

Self-Silencing and Online Learning

Cass R. Sunstein

I.

A number of years ago, I found myself in a group of about eight people, at a large party, with one of the world's great athletes. The conversation flagged a bit, and I chimed in with a question I had long wanted to ask a professional athlete: "If the game is almost over, and if your team is losing, and if all the pressure is on you, and if everything depends on how you perform—*are you having fun?*?" He looked me right in the eye, and then he scanned the small group, and said this: "Absolutely! I was trained for that; I live for that! You ask if I am having fun when the game is on the line—are you kidding? Of course! I love it!"

About an hour later, I was near a corner of the room, all by myself, and not far away there was the athlete, also standing by himself. I called him over, looked up (he's tall), and said very quietly, "Your answer to my question really surprised me. I play a little sport myself, called squash, and I know some of the strongest players, and I don't think they would have given that answer. Is it really true?"

He smiled and said, "You want to hear the real answer? You asked if I am having fun when all the pressure is on me, and the game is on the line, and there's not much time left. Fun? God, no! It's awful. There's a ton of pressure, and it's on my shoulders." Then he paused. "Don't get me wrong. I know what to do. I'm trained for it. I'm good at it. But fun? Not at all!"

(I am not disclosing his name, because he trusted me.)

A second story is from Saudi Arabia. By custom, husbands usually have the final word on whether their wives join the labor force. Not long ago, researchers found that the overwhelming majority of young married men in Saudi Arabia are privately in favor of female labor force participation.¹ At the same time, they think that the level of support among other young men is low (including their own neighbors). The researchers randomly corrected those beliefs. The correction had significant effects. *Four months afterward, the wives of men whose beliefs were corrected were more likely to have applied and interviewed for jobs outside of the home.*

Cass R. Sunstein, Robert Walmsley University Professor, Harvard University. This essay was the basis for the keynote address at a virtual symposium at the University of Miami Law School, Power, Privilege, and Transformation: Lessons from the Pandemic for Online Legal Education, held on August 5, 2020. Readers are asked to make allowances for an essay meant as the foundation for oral remarks.

1. Leonardo Bursztyn et al., *Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia*, 110 AM. ECON. REV. 2997 (2020).

Both of these stories are about self-silencing, and about the pressures imposed by social norms. It is expected, and slightly heroic, for a great athlete to say that he likes the pressure of the last moments of a game. It is not heroic, and might become a newspaper story, if a great athlete says that he dislikes that pressure. If young men in Saudi Arabia learn that most of their peers think that it is just fine if their wives work outside of the home, then they are more likely to say that very thing to their wives. As a result, female labor force participation increases significantly.

Self-silencing is everywhere. It certainly occurs in law school classrooms. Some students might have convictions that they believe to be unpopular. They might be right of center. They might be left of center. They might have strong religious convictions. They might have beliefs that are difficult to map in standard ideological terms. They might be people of color. But because of their belief that their views are out of step with those of their peers, they might silence themselves. In producing censorship, social norms can do some of the work of law. It has become common to speak of “preference falsification,” which occurs when people speak falsely about their wants and desires.² Self-silencing can indeed lead to preference falsification, but it is a much broader concept. People might speak falsely or be silent about their approach to law, their values, their emotional states (very broadly speaking), and their experiences. Christian students might not speak about their faith. Women might not reveal experiences of sexual abuse. Believers in gun rights, or skeptics about affirmative action, might shut up.

2.

Consider a small experiment in democracy that I conducted with some colleagues a number of years ago.³ We brought about sixty American citizens together and assembled them into small groups, generally 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 climate change?

As the experiment was designed, some groups were “liberal” and others were “conservative”—drawn from residents of Boulder and Colorado Springs. (There was no mixing.) It is widely known that Boulder tends to be left of center and that Colorado Springs tends to be right of center. The groups were carefully screened to ensure that their members conformed to these stereotypes. People were asked to state their opinions individually and anonymously (by writing them down in private) both before and after fifteen minutes of group discussion, and also to try to reach a public verdict before making their final anonymous statements as individuals. I was especially interested in what would happen to

2. See, e.g., TIMUR KURAN, *PRIVATE TRUTHS, PUBLIC LIES* (1998).

3. David Schkade et al., *What Happened on Deliberation Day*, 95 CAL. L. REV. 915 (2007).

their anonymous statements of their views. How would those change as a result of a brief period of conversation with like-minded others?

The results were both simple and disturbing. In almost every group, members ended up with more extreme positions after they spoke with one another. They became more extreme both in their public statements during group deliberations, and their private statements, offered anonymously afterward. To offer a little more detail: Discussion made same-sex unions more popular among liberals; discussion made such unions less popular among conservatives. Liberals favored an international treaty to control climate change 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 affirmative action after discussion. Firmly negative about affirmative action before discussion, conservatives became even more negative about affirmative action after discussion.

As a result of the experiment, both liberal and conservative groups became more ideologically homogeneous. 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.

These findings are case studies in *group polarization*, which occurs when like-minded people engaged in discussion with one another end up thinking a more extreme version of what they thought before they started to talk.⁴ Consider some examples of the basic phenomenon, which has been found in over a dozen nations.⁵ (a) A group of moderately pro-feminist women will become more strongly pro-feminist after discussion.⁶ (b) After discussion, citizens of France become more critical of the United States and its intentions with respect to economic aid.⁷ (c) After discussion, whites predisposed to show racial prejudice offer more negative responses to the question whether white racism

4. See, e.g., CASS R. SUNSTEIN, *CONFORMITY: THE POWER OF SOCIAL INFLUENCES* (2019), for an overview.
5. See ROGER BROWN, *SOCIAL PSYCHOLOGY* 200-248 (2d ed. 1983). These examples include the United States, Canada, New Zealand, India, Bangladesh, Germany, and France. See, e.g., David Schkade, Cass R. Sunstein & Reed Hastie, *What Happened on Deliberation Day*, 95 *CALIF. L. REV.* 915 (2007) (U.S.); Johannes A. Zuber et al., *Choice Shift and Group Polarization: An Analysis of the Status of Arguments and Social Decision Schemes*, 62 *J. PERSONALITY AND SOC. PSYCH.* 50 (1992) (Germany); Dominic Abrams et al., *Knowing What to Think by Knowing Who You Are: Self-Categorization and the Nature of Norm Formation, Conformity and Group Polarization*, 29 *BRITISH J. SOC. PSYCH.* 97, 112 (1990) (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.
6. David G. Myers, *Discussion-Induced Attitude Polarization*, 28 *HUM. REL.* 699 (1975).
7. BROWN, *supra* note 5, at 223-24.

is responsible for conditions faced by African Americans in American cities.⁸ (d) After discussion, whites predisposed *not* to show racial prejudice offer more positive responses to the same question.⁹

These findings bear of course on classroom discussions, both in person and online. If students are inclined to a particular point of view, they might end up more unified, more confident, and more extreme. Certainly that is the case if the professor does not in some sense intervene. There is no question that in some classes, again both in person and online, group polarization is the essential pattern.

3.

Some of the most interesting work on social influences, bearing directly on what can happen in the classroom, involves the existence of informational and reputational “cascades.”¹⁰ A starting point is that when individuals lack a great deal of private information (and sometimes even when they have such information), they are attentive to the information provided by the statements or actions of others. If A is unaware whether genetic modification of food is a serious problem, he may be moved in the direction of alarm if B seems to think that alarm is justified. If A and B believe that alarm is justified, C may end up thinking so too, at least if she lacks independent information to the contrary. If A, B, and C believe that genetic modification of food is a serious problem, D will need a good deal of confidence to reject their shared conclusion.

The result of this process can be to produce cascade effects, as large groups of people eventually end up believing something simply because other people seem to believe it too. It should be clear that whether cascade effects occur may depend on seemingly small factors, such as the initial distribution of beliefs, the order in which people announce what they think, and people’s thresholds for abandoning their private beliefs in deference to the views announced by others.

Though the cascade phenomenon has been discussed largely in connection with factual judgments, the same processes are at work for judgments of value; we can easily imagine value cascades (information-induced or otherwise) that may well produce large-scale shifts.¹¹ In such contexts, many people, lacking firm

8. David G. Myers & George D. Bishop, *Discussion Effects on Racial Attitudes*, 169(3947) *SCIENCE* 778 (1970).
9. *Id.* See also Andrea Bohman et al., *Politics and Prejudice: How Political Discussion with Peers is Related to Attitudes About Immigrants During Adolescence*, 4 *FRONTIERS IN SOCIO.* art 70 (2019); V. Paul Poteat & Lisa B. Spanierman, *Modern Racism Attitudes Among White Students: The Role of Dominance and Authoritarianism and the Mediating Effects of Racial Color-Blindness*, 152(6) *J. SOC. PSYCH.* 758 (2012).
10. See Sushil Bikhchandani et al., *A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades*, 100 *J. POL. ECON.* 992 (1992).
11. An intriguing wrinkle is that when a cascade gets going, people might underrate the extent to which those who join it are reacting to the signals of others, and not their own private signals. For that reason, they might see the cascade as containing far more informational content than it actually does. See Erik Eyster & Matthew Rabin, *Naïve Herding in Rich-Information Settings*, 2 *AM. ECON. J.: MICROECONOMICS* 221 (2010); Erik Eyster et al., *An Experiment on Social Mislarning*

convictions of their own, may end up believing what (relevant) others seem to believe. Changes in social attitudes toward smoking, abortion, climate change, same-sex marriage, criminal justice, Second Amendment rights, recycling, and sexual harassment have a great deal to do with these effects. And small differences in initial conditions, in thresholds for abandoning private beliefs because of reputational pressures, and in who hears what when, can lead to major differences in outcomes.

Classrooms are full of cascades of various kinds. Most teachers have had the experience of an initial comment, perhaps a surprising one, shifting a discussion in an unanticipated direction, producing an apparent consensus that could not have been anticipated in advance, that could have easily been otherwise, and that does not, in fact, reflect the actual distribution of views among students.

4.

A number of years ago, I was walking to lunch with two colleagues. One was white; the other was African American. Our topic was police abuse. My white colleague said something like this: “We’re all subject to that. When I was nineteen, a police officer stopped me and asked me for my license. No reason at all. Just because I was young, I guess.”

My African American colleague paused and said, “Once I was driving in town with three Black friends. We were doing nothing wrong. Two police officers stopped us, asked us to get out of the car, shoved us against the trunk, and drew guns on us while they kept us there for ninety minutes. I have four other stories a lot like that. Did that ever happen to you?”

The question was rhetorical. But what kind of story is this? Is it a story of white privilege? Maybe. But to know whether we have a case of privilege, we need a baseline, or a theory of entitlement. It might be a privilege to receive an undeserved tax break. Is it a privilege not to be abused by the police? We do better to speak of inequality, and better still to speak of a system of racial caste. If one group of people receives systematically worse treatment than another from the police, and if what distinguishes the two groups is skin color, then we are dealing with something like a caste system.

To white people, some aspects of that system are not visible. When our African American colleague told his story, and then said that something similar had happened four other times, I did not think, “Oh, I already knew that.” I learned something.

5.

Suppose that group members have a great deal of information—enough to produce the unambiguously right outcome if that information is elicited and properly aggregated. Even if this is so, an obvious problem is that groups will not perform well if they emphasize broadly shared information while neglecting

(2018), <https://epub.ub.uni-muenchen.de/58086/1/73.pdf>. Norm entrepreneurs have a strong interest in promoting this mistake.

information that is held by one or a few members. Unfortunately, countless studies demonstrate that this regrettable result is highly likely.¹²

“Hidden profiles” is the technical term for accurate understandings that groups could but do not reach. Hidden profiles are, in turn, a product of the *common-knowledge effect*, through which information held by all group members has more influence on group judgments than information held by only a few members.¹³ The most obvious explanation of the effect is the simple fact that as a statistical matter, common knowledge is more likely to be communicated to the group. But social influences play a big role as well.

Consider a study of serious errors within working groups.¹⁴ The purpose of the study was to see how groups collaborate to make personnel decisions. Resumes for three candidates, applying for a position of marketing manager, were placed before group members. The experimenters rigged the attributes of the candidates so that one applicant was clearly the best for the job described. Group members would learn who was best if they were able to see, and to consider, all available information. Packets of information were given to the subjects, each containing a subset of information from the resumes, so that each group member had only part of the relevant information. The groups consisted of three people, some operating face to face, some operating online.

Almost none of the deliberating groups made what was conspicuously the right choice! The reason is simple: People failed to share their information in a way that would permit the group to make that choice. Members tended to share positive information about the winning candidate and negative information about the losers. They suppressed negative information about the winner and positive information about the losers. Hence, their statements served to “reinforce the march toward group consensus rather than add complications and fuel debate.”¹⁵

Or consider a simulation of political elections, in which information was parceled out to individual members about three candidates for political office, and in which properly pooled information could have led to what was clearly the best choice, candidate A.¹⁶ In the first condition, each member of the four-person groups was given most of the relevant information (sixty-six percent of the information about each candidate). In that condition, sixty-seven percent of group members favored candidate A before discussion, and eighty-five percent did so after discussion.¹⁷ This is a clear example of sensible aggregation of

12. Garold Stasser & William Titus, *Hidden Profiles: A Brief History*, 14 *PSYCH. INQUIRY* 304 (2003).

13. Daniel Gigone & Reid Hastie, *The Common Knowledge Effect: Information Sharing and Group Judgment*, 65 *J. PERSONALITY & SOC. PSYCH.* 959 (1993).

14. Ross Hightower & Lutfus Sayeed, *The Impact of Computer-Mediated Communication Systems on Biased Group Discussion*, 11 *COMPUTS. HUM. BEHAV.* 33 (1995).

15. PATRICIA WALLACE, *THE PSYCHOLOGY OF THE INTERNET* 82 (1999).

16. Garold Stasser & William Titus, *Pooling of Unshared Information in Group Decision Making: Biased Information Sampling During Discussion*, 48 *J. PERSONALITY & SOC. PSYCH.* 1467 (1985).

17. *Id.* at 1473; See also Stasser & Titus, *Hidden Profiles*, *supra* note 12, at 304.

information. Groups significantly outperformed individuals, apparently because of the exchange of information and reasons. Here, then, is a clear illustration of the possibility that groups *can* aggregate what members know in a way that produces sensible outcomes.

In the second condition, by contrast, the information that favored candidate A was parceled out to various members of the group, so that only thirty-three percent of information about each candidate was shared. Furthermore, as the condition was designed, that shared information favored two unambiguously inferior candidates, B and C. In that condition, fewer than twenty-five percent of group members favored candidate A before discussion—a natural product of the initial distribution of information. But if the unshared information emerged through group discussion, and were taken seriously, the group would end up selecting candidate A. The bad news is that number actually *fell* after discussion, simply because the shared information had disproportionate influence on group members.¹⁸

In other words, groups did worse than individuals, not better, when the information was distributed so that the key material was unshared and could emerge only from discussion. Under those circumstances, the shared information was far more influential than the unshared information, to the detriment of the group's ultimate decision.

From this and many similar studies, the general conclusion is that *shared information has a far larger impact than unshared information*. If everyone in a six-person group has the same information, and if two members have information of their own that everyone else lacks, there is a good chance that the shared information will dominate. More specifically, when most of the unshared information is opposed to the position that is initially the most popular, that unshared information will be omitted from the discussion—and will not have much of an influence on what groups end up choosing.¹⁹ It follows that “group decisions and postgroup preferences reflect the initial preferences of group members *even when the exchange of unshared information should have resulted in substantial shifts in opinion.*”²⁰

These results are best understood as a consequence of the *common-knowledge effect*, by which information held by all group members has a lot more influence on group judgments than information held by one member or a few.²¹ More technically, the “influence of a particular item of information is directly and positively related to the number of group members who have knowledge of that item before the group discussion and judgment.”²² When information is unshared, group judgments have been found to be no more accurate than the

18. Stasser & Titus, *Pooling of Unshared Information*, *supra* note 16, at 1473.

19. *Id.* at 1476.

20. *Id.* (emphasis added).

21. Gigone & Hastie, *supra* note 13.

22. *Id.* at 960.

average of the individual judgments, even though—and this is the central point—the groups have possession of more information than any of the individuals.²³

6.

Classrooms are hardly immune to self-silencing, group polarization, cascade effects, and hidden profiles. *First*: Students might self-silence because they do not want others to see that they are confused or baffled. Or they might self-silence because their views seem, to them, to be out of step with what most people think. They might not want to risk the trouble that might come from conveying those views. Or they might think if they reveal certain values, perspectives, and experiences, they will be at risk or lose something. *Second*: Classrooms might consist largely of like-minded people, or at least like-minded people might do much of the talking, which might lead to unity, confidence, and extremism.

Third: Classrooms often show cascade effects, as a teacher or a student states an initial position; a student agrees; another student agrees; and quickly or eventually, the entire group converges on that position, or at least seems to do that. In that respect, classes can show multiple equilibria. A large group might favor position A rather than position B, even though a small shock (in the form, perhaps, of a dissident view from the right person at the right time) might have swung the group to position B. *Fourth*: Classes might not get the information they could have and need, because those who have different or novel information do not reveal it. In fact, that happens all the time.

Some groups are less equal than others. In some law schools, for example, people who are right of center, or who have strong religious convictions, might think that they need to self-silence. (They might in some sense be right, given what matters to them.) They might believe in originalism in a constitutional law class in which originalism is taboo. They might think that *Roe v. Wade* was wrong, even preposterous, and that affirmative action is unconstitutional; but they do not say what they think. In discussions of the First, Fourth, and Fifth Amendments, students of color might have had perspectives or experiences that bear on Supreme Court decisions, perhaps suggesting that the racial dimensions have been overlooked, or that there is an equal protection dimension to which the Court has been blind. In the face of self-silencing, a cascade might develop, suggesting that the class as a whole believes X when many of its members reject X in favor of Y. Hidden profiles might remain hidden.

7.

Many teachers are aware of all this, whether or not they know the social science. They take steps to allow people to say what they think. They combat group polarization. If a cascade seems to be starting, they elicit opposing views, not because the cascade is necessarily wrong, but to ensure that if it prevails, it is because of its merits, not because of social dynamics. They try to elicit hidden profiles. These techniques are part and parcel of *liberal education*, understood to

23. *Id.* at 973.

entail a commitment to freedom of thought and expression, and an openness to competing views and perspectives (including those that contest liberalism itself). In some circles, the idea of “liberalism” is being caricatured, in such a way as to hold it responsible for all sorts of modern ills. But liberalism entails a respect for diverse opinions and a firm commitment to free discussion, as well as a master commitment to human dignity. It is radically opposed to the illiberal left (as well as the illiberal right).

In these endeavors, even the best teachers sometimes fail. Sometimes, for example, a left-of-center orthodoxy prevails in the classroom. If so, people who are right of center might feel silenced, even if the instructor tries hard to encourage them to speak. Or suppose that the class is predominantly white, and that students of color feel marginal or disrespected. If so, they might self-silence. Good classrooms are full of an *illusion of consensus*. Teachers often learn, a day later, or perhaps a year later, that some students had experiences, perspectives, or beliefs that would have proved valuable and challenging, but that were never voiced.

Is online education better or worse on these counts? There is no simple answer, because everything depends on how teachers choose to teach online. But it is clear that there are both risks and opportunities. One risk: It is easy to make a video of what people are saying, and to make their mistakes, or their departures from the status quo, relatively permanent. Outliers can be “canceled” or lapidated. The fear of recording can deter people from saying anything at all.

At the same time, online teaching can be used to combat some of the baleful effects of self-silencing, group polarization, and cascade effects. Faculty members can take advantage of rapid surveys, which can be anonymous. A teacher can ask some question and see the actual distribution of views within a class; surprises are likely. The fact that people are not in a literal room with one another can be helpful. It might soften some of the social pressures. People might well be able to feel braver, or safer, if they are in a room of their own. A great deal depends on how the teacher manages the show.

8.

Two years ago, in a course on social change, I decided to have a session on #MeToo. Several students asked if they could skip the in-person class and just watch the recorded session. They said that being there would be traumatic for them. They said that they would feel safer watching it alone.

That request raised a deep question: How should we think about the question of “safety”? Here is a very partial answer. People should feel safe to say what they think, so long as it is relevant (and so long as it does not violate basic norms of civility, properly defined²⁴). Online education presents unprecedented opportunities for overcoming self-silencing, and some of the baleful consequences of group polarization, cascade effects, and hidden profiles. It can also strike a

24. I am intentionally bracketing a large question here.

blow, or two, or more, against the caste like features of American society, and against the marginalization of reasonable people of diverse kinds.