Does democratic deliberation change minds?

Gerry Mackie

University of California, San Diego, USA

abstract   Discussion is frequently observed in democratic politics, but change in view is rarely observed. Call this the 'unchanging minds hypothesis'. I assume that a given belief or desire is not isolated, but, rather, is located in a network structure of attitudes, such that persuasion sufficient to change an attitude in isolation is not sufficient to change the attitude as supported by its network. The network structure of attitudes explains why the unchanging minds hypothesis seems to be true, and why it is false: due to the network, the effects of deliberative persuasion are typically latent, indirect, delayed, or disguised. Finally, I connect up the coherence account of attitudes to several topics in recent political and democratic theory.

keywords   deliberation, democracy, persuasion, change in view, coherence

I. Introduction

Discussion, as opposed to mere voting, is frequently observed in parliaments and in natural groups deciding on joint action, but it is just as frequently observed that public deliberation on a pending item seldom seems to change anyone's mind. Call this the 'unchanging minds hypothesis'. As rhetors, we act as if argument were potent, but, as auditors, we act as if it were not. The resolution of this puzzle, I shall argue, has to do with the coherence of any given individual's attitudes (the webbed interdependence among her various beliefs, desires, and actions), which make it such that the effects of persuasion are typically latent, indirect, delayed, or disguised.

Consider the teenaged Bronstein, who at first delighted the hostess of his lodgings by his sensible opposition to her sons' dangerous socialist ideas:

The stronger the pull of the new ideas the more desperately he clung to his assumed conservatism and indifference to politics. His spirit of contradiction and his eagerness to excel in argument did not easily allow him to yield. But his defences and vanity had to give way. In the middle of the school year he suddenly acknowledged his 'defeat', and at once began to argue for socialism with an ardour and acuteness which took aback those who had converted him.

Again and again we shall see this psychological mechanism at work in him: He is confronted with a new idea to which up to a point he is conditioned to respond; yet he resists at first with stubborn haughtiness; his resistance grows with the attraction; and he subdues incipient doubt and hesitation. Then his inner defences crumble, his self-confidence begins to vanish; but he is still too proud or not convinced enough to give any sign of yielding. There is no indication yet of the struggle that goes on in his mind. Then, suddenly, the new conviction hardens in him, and, as if in a single moment, over-comes his spirit of contradiction and his vanity. He startles his erstwhile opponents not merely by his complete and disinterested surrender, but by the enthusiasm with which he embraces their cause, and sometimes by the unexpected and far reaching conclusions which he draws from their arguments.[1]

I quote at length from Deutscher's life of Trotsky because the passage insight-fully describes important features of political persuasion, albeit in an exaggerated instance: strong resistance to persuasion on a specific point, perhaps some private softening on connected points, exposure to further persuasion on more or less connected points and further resistance; possibly the emergence of alternative attitude structures, private wavering between the old structure and new structure; seemly delay, rehearsal of excuses, justifications, and apologies; commitment by public declaration, and, finally, if need be, rearrangement of personal affiliations.

The basic premise of this article is that a given belief or desire is not isolated, but, rather, is located in a network structure of attitudes, such that persuasion suf-ficient to change an attitude in isolation is not sufficient to change the attitude as supported by its network. The network structure of attitudes explains why the unchanging minds hypothesis *seems* to be true, and why it is *false*. After this introduction, the second step in the argument is an account of consistency in beliefs and desires. Perfect consistency is assumed in canonical rational choice theory, but I argue that consistency is imperfect and, practically, is motivated by several aims; most importantly with respect to deliberation, consistency is valued as evidence of credibility in the democratic forum. Third, if attitudes are in a network structure, then we would only rarely observe any *manifest*, *direct*, and *immediate* effects of persuasion on attitude, either in natural or experimental set-tings, which contributes to the illusion of unchanging minds. Nor, typically, would we observe *latent*, *indirect*, and *delayed* effects of persuasion in natural settings, which again contributes to the illusion. If, however, attitudes are in a network structure, then we should be able to observe the non-direct effects of persuasion in experimental settings. Indeed, these effects are shown by a line of controlled experiments on group judgments carried out by the minority-influence school in social psychology. Fourth, even if one is persuaded by an argument, the

need for consistency as evidence of credibility in the democratic forum motivates one to disguise attitude change, and this further contributes to the illusion of unchanging minds. Additional experiments by the same line of researchers show that one's consistency and credibility are optimized by clinging to one's principles but agreeing to an unpleasant and unavoidable compromise in the end. Compromise is how individual consistency is made consistent with collective consistency, and such compromise is what we observe in practical settings where contesting viewpoints must resolve on joint action.

Fifth, the network structure of attitudes also explains the mechanics of minor and major attitude change in response to persuasion, as illustrated in the Trotsky example. A simple apparatus borrowed from connectionist psychology allows for an attitude network and its deliberative effects to be modeled and measured. It can be shown that an accumulation over time of minor and even latent changes in single attitudes can add up to sudden change in some large region of the network of attitudes. Persuasion seems to have no effect over several years, and then has a huge effect. The American politician Gary Hart said that you cannot turn people around only by one shoulder. The so-called 'parallel constraint satisfaction network' also better explains phenomena such as cognitive dissonance, adaptive preferences, inference to the best explanation, and, I suggest, public deliberation. Sixth, in the concluding section, I connect this approach to a few issues in political thought: Leiter's expressivism, Rawls's reflective equilibrium, Gaus's account of political concepts and political theory, Schmitt's anti-parliamentarism, Dryzek's contestation of discourses, Habermas's two-track theory of democracy, Manin's deliberative legitimacy, and the measurement of deliberative effects. I conclude that public deliberation is not futile.

II. Credibility and consistency

Rational choice theory assumes that actors possess consistent beliefs and desires.[2] Rationality is such consistency. The main defense of the assumption is instrumental: inconsistency is claimed to be self-defeating. Inconsistent beliefs about the world must be inaccurate. Short of ideal knowledge, however, even the best theory might contain inconsistencies or be inconsistent with another best theory. Classical physics seemed to explain all observations with a few simple laws and quantities – except for discovery of the puzzling constancy of the speed of light. Nowadays, quantum theory and general relativity is each the best theory in its own domain, but the two theories are not fully consistent with one another.[3] The money-pump argument against intransitive desires is not valid – for one reason, because an individual with inconsistent desires might refrain from acting on them,[4] for another, because a person with inconsistent desires could also operate as a money magnet.[5] Even if inconsistency of desire were individually self-defeating, it would only be so in realms of choice over which the individual is decisive;[6] over those realms of social choice where the individual

typically lacks decisiveness, there is no individual goal-attaining incentive to maintain consistency. Some individual inconsistency of desire is self-defeating, as when weakness of will has us wanting to quit smoking, but smoking nonetheless. But some inconsistency (such as an unpredictably erratic flight from predators, humor, or creativity) is self-advancing.[7] Surprisingly, individuals vary in their preference for consistency, and in one survey half of the subjects felt no special strain toward consistency.[8] The point of this brief review is that consistency is a means, not an end. Wood suggests a tripartite scheme of consistency motivations: accuracy motives, for survival in the world; integrity motives, relating to self-conception and expression of values; and social-relations motives, to facilitate cooperation and competition with others.[9]

Consistency was the dominant concept in the psychology of attitudes in the 1950s, 1960s, and 1970s. The various theories of cognitive consistency postulated a motive to maintain consistency among one's beliefs, desires, and actions, the most well known being Festinger's theory of cognitive dissonance.[10] In the seminal 'forced compliance' experiment, people were put to work at a boring task, and then paid either $1 or $20 to tell a waiting subject, actually a confederate of the experimenter, that the upcoming task would be exciting. Afterward, they were asked how interesting and enjoyable the boring task had been. At that point, the subjects could not change what they had told the confederate, but they could change their evaluation of the task. Taking $20 for telling the confederate that the task is exciting, and evaluating the task as boring, is more consonant; taking $1 for telling the confederate that the task is exciting, and evaluating the task as boring, is more dissonant. Standard incentive theory predicted that those who made the $20 would be more favorable to the task. If consonance is a motive, then those taking $1 would reduce dissonance by becoming more favorable to the task. Subjects' evaluations of the task were consistent with cognitive dissonance theory. This astounding finding was not easily replicated, but multiple researchers eventually identified the conjunction of conditions leading to dissonance arousal and attitude change: that the subject be personally responsible for aversive consequences.[11] The early impression-management reinterpretation of cognitive dissonance effects, that the motive behind consistency is 'the individual's need to maintain credibility for purposes of social influence',[12] does not contradict the later experimental refinement of conditions of the effect.

Consistency must be motivated. Motivations for consistency in democratic deliberation include accuracy of beliefs about the world and the avoidance of collective choices that are self-defeating because of inconsistency. Further, I value my consistency because it makes me credible in talk and action, and because it supports my integrity and self-worth. I value your consistency because it evidences your credibility and integrity to me. Credibility *is* consistency, in the circumstances of the democratic forum,[13] but this creates a problem. Those whose views change easily, from one position to another, on only the mildest breezes of persuasion, seem to us to lack credibility and integrity. Parliament-

arians express contempt for their weathervane colleagues – if not opportunist, their changeability indicates a shallowness of thought and corresponding lack of insight. At the same time, however, the deliberative ideal counsels us to be flexible and open to the views of others. How can these imperatives be reconciled? I suggest that they are best reconciled by aiming, as we do in practical politics, for compromise rather than consensus. Compromise accomplishes joint action, but preserves the credibility and integrity of each of the individuals who assemble it. If the compromise is successful, then as time goes by old actors may slowly adjust their self-presentations or views to the new order, and new actors may accept it as consensually settled.

If there were no problems of integrity and credibility, would it then be advisable that deliberators' views be maximally open to one another's? Not necessarily. Suppose that our collective task is the measurement or identification of some objective feature of the landscape, for example, we are navigating a ship at night and attempting to determine from sparse evidence (radar, sound, or moving lights) whether another ship is moving away from us (safe) or toward us (dangerous). Different pieces of evidence are reported and then evaluated by the navigation team. In a computer simulation, Hutchins modeled individuals each as a computer simulation of a neural network (these 'parallel constraint satisfaction networks' will assume an important role later in the article), each set up to choose one or another interpretation of the evidence, let us say, whether the other ship is moving away or coming closer.[14] Each individual has the same access to information, each has the same inner structure, but each differs in his or her initial expectations. Six of these individuals are linked together, and the persuasiveness of communication from one to another is varied from zero to one. With no persuasiveness each individual settles into an interpretation determined by his or her initial predispositions – three on one interpretation and three on another. With some persuasiveness, four settle on one interpretation and two on another. Notice that the lack of unanimity makes it such that the group would be sensitive to new evidence. With high persuasiveness, however, all six rapidly settle on one interpretation. We have consensus, but we also have a system with extreme confirmation bias: unless it is overwhelming, contradictory new evidence will be rejected as false, even if it is true. What is happening is that each actor is giving more credence to the evidence reported by other actors than the credence she gives to the evidence she receives directly from the environment. One can see such deluded consensus in stock-market bubbles or mass religious frenzies. It must not be, then, that actors should be maximally open to the views of others – not so open as to overwhelm the perspective of each. Rather, actors should be open to one another's views to some optimal extent, one that preserves what each has to contribute.

Consistency is more or less imperfect and, practically, is motivated by aims of accuracy, unwasted effort, integrity, and, in the democratic forum, credibility.

III. Compliance and conversion

Picture attitudes as related in a network structure. A belief, for example, is never isolated, but depends directly on several other beliefs, positively and negatively, and the other beliefs, in turn, depend directly on others, in complex, but not obscure, patterns. At the same time, a belief is stronger or weaker, for instance, one that is based largely on many direct perceptions, such as that it is cold in the winter in Chicago, is likely to be very strong. A belief based on a fragile scaffolding of inferences, such as that Ross 154 is the eleventh nearest star (true – probably), is weaker. New information contained in a persuasive message challenges not only the strength of the target belief, but also the beliefs directly and indirectly related to the target belief. Therefore, evidence against the target belief alone, that would be enough to revise the target belief taken in isolation, may not be enough to trigger revision because of its connection to related, but unchallenged, beliefs. Members of the antiwar new left, for example, rejected refugee claims about the Khmer Rouge massacre as anticommunist propaganda until those claims were later corroborated by invading, communist Vietnamese. Although the refugee claims were good evidence, accepting them disturbed too many other beliefs, not all of them simply procommunist, such as that nationalist leaders do not massacre their own populations. Anticommunists immediately credited the refugees' true testimony.

Of course, a person's beliefs are not all completely linked to one another. A particular belief may inhabit a more or less dense belief structure, and may be more or less central to the structure it inhabits, clumps of beliefs may be more or less connected to one another, and so on. Even if in some ideal sense all of one's beliefs should cohere with one another, as practical creatures we rely on one clump of beliefs in one domain and another in another. What I say about cognition also applies to motivation: that more emotionally charged desires are more stable, and that persuasion is less powerful against more important desires, is confirmed by experimental research. Desires are not only extrinsically constrained by beliefs and resources, they also may conflict intrinsically, such as wanting to sleep and wanting to eat at the same time. Desires also may be linked – one may be instrumental to the next, the next one to a further one, and the further one to some ultimate end. Similarly, for actions, the commitments that make up a person's private integrity and public reputation are variably important, variably interdependent, and variably linked. Finally, beliefs, desires, and actions are appropriately interconnected so as to compose attitudes.[15]

The hypothesis of unchanging minds appeals to the frequently observed fact that a persuasive challenge by Rhetor *R* against a targeted attitude *T* held by Auditor *A* does not result in *immediate evidence* of *A* changing her view of *T*. If attitudes are in a network structure, then a challenge to *T* sufficient to overcome *T* in isolation is likely insufficient to overcome *T* as supported by the network. One would expect, though, that as *R*'s challenge resonates through the network

of attitudes, the challenge would have some effect, if not a decisive one, on T, some effect on networked attitudes, and perhaps a delayed effect on T as the network readjusts over time to the challenge.

Alternatively, the hypothesis of unchanging minds could be explained by non-rational conformism. That would be a disappointment to the democrat, who hopes that the collective judgment about the common good would, instead, be based on good reasons arising from contestation of independent individual judgments. The conformity explanation seems to be supported by the widely known social psychology experiments of Sherif and Asch. A fixed point of light shown in completely dark surroundings, so that there are no visual references other than the light, will appear to the viewer to move. This is known as the 'autokinetic effect', first noticed by astronomers. Sherif, reporting in 1935, put to work the ambiguous cue of the autokinetic effect in a series of experiments on the topic of normalization, the convergence of individual judgments to a single group judgment. People were seated 15 feet (4.57 m) away from a source illuminated for five seconds, and were asked to estimate how far the light moved. They viewed the light either, first, alone and then as part of a small group of two or three, or, first, with the small group and then alone. Those who viewed the light first alone diverged in judgment, but, when repeating the task in a group, individual judgments converged. Those who began in a group converged in judgment, and stayed with the group norm upon repeating the task as individuals.[16]

Asch originally thought that an objective task, for example, matching a standard line to an unambiguously correct one of three comparison lines, would not involve social influence as had Sherif's subjective task. Asch's experiments, reported in 1956, pitted a minority of one against an otherwise unanimous majority. The scene was arranged so that one naive subject would publicly declare his judgment following the unanimous public declaration of judgment by each of seven to nine confederates of the experimenter. On the first, second, and fifth of nine trials (the sequence doubled, for a total of 18 trials), the confederates unanimously declared the correct judgment, otherwise they unanimously agreed on an incorrect judgment. About one in four of the subjects, and one-third of subjects' responses, conformed to the obviously mistaken majority. Most of the naive individuals were disconcerted by the experiment; most of the conformists reported a social rather than perceptual motive for their responses; but a few of the conformists said they were neither troubled by the experience nor aware of majority error. Increasing the size of the unanimous majority above three does not change the result, but the addition of a second naive subject almost obliterates both the tension and the conformism. Private response reduces, but does not eliminate conformism, it is the consistency among independent judges that more motivates the conformist response than does a desire for social approval.[17]

The impressive disconfirmation of Asch's initial hypothesis promoted a paradigm of functionalist conformity in research on the social psychology of groups. However, if groups function to maintain the conformity of members, then how

can they respond to a changed environment? Why do groups change when they do? Why do some members still dissent? Moscovici countered functionalism with an interactionist approach: influence is reciprocal among members, not between a group and its members; social change is as much an objective of influence as social control; in addition to the process of conformity and its dubious functional opposite, deviation, there is also the process of normalization or compromise, as well as the process of innovation by means of minority influence.[18] Moscovici and his colleagues made use of another visual judgment task, the blue–green procedure, to study minority influence. Now we have a group of four naive subjects and two confederates of the experimenter, who view blue slides and say out loud the simple color they see. In the control condition, six naive subjects uniformly said blue in response to viewing the blue slides. In a first experimental condition, the two confederates both claimed to see green on every trial, showing both interindividual and intra-individual consistency; in a second experimental condition, the two confederates said green for two-thirds of the trials, showing interindividual, but not intra-individual, consistency; and, in a third, the two confederates gave wholly inconsistent responses. The minority had no influence on the naive subjects, except under the first condition of strong consistency, when about one in three naive subjects said green at least once and 8 percent of the subjects' total responses were green. Next, in privacy, the same subjects looked at a series of slides changing gradually from blue to green and said blue or green to each so as to disclose the threshold of change. Compared to the control group, the experimental subjects in the strong consistency condition said green earlier in the series; a subject in this condition said green even earlier if he or she had *not* been one of those who had ever said green in the public phase of the ordeal. Thus, the minority had a *latent* influence, an influence that was stronger on those who had more manifestly resisted! As with the Asch experiments, subjects remained disturbed by the experience even after they had been told what it was all about.[19]

Moscovici's school maintains that majority influence and minority influence differ in process and result. Majority views tend to activate a comparison process; the individual compares his response to the majority's response; since the majority is likely to be correct or not worth contestation, the result tends to be individual *compliance* and public agreement with the majority view regardless of individual acceptance. Minority views tend to activate a validation process; the minority is presenting an innovation that must be compared to the reality in question; since the innovative minority view must be thought through to be understood, the result tends to be individual *conversion* and a shading toward the minority view regardless of individual awareness. In order to determine whether the subjects' green responses were merely verbal and a matter of public compliance or also perceptual and a matter of private conversion, Moscovici revised the experiment. If one stares at a color for a bit and then at a white screen, the afterimage one sees is the complementary color. The complement of blue is yellow-orange and the complement of green is red-purple. Subjects were paired with a

confederate who they were led to believe represented majority views in one condition and minority views in another condition. First, in the influence phase, the two viewed the blue slides together, the confederate consistently saying green out loud. Next, in the public phase, they viewed the slides together and privately wrote down the color of the image and afterimage; then the confederate left on a pretext and in the private phase the subject viewed the slides and wrote down the color of the image and afterimage. Naive subjects responded with green for the image about 5 percent of the time. However, those exposed to the 'minority' influence responded with the complement of green for the afterimage significantly more than did those exposed to the 'majority' influence, indicating some conversion in response to minority influence.[20] Again, minority influence has a latent effect.

A similar experiment added the twist that the subject and confederate underwent 45 minutes of sensory deprivation after the first influence phase, which seemed to amplify all effects, particularly enhancing the impact of publicity or privacy.

Table 1

Responding green (%)	Public phase	Private phase
Minority influence	9	20
Majority influence	27	23

As can be seen from Table 1, minority influence is weaker in public than in private; majority influence is stronger in public than in private; in public, majority influence is much stronger than minority influence; and in private, majority and minority influence are almost the same.[21]

The abortion and contraception attitude experiments of Mugny and Perez illustrate the *indirect* effect. Young people in Spain, when abortion was illegal and controversial, read a text supposedly by an unspecified 'minority group' ardently endorsing completely legal and state-funded abortion. Later, the subjects repeated a 25-item questionnaire on attitudes toward abortion and contraception. Exposure to minority influence favoring legal abortion did not change subjects' attitudes to abortion, but did favorably change their attitudes to related, but unmentioned, contraception; in similar experiments, a *delayed* minority influence effect was also demonstrated.[22]

Latent, indirect, and delayed effects of persuasion were also demonstrated by one of the major cognitive consistency theorists, McGuire, in quite a different research paradigm. At a first session, subjects rated both the desirability and probability of 48 statements, not being told that the statements had been

scrambled from 16 sets of three syllogistically related propositions. At the second session a week later they were exposed to believably persuasive messages on statements relating to some of the underlying syllogisms, and then asked to rate desirability and probability again. At the third session another week later they rated yet again. A persuasive message was aimed at an *explicit* issue, and change in attitudes measured on both the explicit issue and on logically related, but unmentioned, *remote* issues. The result was that change on an explicit issue also produced change on remote issues so as to maintain internal consistency of beliefs. The changes on remote issues, however, were less than would be required for complete consistency with the change on the explicit issue, indicating inertia within attitude structures; and changes on the remote issue continued in the logically required direction a week after receipt of the persuasive message. Immediately after the persuasive message, change on explicit issues was high and positive, and change on remote issues low and positive; a week later, change on the explicit issue had decayed by almost half and change on remote issues had decayed slightly, such that explicit and remote issues approached mutual consistency.[23] In a related experiment, McGuire showed that dissonance-increasing persuasive messages (analogous to minority influence), as compared to dissonance-reducing persuasive messages (analogous to majority influence), resulted in less change on explicit (direct) issues and more change on remote (indirect) issues.[24]

Finally, experimental research testing Rosenberg and Abelson's theory of cognitive balancing found the subjects adjusting weights on the explicit beliefs and desires given in a role-playing exercise so as to take account of goals only implicit to the given attitude structure: 'subjects seek not only the attainment of cognitive balance and consistency but they seek also to alter their beliefs and evaluations in ways that maximize expected gain and minimize expected loss'.[25]

According to these experiments, persuasion has latent, indirect, and delayed effects, consistent with a network structure of attitudes. In ordinary experience, R is typically in the position to observe that her challenge has no manifest, direct, and immediate effect on T, and is typically not in the position to observe the latent, indirect, and delayed effects of her challenge in A. This contributes to the *illusion* of unchanged minds. Another mechanism contributes to that illusion. Suppose, now, that A is directly persuaded to change her view of T. Nevertheless, credibility as consistency in the public forum would require that in the short run A *disguises* that she has been persuaded.

IV. Consensus and compromise

Why disguise attitude change? Consistency is ordinarily the leading component of influence – in the unquestioned majority, by way of interindividual consistency and, in the resolute minority, by way of both interindividual and intraindividual consistency. But for contending minorities to compose a majority

sufficient for action, each minority must sacrifice its factional consistency to a majority plan of group action. How then can a faction maximize credibility? A faction could be completely consistent and refuse all compromise: although not publicly successful at the moment, an accumulation of latent and private changes may accomplish conversions in the long run. Alternatively, a faction could quickly acquiesce to a consensus: unless factions converge on a norm, the joint action necessary to wield some portion of public influence may never come about. Then again, a faction could maximize both private and public influence, by insisting on its position to the end, and compromising on joint action only as a tactical concession. Compromise is what we usually see in politics, and there is laboratory evidence that the compromise strategy optimizes both types of influence.

People were asked to participate in mock jury deliberations on compensating a victim damaged by the mid-air collapse of a ski-lift. The victim asked for $500,000. Most subjects thought he was due $150,000–300,000, and only they were invited to participate in the experiment. Three naive individuals whose individual positions were not more than $50,000 apart were placed with a confederate of the experimenter who held that the victim was due only $50,000. Stubborn *consistency* was operationalized as the confederate holding to $50,000 through 10 rounds of stylized discussion, *early compromise* as the confederate changing to $100,000 on the second and following rounds, and *late compromise* as the confederate changing to $100,000 on the ninth and tenth rounds; the *control* was three naive subjects deliberating on their own. The next day subjects returned and deliberated a similar case involving a negligent corporation without the confederates. On the first day's confederate-influence case, early compromise and late compromise had the most influence (see Table 2). On the second day's similar case, the most influence was found among subjects who had the day before been exposed to the late-compromising confederate and the least influence was found among subjects who had been exposed to the early-compromising confederate.[26] Overall, late compromise had the most influence.

Table 2

Influence[a] (Higher number means more influence)	Consistent	Late compromise	Early compromise
Influence case	28,500	56,945	59,445
Later similar case	76,810	103,881	47,692

Note: Influence = (mean control) – (mean position, last round).

In another experiment, people were asked to name the simple color of a disk

which could be seen as either blue or green. One subject was placed with one confederate, and the confederate gave the response opposite to the subject's. In the negotiative consistency condition, the confederate and subject discussed the choice before each trial until the subject agreed or until 20 trials were completed; in the repetitive consistency condition, the confederate repeated the same choice for 20 trials; and in the abjuration condition, after five consistent choices the confederate waved his or her hand and unconvincingly said, 'Okay, it must be [the subject's response]', and agreed with the subject for five more choices. Abjuration had the strongest delayed influence. In an additional experiment, abjuration was compared to disavowal (agreeing solely to reduce interpersonal conflict) and to renunciation (conforming without saying anything). Again, abjuration had the strongest delayed influence.[27]

To generalize, it seems that one's display of consistency, and thus one's credibility, is optimized by clinging to one's principles, but reluctantly agreeing to an unpleasant, but unavoidable, compromise in the end, which is precisely the rhetoric we see associated with the passage of major legislation or a controversial committee decision. Such compromise is how individual consistency is made consistent with collective consistency. To demand complete consensus as the goal of group deliberation not only humiliates the minority, but also, since persuasion is slow, both prematurely and permanently forecloses the possibility of minority influence and change in the majority position.[28] If someone is quickly convinced by the arguments of another, nevertheless, for the sake of his long-term credibility, he should not immediately concede. What he can do, perhaps, is be quiet if the issue arises again. If, rather, someone is slowly convinced by an alternative, then her credibility remains intact. Either way, persuasion takes time for results.

Habermas (if not his followers) construes political compromise as self-interested bargaining, a begrimed second-best to the pious first-best ideal of consensual agreement.[29] I have argued, first, that pressure to consensus, rather than acceptance of compromise, might lead a group into a deluded consensus that overwhelms the perspectives of its members. Second, the reconciliation of credibility and integrity as consistency with agreement on joint action requires formulation of the agreement as a compromise rather than a consensus. Compromise can be a more noble and more motivationally intricate process than the portrayal of it as a poor approximation of some fully consensual ideal.

If attitude change is disguised, then R again sees no immediate evidence of attitude change in A. Latent, indirect, delayed, and disguised effects of persuasion contribute to the illusion of unchanged minds. We have considered minor and briefly-delayed change in attitude. Next, we will consider major and long-delayed change in attitudes. A gradual accumulation of minor and even merely latent changes in single attitudes can add up to a sudden change in some large and densely connected region of the network of attitudes, and such changes can be modeled and measured.

V. Coherence

Despite the obvious importance and appeal of the notion of coherence in under-standing human action, and the fruitful harvest of novel experimental findings, theories of cognitive consistency fell into disuse. That seems to be because without more definite specifications of consistency and its operations, nothing of much usefulness remained to be said.[30] Likewise, it has been said that theories of deliberative democracy lack a formal framework as compared to the canonical decision theory that the observations of deliberative attitude change disturb. Both deficiencies are remedied by the development of parallel constraint satisfaction network models used recently to explain cognitive dissonance, inference to the best explanation, individual decision-making, analogical reasoning, political pub-lic opinion change, and attributions in dyadic interpersonal relationships, among other things.[31] The surprisingly simple apparatus of these coherence models is capable of explaining and predicting not only changes in beliefs, but also changes in desires, both gradual and sudden, as well as ambivalence, local inconsistencies, and framing effects, and could be used to explain latent, indirect, delayed, and dis-guised effects of persuasion. If successfully developed, this symbolic connection-ist approach could subsume the decision theory of fixed desires as a special case.[32]

How does parallel constraint satisfaction work? Suppose that attitudes are related in a network structure. An individual's various beliefs, desires, and actions can be represented as nodes in the network, and the relations among them as links between the nodes. The nodes differ in initial strength. The links differ in both sign (positive, null, or negative) and in initial strength. Activation by a persuasive input spreads 'around the network in parallel until the activation of nodes asymptotes. The final activation of a node is a function of its initial acti-vation, the activation of the nodes to which it is linked, and the sign and strength of the links to other nodes', the parallel constraints.[33] The final activation of the nodes is interpreted as a comparative degree of acceptability to the individual. A belief, for example, that is mostly supported by other beliefs will be more positively activated and more acceptable, one that is less supported will be less positively activated and less acceptable, and one that is mostly contradicted by other beliefs will be negatively activated and unacceptable. As with learning something such as simple game theory, it takes the working out of a few examples to get a feel for the model beyond what must seem to the reader to be a highly abstract description.

Imagine a persuasive challenge as a jolt of negative energy to the targeted atti-tude in a *network*. From node *T* that energy spreads in all directions at once (in *parallel*), weakening along positive links and strengthening along negative links (the *constraints*). Linked nodes update one another until, eventually, the system settles into a new equilibrium (*satisfaction*), with most nodes at new values. Standard decision theory, in contrast, models change in only one constraint at a time.

Recall the famous 'forced compliance' experiment demonstrating reduction of cognitive dissonance. A second-generation replication of the experiment assigned subjects to choice and no-choice conditions. As expected, subjects who had a *choice* about making the counter-attitudinal statement and who received the *lower* payoff showed a dissonance-reducing change in *favor* of the attitude. Subjects who had *no choice* about making the counter-attitudinal statement (remember that dissonance reduction was eventually shown to be limited to situations of personal responsibility for aversive consequences) and who received the *higher* payoff showed an incentive-type change in *favor* of the attitude. Shultz and Lepper simulate the experiment with an extremely simple parallel constraint satisfaction model, which predicts *both* the choice result, as does cognitive dissonance theory, *and* the no-choice result, as cognitive dissonance theory does not.

Even more impressive, in my opinion, is the performance of Shultz and Lepper's consonance model with respect to 'sour grapes', a concept that grew out of Festinger's dissonance-reduction theory. Choosing one alternative over another sets up dissonance between the chosen and the foregone alternatives – the fox who does not get the grapes calls them sour.[34] In the original free-choice experiment, subjects were asked to rate eight different small appliances. Then they were given either a difficult *High–High* choice between two high-rated objects or an easy *High–Low* choice between a high-rated object and a low-rated object, and what they chose they got to keep. Then they rated the appliances again. Separation was defined as the post-choice rating minus the pre-choice rating on each object. Cognitive dissonance theory predicted greater separation in the High–High choice condition than in the High–Low choice condition, which is what was observed. Also observed, but not predicted by cognitive dissonance theory, was that most of the separation was due to down-evaluation of the rejected alternative in the High–High choice condition. Shultz and Lepper's simple parallel constraint satisfaction model predicts *both* observations. Moreover, Shultz and Lepper replicated the free-choice experiment with the addition of a new, difficult *Low–Low* condition, a choice between two low-rated objects. Both cognitive dissonance theory and the Shultz and Lepper consonance model would predict greater separation in the difficult Low–Low condition than in the easy High–Low condition; the consonance model would also predict that most of the separation would be due to up-evaluation of the chosen alternative in the Low–Low choice condition, which is what was observed in the experiment.[35]

Thagard's theory of explanatory coherence, and its computer-program implementation, is another noteworthy application of the parallel constraint satisfaction model to high-level cognition. Parallel constraint satisfaction models were first developed for the study of visual perception. Eventually, they were suggested as an explanation for the gestalt shift in perception of ambiguous visual stimuli, particularly of the Necker cube illusion, in which a two-dimensional line drawing is seen as depicting either one or the other of two distinct

three-dimensional cubes, but never both. A connectionist network is constructed; the nodes of the network represent the corners of the drawing; the links between nodes are relations of compatibility and incompatibility between interpretations; and parallel constraint satisfaction converges on one interpretation or the other, but never both, just as in perception. Holyoak realized that the connectionist model could be applied to analogical reasoning, and Thagard that it could model the scientific process of inference to the best explanation, reconciling the accretion theory of scientific change with the gestalt-shift theory of scientific change.[36]

The addition or change of a piece of empirical data may not cohere with the reigning theory, but that does not overturn the theory, as the theory is otherwise supported by many other pieces of empirical data. In a parallel constraint satisfaction network, the anomalous observation would receive a low value – suggesting that the observation might be mistaken. A number of suspect anomalies might accumulate; as they do, the value on the reigning theory marginally declines and that on an emergent theory marginally increases. A point may come when the anomalies accumulate to a critical mass, and the network flips to a positive, accepting value on the emergent theory and a negative, rejecting value on the formerly reigning theory (as well as overcoming suspicion of the anomalous observations, changing their values from low to high). Yes, a massive piece of evidence (the smashing of the Titanic into an iceberg, Hitler's invasion of Poland, or the attack on the Twin Towers) could quickly overwhelm any theory inconsistent with it and motivate joint action. For many social choices, however, evidence is vague, incomplete, and contradictory, and a variety of plausible theories compete to explain that imperfect evidence. Many bad theories might be eliminated, but several more than one good theory might remain. Thagard's computer implementation tends to select one hypothesis if the absolute value of the ratio of excitatory weights to inhibitory weights is low; but can activate more than one contradictory hypothesis if that ratio is high. As observations about the physical world are much less likely to conflict with one another than are the judgments expressed in politics, this parameter could account for the tendency to one reigning theory in physical science, but several competing ideologies in politics.

Thagard has suggested that his coherence model is apt for epistemic justification, mathematics, logical justification, ethical justification, legal justification, practical reasoning, perception, discourse comprehension, analogy, cognitive dissonance, impression formation, and democratic deliberation.[37] He first introduced his theory of explanatory coherence to explain shift in conceptual systems, to explain the scientific revolutions brought to scholarly attention by Kuhn.[38] He defines explanatory coherence as seven explicit principles, which are implemented in a computer program that takes as input hypotheses data and relations of coherence and incoherence. Activation spreads from the data units through the network, which updates in parallel until settling on stable activation values. More highly activated hypotheses enjoy more explanatory coherence. Thagard applied

his theory of explanatory coherence to the major scientific revolutions: the Darwinian, the geological (plate tectonics), the Copernican, the Newtonian, the Einsteinian, the quantum, the behaviorist, the cognitivist, and the emergence of Lavoisier's oxygen theory in chemistry.

Inference to the best explanation in science is analogous to deliberative opinion change in a democratic assembly.[39] In science, says Thagard, persuasion is not instantaneous:

> The major mechanism by which [a new conceptual system *publicly* overcomes an older one] is scientific *argument*. It would be naive to suppose that arguments directly convince people. Rarely on an issue of any complexity and importance can you simply say to someone: Here are premises you accept, from which the conclusion follows, so accept it. There are always responses [available to] arguments. But this does not mean that argument is futile, for the process of argument and later reflection on it can lead to revision of conceptual links, enabling an alternative system to come to the fore . . . Chemists resistant to Lavoisier's ideas nevertheless repeated his experiments, thereby acquiring parts of his conceptual system. Perrin reports that it typically took several years for people to pass from opposition to Lavoisier's views to their acceptance. On my account, these years were spent both building up the new system and strengthening its links to where the new system seemed more coherent than the old.[40]

With respect to democratic deliberation, Chambers suggests that consensual will formation is not the instantaneous result of a single conversation, but is the cumulative product of many criss-crossing conversations over time:

> We do not like to admit that we are wrong even in the face of evidence; we are very attached to our views; we often enter conversations with set opinions and leave with the same set opinions . . . That a single . . . conversation, especially on a highly charged subject, appears much more likely to end in disagreement than agreement is not strong evidence against the power of rational argumentation. If we step back from the model of the single conversation, we see that people do in fact change their minds; they do find new arguments, positions, and perspectives more convincing than old ones; they are swayed by argumentation. This process goes on over time, however . . . We often reevaluate our position between conversations rather than within them. We are sometimes not even aware that our position has subtly shifted in response to and reflection upon a criticism or challenge. Not only is the process gradual but it is fragmentary and partial . . . that single conversations often do not end in agreement does not mean that people are not swayed by argumentation.[41]

Because of the network structure of attitudes, major attitude change is a delayed, gestalt-shift-type process, a process of rational resistance in the short run and rational conversion, if appropriate, in the long run. That explains the features of political persuasion as seen in the case of Trotsky. Moreover, that helps explain why, as rhetors, we act as if argument were potent, but, as auditors, we act as if it were not.

VI. Conclusion

The fact that attitudes are not isolated, but are located in a network structure, explains several features of persuasion and attitude change, notably, the puzzle of unchanging minds. That the network and its effects can be efficiently represented by a simple connectionist model strengthens the explanations, and, in turn, suggests fruitful new reflections. Here, I will mention a few possible insights into political thought.

At the level of ordinary, everyday life, it is a relief to know that people can change their minds in the long run and, thus, that argument is not generally pointless. Legal theorist Brian Leiter, who broadcasts persuasive arguments from his blog almost daily, bewails that, 'When it comes to politics . . . reasons and evidence appear to play almost no role in changing anyone's views . . . it is quite rare to persuade anyone by a careful, reasoned argument.'[42] In a classic instance of sour grapes (itself explainable within the connectionist approach), Leiter declares that his goal is not to persuade, but, rather, to bolster the like-minded and to express their joint outrage. The expressive value of persuasion would be nil, however, if there were no possibility of instrumental effect, or so it seems to me.

Inference to the best explanation of a scientific theory in regard to observations, even to the extent that the best theory marks anomalous observations as doubtful, closely resembles Rawls's reflective equilibrium between, on the one hand, a few and simple principles of justice and, on the other hand, our many and complex considered judgments of justice. Rawls could be describing the activation and adjustment to equilibrium of a parallel constraint satisfaction network:

> By going back and forth, sometimes altering the conditions of the contractual circumstances [generator of the simple principles], at others withdrawing our judgments and conforming them to principle . . . eventually we shall find . . . a reflective equilibrium. It is an equilibrium because at last our principles and judgments coincide . . . for the time being we have done what we can to render coherent and to justify our convictions of social justice.[43]

Rawls's idea that our more confident convictions on the injustice of religious intolerance and racial discrimination provide support to an overarching conception of justice, which in turn 'upvalues' our heretofore weaker convictions about the distribution of wealth and authority, is entirely in the connectionist spirit. Further, for Rawls, it is not that his conception of justice is true and others false, rather his two principles of justice are only better supported than other known alternatives such as utilitarianism and intuitionism:

> Now each of these conceptions presumably has its assets and liabilities; there are reasons for and against any alternative one selects. The fact that a conception is open to criticism is not necessarily decisive against it, nor are certain desirable features always conclusive in its favor . . . when everything is tallied up, it may be perfectly clear where the balance of reason lies . . . The argument for it is not strictly speaking a proof . . . but . . . it may present considerations capable of determining the intellect.[44]

In *Political Liberalism*, Rawls is concerned to explain how conscientious persons with full powers of reason, even after free discussion, might reasonably disagree. Many political judgments are vague, perspectival, and conflicting – what he calls the burdens of judgment. In such circumstances, a number of reasonable comprehensive doctrines would persist.[45] In Thagard's explanatory coherence network, a high ratio of excitatory weights over inhibitory weights settles into a middling positive valuation of several overall theories, rather than just one. Departing from the specifics of *Political Liberalism*, on a scale from –1 to +1, an established theory in physics could enjoy a +0.9 coherence rating, and outmoded theories each a negative coherence rating. A mature moderate in political theory, however, might reckon that left-liberalism is +0.65 coherent concerning our convictions about justice, and that right-liberalism is +0.45 coherent, or the reverse (and that communism and fascism rate negatively). The political theorist appreciates the appeal of the competing system, believes that her favored system does more, but rejects some systems as mistaken.

The connectionist apparatus might also improve our understanding of the relationship of political concepts and political theories. One popular view, after Gallie, is that political concepts are essentially contested. A complex evaluative concept (Gallie's example is what it means to be a champion in a sport) can be interpreted by different parties each in her own way, and, in principle, it is impossible to resolve the dispute among them. Gaus challenges this unconstrained view of the political concept. He defines a conception as an organized interpretation of a concept, and, in turn, a political theory as a coherent system of conceptions of liberty, power, equality, justice, authority, and the like. If Gallie is correct, asks Gaus, why do we fight so hard for our own conception and why do we resist an opponent's? 'Why argue when no one can be right?'

> A person's favored conception of, say, liberty, is not freestanding, but linked to her favored conceptions of equality, justice, and so on. Consequently debates about one concept lead to our interpretations of others; to give way to our opponent on one concept may lead to undermining our entire political outlook.[46]

A concept in isolation will be floppy, open to many interpretations, with no way to decide among them, but a system of conceptions constrains the concepts it subsumes. The popular notion of political concepts as essentially contested could be a mistaken artifact of the analytic prejudice for taking concepts in isolation.[47]

If representative government or parliamentarism is government by discussion, and if parliamentary discussion is an empty and trivial formality, in other words, if argument is without effect, then parliamentarism loses its justification. In 1923, Carl Schmitt warned that, 'on the European continent . . . the smallest number still believe that just laws and the right politics can be achieved through newspaper articles, speeches at demonstrations, and parliamentary debates'.[48] The work of Habermas could be said to be one long answer to Schmitt's disastrous error. Perhaps in response to the apparent lack of persuasive effect in parliaments,

but for good independent reasons as well, Habermas and other deliberative democrats highlight discursive processes in civil society. Habermas's two-track model of democracy contains strong publics and the state, especially the legislature, involved in both opinion formation and decision-making, and weak publics and civil society involved in opinion formation and unconstrained by the responsibility of decision. Communicative power suffuses both. Weak publics raise problems and solutions, and the strong public decides among the competing solutions.[49] It may not be necessary to flee so far into civil society if, contra Schmitt, parliament does indeed change its mind.

Democracy and democratization are not just in the state, but also in civil society, according to John Dryzek. Generally, the locus of democratization depends on historical and comparative factors. But often, civil society is more attractive than the state, because it is not constrained by system imperatives or by pursuit of strategic advantage. Dryzek interprets civil society as a contestation of discourses. For example, in reconstructive empirical work, he identifies 11 discourses in environmental politics, ranging from survivalism to economic rationalization to ecological democracy.[50] Dryzek says that 'shifts in the relative weight of discourses or discourse positions can have real consequences', providing examples on environmental policy in international civil society.[51] Discourses and their contestation can be understood in connectionist terms. Each discourse is a more or less coherent organization of empirical observations, moral intuitions, and affect. Just as scientists attend to a few theories rather than millions of observations, so citizens attend to a few discourses. Individuals, due to the burdens of judgment, and due to error, variably weight the justificatory power of the several discourses from –1 to +1. There is also a vague weighting by the general public of the discourses (for example, in Europe, the discourse of liberalism is positive and monarchism, communism, fascism, and Islamism increasingly negative). Discourses can be distorted by power, for example, propagandistic repetition of its elements and their false association with primal needs can wrongly 'upvalue' a discourse. The contestation of discourses can result in revolutionary upheavals. A new discourse of superior coherence can displace older discourses of inferior coherence. Conclusions from one discourse can enter a second and disrupt its coherence.

For Manin, democracy is not legitimated by the unanimous will of all, but by the process of deliberation of all. Public deliberation reforms the uninformed, incomplete, intransitive, and unjustified pre-political preferences of the individual citizens. Political argumentation is not logic, it does not proceed from true premises to a true conclusion. Argumentation assumes premises generally accepted by the addressed public, and need not proceed by deduction. Arguments offered are neither true nor false, but stronger or weaker, and the conclusion of argumentation is not true or false, but more or less supported by argument. Conclusions are not demonstrated, they can only be justified. Competition among candidates for representative office motivates them to offer the more general

viewpoint. The deliberative process is brought to a close by a majority vote, which reflects the greater strength of one set of reasons over another. The majority view is not true, and the minority view is not false. The minority *also* has good reasons, just not reasons judged as strong as those of the majority. The result is legitimate because everyone was able to take part, all views were taken into consideration, and each was free to accept or reject arguments and conclusions.[52] To go beyond Manin, the minority improves public judgment by activating a validation process in the otherwise conformist majority. Respecting the minority discourages deluded consensus, and compromise protects the integrity and credibility of the contestants.

Delli Carpini et al.'s 2004 review of the empirical literature on the presumed individual and collective benefits of public discourse concludes that the results are suggestive and promising, but inconclusive. Americans do engage in public talk, and that is good to know. Otherwise, deliberation *can* lead to (but does not guarantee and does not tend to deliver) supposed benefits, but its impact is highly context dependent, 'rife with opportunities for going awry'.[53] Much of that literature, however, is made up of one-shot studies which fail to consider the cumulative and delayed effects of deliberation. My claim is that casual observations and systematic studies which assert an absence of deliberative effect may be mismeasuring persuasion. They are neglecting to consider the networked structure of attitudes, and are neglecting to measure for latent, indirect, delayed, and disguised attitude change.

I am not making an unfalsifiable claim – it is possible to conceptualize networked attitudes and to measure for changes in the network. A first simple step in this direction was taken by a study which surveyed and modeled changes in view from the beginning of the Gulf War in 1991 to two weeks later.[54] Some six attitudinal constructs were extracted from survey answers provided by 122 students: general support for US action, pacifism, the legitimacy of US action, isolationism, Saddam Hussein's threat, and the possibility of terrorist retaliation. Opposition to US action changed to support for US action over the two weeks. The researchers propose that external evidence had caused respondents to see Saddam as more of a threat, and that this had changed opposition to support for intervention. Those who consider attitudes only in isolation would assume that a pacifist attitude and judgment of illegitimacy of intervention would be stable regardless of variation in threat from Saddam. The researchers' parallel constraint satisfaction model of attitude structure, however, in response to increased threat from Saddam, predicts both reversal from opposition to support for intervention and reversal in the endorsement of a pacifist attitude and a judgment of illegitimacy, which is what was observed in survey answers. Studies of this kind, and, probably better, long-range studies of public opinion, or historical studies of the contestation of discourses over a good number of years, are more apt to discern variations in deliberative effect.

Deliberation is 'discussion aimed at producing reasonable, well-informed

opinions in which participants are willing to revise preferences in light of discussion, new information, and claims made by fellow participants'.[55] If the unchanging minds hypothesis were widely true, then deliberation would have little effect. Whether or not deliberation is desirable, it would be futile.[56] This article, if successful, defeats the unchanging minds objection to deliberative democracy.

I hope that sooner or later you will come to agree with my idea that the general will can be productively understood in terms of the parallel constraint satisfaction network model.

notes

1. Isaac Deutscher, *The Prophet Armed* (New York: Oxford University Press, 1954), p. 22. The example is borrowed from Ted Goertzel, *Turncoats and True Believers: The Dynamics of Political Belief and Disillusionment* (Buffalo, NY: Prometheus Books, 1992).

2. Jon Elster, 'Introduction', in *Rational Choice*, edited by Jon Elster (New York: New York University Press, 1986).

3. P. Thagard, C. Eliasmith, P. Rusnock and C.P. Shelley, 'Knowledge and Coherence', in *Common Sense, Reasoning and Rationality*, Vol. 11, edited by R. Elio (New York: Oxford University Press, 2002).

4. See Wlodek Rabinowicz, 'Money Pump with Foresight', URL: http://socpol.anu.edu.au/pdf-files/rabinowicz.pdf (Lund: Department of Philosophy, Lund University, 2002).

5. Gerry Mackie, 'The Money Pump is Dry', Social and Political Theory Program Brownbag Seminar, Australian National University, 6 February 2001.

6. See Geoffrey Brennan and Loren Lomasky, *Democracy and Decision* (Cambridge: Cambridge University Press, 1993).

7. Geoffrey F. Miller, 'Protean Primates: The Evolution of Adaptive Unpredictability in Competition and Courtship', in *Machiavellian Intelligence II*, edited by Andrew Whiten and Richard W. Byrne (Cambridge: Cambridge University Press, 1997).

8. Robert B. Cialdini, Melanie R. Trost and Jason T. Newsom, 'Preference for Consistency: The Development of a Valid Measure and the Discovery of Surprising Behavioral Implications', *Journal of Personality and Social Psychology* 69 (1995): 318–28.

9. Wendy Wood, 'Attitude Change: Persuasion and Social Influence', *Annual Review of Psychology* 51 (2000): 539–70.

10. The concept of cognitive consistency peaked with Robert P. Abelson, Elliot Aronson, William J. McGuire, Theodore M. Newcomb, Milton J. Rosenberg, and Percy H. Tannenbaum (eds), *Theories of Cognitive Consistency: A Sourcebook* (Chicago, IL: Rand McNally, 1968), an 84-chapter anthology.

11. Alice H. Eagly and Shelly Chaiken, *The Psychology of Attitudes* (Fort Worth, TX: Harcourt Brace Jovanovich, 1993), pp. 499–556.

12. James T. Tedeschi, Barry R. Schlenker and Thomas V. Bonoma, 'Cognitive Dissonance: Private Ratiocination or Public Spectacle?' *American Psychologist* 26 (1971): 691.

13. Gerry Mackie, 'All Men Are Liars: Is Democracy Meaningless?', in *Deliberative Democracy*, edited by Jon Elster (Cambridge: Cambridge University Press, 1998).

14. Edwin Hutchins, *Cognition in the Wild* (Cambridge, MA: MIT Press, 1995), pp. 240–62.

15. This paragraph is based in part on Eagly and Chaiken, *The Psychology of Attitudes*, pp. 580–9, 668–70.

16. Muzafer Sherif, 'A Study of Some Social Factors in Perception', *Archives of Psychology* 22 (1935).

17. Solomon Asch, 'Studies of Independence and Conformity: A Minority of One Against a Unanimous Majority', *Psychological Monographs: General and Applied* 70 (1956); Serge Moscovici, 'Social Influence and Conformity', in *Handbook of Social Psychology*, 3rd edn, edited by Gardner Lindzey and Elliot Aronson (New York: Random House, 1985), p. 363; J. Richard Eiser, *Social Psychology* (Cambridge: Cambridge University Press, 1986), pp. 35–8.

18. Serge Moscovici, *Social Influence and Social Change* (London: Academic Press, 1976).

19. Moscovici, 'Social Influence and Conformity', pp. 386–7; Genevieve Paicheler, *The Psychology of Minority Influence* (Cambridge: Cambridge University Press, 1988), pp. 129–30.

20. Interpretation and replication of this charming result is a matter of controversy. See, for example, Robin Martin, 'Majority Influence and Minority Influence: A Replication with an Unambiguous Blue Slide', *European Journal of Social Psychology* 25 (1995): 373–81. But the basic results of the minority-influence school (that majority impact is greater on public and on direct private change, and that minority impact is greater on indirect private change and that consistency is important to minority impact) are robust, according to Wendy Wood, Sharon Lundgren, Judith A. Ouellette, Shelly Busceme and Tamela Blackstone, 'Minority Influence: A Meta-Analytic Review of Social Influence Processes', *Psychological Bulletin* 115 (1994): 323–45.

21. Moscovici, 'Social Influence and Conformity', p. 396.

22. Gabriel Mugny and Juan A. Perez, *The Social Psychology of Minority Influence* (Cambridge: Cambridge University Press, 1991).

23. William J. McGuire, 'Cognitive Consistency and Attitude Change', *Journal of Abnormal and Social Psychology* 60 (1960): 345–53. McGuire purposely used low academic achieving subjects, among whom he measured a 'Socratic effect', absent persuasion, that is, merely eliciting opinions on logically related issues increased consistency among the same opinions when elicited a week later. Ronald C. Dillehay, Chester A. Insko and M. Brewster Smith, 'Logical Consistency and Attitude Change', *Journal of Personality and Social Psychology* 3 (1966): 646–54 replicated McGuire's 'Cognitive Consistency and Attitude Change' with high academic achieving subjects, and confirmed all but the Socratic effect: the opinions of the more educated subjects were probably more consistent at the outset.

24. William J. McGuire, 'Direct and Indirect Persuasive Effects of Dissonance-Producing Messages', *Journal of Abnormal and Social Psychology* 60 (1960): 354–8. See also William J. McGuire, 'Dynamic Operations of Thought Systems', *American Psychologist* 45 (1990): 504–12.

25. Milton J. Rosenberg and Robert P. Abelson, 'An Analysis of Cognitive Balancing',

in *Attitude Organization and Change*, edited by Milton J. Rosenberg et al. (New Haven, CT: Yale University Press, 1960).

26. Charlan Nemeth, 'Compromising Public Influence for Private Change', in *Perspectives on Minority Influence*, edited by Serge Moscovici, Gabriel Mugny and Eddy Van Avermaet (Cambridge: Cambridge University Press, 1985).

27. Bernard Personnaz and Michel Guillon, 'Conflict and Conversion', in *Perspectives on Minority Influence*, edited by Serge Moscovici, Gabriel Mugny and Eddy Van Avermaet (Cambridge: Cambridge University Press, 1985).

28. Aside from merely advancing its position, minority dissent also demonstrably stimulates the majority to more accurate and creative thinking. See Charlan Nemeth, 'The Value of Minority Dissent', in *Minority Influence*, edited by Serge Moscovici, Angelica Mucchi-Faina and Anne Maass (Chicago, IL: Nelson-Hall, 1994).

29. Jürgen Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (Cambridge, MA: MIT Press, 1996). See Jon Elster, 'Arguing and Bargaining in Two Constituent Assemblies', *University of Pennsylvania Journal of Constitutional Law* 2 (2000): 345.

30. See for commentary Stephen J. Read and Lynn Carol Miller, 'Dissonance and Balance in Belief Systems: The Promise of Parallel Constraint Satisfaction Processes and Connectionist Modeling Approaches', in *Beliefs, Reasoning, and Decision Making: Psycho-Logic in Honor of Bob Abelson*, edited by Roger C. Schank and Ellen Langer (Hillsdale: Lawrence Erlbaum, 1994), pp. 211–3; Elijah Millgram and Paul Thagard, 'Deliberative Coherence', *Synthese* 108 (1996): 63–88; Thomas R. Shultz and Mark R. Lepper, 'Cognitive Dissonance Reduction as Constraint Satisfaction', *Psychological Review* 103 (1996): 219–40.

31. Shultz and Lepper, 'Cognitive Dissonance Reduction as Constraint Satisfaction'; Paul Thagard, *Conceptual Revolutions* (Princeton, NJ: Princeton University Press, 1992); Millgram and Thagard, 'Deliberative Coherence'; Keith J. Holyoak and Paul Thagard, *Mental Leaps: Analogy in Creative Thought* (Cambridge, MA: MIT Press, 1995); Barbara A. Spellman and Keith J. Holyoak, 'If Saddam is Hitler, Then Who is George Bush?' *Journal of Personality and Social Psychology* 62 (1992): 913–33; Barbara A. Spellman, Jodie B. Ullman and Keith J. Holyoak, 'A Coherence Model of Cognitive Consistency: Dynamics of Attitude Change During the Persian Gulf War', *Journal of Social Issues* 49 (1993): 147–65; Stephen J. Read and Lynn Carol Miller (eds), *Connectionist Models of Social Reasoning and Social Behavior* (Mahwah: Lawrence Erlbaum, 1998). For overviews, see Paul Thagard, *Coherence in Thought and Action* (Cambridge, MA: MIT Press, 2000); Read and Miller, 'Dissonance and Balance in Belief Systems: The Promise of Parallel Constraint Satisfaction Processes and Connectionist Modeling Approaches'.

32. 'For classical decision theory, and the microeconomics that is so heavily based on it, preferences are basic and mysterious. In contrast, the theory of deliberative coherence [and its computer implementation] are intended to explain why we have the preferences we do.' See Paul Thagard and Elijah Millgram, 'Inference to the Best Plan: A Coherence Theory of Decision', in *Goal-Driven Learning*, edited by A. Ram and D.B. Leake (Cambridge, MA: MIT Press, 1995), p. 450.

33. This paragraph borrows closely from Read and Miller, 'Dissonance and Balance in Belief Systems: The Promise of Parallel Constraint Satisfaction Processes and Connectionist Modeling Approaches', p. 215. Additionally, see ibid., p. 218:

The following equation is used for updating the activation of the nodes:

$$a_j(t + 1) = a_j(t)(1 - d) + enet_j(max - a_j(t)) + inet_j(a_j(t) - min)$$

where d is a decay parameter, $enet_j$ is the net excitatory input and $inet_j$ is the net inhibitory input. Note that $enet_j$ is equal to $\Sigma w_{ij}a_i(t)$ for $w_{ij} > 0$, where w_{ij} is the weight between nodes i and j and $inet_j$ is equal to $\Sigma w_{ij}a_i(t)$ for $w_{ij} < 0$. Note also that min is the minimum activation value possible -1.0 and max is the maximum activation value $+1.0$. The spread of activation stops when all nodes asymptote.

34. Jon Elster, *Sour Grapes* (Cambridge: Cambridge University Press, 1983).
35. Shultz and Lepper, 'Cognitive Dissonance Reduction as Constraint Satisfaction'. I have renamed the conditions for the sake of clarity.
36. Holyoak and Thagard, *Mental Leaps: Analogy in Creative Thought*, pp. 240–7; Thagard, *Conceptual Revolutions*, pp. 47–50.
37. Paul Thagard and Karsten Verbeurgt, 'Coherence as Constraint Satisfaction', *Cognitive Science* 22 (1998): 1–24.
38. Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago, IL: University of Chicago Press, 1962).
39. Millgram and Thagard, 'Deliberative Coherence' and Thagard and Millgram, 'Inference to the Best Plan: A Coherence Theory of Decision' sketch a coherence model of individual practical reason. Paul Thagard, 'Ethical Coherence', *Philosophical Psychology* 11 (1999): 405–22 has sketched a model of ethical coherence, which includes consideration of deductive, explanatory, deliberative, and analogical coherences. A model of democratic deliberation remains to be developed.
40. Thagard, *Conceptual Revolutions*, p. 59. The second square-bracketed phrase in the quote reverses an apparent transposition in the original.
41. Simone Chambers, *Reasonable Democracy: Jürgen Habermas and the Politics of Discourse* (Ithaca, NY: Cornell University Press, 1996), pp. 169–70.
42. See Brian Leiter, URL: http://leiterreports.typepad.com/blog/2005/06/on_rhetoric_civ.html.
43. John Rawls, *A Theory of Justice*, revised edn (Oxford: Oxford University Press, 1999), pp. 17–19. Neither inference to the best explanation nor reflective equilibrium are merely coherentist: the first is founded on empirical observations and the second on considered convictions of justice.
44. Ibid., p. 108.
45. John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), pp. 54–8.
46. Gerald Gaus, *Political Concepts and Political Theories* (Boulder, CO: Westview Press, 2000), pp. 41–3.
47. In a similar spirit, Ian Shapiro complains that political theorists frequently and wrongly think in terms of fruitlessly debated gross concepts, neglecting both the constraining relational nature of political concepts and their further constraint by empirical considerations. See Ian Shapiro, 'Gross Concepts in Political Argument', in *The Flight from Reality in the Human Sciences* (Princeton, NJ: Princeton University Press, 2005).
48. Carl Schmitt, *The Crisis of Parliamentary Democracy* (Cambridge, MA: MIT Press, 1985), p. 50.

49. Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*.
50. John Dryzek, *The Politics of the Earth: Environmental Discourses* (Oxford: Oxford University Press, 1997).
51. John Dryzek, *Deliberative Democracy and Beyond: Liberals, Critics, Contestations* (Oxford: Oxford University Press, 2000), pp. 124–9.
52. Bernard Manin, 'On Legitimacy and Political Deliberation', *Political Theory* 15 (1987): 338–68. Habermas happened on ideas similar to Manin's: 'Popular Sovereignty as Procedure' is reprinted as Appendix I of Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, pp. 474–7.
53. Michael X. Delli Carpini, Fay Lomax Cook and Lawrence R. Jacobs, 'Public Deliberation, Discursive Participation, and Citizen Engagement: A Review of the Empirical Literature', *Annual Review of Political Science* 7 (2004): 328.
54. Spellman, Ullman and Holyoak, 'A Coherence Model of Cognitive Consistency: Dynamics of Attitude Change During the Persian Gulf War'.
55. Simone Chambers, 'Deliberative Democratic Theory', *Annual Review of Political Science* 6 (2003): 307–26.
56. The claim of futility (along with perversity and jeopardy) is a frequent tactic in political discourse, according to Albert O. Hirschman, *The Rhetoric of Reaction* (Cambridge, MA: Harvard University Press, 1991).