The Law of Group Polarization*

Cass R. Sunstein

Law and Political Science, University of Chicago

CONSIDER the following events:

- Affirmative action is under attack in the state of Texas. A number of professors at a particular branch of the University of Texas are inclined to be supportive of affirmative action; they meet to exchange views and to plan further action, if necessary. What are these professors likely to think, and to do, after they talk?
- After a nationally publicized shooting at a high school, a group of people in the community, most of them tentatively in favor of greater gun control, come together to discuss the possibility of imposing new gun control measures. What, if anything, will happen to individual views as a result of this discussion?
- A jury is deciding on an appropriate punitive damage award in a case of recklessly negligent behavior by a large company; the behavior resulted in a serious injury to a small child. Before deliberating as a group, individual jurors have chosen appropriate awards, leading to an average of $1.5 million and a median of $1 million. As a statistical generalization, how will the jury's ultimate award tend to compare to these figures?
- A group of women are concerned about what they consider to be a mounting "tyranny of feminism." They believe that women should be able to make their own choices, but they also think that men and women are fundamentally different, and that their differences legitimately lead to different social roles. The group decides to meet every two weeks to focus on common concerns. After a year, is it possible to say what its members are likely to think?

Every society contains innumerable deliberating groups. Church groups, political parties, women's organizations, juries, dissident organizations, legislative bodies, regulatory commissions, multimember courts, faculties, student organizations, those participating in talk radio programs, Internet discussion groups, and others engage in deliberation. It is a simple social fact that sometimes people enter discussions with one view and leave with another, even on political and moral questions. Emphasizing this fact, many recent observers have embraced the

traditional American aspiration to "deliberative democracy," an ideal that is designed to combine popular responsiveness with a high degree of reflection and exchange among people with competing views.[1] But for the most part, the resulting literature has not been empirically informed. It has not dealt much with the real-world consequences of deliberation, and with whether generalizations hold in actual deliberative settings, with groups of different predispositions and compositions.

My principal purpose is to investigate a striking but largely neglected statistical regularity—that of *group polarization*—and to relate this phenomenon to underlying questions about the role of deliberation in the "public sphere" of a heterogeneous democracy. In brief, group polarization means that *members of a deliberating group predictably move toward a more extreme point in the direction indicated by the members' predeliberation tendencies.*[2] Thus, for example, members of the first deliberating group are likely to become more firmly committed to affirmative action; the second group will probably end up favoring gun control quite enthusiastically; the punitive damages jury will likely come up with an award higher than the median, perhaps higher than the mean as well, and very possibly as high as or higher than that of the highest predeliberation award of any individual member; the group of women concerned about feminism is likely to become very conservative indeed on gender issues. Notably, groups consisting of individuals with extremist tendencies are more likely to shift, and likely to shift more (a point that bears on the wellsprings of violence and terrorism); the same is true for groups with some kind of salient shared identity (like Republicans, Democrats, and lawyers, but unlike jurors and experimental subjects). When like-minded people are participating in "iterated polarization games"—when they meet regularly, without sustained exposure to competing views—extreme movements are all the more likely.

Two principal mechanisms underlie group polarization. The first points to social influences on behavior and in particular to people's desire to maintain their reputation and their self-conception. The second emphasizes the limited "argument pools" within any group, and the directions in which those limited

pools lead group members. An understanding of the two mechanisms provides many insights into deliberating bodies. Such an understanding illuminates a great deal, for example, about likely processes within multimember courts, juries, political parties, and legislatures—not to mention ethnic groups, extremist organizations, terrorists, criminal conspiracies, student associations, faculties, institutions engaged in feuds or "turf battles," workplaces, and families. At the same time, these mechanisms raise serious questions about deliberation from the normative point of view. If deliberation predictably pushes groups toward a more extreme point in the direction of their original tendency, whatever it may be, do we have any reason to think that deliberation is producing improvements? A sensible answer would emphasize the importance of paying far more attention to the circumstances and nature of deliberation, not merely to the fact that it is occurring.

One of my largest purposes is to cast light on *enclave deliberation*, a process that I understand to involve deliberation among like-minded people who talk or even live, much of the time, in isolated enclaves. I will urge that enclave deliberation is, simultaneously, a potential danger to social stability, a source of social fragmentation or even violence, and a safeguard against social injustice and unreasonableness. As we will see, group polarization helps to cast new light on the old idea that social homogeneity can be quite damaging to good deliberation. When people are hearing echoes of their own voices, the consequence may be far more than support and reinforcement. But there is a point more supportive of enclave deliberation: participants in heterogeneous groups tend to give least weight to the views of low-status members[3]—in some times and places, women, African-Americans, less-educated people. Hence enclave deliberation might be the only way to ensure that those views are developed and eventually heard. Without a place for enclave deliberation, citizens in the broader public sphere may move in certain directions, even extreme directions, precisely because opposing voices are not heard at all. An ambivalent lesson is that deliberating enclaves can be breeding grounds for *both* the development of unjustly suppressed views and for unjustified extremism, indeed fanaticism. A less ambivalent lesson involves the need, not to celebrate or to challenge deliberation as such, but to design institutions so as to ensure that when individuals and groups move, it is because of the force of the arguments, not because of the social dynamics that I will emphasize here.

I. HOW AND WHY GROUPS POLARIZE

A. THE BASIC PHENOMENON

Group polarization is among the most robust patterns found in deliberating bodies, and it has been found all over the world and in many diverse tasks.

---

[3]Caryn Christenson and Ann Abbott, "Team medical decision making," *Decision Making in Health Care*, ed. Gretchen Chapman and Frank Sonnenberg (New York: Cambridge University Press, 2000), pp. 267, 273–6.

The result is that groups often make more extreme decisions than would the typical or average individual in the group (where "extreme" is defined solely internally, by reference to the group's initial dispositions). Note that in the experimental work, both extremism and tendencies are measured not by reference to anything external, or to a normative standard, but by reference to the particular scale that is brought before the individuals who compose the group. Thus, for example, people might be asked, on a scale of -5 to 5, how strongly they agree or disagree with a particular statement (white racism is responsible for the disadvantages faced by African–Americans, the government should increase regulation of nuclear power, America should increase foreign aid). We shall see that the experimental literature is closely connected to real-world phenomena.

Though standard, the term "group polarization" is somewhat misleading. It is not meant to suggest that group members will shift to the poles, nor does it refer to an increase in variance among groups, though this may be the ultimate result. Instead the term refers to a predictable shift *within* a group discussing a case or problem. As the shift occurs, groups, and group members, move and coalesce, not toward the middle of antecedent dispositions, but toward a more extreme position in the direction indicated by those dispositions. The effect of deliberation is both to decrease variance among group members, as individual differences diminish, and also to produce convergence on a relatively more extreme point among predeliberation judgments.

Consider some examples of the basic phenomenon, which has been found in over a dozen nations.[4] (a) A group of moderately profeminist women will become more strongly profeminist after discussion.[5] (b) After discussion, citizens of France become more critical of the United States and its intentions with respect to economic aid.[6] (c) After discussion, whites predisposed to show racial prejudice offer more negative responses to the question whether white racism is responsible for conditions faced by African–Americans in American cities.[7] (d) After discussion, whites predisposed not to show racial prejudice offer more positive responses to the same question.[8] As statistical regularities, it should follow, for example, that that those moderately critical of an ongoing war effort will, after discussion, sharply oppose the war; that those who believe that global warming is

[4] R. Brown, *Social Psychology*, 2nd edn (New York: Free Press, 1985), p. 222. These include the United States, Canada, New Zealand, Germany, and France. See, e.g., Johannes A. Zuber et al., "Choice shift and group polarization," *Journal of Personality and Social Psychology*, 62 (1992), 50–61 on Germany; Dominic Abrams et al., "Knowing what to think by knowing who you are," *British Journal of Social Psychology*, 29 (1990), 97–119 at p. 112 on New Zealand. Of course it is possible that some cultures would show a greater or lesser tendency toward polarization; this would be an extremely interesting area for empirical study.

[5] See D. G. Myers, "Discussion-induced attitude polarization," *Human Relations*, 28 (1975), 699.

[6] R. Brown, *Social Psychology*, 2nd edn (New York: Free Press, 1985), p. 224.

[7] D. G. Myers and G. D. Bishop, "Enhancement of dominant attitudes in group discussion," *Journal of Personality and Social Psychology*, 20 (1971), 386–91.

[8] See ibid.

a serious problem are likely, after discussion, to hold that belief with considerable confidence; that people tending to believe in the inferiority of a certain racial group will become more entrenched in this belief as a result of discussion; that those tending to condemn the United States will, as a result of discussion, end up condemning the United States with some intensity.

There have been two main explanations for group polarization, both of which have been extensively investigated.

1. *Social comparison*. The first, involving social comparison, begins with the claim that people want to be perceived favorably by other group members, and also to perceive themselves favorably. Once they hear what others believe, they adjust their positions in the direction of the dominant position. The result is to press the group's position toward one or another extreme, and also to induce shifts in individual members. People may wish, for example, not to seem too enthusiastic or too restrained in their enthusiasm for affirmative action, feminism, or an increase in national defense; hence their views may shift when they see what other group members think. The result will be group polarization.

The dynamic behind the social comparison explanation is that most people may want to take a position of a certain socially preferred sort—in the case of risk-taking, for example, they may want to be perceived (and to perceive themselves) as moderate risk-takers, and their choice of position is partly a product of this desire. No one can know what such a position would be until the positions of others are revealed. Thus individuals move their judgments in order to preserve their image to others and their image to themselves.

2. *Persuasive arguments*. The second explanation, emphasizing the role of persuasive arguments, is based on a common sense intuition: that any individual's position on an issue is partly a function of which arguments presented within the group seem convincing. The choice therefore moves in the direction of the most persuasive position defended by the group, taken as a collectivity. Because a group whose members are already inclined in a certain direction will have a disproportionate number of arguments supporting that same direction, the result of discussion will be to move individuals further in the direction of their initial inclinations. The key is the existence of a limited argument pool, one that is skewed (speaking purely descriptively) in a particular direction. Members of a group will have thought of some, but not all, of the arguments that justify their initial inclination; consider the question whether to fear global warming or not to do so. In discussion, arguments of a large number of individuals are stated and heard, but the total argument pool will be tilted in one or another direction, depending on the predispositions of the people who compose the group. Hence there will be a shift in the direction of the original tilt.

There is a related possibility, not quite reducible to either of the two standard arguments, but using elements of each. In their individual judgments, people are

averse to extremes; they tend to seek the middle of the relevant poles.[9] It is possible that when people are making judgments individually, they err on the side of caution, expressing a view in the direction that they really hold, but stating that view cautiously, for fear of seeming extreme. Once other people express supportive views, the relevant inhibition disappears, and people feel free to say what, in a sense, they really believe. There appears to be no direct test of this hypothesis, but it is reasonable to believe that the phenomenon plays a role in group polarization and choice shifts.

B. REFINEMENTS—AND DEPOLARIZATION

I now turn to some refinements, complicating the basic account of group polarization. For purposes of understanding the relationship between that phenomenon and democracy, the central points are twofold. *First*, it matters a great deal whether people consider themselves part of the same social group as the other members; a sense of shared identity will heighten the shift, and a belief that identity is not shared will reduce and possibly eliminate it. *Second*, deliberating groups will tend to depolarize if they consist of equally opposed subgroups and if members have a degree of flexibility in their positions. Both of these findings have great relevance to any account of the relationship between group polarization and democratic institutions, as we will see.

1. *Statistical regularities*. Of course not all groups polarize; some groups end up in the middle, not toward either extreme. Nor is it hard to understand why this might be so. If the people defending the original tendency are particularly unpersuasive, group polarization is unlikely to occur. If the outliers are especially convincing, groups may even shift away from their original tendency and in the direction held by few or even one.

Sometimes, moreover, external constraints or an external "shock" may prevent or blunt group polarization. Group members with well-defined views on a certain issue (gun control, separation of church and state, intervention in foreign nations) may be prone to polarize, but in order to maintain political effectiveness, even basic credibility, they will sometimes maintain a relatively moderate face, publicly or even privately. Groups that have started to polarize in an extreme direction may move toward the middle in order to promote their own legitimacy or because of new revelations of one kind of another. Readers are likely to be able to identify their own preferred examples.

2. *Affective factors, identity, and solidarity*. Affective factors are quite important in group decisions, and when manipulated, such factors will significantly increase or decrease polarization. If group members are linked by

---

[9]See, e.g., Mark Kelman et al., "Context dependence in legal decision making," *Behavioral Law and Economics*, ed. Cass R. Sunstein (Cambridge: Cambridge University Press, 2000), pp. 61, 71–6.

affective ties, dissent is significantly less frequent.[10] The existence of affective ties thus reduces the number of divergent arguments and also intensifies social influences on choice. Hence people are less likely to shift if the direction advocated is being pushed by unfriendly group members; the likelihood of a shift, and its likely size, are increased when people perceive fellow members as friendly, likeable, and similar to them.[11] 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."

In a refinement of particular importance to social deliberation and the theory of democracy, it has been found to matter whether people think of themselves, antecedently or otherwise, as part of a group having a degree of solidarity. If they think of themselves in this way, group polarization is all the more likely, and it is likely too to be more extreme.[12] Thus when the context emphasizes each person's membership in the social group engaging in deliberation, polarization increases. This finding is in line with more general evidence that social ties among deliberating group members tend to suppress dissent and in that way to lead to inferior decisions.[13] This should not be surprising. If ordinary findings of group polarization are a product of social influences and limited argument pools, 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, sex) unites them, group polarization will be heightened.

3. *Depolarization and deliberation without shifts*. Is it possible to construct either groups that will depolarize—that will tend toward the middle—or groups whose members will not shift at all? Both phenomena seem to be real in actual deliberating bodies. In fact the persuasive arguments theory implies that there will be depolarization if and when new persuasive arguments are offered that are opposite to the direction initially favored by group members. There is evidence for this phenomenon.[14] Depolarization, rather than polarization, will also be found when the relevant group consists of

[10]See Brooke Harrington, "The pervasive effects of embeddedness in organizations," p. 24 (unpublished manuscript 2000).

[11]See Hermann Brandstatter, "Social emotions in discussion groups," *Dynamics of Group Decisions*, ed. Hermann Brandstatter et al. (Beverly Hills: Sage Publications, 1978).

[12]See Russell Spears, Martin Lee, and Stephen Lee, "De-individuation and group polarization in computer-mediated communication," *British Journal of Social Psychology*, 29 (1990), 121–34; Dominic Abrams et al., "Knowing what to think by knowing who you are," *British Journal of Social Psychology*, 29 (1990), 97–119 at p. 112; Patricia Wallace, *The Psychology of the Internet* (Cambridge: Cambridge University Presss, 1999), pp. 73–6.

[13]See Lee Roy Beach, *The Psychology of Decision Making in Organizations* (Thousand Oaks, Calif.: Sage Publications, 1997); Harrington, "The pervasive effects of embeddedness in organizations."

[14]A similar possibility is that hearing other similar opinions produces greater confidence in individual positions, opening members to a more extreme judgment in the same direction. This has been raised recently by Chip Heath and Richard Gonzales, "Interaction with others increases decision confidence but not decision quality: evidence against information collection views of interactive decision making," *Organizational Behavior and Human Decision Processes*, 61 (1997), 305–26.

individuals drawn equally from two extremes.[15] And "familiar and long-debated issues do not depolarize easily."[16] With respect to such issues, people are simply less likely to shift at all. And when one or more people in a group know the right answer to a factual question, the group is likely to shift in the direction of accuracy.

C. Actual Deliberation Within Identifiable Groups: Iterated "Polarization Games"

Studies of group polarization involve one-shot experiments. We will turn shortly to group polarization in the real world. But let us notice an intriguing implication of the experiments, an implication with special importance for democratic deliberation involving people who meet with each other not once, but on a regular basis.

If participants engage in repeated discussions—if, for example, they meet each month, express views, and take votes—there should be repeated shifts toward, and past, the defined pole. Thus, for example, if a group of citizens is thinking about genetic engineering of food, or the minimum wage, or the World Trade Organization, the consequence of their discussions, over time, should be to lead in quite extreme directions. In these iterated "polarization games," deliberation over time should produce a situation in which individuals hold positions more extreme than those of any individual member before the series of deliberations began. In fact the idea of iterated polarization games seems far more realistic than the processes studied in one-shot experiments. There appears to be no study of such iterated 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.

D. Rhetorical Asymmetry and the "Severity Shift": A Pervasive Phenomenon?

In a noteworthy qualification of the general literature on group polarization, the previously mentioned study of punitive damage awards by juries found a striking pattern for dollar awards.[17] For *any* dollar award above zero, the general effect of deliberation was to increase awards above those of the median voter. This was a kind of "severity shift." Dollar awards did not simply polarize; while higher

---

[15]See H. Burnstein, "Persuasion as argument processing," *Group Decision Making*, ed. H. Brandstatter, J. H. Davis and G. Stocker-Kreichgauer (London: Academic Press, 1982).

[16]Roger Brown, *Social Psychology*, 2nd edn (New York: Free Press, 1985), p. 226.

[17]See David Schkade, Cass R. Sunstein, and Daniel Kahneman, "Deliberating about dollars," *Columbia Law Review*, 100 (2000), 1139–75.

awards increased dramatically, as compared to the median of predeliberation votes, low awards increased as well. Why is this?

Both the original experiment and a follow-up experiment suggest that the severity shift is a product of a "rhetorical asymmetry" that favors, other things being equal and in any contest, the person or persons urging higher awards. In our culture, and in light of existing social norms, the person favoring the higher amount for punitive damages appears likely to be more convincing than the person favoring the lower amount. It is important to emphasize that this asymmetry operates independently of any facts about the individual case. The reason appears to be that with respect to dollar awards involving a corporate defendant, stronger arguments—"we need to deter this kind of conduct," "we need to send a powerful signal," "we need to attract their attention"—tend to have comparatively greater weight.

Undoubtedly there are many other contexts containing rhetorical asymmetry, and undoubtedly the asymmetry can affect outcomes in democratic institutions, as it did in the jury study. Legislative judgments about criminal punishment may, for example, involve an asymmetry of exactly this kind; those favoring greater punishment for drug-related offenses appear to be at a systematic advantage over those favoring lesser punishment. In certain settings, those favoring lower taxes, or more aid for scholarship students, or greater funding for environmental protection, may have a similar rhetorical advantage. Much remains to be explored here. For present purposes the point is that when there is an initial distribution of views in a certain direction, and when existing norms give a more extreme movement in that direction a rhetorical advantage, quite extreme shifts can be expected.

II. POLARIZATION AND DEMOCRACY

In this section I discuss evidence of group polarization in legal and political institutions, and I trace some implications of that evidence for participants in a deliberative democracy. I will deal with normative issues below; my purpose here is to cast a new light on social practices.

A. POLARIZING EVENTS AND POLARIZATION ENTREPRENEURS

Group polarization has a large effect on many deliberating groups and institutions; its effects are hardly limited to the laboratory. Consider, for example, the political and social role of religious organizations. Such organizations tend to strengthen group members' religious convictions, simply by virtue of the fact that like-minded people are talking to one another. Religious groups amplify the religious impulse, especially if group members are insulated from other groups, and on occasion, the result can be to lead people in quite bizarre directions. Whether or not this is so, political activity by members of

religious organizations is undoubtedly affected by cascade-like effects and by group polarization. In a related vein, survey evidence shows that dramatic social events, like the assassination of Martin Luther King and civil rights disturbances, tend to polarize attitudes, with both positive and negative attitudes increasing within demographic groups.[18] The point emphatically holds for the terrorist attacks in New York City and Washington DC on September 11, 2001.

In fact it is possible to imagine "professional polarizers," or "polarization entrepreneurs," that is, political activists who have, as one of their goals, the creation of spheres in which like-minded people can hear a particular point of view from one or more articulate people, and also participate, actually or vicariously, in a deliberative discussion in which a certain point of view becomes entrenched and strengthened. For those seeking to promote social reform, an extremely promising strategy is to begin by promoting discussions among people who tend to favor the relevant reform; such discussions are likely to intensify the underlying convictions and concerns. Social reformers of all stripes may qualify as polarization entrepreneurs; the category includes those fighting communism in Eastern Europe and apartheid in South Africa, as well as terrorist leaders and those involved in criminal conspiracies of many kinds.

B. "Outgroups"

Group polarization has particular implications for insulated "outgroups" and (in the extreme case) for the treatment of conspiracies. Recall that polarization increases when group members identify themselves along some salient dimension, and especially when the group is able to define itself by contrast to another group. Outgroups are in this position—of self-contrast to others—by definition. Excluded by choice or coercion from discussion with others, such groups may become polarized in quite extreme directions, often in part because of group polarization. It is for this reason that outgroup members can sometimes be led, or lead themselves, to violent acts.

The tendency toward polarization among outgroups helps explain special concern about "hate speech," where group antagonisms can be heightened, and it simultaneously raises some questions about the idea that certain group discussions produce "consciousness raising." It is possible, at least, that the consequence of discussion is not only or mostly to raise consciousness (an ambiguous idea to be sure), but to produce group polarization in one direction or another—and at the same time to increase confidence in the position that has newly emerged. This does not mean that consciousness is never raised; undoubtedly group discussion can identify and clarify problems that were previously repressed, or understood as an individual rather than social product.

---

[18]See R. T. Riley and T. F. Pettigrew, "Dramatic events and attitude change," *Journal of Personality and Social Psychology*, 34 (1976), 1004–15.

But nothing of this sort is established by the mere fact that views have changed and coalesced, and are held, post-discussion, with a high degree of confidence.

C. FEUDS, ETHNIC AND INTERNATIONAL STRIFE, AND WAR

Group polarization is inevitably at work in feuds, ethnic and international strife, and war. One of the characteristic features of feuds is that members of feuding groups tend to talk only to one another, fueling and amplifying their outrage, and solidifying their impression of the relevant events. Informational and reputational forces are very much at work here, producing cascade effects, and group polarization can lead members to increasingly extreme positions. It is not too much of a leap to suggest that these effects are sometimes present within ethnic groups and even nations, notwithstanding the usually high degree of national heterogeneity. In America, sharp divergences between whites and African–Americans, on particular salient events or more generally, can be explained by reference to group polarization. The same is true for sharp divergences of viewpoints within and across nations. Group polarization occurs within Israel and among the Palestinian Authority; it occurs within the United States and among those inclined to support, or at least not to condemn, terrorist acts. A large part of the perennial question, ''Why do they hate us?'' lies not in ancient grievances or individual consciences but in the social influences emphasized here. Of course the media play a large role, as we shall now see.

D. THE INTERNET, COMMUNICATIONS POLICY, AND MASS DELIBERATION

Many people have expressed concern about processes of social influence on the mass media and the Internet. The general problem is said to be one of fragmentation, with certain people hearing more and louder versions of their own pre-existing commitments, thus reducing the benefits that come from exposure to competing views and unnoticed problems. With greater specialization, people are increasingly able to avoid general interest newspapers and magazines, and to make choices that reflect their own predispositions. The Internet is making it possible for people to design their own highly individuated communications packages, filtering out troublesome issues and disfavored voices. 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.[19] New communications technologies may increase this phenomenon.

An understanding of group polarization explains why a fragmented communications market may create problems.[20] A ''plausible hypothesis is that the Internet-like setting is most likely to create a strong tendency toward group

[19]Ronald N. Jacobs, *Race, Media, and the Crisis of Civil Society* (New York: Cambridge University Press, 2000).
[20]See Cass R. Sunstein, *Republic.com* (Princeton, N.J.: Princeton University Press, 2001).

polarization when the members of the group feel some sense of group identity."[21] If certain people are deliberating with many like-minded others, views will not be reinforced but instead shifted to more extreme points. 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.

III. DELIBERATIVE TROUBLE

I now turn to normative issues, involving the relationship among group polarization, democratic theory, and legal institutions. I focus in particular on the implications of group polarization for institutional design, with special reference to the uses of heterogeneity and the complex issues presented by deliberation inside particular "enclaves." Should enclave deliberation count as deliberation at all? If deliberation requires a measure of disagreement, this is a serious question. But even like-minded people will have different perspectives and views, so that a group of people who tend to like affirmative action, or to fear global warming, will produce some kind of exchange of opinion. I will urge that in spite of this point, enclave deliberation raises serious difficulties for the participants and possibly for society as a whole. But there are many complexities here. In some cases, enclave deliberation will be a defective form for the participants, but will serve to foster a diversity of views for the wider public, and will therefore be desirable from the social point of view.

The central problem is that widespread error and social fragmentation are likely to result when like-minded people, insulated from others, move in extreme directions simply because of limited argument pools and parochial influences. As an extreme example, consider a system of one-party domination, which stifles dissent in part because it refuses to establish space for the emergence of divergent positions; in this way, it intensifies polarization within the party while also disabling external criticism. In terms of institutional design, the most natural response is to ensure that members of deliberating groups, whether small or large, will not isolate themselves from competing views—a point with implications for multimember courts, open primaries, freedom of association, and the architecture of the Internet. Here, then, is a plea for ensuring that deliberation occurs within a large and heterogeneous public sphere, and for guarding against a situation in which like-minded people are walling themselves off from alternative perspectives.

But there is a difficulty with this response: A certain measure of isolation will, in some cases, be crucial to the development of ideas and approaches that would not otherwise emerge and that deserve a social hearing. Members of low-status groups are often quiet within heterogeneous bodies, and thus deliberation, in

---

[21]See Patricia Wallace, *The Psychology of the Internet*, pp. 73–84.

such bodies, tends to be dominated by high-status members. Any shift—in technology, norms, or legal practice—that increases the number of deliberating enclaves will increase the diversity of society's aggregate "argument pool" while also increasing the danger of extremism and instability, ultimately even violence. Terrorism itself is a product, in part, of group polarization. Shifts toward a general "public sphere," without much in the way of enclave deliberation, will decrease the likelihood of extremism and instability, but at the same time produce what may be a stifling uniformity. And shifts toward more in the way of enclave deliberation will increase society's aggregate "argument pool," and hence enrich the marketplace of ideas, while also increasing extremism, fragmentation, hostility, and even violence.

No algorithm is available to solve the resulting conundrums. But some general lessons do emerge. It is important to ensure social spaces for deliberation by like-minded persons, but it is equally important to ensure that members of the relevant groups are not isolated from conversation with people having quite different views. The goal of that conversation is to promote the interests of those inside and outside the relevant enclaves, by subjecting group members to competing positions, by allowing them to exchange views with others and to see things from their point of view, and by ensuring that the wider society does not marginalize, and thus insulate itself from, views that may turn out to be right, or at least informative.

A. DOUBTS AND QUESTIONS

1. Why Deliberate?

If the effect of deliberation is to move people toward a more extreme point in the direction of their original tendency, why is it anything to celebrate? The underlying mechanisms do not provide much reason for confidence. If people are shifting their position in order to maintain their reputation and self-conception, before groups that may or may not be representative of the public as a whole, is there any reason to think that deliberation is making things better rather than worse? To the extent that shifts are occurring as a result of partial and frequently skewed argument pools, the results of deliberative judgments may be far worse than the results of simply taking the median of predeliberation judgments.

The most important point here is that those who emphasize the ideals associated with deliberative democracy tend to emphasize its preconditions, which include political equality, an absence of strategic behavior, full information, and the goal of "reaching understanding."[22] In real-world

---

[22]See Jürgen Habermas, *A Theory of Communicative Action*, trans. Thomas McCarthy (Boston: Beacon Press, 1984), p. 99. Thus Habermas distinguishes between strategic and communicative action and stresses "the cooperatively pursued goal of reaching understanding"; compare the treatment in Gutmann and Thompson, *Democracy and Disagreement,* pp. 52–94, emphasizing the idea of reciprocity, which emphasizes the desire to justify one's position by reference to reasons.

deliberations, behavior is often strategic, and equality is often absent in one or another form. But the existence of a limited argument pool, strengthening the existing tendency within the group, will operate in favor of group polarization even if no individual behaves strategically. By itself this will produce group polarization whether or not social influence is operating. On the other hand, the social context of deliberation can make a large difference, and under certain conditions, group polarization need not occur. The nature of the deliberative process, and the characteristics of the deliberating participants, can matter a great deal. I will return to this issue below.

In any case social influences need not be inconsistent with the effort to produce truth and understanding; when people attempt to position themselves in a way that fits with their best self-conception, or their preferred self-presentation, nothing has gone wrong, even from the standpoint of deliberation's most enthusiastic defenders. Perhaps group polarization could be reduced or even eliminated if we emphasized that good deliberation has full information as a precondition; by hypothesis, argument pools would not be limited if all information were available. But that requirement is extremely stringent, and if there is already full information, the role of deliberation is greatly reduced. In any case the group polarization phenomenon suggests that in real-world situations, deliberation is hardly guaranteed to increase the likelihood of arriving at truth. The trick is to produce an institutional design that will increase the likelihood that deliberation will lead in sensible directions, so that any polarization, if it occurs, will be a result of learning, rather than group dynamics.

2. Movements Right and Wrong

Of course we cannot say, from the mere fact of polarization, that there has been a movement in the wrong direction. Perhaps the more extreme tendency is better; recall that group polarization is likely to have fueled the antislavery movement and many others that deserve to meet with widespread approval. In the context of punitive damage awards, perhaps a severity shift produces good outcomes. Extremism should hardly be a word of opprobrium; everything depends on what extremists are arguing *for*. In addition, group polarization can be explained partly by reference to the fact that people who are confident are likely to be persuasive; and it is sensible to say that as a statistical matter, though not an invariable truth, people who are confident are more likely to be right. But when group discussion tends to lead people to more strongly held versions of the same view with which they began, and when social influences and limited argument pools are responsible, there is little reason for great confidence in the effects of deliberation.

We can go further. If it is possible to identify a particular viewpoint as unreasonable, it is also possible to worry about group discussion among people who share that viewpoint. If the underlying views are unreasonable, it makes sense to fear that these discussions may fuel increasing hatred and extremism

(used here in an evaluative sense). This does not mean that the discussions can or should be regulated in a system dedicated to freedom of speech. But it does raise questions about the idea that ''more speech'' is necessarily an adequate remedy.

B. THE VIRTUES OF HETEROGENEITY

The simplest lesson here involves both individual susceptibility and institutional design. For many people, mere awareness of the role of limited argument pools and social influences might provide some inoculation against inadequately justified movements of opinion within groups. More important, institutions might well be designed to ensure that when shifts are occurring, it is not because of arbitrary or illegitimate constraints on the available range of arguments. This is a central task of constitutional design, and in this light a system of checks and balances might be defended, not as an undemocratic check on the will of the people, but as an effort to protect against potentially harmful consequences of group discussion.

To explore some of the advantages of heterogeneity, imagine a deliberating body consisting of all citizens in the relevant group; this may mean all citizens in a community, a state, a nation, or the world. By hypothesis, the argument pool would be very large. It would be limited only to the extent that the set of citizen views was similarly limited. Social influences would undoubtedly remain. Hence people might shift because of a desire to maintain their reputation and self-conception, by standing in a certain relation to the rest of the group. But to the extent that 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 an accidentally skewed group sample.

This thought experiment does not suggest that the hypothesized deliberating body would be ideal. Perhaps all citizens, presenting all individual views, would offer a skewed picture from the normative point of view; in a pervasively unjust society, a deliberating body consisting of everyone may produce nothing to celebrate. Perhaps weak arguments would be made and repeated and repeated again, while good arguments would be offered infrequently. As we will see below, it is often important to ensure enclaves in which polarization will take place, precisely in order to ensure the emergence of views that are suppressed, by social influences or otherwise, but reasonable or even right. But at least a deliberating body of all citizens would remove some of the distortions in the group polarization experiments, where generally like-minded people, not exposed to others, shift in large part because of that limited exposure. Hence the outcomes of these deliberations will not be a product of the arbitrariness that can be introduced by skewed argument pools.

C. Enclave Deliberation and Suppressed Voices

The discussion has yet to focus on the potential vices of heterogeneity and the potentially desirable effects of deliberating "enclaves," consisting of groups of like-minded individuals. It seems obvious that such groups can be extremely important in a heterogeneous society, not least because members of some demographic groups tend to be especially quiet when participating in broader deliberative bodies. In this light, a special advantage of what we might call "enclave deliberation" is that it promotes the development of positions that would otherwise be invisible, silenced, or squelched in general debate. While this is literally dangerous in numerous contexts, it can also be a great advantage; many desirable social movements have been made possible through this route. The efforts of marginalized groups to exclude outsiders, and even of political parties to limit their primaries to party members, can be justified in similar terms. Even if group polarization is at work—indeed *because* group polarization is at work—enclaves can provide a wide range of social benefits, not least because they greatly enrich the social "argument pool."

The central empirical point here is that in deliberating bodies, high-status members tend to initiate communication more than others, and their ideas are more influential, partly because low-status members lack confidence in their own abilities, partly because they fear retribution.[23] For example, women's ideas are often less influential and sometimes are "suppressed altogether in mixed-gender groups,"[24] and in ordinary circumstances, cultural minorities have disproportionately little influence on decisions by cultural mixed groups.[25] Interestingly, there is evidence that with changes in gender norms, some tasks show no gender differences in influence on groups; this evidence confirming the claim that people's role in group deliberation will be influenced by whether social norms produce status hierarchies. In these circumstances, it makes sense to promote deliberating enclaves in which members of multiple groups may speak with one another and develop their views.

But there is a serious danger in such enclaves. The danger is that through the mechanisms of social influence and persuasive arguments, members will move to positions that lack merit but are predictable consequences of the particular circumstances of enclave deliberation. In the extreme case, enclave deliberation may even put social stability at risk (for better or for worse). 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 for its own members.

There is no simple solution to the dangers of enclave deliberation. Sometimes the threat to social stability is desirable. From the standpoint of institutional

[23]See Christenson and Abbott, "Team medical decision making," p. 273.
[24]Ibid., p. 274.
[25]C. Kirchmeyer and A. Cohen, "Multicultural groups: their performance and reactions with constructive conflict," *Group and Organization Management*, 17 (1992), 153.

design, the problem is that any effort to promote enclave deliberation will ensure group polarization among a wide range of groups, some necessary to the pursuit of justice, others likely to promote injustice, and some potentially quite dangerous. In this light we should be able to see more clearly the sense in which Edmund Burke's conception of representation—rejecting "local purposes" and "local prejudices" in favor of "the general reason of the whole"[26]—is not contingently but is instead *essentially* conservative (speaking purely descriptively, as a safeguard of existing practices). The reason is that the submersion of "local purposes" and "local prejudices" into a heterogenous "deliberative assembly" will inevitably tend to weaken the resolve of groups—and particularly low-status or marginalized groups—whose purely internal deliberations would produce a high degree of polarization.

Hence James Madison—with his fear of popular passions producing "a rage for paper money, for an abolition of debts, for an equal division of property, or for any other improper or wicked project"[27]—would naturally be drawn to a Burkean conception of representation, favoring large election districts and long length of service[28] to counteract the forces of polarization. By contrast, those who believe that "destabilization" is an intrinsic good, or that the status quo contains sufficient injustice that it is worthwhile to incur the risks of encouraging polarization on the part of diverse groups, will, or should, be drawn to a system that enthusiastically promotes insular deliberation within enclaves.

In a nation in which most people are confused or evil, enclave deliberation may be the only way to develop a sense of clarity or justice, at least for some. But even in such a nation, enclave deliberation is unlikely to produce change unless its members are eventually brought into contact with others. In democratic societies, the best response is to ensure that any such enclaves are not walled off from competing views, and 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 deadly) combination of extremism with marginality.

D. THE PUBLIC SPHERE AND APPROPRIATE HETEROGENEITY

1. *The Public Sphere*

A reasonable conclusion would return to the need for full information, not only about facts but also about relevant values and options, and suggest that for a

---

[26]See "Speech to the electors (Nov. 3, 1774)," reprinted in *Burke's Politics*, ed. R. J. S. Hoffman and P. Levack (New York: A. A. Knopf, 1949), p. 116.

[27]See *The Federalist*, No. 10.

[28]See Cass R. Sunstein, "Interest groups in American public law," *Stanford Law Review*, 38 (1985), 29–87 at p. 42.

designer or leader of any institution, it makes sense to promote ample social space both for enclave deliberation and for discussions involving a broad array of views, including those who have been within diverse enclaves. The idea of a "public sphere," developed most prominently by Jürgen Habermas, can be understood as an effort to ensure a domain in which multiple views can be heard by people with multiple perspectives.[29] Of course any argument pool will be limited. No one has time to listen to every point of view. But an understanding of group polarization helps show that heterogeneous groups are often a far better source of good judgments, simply because more arguments will be made available.

2. A New Look at Group Representation

The point very much bears on the continuing debate over proportional or group representation.[30] On one approach, political groups should be allowed to have representation to the extent that they are able to get more than a minimal share of the vote. On another approach, steps would be taken to increase the likelihood that members of disadvantaged or marginal groups—perhaps African–Americans, religious minorities, gays and lesbians, women—would have their own representatives in the deliberating body. The decision whether to move in one or another direction depends on many factors, and an understanding of group polarization is hardly sufficient. But at least it can be said that proportional or group representation draws strength from the goal of ensuring exposure to a diverse range of views. On the one hand, group representation should help counteract the risks of polarization, and susceptibility to cascade effects, that come from deliberation among like-minded people. On the other hand, group representation should help reduce the dangers that come from insulation of those in the smaller enclave, by subjecting enclave representations to a broader debate. For these purposes, it might well be insufficient that representatives, not themselves members of any enclave, are electorally accountable to constituents who include enclave members. The point of group representation is to promote a process in which those in the enclave hear what others have to say, and in which those in other enclaves, or in no enclaves at all, are able to listen to people with very different points of view.

3. Appropriate Heterogeneity

The principal qualification here is that the real question is how to ensure *appropriate* heterogeneity. For example, it would not make sense to say that in a deliberating group attempting to think through issues of affirmative action, it is

---

[29]See Jürgen Habermas, *The Structural Transformation of the Public Sphere*, trans. Thomas Burger and Frederick Lawrence (Oxford: Polity; 1989), pp. 231–50.

[30]See Anne Phillips, *The Politics of Presence* (Oxford: Oxford University Press, 1995); Iris Young, *Justice and the Politics of Difference* (Princeton, N.J.: Princeton University Press, 1990), pp. 183–91; Cass R. Sunstein, "Beyond the republican revival." *Yale Law Journal*, 97 (1988), 1539 at pp. 1585–9.

important to allow exposure to people who think that slavery was good and should be restored. The constraints of time and attention call for limits to heterogeneity; and—a separate point—for good deliberation to take place, some views are properly placed off the table, simply because time is limited and they are so invidious, implausible, or both. This point might seem to create a final conundrum: To know what points of view should be represented in any group deliberation, it is important to have a good sense of the substantive issues involved, indeed a sufficiently good sense as to generate judgments about what points of view must be included and excluded. But if we already know that, why should we not proceed directly to the merits? If we already know that, before deliberation occurs, does deliberation have any point at all?

The answer is that we often do know enough to know which views count as reasonable, without knowing which view counts as right, and this point is sufficient to allow people to construct deliberative processes that should correct for the most serious problems potentially created by group deliberation. What is necessary is not to allow every view to be heard, but to ensure that no single view is so widely heard, and reinforced, that people are unable to engage in critical evaluation of the reasonable competitors.

E. THE DELIBERATIVE OPINION POLL: A CONTRAST

In an interesting combination of theoretical and empirical 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 to deliberate about various issues. Deliberative opinion polls have now been conducted in several nations, including the United States, England and Australia. Fishkin finds some noteworthy shifts in individual views; but he does not find a systematic tendency toward polarization. In his studies, individuals shift both toward and away from the median of predeliberation views.

In England, for example, deliberation led to reduced interest in using imprisonment as a tool for combatting crime.[31] The percentage believing that "sending more offenders to prison" is an effective way to prevent crime went down from 57 per cent to 38 per cent; the percentage believing that fewer people should be sent to prison increased from 29 per cent to 44 per cent; belief in the effectiveness of "stiffer sentences" was reduced from 78 per cent to 65 per cent. Similar shifts were shown in the direction of greater enthusiasm for procedural rights of defendants and increased willingness to explore alternatives to prison. In other experiments with the deliberative opinion poll, shifts included a mixture of findings, with larger percentages of individuals concluding that legal pressures should be increased on fathers for child support (from 70 per cent to 85 per cent)

[31]See James Fishkin, *The Voice of the People* (New Haven, Conn.: Yale University Press, 1995), pp. 206–7.

and that welfare and health care should be turned over to the states (from 56 per cent to 66 per cent). Indeed, on many particular issues, the effect of deliberation was to create an increase in the intensity with which people held their pre-existing convictions. These findings are consistent with the prediction of group polarization. But this was hardly a uniform pattern, and on some questions deliberation increased the percentage of people holding a minority position (with, for example, a jump from 36 per cent to 57 per cent of people favoring policies making divorce "harder to get"). These are not the changes that would be predicted by group polarization.

There appear to be several factors distinguishing the deliberative opinion poll from experiments on group polarization. First, Fishkin's deliberators did not vote as a group, and opinions were collected both individually and confidentially; while group polarization is observed when no group decision is expected, the extent of polarization is likely to decrease, simply because members have not been asked to sign onto a group decision as such. Second, Fishkin's experiments involved random sampling of the population, designed to create a diverse and representative microcosm (note, though, that group polarization can and does occur within diverse groups). Third, the relevant experiments contained balanced panels of experts, able to respond to questions from small group discussions. Fourth, Fishkin's groups were overseen by moderators, trained to make sure that no one dominates the discussion, to ensure general participation, and to ensure a level of openness likely to alter some of the dynamics discussed here. Fifth, Fishkin's studies presented participants with a set of written materials that attempted to be balanced and that contained detailed arguments on both sides. The likely consequence 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, the very effort to produce balance should be expected to shift large majorities into small ones, pressing both sides closer to 50 per cent representation; and this is in fact what was observed in many of the outcomes in deliberative opinion polls.

In short, the external materials and expert panels shift the argument pool available to the deliberators and are also likely to have effects on social influence. Once certain arguments are on the table, it is harder to say how one or another position will affect a group member's reputation. Taken as a whole, a significant amount of Fishkin's data nonetheless seems consistent with the group polarization hypothesis. What does not is probably a product of some combination of moderator behavior, effects of external presentation, and deviations produced by convincing arguments from members of the particular groups involved. The most sensible conclusion is that the existence of monitors, an absence of a group decision, the great heterogeneity of the people involved in Fishkin's studies, together with the external arguments, makes the deliberative opinion poll quite different from the group polarization studies, in which small

groups of deliberators have relatively clear antecedent tendencies in one or another direction.

There are large lessons here about appropriate institutional design for deliberating bodies. Group polarization can be heightened, diminished, and possibly even eliminated with seemingly small alterations in institutional arrangements. To the extent that limited argument pools and social influences are likely to have unfortunate effects, correctives can be introduced, perhaps above all by exposing group members, at one point or another, to arguments to which they are not antecedently inclined. To the extent that institutional proposals are intended to increase public participation by promoting deliberation by ordinary people, they would do well to incorporate an understanding of these facts, which are sometimes neglected. The value of deliberation, as a social phenomenon, depends very much on social context—on the nature of the process and the nature of the participants. Here institutions are crucial. One of the most important lessons is among the most general: It is desirable to create spaces for enclave deliberation without insulating enclave members from those with opposing views, and without insulating those outside of the enclave from the views of those within it.