican appointees, on three-judge panels, show especially conservative voting patterns when they sit only with fellow Republican appointees; Democratic appointees show especially liberal voting patterns when they sit only with fellow Democratic appointees.¹²

The phenomenon of group polarization has conspicuous importance for social media and the Internet more generally, at least to the extent that groups with distinctive identities engage in within-group discussion. Effects of the kind just described should be expected with terrorist and hate groups as well as less extreme organizations of all sorts. If the public is balkanized, and if different groups are designing their own preferred communications packages, the consequence will be not merely the same but still more balkanization, as group members move one another toward more

extreme points in line with their initial tendencies. At the same time, different deliberating groups, each consisting of like-minded people, will be driven increasingly far apart, simply because most of their discussions are with one another.

It is true, of course, that most of us do not use the power to filter so as to become walled off from other points of view. (Even so, some people will do, and are doing, exactly that. I will turn to empirical issues in due course, but for now, let's just reiterate that on Facebook and Twitter, many people create something like echo chambers; they want their own views to be confirmed. Something similar is true for the millions of people who select newspapers, radio stations, and television channels because they can hear some version of their own voices.)

This is sufficient for significant polarization to occur, and cause serious social risks. In general, *it is precisely the people most likely to filter out opposing views who most need to hear them*. Social media make it easier for people to surround themselves (virtually) with the opinions of like-minded others and insulate themselves from competing views. For this reason alone, they are a breeding ground for polarization, and potentially dangerous for both democracy and social peace.

WHY POLARIZATION?

There have been three main explanations for group polarization. Massive evidence now supports all of them.

Persuasive arguments and information: The first explanation emphasizes the role of persuasive arguments and information. The intuition here is simple. Any individual's position on any issue is a function, at least in part, of which arguments seem convincing. On balance, that is good news. People *should* pay attention to arguments. If you are like most people, you are likely to notice the information held and revealed by those with whom you interact. And if your position is going to move as a result of group discussion, it is likely to move in the direction of the most persuasive position

defended within the group, taken as a whole. The most persuasive position will be defined, in large part, by the reasonableness and number of arguments offered in its favor.

Here's the central point: if the group's members are already inclined in a certain direction, they will offer a disproportionately large number of arguments tending in that same direction, and a disproportionately small number of arguments tending the other way. The consequence of discussion will naturally be to move people further in the direction of their initial inclinations. Thus, for example, a group whose members lean in favor of the nation's current leader will, in discussion, provide a wide range of arguments in that leader's favor, and the arguments made in opposition to them will be both fewer and weaker. The group's members, to the degree that they shift, will shift toward a more extreme position in favor of the current leader. And the group as a whole, if a group decision is required, will move not to the median position but instead to a more extreme point.

On this account, the central factor behind group polarization is the existence of a *limited argument pool*—one that is skewed (speaking purely descriptively) in a particular direction. It is easy to see how shifts might happen with discussion groups online; consider a group of Democrats or Republicans, terrorists, or environmentalists. The point helps to explain what happens every day on social networks. If your Twitter feed consists of people who think as you do, or if your Facebook friends share your convictions, the argument pool will be sharply limited. Indeed, shifts should occur with individuals not engaged in discussion but instead consulting only ideas—on radio, television, or the Internet—to which they are pre-disposed. The tendency of such consultations will be to entrench and reinforce preexisting positions—often resulting in extremism. If people who watch Fox News are drawn further in a conservative direction, or if people who watch Russian state television end up with less enthusiasm for the United States, the relevant argument pool is probably playing a large role.

Reputational considerations. The second mechanism, involving people's concern for their reputations, begins with the reasonable

suggestion that people want to be perceived favorably by other group members and also perceive themselves favorably. Once they hear what others believe, they frequently adjust their positions in the direction of the dominant position. The German sociologist Elisabeth Noell-Neumann has used this idea as the foundation for a general theory of public opinion, involving a "spiral of silence" in which people with minority positions silence themselves, potentially exciting those positions from society over time.¹³ That happens in authoritarian societies, but it can also take place in democracies. Sometimes it is a good thing: people who believe that the sun goes around the earth, or that slavery was a fine idea, may end up self-silencing, as societies converge on scientific and moral truths, making false or invidious beliefs disappear (or nearly so). But the spiral of silence is not always benign. For present purposes, the central point is that when people care about their reputations, what they say within group discussions will be affected. As a result, groups can become more extreme.

Suppose, for instance, that people in a certain group tend to be sharply opposed to a certain war, continued reliance on fossil fuels, and gun ownership, and that they also want to be seen as sharply opposed to all these policies. If they are in a group whose members are also sharply opposed to these things, they might well shift in the direction of even stronger opposition after they see what other group members think. In countless studies, exactly this pattern is observed. Of course people will not shift if they have a clear sense of what they think and will not let the opinions of others move them. But most people, most of the time, are not so fixed in their views.

The point offers an account of the likely effects of exposure to ideas and claims on television, radio, and social media—even in the absence of a chance for interaction. Note that group polarization occurs *merely on the basis of exposure to the views of others*. Discussion is not necessary. The "mere exposure" effect means that polarization is likely to be a common phenomenon in a balkanized speech market.

Imagine that conservatives are visiting conservative sites; that liberals are visiting liberal sites; that environmentalists are visiting sites dedicated to establishing the risks of genetic engineering and climate change; that critics of environmentalists are visiting sites whose purpose is to expose frauds allegedly perpetrated by environmentalists; and that people inclined to racial hatred are visiting sites that express racial hatred. To the extent that these exposures are not complemented by exposure to competing views, group polarization will be the inevitable consequence.

✓ **Confidence, extremism, and corroboration.** The most intriguing explanation of group polarization stresses the close links among confidence, extremism, and corroboration by others.¹⁴ On many issues, including political ones, people are really not sure what they think, and their lack of certainty inclines them toward the middle. We might feel tentative about complex matters—the effects of an increase in the minimum wage, the proper approach to climate change, or what to do about some dangerous international situation. Our views are moderate and provisional.

It is only as people gain confidence that they become more extreme in their beliefs. For better or worse, they can be radicalized, even if the ultimate conclusion is not all that radical. Agreement from others tends to increase confidence, and for this reason like-minded people, having deliberated with one another, become more convinced that they are right—and hence more extreme. Even in mundane contexts, involving the attractiveness of people in slides and the comfort of chairs, the opinions of ordinary people in experiments become more extreme simply because their views have been corroborated, and because they become more confident after learning that others share their views.¹⁵

This is a quite-fundamental point, and it helps explain what happens on Twitter every day. If you learn that other “people like you” like a certain band, a particular movie, an identifiable political position, or a specific candidate, you might well follow their lead. Indeed, if you learn that “people like you” tend to have a certain position on national security or Social Security reform, you

might well end up adopting their position, and perhaps doing so with great confidence, even if you haven’t much thought about the question independently. When people find that others share their initial inclination, they often become more confident and therefore more extreme. Consider in this regard the effects of social media platforms on which people’s views end up being constantly reaffirmed by like-minded types.

I am using the idea of radicalization pretty loosely here. You can become radicalized in the sense that you come to believe, firmly, a position that is within the political mainstream—for example, that your preferred political candidate is not just the best but immeasurably better than the alternatives, and that any other choice would be catastrophic. On both sides, this happened in the 2016 campaign between Donald Trump (#CrookedHillary) and Hillary Clinton (#NeverTrump). Of course you can become radicalized in more disturbing ways; we will get to that in due course.

THE ENORMOUS IMPORTANCE OF GROUP IDENTITY

With respect to polarization, perceptions of identity and group membership are important, both for communications in general and social media in particular. Group polarization will significantly increase if people think of themselves as part of a group having a shared identity and a degree of solidarity. If they think of themselves in this way, group polarization is both more likely and more extreme.¹⁶ If, say, a number of people in an online discussion group think of themselves as opponents of high taxes, advocates of animal rights, or critics of the Supreme Court, their discussions are likely to move them in extreme directions, simply because they understand each other as part of a common cause. Similar movements should be expected for those who listen to a radio show known to be conservative, or who watch a television program dedicated to traditional religious values or exposing white racism. A lot of evidence so suggests.¹⁷

Group identity matters in another way. Suppose that you are participating in an online discussion, but you think that other

group members are significantly *different* from you. If so, you are less likely to be moved by what they say. If, for example, other group members are styled “Republicans,” and you consider yourself a Democrat, you might not shift at all—even if you would indeed shift as a result of the same arguments if you were all styled “voters,” “jurors,” or “citizens.” Thus a perception of shared group identity will heighten the influence of others’ views on your own, whereas a perception of unshared identity and relevant differences will reduce that effect, and possibly even eliminate it.

These findings should not be surprising. Recall that in ordinary cases, group polarization is a product of limited argument pools, reputational considerations, and the effects of corroboration. If this is so, it stands to reason that when group members think of one another as similar along a salient dimension, or if some external factor (politics, geography, race, or sex) unites them, group polarization will be heightened. If identity is shared, persuasive arguments are likely to be still more persuasive; the identity of those who are making those arguments gives them a kind of credential or boost. If identity is shared, social influences will have even greater force. People do not like their reputations to suffer in the eyes of those who seem most like them. And if you think that group members are in some relevant sense different from you, their arguments are less likely to be persuasive, and social influences may not operate as much or at all. If “people like you” support your initial inclination, you will become more confident. But if “people not like you” support that inclination, you might become less confident and start to rethink your position. If your political opponents—those whom you think most confused and destructive—think that your position is right, you might end up thinking that it is wrong.

GROUP POLARIZATION ONLINE

Group polarization is unquestionably occurring online. From the evidence discussed so far, it seems plain that the Internet is serving, for many, as a breeding ground for extremism, precisely because

like-minded people are connecting with greater ease and frequency with one another, and often without hearing contrary views. Repeated exposure to an extreme position, with the suggestion that many people hold it, will predictably move those exposed, and likely predisposed, to believe in it.

One consequence can be a high degree of fragmentation, as diverse people, not originally fixed in their views and perhaps not so far apart, end up in extremely different places simply because of what they are reading and viewing. (Recall the Colorado experiment.) Another consequence can be a high degree of error and confusion. YouTube is a lot of fun, and in a way it is a genuine democratizing force, but there is a risk that isolated clips, taken out of context, will lead like-minded people to end up with a distorted understanding of some issue, person, or practice.

Call this the “Jon Stewart strategy.” Stewart, once a nightly television show host, is immensely talented and even a kind of genius. As a comedian, one of his most successful approaches has been to display some brief clip in which the speaker seems evil, ugly, foolish, or (most often) idiotic. If everyone is laughing together (at, say, a famous politician), the clip will seem to capture the person; it will show his essence. That might be fair—but it usually isn’t (except perhaps on a comedy show). In some contexts, it’s completely unfair—a kind of violation.

If you follow any human being on video for a lengthy period of time, you will almost certainly be able to uncover a clip in which he appears to be evil, ugly, foolish, or idiotic—all the more so if the clip is played over and over again. Repetition can make anyone seem ridiculous. That’s a terrific comic strategy, but it can be used for political purposes as well—and it certainly is. If you want to get people to dislike or ridicule a political opponent, here’s a clue: use the Jon Stewart strategy. But if you do, you should not be proud of yourself.

A number of studies have shown group polarization in online settings or those that mimic them. An especially interesting experiment finds particularly high levels of polarization when group members meet relatively anonymously and group identity is

emphasized.¹⁸ From this experiment, it is reasonable to conclude that polarization is highly likely to occur, and to be extreme, under circumstances in which group membership is made salient and people have a high degree of anonymity. These are, of course, frequent features of online deliberation.¹⁹

Consider in this regard a revealing study not of extremism but of serious errors within working groups, both face-to-face and online. The purpose of the study was to see how groups might collaborate to make personnel decisions. Résumés for three candidates applying for a marketing manager position were placed before several groups. The experimenters rigged attributes of the candidates so that one applicant was clearly best matched for the job described. Packets of information, each containing only a subset of information from the résumés, were given to the subjects, so that each group member had only part of the relevant information. The groups consisted of three people, some operating face-to-face, and others operating online.

Two results were especially striking. First, group polarization was common, in the sense that groups ended up in a more extreme position in line with members' pre-deliberation views. Second, almost none of the deliberating groups made what was conspicuously the right choice!

The reason is that they failed to share information in a way that would permit the group to make the correct decision. In online groups, the level of mistakes was especially high, for the simple reason that members tended to share positive information about the emerging winning candidate and negative information about the losers, while also suppressing negative information about emerging winners and positive information about the emerging losers. These contributions served to "reinforce the march toward group consensus rather than add complications and fuel debate."²¹ In fact, this tendency was *twice* as large within the online groups. There is a warning here about the consequences of the Internet for democratic deliberation.

It is true that many people go online and use social media to learn about alternative positions, not merely to reinforce their

existing tendencies. It is certainly possible for more information and less polarization to result from the increase in available sources. This happens every day. But the study just described offers a clear warning. When people deliberate together, they often give disproportionate weight to "common knowledge"—information that they all share in advance. By contrast, they frequently give too little weight to unshared information—information that is held by one or only a few people. There is every reason to think that the same asymmetry is occurring online.

HASHTAG NATION AND HASHTAG ENTREPRENEURS

Consider #Syria, #BlackLivesMatter, #BobDylan, #ObamaIsAMuslim, #StarWars, and #Republic. All those are convenient sorting mechanisms. Do hashtags contribute to polarization?

The history here is illuminating. Hashtags were originally used in the Internet Relay Chats of the early 1990s, as a way to organize groups within chats. Twitter was the first social media site to use them. In 2007, just a year after the service launched, hashtags were proposed by Chris Messina, an open-source advocate, as an ad hoc strategy for sorting through conversations. He expanded on the proposal in a blog post a few days later, clarifying that hashtags should not be used for creating groups on Twitter ("I'm not at all convinced that groups . . . are ultimately a good idea or a good fit for Twitter") but instead for "improving *contextualization*, *content filtering* and *exploratory serendipity* within Twitter."²² The mechanism proved both popular and useful, and it has been incorporated into many other social media platforms, including Instagram, Tumblr, and Vine. It is also used in e-mail. It is instructive that Messina referred at once to two potentially contradictory ideas—content filtering and exploratory serendipity—though it is not exactly clear what he meant by the latter.²³

For my purposes, the question is whether hashtags produce one or the other, or instead both. Thus far, content filtering clearly seems to be the dominant effect. Hashtags typically signal subject

matter, and they may also signal a point of view. If a hashtag says #DemocratsAreCommunists or #RepublicansAreFascists, people know the kind of thing that they will find. But if a hashtag says #AffirmativeAction, #Polygamy, or #TurkeyCoup, you know the topic, but you cannot be sure what perspective you will discover. When serendipity occurs, it is typically because a hashtag leads users to perspectives and points of view that they did not expect. The empirical literature is constantly expanding. Consider here a few representative examples.

In a 2016 study, Deen Freelon and his colleagues gathered 40.8 million tweets that included #BlackLivesMatter and related terms and hashtags (generally consisting of the full and hashtagged names of twenty African Americans killed by police over the relevant one-year period).²⁴ One of their central findings is that activists used hashtags, above all #BlackLivesMatter, for the purposes of both education and amplification, usually seeking to draw attention to what they saw as a form of structural racism. As one activist put it, “Getting something on Twitter means that people are talking, they are conscious. And that consciousness can lead to action.”²⁵ Freelon and his coauthors found that social media posts played a large role in spreading identifiable narratives and accounts of killings by the police. The Black Lives Matter movement had a significant impact on both opinion and action in many cities as well as at the national level. #BlackLivesMatter mattered.

Similar polarization can be seen with the use of #AllLivesMatter, a hashtag whose purpose was to offer a competing narrative to that reflected in #BlackLivesMatter, to the effect that it is partisan or parochial, or even racist, to single out “black lives” for special emphasis. The use of #AllLivesMatter is for identifiable purposes (usually conservative), and it appeals to people with identifiable views, critical of #BlackLivesMatter.²⁶ Research has found that on social media, “the only other lives that were significantly discussed within #AllLivesMatter are the lives of law enforcement officers, particularly during times in which there is heavy protesting.”²⁷ It is clear

that the #AllLivesMatter hashtag arose to create an ideologically defined narrative, clearly showing polarization.

An influential study by Sarida Yardi and Dana Boyd explored thirty thousand tweets about the 2009 shooting of George Tiller, a late-term abortion doctor, and the subsequent conversations among anti-abortion and pro-choice advocates. They found that many users adopted hashtags that signaled a specific view about the debate. Importantly, users with the same ideologies were most likely to interact with each other, but not with those with competing views. Moreover, and consistent with my concerns here, Yardi and Boyd demonstrated that “replies between like-minded individuals strengthen group identity whereas replies between different-minded individuals reinforce ingroup and outgroup affiliation.”²⁸

Nonetheless, serendipitous encounters did happen, in the sense that those searching through various hashtags were likely to come across a diverse set of viewpoints. But meaningful discussion across ideological lines remained extremely rare; people with differing points of view were more likely to talk over or past each other than to engage in substantive conversation. In light of the character limit on Twitter, that is not exactly surprising.

An especially interesting line of research explores how members of Congress are using hashtags, often framing issues in their preferred ways and promoting echo chambers that serve their interests.²⁹ A central finding is that while Democrats and Republicans discuss overlapping issues, they use notably different hashtags. Among Democrats, the most popular issues in the relevant period include health care (#ACA, for the Affordable Care Act), student loans (#DontDoubleMyRate), and employment (#JOBS). The Republicans’ top issues are not so different: employment (#4jobs), themselves (#rcot), and health care (#Obamacare). But the two parties do use radically different frames. The term #ACA, preferred by Democrats, has a positive or neutral valence about the Affordable Care Act, whereas #Obamacare and #Fullrepeal, favored by Republicans, are clearly meant to be negative.

Like political activists, members of Congress can be seen as hashtag entrepreneurs. They choose a particular frame: #AllLivesMatter, #TheSystemIsRigged, or #CorruptHillary. They hope that it will attract widespread interest, helping to construct both emotions and beliefs. Hashtag entrepreneurship is increasingly central to modern political life.

There is no question that we will learn much more over time. But two conclusions seem plain. First, hashtags work as engines for group polarization (and also cybercascades; see chapter 4). Second, hashtags create communities of interest around identifiable subjects, and those communities can include diverse views. Both of these effects involve a high degree of sorting and filtering, but the second need not produce polarization, and it might result in encounters with widely diverse points of view.

A GLANCE AT POLITICIANS, INCLUDING DONALD TRUMP

All over the world, politicians have been using social media, often to create the conditions for polarization effects. A full discussion would require a book all its own, so consider just two prominent examples. In the United States, Barack Obama was the first president to use social media to promote his campaign. The numbers from the 2008 general election tell the tale: on every social media platform, Obama did far better than his rival, John McCain. Obama had more than 2 million Facebook supporters, compared to 600,000 for McCain. Obama counted 112,000 active Twitter followers, as opposed to 4,600 for McCain. On YouTube, Obama had more videos, subscribers, and video views than did McCain by about four to one.³⁰ In developing his social media campaign, Obama hired tech entrepreneurs, including Chris Hughes, a cofounder of Facebook.³¹ At the height of his campaign, Obama employed 100 staff members devoted to social media.³² With the help of social media, podcasting, and mobile messaging, Obama managed to capture 70 percent of the group of eighteen- to twenty-five-year-old voters—the highest margin since exit polling started in 1976.³³ Of course we cannot

say that his success is attributable to his use of social media. But the Obama campaign used Facebook and Twitter in a way that undoubtedly spurred a high degree of group polarization (with greater enthusiasm among his supporters).

In the 2012 campaign, Obama mobilized social media resources against Mitt Romney. Again, he outperformed his rival by wide margins. He spent ten times more on his digital campaign spending (\$47 million versus \$4.7 million), garnered twice as many Facebook “likes,” and had twenty times more retweets.³⁴ True, none of that is the best measure of popularity, but it does show far better use of social media.

In the 2016 election, Donald Trump took a unique path. Perhaps because of his background in entertainment, he was able to use social media to excellent effect, with his own preferred hashtags, above all #MakeAmericaGreatAgain but also #CrookedHillary. Even before the Republican convention, Trump had about ten million likes on Facebook, and his YouTube channel had over six million views and forty thousand subscribers.³⁵ Trump maintained a presence on Vine, Periscope, and Instagram, but his use of Twitter was a defining feature of his campaign, including insults that went viral (“Little Marco,” “Lyn’ Ted,” and again, “Crooked Hillary”). He live-tweeted himself, and in a relatively short period, he came to have more than eleven million followers. That number understates the impact of his tweets, many of which received attention in the national media and hence were greatly amplified. There is no doubt that his activity on Twitter put him at the center of what was, for many, an engine for group polarization—and helped vault him to the presidency.

FRAGMENTATION, POLARIZATION, RADIO, AND TELEVISION

An understanding of group polarization casts light not only on online behavior but also on the potential effects of contemporary radio and television, at least if stations are numerous and many take a well-defined point of view. Recall that mere exposure to the positions of others creates group polarization. It follows that this effect will be

at work for nondeliberating groups, in the form of collections of individuals whose communications choices go in the same direction, and who do not expose themselves to alternative positions. The same process is likely to occur for newspaper choices. If some people are reading the liberal newspaper, and others are reading the conservative one, polarization is inevitable. When they are working well, general-interest intermediaries have a distinctive role here by virtue of their effort to present a wide range of topics and views. Polarization is far less likely to happen when such intermediaries dominate the scene. A similar observation can be made about the public forum doctrine. When diverse speakers have access to a heterogeneous public, individuals and groups are less likely to be able to insulate themselves from competing positions and concerns. Fragmentation is correspondingly less likely.

Group polarization also raises more general issues about communications policy. Consider the “fairness doctrine,” now largely abandoned, but once requiring radio and television broadcasters to devote time to public issues and allow an opportunity for opposing views to speak. The latter prong of the doctrine was designed to ensure that listeners would not be exposed to any single view. If one view was covered, the opposing position would have to be allowed a right of access.

When the Federal Communications Commission abandoned the fairness doctrine, it did so on the ground that much of the time, this second prong led broadcasters to avoid controversial issues entirely and to present views in a way that suggested a bland uniformity. Subsequent research has indicated that the commission was right. The elimination of the fairness doctrine has indeed produced a flowering of controversial substantive programming, sometimes involving extreme views of one kind or another; consider talk radio.³⁶

Typically this is regarded as a story of wonderfully successful de-regulation, and in general that is correct. The effects of eliminating the fairness doctrine were precisely what was sought and intended. Those effects are indeed good, and they should be celebrated. But

if we attend to the problem of group polarization, the evaluation is a bit more complicated. On the good side, the existence of diverse pockets of opinion enriches society’s total argument pool, potentially to the benefit of all of us. At the same time, the growth of a wide variety of issue-oriented programming—expressing strong, often extreme perspectives, and appealing to dramatically different groups of listeners and viewers—undoubtedly creates group polarization. All too many people are now exposed mostly to louder echoes of their own voices, resulting, on too many occasions, in misunderstanding and enmity. On one view, it is better for people to hear fewer controversial positions than for them to hear a single such view, stated over and over again.

I do not suggest or believe that the fairness doctrine should be restored. Law professor Heather Gerken has rightly drawn attention to “second-order diversity”—the kind of diversity that comes when society consists of many institutions and groups, some of which have little in the way of internal diversity.³⁷ As Gerken has shown, we all benefit from a decentralized system in which different groups have different predispositions and sometimes go to different extremes. Instead of seeking diversity *within* each group, we might want diverse groups, even if many or most demonstrate little internal diversity.

The same goes for communications outlets (and social media as well). If some radio shows press quite conservative arguments, and others press quite liberal ones, we might all be able to benefit from what emerges. It’s a strong argument. But at the very least, there is a risk in the current situation that too many people will choose to insulate themselves from exposure to views that are more moderate, extreme in another direction, or in any case different from their own.

IS GROUP POLARIZATION BAD? OF ENCLAVE DELIBERATION

Notwithstanding the tenor of the discussion thus far, we cannot always say, from the mere fact of group polarization, that there has been a movement in the *wrong* direction. In some cases, the more

extreme tendency is better rather than worse. It might even be much better. Indeed, group polarization has helped fuel many movements of great value—including, for example, the civil rights movement, the antislavery movement, the disability rights movement, the movement for equality between men and women, and the movement for same-sex marriage. All these movements were extreme in their time, and within-group discussion certainly bred greater extremism, but extremism need not be a word of opprobrium. If greater communications choices produce greater extremism, society may be better off as a result.

One reason is that when many different groups are deliberating with one another, society will hear a far wider range of views; recall the idea of second-order diversity. Even if the “information diet” of many individuals is homogeneous or insufficiently diverse, society as a whole might have a richer and fuller set of ideas. This is another side of the general picture of social fragmentation. It suggests some large benefits from pluralism and diversity—benefits even if individuals customize and cluster in groups. Another benefit of clustering is that it can counteract “epistemic injustice,” in which people lack a sufficient ability to interpret their experiences, and in which they cannot obtain a hearing for those experiences.³⁸ Social media can counteract that injustice. In fact, they do that every day.

We might define *enclave deliberation* as that form of deliberation that occurs within more or less insulated groups, in which like-minded people speak mostly to one another. The Internet, including social media, makes it much easier (and less costly) to engage in enclave deliberation. Your Facebook page might itself allow for a form of such deliberation. It is obvious that enclave deliberation can be extremely important in a heterogeneous society, not least because members of some groups tend to be especially quiet when participating in broader deliberative bodies.

In this light, a special advantage of enclave deliberation is that it promotes the development of understanding, knowledge, and positions that would otherwise be invisible, silenced, or squelched in

general debate. The efforts of marginalized groups to exclude outsiders, and even of political parties to limit their primaries to party members, might be justified in similar terms. Even if group polarization is at work—perhaps *because* group polarization is at work—enclaves, emphatically including those produced by social media, can provide a wide range of social benefits, not least because they greatly enrich the social “argument pool.” There is no question that Twitter in particular is doing exactly that.

A central empirical point here is that in deliberating bodies, high-status members tend to speak more than others, and their ideas are more influential—partly because low-status members sometimes lack confidence in their own abilities, and partly because they fear retribution.³⁹ For example, women’s ideas are often less influential and sometimes are “suppressed altogether in mixed-gender groups.”⁴⁰ In many circumstances, cultural minorities have disproportionately little influence on decisions by culturally mixed groups. In light of the inevitable existence of some status-based hierarchies, it makes sense to be receptive to deliberating enclaves in which members of multiple groups may speak with one another and develop their views. Online communication is especially valuable insofar as it makes this easier.

But there is also a serious danger in such enclaves. The danger is that members will move to positions that lack merit but are predictable consequences of the particular circumstances of enclave deliberation. In extreme cases, enclave deliberation may even put social stability at risk. And it is impossible to say, in the abstract, that those who sort themselves into enclaves will generally move in a direction that is desirable for society at large or even its own members.

It is easy to think of illustrations to the contrary—as, for instance, in the rise of Nazism, hate groups, terrorism, and cults of various sorts. If we take the idea of cults broadly enough, we can find them all over social media, such as people who believe that Barack Obama was not born in the United States, Israel was responsible for the attacks of 9/11, vaccines cause autism, or Elvis

Presley is still alive. And of course terrorist organizations present special challenges on this count—a point to which I will return.

ENCLAVES AND A PUBLIC SPHERE

Whenever group discussion tends to lead people to more strongly held versions of the same view with which they began, there is legitimate reason for concern. Certainly this does not mean that the discussions can or should be regulated. But it does raise questions about the idea that “more speech” is necessarily an adequate remedy for bad speech—especially if many people are inclined and increasingly able to wall themselves off from competing views. In democratic societies, a possible response is suggested by the public forum doctrine, whose most fundamental goal is to increase the likelihood that at certain points, there is an exchange of views between enclave members and those who disagree with them. It is total or near-total self-insulation, rather than group deliberation as such, that carries with it the most serious dangers, often in the highly unfortunate (and sometimes literally deadly) combination of extremism with marginalization.

To explore some of the advantages of heterogeneity, let us engage in a thought experiment. Imagine a deliberating body consisting not of a subset of like-minded people but instead all citizens in the relevant group; this may mean all citizens in a community, a state, a nation, or even the world. Imagine that through the magic of the computer, everyone can talk to everyone else. By hypothesis, the argument pool would be enormous. It would be limited only to the extent that the set of citizen views was similarly limited. Of course, reputational influences would remain. If you are one of a small minority of people who deny that global warming is a serious problem, you might decide to join the crowd. But when deliberation revealed to people that their private position was different in relation to the group from what they thought it was, any shift would be in response to an accurate understanding of all relevant citizens, and not a product of a skewed sample. And in fact, we can think of some online efforts as attempting to approximate

this thought experiment. Wikipedia, for example, allows anyone to be an editor (within limits), and the theory is that countless people can contribute their dispersed information to produce a resource offering an immense amount of human knowledge. This largely successful effort, resulting in a single product to which all can contribute, might be compared with deliberating enclaves of like-minded people.

The thought experiment, or the Wikipedia example, does not suggest that a fragmented or balkanized speech market is always bad, or that the hypothesized, all-inclusive deliberating body would be ideal. It would be foolish to suggest that all discussion should occur, even as an ideal, with all others. The great benefit of deliberating enclaves is that positions may emerge that otherwise would not, and they deserve to play a larger role within both the enclave and the heterogeneous public. Properly understood, the case for deliberating enclaves is that they will improve social deliberation, democratic and otherwise, precisely because enclave deliberation is often required for incubating new ideas and perspectives that will add a great deal to public debate. Social media, including Facebook, Twitter, Instagram, and Snapchat, can be exemplary here.

But for these improvements to take place, members must not insulate themselves from competing positions. At the very least, any such attempt at insulation must not be a prolonged affair. The phenomenon of group polarization suggests that with respect to communications, consumer sovereignty might well produce serious problems for individuals and society at large—and these problems will occur by a kind of iron logic of social interaction.

NO POLARIZATION AND DEPOLARIZATION

It is not exactly news that political candidates and their supporters are using the Internet, including social media, to their advantage. What is perhaps more interesting is that candidates for public office and their supporters have also been using the Internet, including social media, in a way that shows an intuitive understanding

of group polarization. Their sites operate as forums in which like-minded people congregate and adopt shared positions about policies, adversaries, and their candidates. Candidates try to produce echo chambers—their own online version of the Colorado experiment in which social interactions produce more consensus and enthusiasm—eventually yielding both time and money. The mechanisms discussed here may or may not cause harm, but those who are aware of them can certainly use them strategically. In the 2016 campaign for the presidency, Donald Trump showed a keen working knowledge of social influences and group polarization, constantly emphasizing how popular he was, and pointing constantly to the polls as evidence.

In certain circumstances, however, polarization can be decreased or even eliminated. No shift should be expected from people who are quite confident about what they think, and who are simply not going to be moved by what they hear from other people. If you are entirely sure of your position with respect to nuclear power—if you are confident not only of your precise view but of the degree of confidence with which you ought to hold it too—the positions of other people will not affect you. People of this sort will not shift by virtue of any changes in the communications market.

I have mentioned that federal judges are prone to polarization: Republican appointees show especially conservative voting patterns when sitting with fellow Republican appointees, and Democratic appointees show especially liberal voting patterns when sitting with fellow Democratic appointees. But on two issues, federal judges appear to be uninfluenced by their peers: abortion and capital punishment.⁴¹ On these issues, federal judges show essentially the same voting patterns regardless of whether they are sitting with zero, one, or two judges appointed by a president of the same political party as the president who appointed them. Apparently there is no polarization on abortion and capital punishment simply because judges' views are deeply held and entrenched. We can easily imagine other issues about which ordinary people are similarly unlikely to be affected by group members. In the political

domain, polarization finds its limits here, whatever candidates attempt to do.

With artful design of deliberating groups, moreover, it is possible to produce *depolarization*—shifts toward the middle, away from the extremes. Suppose, for example, that a group of twelve people is constructed so as to include six people who have one view and six people who think the opposite—say, half the group's members believe that air pollution from particulate matter is a serious problem, while the other half think that it is not. If most of the members do not have entirely fixed positions, there is likely to be real movement toward the middle. The persuasive arguments view helps explain why this is so. By hypothesis, the "argument pool" includes an equal number of claims both ways. If people are willing to dismiss those who disagree with them, depolarization might not occur. A group consisting of three Israelis and three members of Hamas might not depolarize; group members might simply dismiss the views of those who disagree. But for many questions, people are likely to listen to one another and hence depolarization is possible.

Of course mixed groups are no panacea. Typically group members end up at a more extreme point in line with the predeliberation tendency. No less than like-minded groups, mixed groups can polarize.⁴² More generally, confronting opposing positions can dampen political participation, in part because people who become more ambivalent and more uncertain about their own views might simply stand to one side.⁴³

But mixed groups have been shown to have two desirable effects. First, exposure to competing positions generally increases political tolerance.⁴⁴ After hearing a variety of views, including those that diverge from their own, many people are more respectful of alternative positions, and more willing to consider them to be plausible or legitimate. An important result of seeing a political conflict as legitimate is a "greater willingness to extend civil liberties to even those groups whose political views one dislikes a great deal."⁴⁵

Second, mixing increases the likelihood that people will be aware of competing rationales and see that their own arguments might

be met with plausible counterarguments.⁴⁶ This effect is especially pronounced for those who start with a “civil orientation toward conflict,” in the sense that they are committed to a degree of social harmony and are willing to acknowledge in advance that dissenting views should be expressed.⁴⁷ These desirable effects of deliberation within mixed groups will not be realized in any deliberative process in which people are sorted or sort themselves into politically homogeneous groups.

There is a valuable lesson about possible uses of communications technologies to produce convergence and possibly even learning among people who disagree with one another. If people hear a wide range of arguments, they are more likely to be moved in the direction of those who disagree with them, at least if the arguments are reasonable, and if those who disagree cannot easily be dismissed as untrustworthy or unreliable.

BALANCED PRESENTATIONS, UNBALANCED VIEWS

Unfortunately, a diverse collection of studies also establishes that when people are exposed to balanced information, movements toward the middle might not occur. For over three decades, it has been well known that such information might not produce consensus, even if it appears directly to address the concerns that led in the first place to divided views. The underlying phenomenon is typically described as *biased assimilation*.⁴⁸ The basic idea is that people assimilate information in a way that is skewed in the direction of support for their prior beliefs.⁴⁹

The initial studies involved capital punishment.⁵⁰ People were asked to read several studies arguing both in favor of and against the view that capital punishment deterred crime. A key finding was that both supporters and opponents of the death penalty were far more convinced by the studies supporting their own beliefs than by those challenging them. After reading the opposing studies, both sides reported that their beliefs had shifted toward a stronger commitment to what they originally thought.

One consequence is that the two sides were more polarized than they were before they began to read. That looks a lot like group polarization—but it happened in response to balanced presentations, not echo chambers.

Similar findings have been made in many contexts.⁵¹ For example, experiments provided people with competing views on the questions whether sexual orientation has a genetic component and whether same-sex couples are likely to be good parents. After receiving information on both sides of those issues, people’s preexisting beliefs were *strengthened*—and there was greater, not less, polarization on the issue of same-sex relationships.⁵² In studies of this kind, people are provided with “pro” and “con” arguments, and at least under certain conditions, provision of such arguments leads to an increase in polarization.

All this complicates the story I am telling here, and it has clear implications for social media: if people online encounter a wide range of views, they might not depolarize at all—at least if they listen to people with whom they are inclined to agree and dismiss everybody else.

WHEN CORRECTIONS BACKFIRE

Suppose that a society is divided on some proposition. The first group believes *A*, and the second group believes not-*A*. Suppose that the first group is entirely correct, and that the second group is full of nonsense. Finally, suppose that truthful information is provided, not from members of the first group, but from some independent source, in support of *A*. It would be reasonable—you might think—to suppose that the second group would come to believe *A*.

But in important settings, the opposite happens. The second group continues to believe not-*A*, and even more firmly than before. The result of the correction is to increase polarization.

In one experiment, people were exposed to a mock news article in which President George W. Bush defended the Iraq War, in part by suggesting (as he in fact did) that there “was a risk, a real risk,

that Saddam Hussein would pass weapons or materials or information to terrorist networks.⁵³ After reading this article, people in the experiment read about the Duelfer Report, which documented the lack of weapons of mass destruction in Iraq. People were then asked to state their agreement, on a five-point scale (from “strongly agree” to “strongly disagree”), with the statement that Iraq “had an active weapons of mass destruction program, the ability to produce these weapons, and large stockpiles of WMD.”

The effect of the correction greatly varied by political ideology. For very liberal subjects, there was a modest shift in favor of disagreement with this statement; the shift was not significant, because these subjects already tended to disagree with it. But for those who characterized themselves as conservative, there was a statistically significant shift in the direction of *agreeing* with the statement. “In other words, the correction backfired—conservatives who received a correction telling them that Iraq did not have WMD were more likely to believe that Iraq had WMD than those in the control condition.”⁵⁴ It follows that the correction had a polarizing effect; it divided people more sharply, on the issue at hand, than they had been divided before.

An independent study confirmed the more general effect. People were asked to evaluate the proposition that cutting taxes is so effective in stimulating economic growth that it actually increases government revenue. They were then asked to read a correction from either the *New York Times* or foxnews.com. The correction turned out to increase people’s commitments to the proposition in question: “Conservatives presented with evidence that tax cuts do not increase government revenues ended up believing this claim more fervently than those who did not receive a correction.”⁵⁵

Liberals are hardly immune to this effect.⁵⁶ In 2005, many liberals believed, wrongly, that President Bush imposed a ban on stem cell research. Presented with a correction from the *New York Times* or foxnews.com, liberals continued to believe what they did before. By contrast, conservatives accepted the correction. Hence the correction produced an increase in polarization.

importantly but not surprisingly, it mattered, in terms of the basic effect, whether the correction came from the *New York Times* or Fox News: conservatives distrusted the former more, and liberals distrusted the latter more.

Think in this light about social media. If your Twitter feed is insisting that a particular scandal happened, or that some public official was reckless or worse, how likely is it that a correction will move you? How likely is it that people will see through fake news?

UNFAMILIAR ISSUES

What if the underlying issue is not familiar? In that event, will balanced information produce polarization or consensus? A measure of agreement might well be expected, if only because people do not begin with strong prior convictions. Online, we often encounter issues that we know nothing about. Maybe the balance of arguments will determine our answers? A study of nanotechnology casts light on that question.⁵⁷

A large set of Americans was divided into two groups. In the “no-information” condition, people were simply told that nanotechnology is a process for producing and manipulating small particles. In that condition, people did not divide along ideological lines about the costs and benefits of using nanotechnology. There were no evident splits between conservatives and liberals, or Republicans or Democrats. Apparently the issue seemed highly technical, and the mere name and description did not split people along any relevant lines.

In the “information-exposed” condition, people were given factual material on the potential risks and benefits of nanotechnology. Notably, exposure to information had essentially no effect on people’s views about those risks and benefits. But such exposure did split people in accordance with their preexisting political orientations. Those who tended to like free markets and distrust government interference ended up more favorably disposed toward the use of nanotechnology. Those who tended to favor social equality

and trust government to promote social goals ended up less favorably disposed toward that use.

In the no-information condition, there was essentially no difference in vision between the two groups in their belief that the benefits of nanotechnology outweighed the risks. By a small majority (61 percent), both groups tended to accept that belief. But after exposure to balanced information, the split grew from 0 to 68 percent with 86 percent of free market enthusiasts believing that the benefits outweighed the costs, and only 23 percent of egalitarians so believing.

For the idea of online learning, that's a definite problem.

MAKING SENSE

Here's a way to understand some of these studies. When people start with strong convictions and really know what they think, they're a lot less likely to be moved by contrary arguments. One reason is that they can just dismiss them, given what they already know. If you believe that the Holocaust happened, you won't be much affected by some report suggesting that the whole idea is a concoction of Jewish historians.

Suppose that contrary arguments have no effect on you, but that supportive arguments and new information in line with what you already think strengthen your convictions. That's certainly been known to happen. Sure enough, balanced presentations should end up fortifying those convictions. And if such presentations involve an unfamiliar subject, such as nanotechnology, something similar should happen, at least if the presentations end up triggering your long-standing concerns.

A second factor involves emotions, not knowledge. If you are strongly committed to a certain belief—say, that climate change is a serious problem—a contrary argument might do little to inform you. It might just make you mad. And if you're mad, you might more strongly hold the views with which you began. It is for this reason, in part, that corrections might not move you. And you

might also wonder, Why would they issue that correction, if the underlying claim weren't true? Aren't they hiding something?

These points suggest that the findings of the various studies are important, but only in certain settings, and they do not capture how most people deal with most of what they see in print or online. True, if people begin with strong commitments, they aren't easy to move, and contrary arguments might backfire—at least if they do not come from a reliable source. That clarifies why the phenomenon of group polarization is paralleled by that of biased assimilation, which in turn means that if people get diverse sources of information online, they might polarize too. But for many issues, people aren't all that sure what they think. They start with a degree of open-mindedness. They're searching. They don't begin with intensely held convictions, and even if they tend to know what they think, they're willing to listen.

It follows that on a wide assortment of issues, it's a good thing if people do not sort themselves into communities of like-minded types. Let's not be too optimistic—but it's even possible that the truth will emerge.