Rational Debate and One-Dimensional Conflict
Author(s): David Spector
Source: *The Quarterly Journal of Economics*, Vol. 115, No. 1 (Feb., 2000), pp. 181-200
Published by: Oxford University Press
Stable URL: http://www.jstor.org/stable/2586938

RATIONAL DEBATE AND ONE-DIMENSIONAL CONFLICT*

DAVID SPECTOR

This paper studies repeated communication regarding a multidimensional collective decision in a large population. When preferences coincide but beliefs about the consequences of the various decisions diverge, it is shown, under some specific assumptions, that public communication causes the disagreement between beliefs either to vanish or to become one-dimensional at the limit. Multidimensional disagreement indeed allows for many directions of communication, including some that are orthogonal to the conflict, along which agents can communicate credibly. The possible convergence toward a one-dimensional conflict where no further communication takes place may be related to the empirically observed geometry of the political conflict in many countries.

I. INTRODUCTION

This paper develops a model of repeated communication between agents facing a collective decision. We consider a situation where agents' interests coincide, while their beliefs about the uncertain payoffs of the collective decision diverge, and we analyze the dynamics of beliefs induced by strategic communication regarding this uncertainty. The main result states that the disagreement cannot remain multidimensional forever, as beliefs all converge either toward the truth, or toward some one-dimensional disagreement. We model communication as a cheap-talk game (see Crawford and Sobel [1982]): in every period a randomly selected agent observes a signal about the state of the world and talks about it publicly, which causes all agents to update their beliefs. Then, another agent is randomly selected to make the collective decision. The information structure allows us to ascribe the possible persistence of some disagreement to a communication failure, as the infinite sequence of signals would be enough, if known to all agents, to let them learn the truth.

The intuition of the main result has a simple geometric expression: a one-dimensional conflict can be a steady state, because there are only two "directions" of communication, so that a "left-wing" agent always wants to report a left-wing signal, and conversely, making communication impossible. The same logic implies that, as long as the disagreement is multidimensional,

* This paper owes much to Thomas Piketty, who took an active part in an earlier version. I am also grateful to Abhijit Banerjee, Michael Kremer, Augustin Landier, and two anonymous referees for their helpful comments. All errors are mine.

The Quarterly Journal of Economics, February 2000

agents manage to communicate credibly along a more "neutral" direction, orthogonal to their conflict, and beliefs keep changing. These two arguments together imply that, starting from a multidimensional disagreement, repeated communication causes beliefs to change and converge toward a steady state, characterized by full agreement or one-dimensional disagreement.

However general the intuition seems, its formal modeling requires very specific assumptions: agents are infinitely impatient, communication is bound to be public, there is a continuum of agents, the payoffs of the collective decision are not observed, and the model only considers a particular utility function and signal structure. Under these assumptions, we show that for a positive measure set of initial beliefs, convergence to a one-dimensional conflict is a positive probability event.

We believe that this result may be of interest for two main reasons. From a theoretical viewpoint, we know relatively little about communication in games with common knowledge of diverging priors, of which this model is a case. More important in our view, our results might be related to an empirical fact generally considered a puzzle: contrary to the predictions of the most natural economic reasoning (see, for example, Arrow [1963]), political conflicts tend be organized along a one-dimensional axis [Poole and Rosenthal 1991; Snyder 1996]. Such considerations of dimensions and axis of conflict may be quite important for a number of economic questions. For example, an important difference between the political economy of redistribution in the United States and in Continental Europe seems to be that the axis of conflict differs: the debate focuses mainly on the level of fiscal transfers in the United States, while in Europe the level of the minimum wage, and in general the amount of direct price and quantity controls, are more central.

We do not claim that our model should be viewed as an explanation, given the difficulty of establishing a link between the real world and our highly abstract model. However, we believe that taking seriously the idea, explored in a growing literature (for example, Banerjee and Somanathan [1998] and Piketty [1995]), that political conflicts are as much about beliefs as about interests, may help to address this puzzle. Stating that "it takes a Nixon to go to China" has become a common way of saying that the credibility of information transmission regarding collective choices depends on people's perceptions of politicians' prior preferences. A recent literature (for example, Cukierman and Tommasi [1998]) has investigated this idea formally, and our model explores it further, with a focus on the long run and on the issue of conflict dimensionality.

The paper is organized as follows: after an informal overview of the main results in Section II; Section III presents the model and discusses its assumptions; Section IV analyzes its steady states; Section V describes its convergence and stability properties; and Section VI concludes.

II. Overview of the Main Results

II.1. The Setting

We summarize here the main ingredients of the model, displayed in full in the next section. In order to focus on belief heterogeneity in collective choice problems, we consider a large population (a continuum of agents) characterized by identical preferences about some collective decision but different initial beliefs about some state of the world relevant for the collective decision. We investigate the transmission of new information about this state of the world, and how it affects the distribution of beliefs in the long run. We assume the population to be initially partitioned into two groups, labeled 1 and 2, differing only with respect to their initial beliefs. We study the following game, repeated in every period:

(i) A random signal is drawn, providing some imprecise information about the true state of the world.
(ii) A randomly selected agent (the speaker) observes the signal.
(iii) The agent who observed the signal sends a message to the whole population, which causes all agents to update their beliefs.
(iv) A randomly selected agent (the receiver), belonging to the group other than the speaker's, takes the collective decision.

This setting has the following implications:

- Almost all beliefs are modified only by the messages sent by the finite number of agents who directly observed a signal. Therefore, all the learning that takes place is through communication about signals.
- Infinitely many imprecise signals add up to very precise information, so the conflict would vanish in the long run if

signals were public. Therefore, any failure to have everyone learn the true state of the world in the long run should be ascribed to the fact that, for strategic reasons, signals are not fully reported.

The core of the paper is the analysis of step (iii), which we model as a cheap-talk game: agents rationally update their beliefs given the speaker's strategy, which is optimal given updating rules. Step (iv) implies that the speaker's goal is to the other group's belief as close as possible to his new belief. Initial belief heterogeneity limits communication; on the other hand, the information conveyed by the signal is relevant for all agents, which should allow for some communication. Our results are driven by the interplay of these two opposing forces.

II.2. The Results

We are interested in the sequence of both groups' beliefs $(\mu^{1,t},\mu^{2,t})$. We show that it converges almost surely toward a steady state, that is, a pair of beliefs such that further communication is impossible. Therefore, the main results deal with the characterization of steady states (Section IV). They are expressed in terms of the dimensionality of the conflict between groups.

Beyond this specific model, the meaning of the distinction between one-dimensional and multidimensional conflict is quite general: a conflict is multidimensional if it is possible for group A's beliefs to become closer to group B's on some issues, and farther from them on some other issues. If, on the contrary, people's beliefs reflect uncertainty between a "conservative" view of the world spanning all relevant issues and an opposite "progressive" view, but rule out beliefs that are conservative on some issues and progressive on some other, then the conflict is one-dimensional.

The main result states that a steady state is characterized either by no conflict at all (both groups agree on the truth) or by a one-dimensional conflict. The argument breaks down into two parts:

- There exist steady states characterized by a one-dimensional conflict of beliefs (Proposition 1).
- As long as the disagreement is multidimensional, further communication is possible (Proposition 2).

The first result (illustrated in Figure I) relies on the idea that if the conflict is already polarized along some "left-right" axis, then

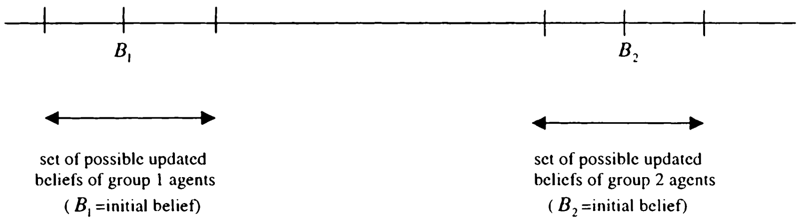

FIGURE I
One-Dimensional Conflicts

information transmission may be impossible. If both groups' beliefs are far away from each other and sufficiently extreme, they are little affected by new information, so that a left-wing speaker always wants to move a "right-wing" agent to the left even after observing a right-wing signal (and conversely), and he cannot be credible.

The second result—the most important of the paper—relies on a geometric argument illustrated in Figure II (in the sketch of the proof of Proposition 2 below): if there exist changes in group A's beliefs bringing them neither closer to nor farther from group B's beliefs (geometrically, the change of beliefs is orthogonal to the conflict between groups), then an agent in group B is a priori indifferent about such belief changes. Therefore, he will want to induce a change of group A's belief along a neutral direction only if he really observed a signal along that direction. This means he can credibly communicate along this direction. Intuitively, this neutral direction means that it is possible to convey information by expressing a view that is left-wing on some issues and right-wing on some other.

Therefore, repeated communication reduces the dimensionality of the disagreement until it becomes at most one-dimensional. Section V describes the dynamics in further detail, and we show in particular that convergence to a one-dimensional conflict (rather than no conflict at all) occurs with a positive probability for a positive-measure set of initial belief distributions.

III. THE MODEL

III.1. Players and Preferences

We consider a continuum of agents partitioned into two groups (1 and 2), and engaged into repeated communication over a discrete infinite horizon $t = 0,1,\ldots$ In every period t a collective decision $M_t = (x_t, y_t) \in \Re^2$ is to be taken. Agents all have the same state-contingent utility function $U(M,s)$, depending on the decision M and some uncertain state of the world s. U is given by

$$U(M,s) = -\|M - M_s\|^2,$$

where $M_s = (x_s, y_s)$ is the optimal decision when the state of the world is s. The state of the world is unknown to the agents, and remains the same in all periods. Agents have an infinite rate of time preference: in period t they maximize the expectation of $U(M_t)$.

Along each of the two dimensions, there are two possible ideal decisions: x_s and y_s can each be equal to either 0 or 1. There are therefore four possible optimal decisions (0,0), (0,1), (1,0) and (1,1). Since the state of the world is relevant only through the optimal decision, we will identify them and write (i,j) for the state of the world where the optimal decision is (i,j).

A belief μ about the state of the world is an element $(\mu_{00}, \mu_{10}, \mu_{01}, \mu_{11})$ of the three-dimensional simplex Δ.

We assume that all agents belonging to group i have identical beliefs in period zero, denoted $\mu^{i,0}$. The difference between beliefs across groups is the only heterogeneity between agents.

We first characterize the indirect preferences induced by some belief μ. The quadratic state-contingent utility function yields the following simple result: an agent's indirect preferences given a belief μ are represented by the indirect utility function,

$$V(M) = -\|M - M*(\mu)\|^2,$$

with $M*(\mu) = E(M_s|\mu) = (\mu_{10} + \mu_{11}, \mu_{01} + \mu_{11})$.

Remark. A belief μ can be any element of the three-dimensional simplex Δ, but as far as the induced preferences are concerned, only $M*(\mu)$, which varies in a two-dimensional set, matters. The information that is conveyed by the belief μ in addition to that conveyed by $M*(\mu)$ is related to the correlation between the agent's beliefs about x_s and about y_s: infinitely many beliefs μ correspond to the same $M*(\mu)$. Although this additional information is irrelevant for the induced preferences given quadratic preferences, it is relevant for the analysis of the communication game, since it affects the way agents update new information.

Notation. Throughout the paper we are going to write Q for the square $[0,1] \times [0,1]$. Clearly, the most preferred decision induced by any belief belongs to Q.

III.2. The Communication Game

The following sequence of events is infinitely repeated:

(i) A signal is drawn randomly according to a probability distribution depending on the true state of the world.

(ii) A randomly selected agent observes the signal, and everyone learns "what the signal is about" (see subsection III.3 below).

(iii) The agent who observed the signal speaks.

(iv) A randomly selected agent belonging to the group other than the speaker's takes the collective decision.

To keep the repeated game simple enough, we assume that in every odd (respectively, even) period, prior to decision-making, an agent (the speaker) is randomly chosen in group 1 (respectively, group 2) according to the uniform distribution. The randomization is independent across periods. The speaker then observes a signal σ_t providing some information about the state of the world, updates his beliefs, and can attempt to communicate to other agents by sending a message. An agent in the other group (the receiver) is then randomly chosen to be a dictator in period t and take the collective decision. We assume that communication is public: when an agent talks, he has to talk to everybody, and not only to the agents of his own group. We are going to write $s(t)$ and $r(t)$, respectively, for the speaker's and the receiver's group in period t, so that $(s(t),r(t))$ is $(1,2)$ if t is odd, and $(2,1)$ otherwise.

The assumptions of a continuum of agents and of public communication imply that, except for the finite number of agents who directly observed a signal, all the others received the same information, coming exclusively from publicly sent messages. Therefore, in period t almost all agents in group i $(i = 1,2)$ have the same belief $\mu^{i,t}$. With probability 1, neither the agent who observes a signal in period t nor the dictator directly observed a signal in an earlier period. This implies that the receiver's decision following a message m_t sent by the speaker is $M * (B(\mu^{r(t),t},m_t))$, where $B(\mu,\alpha)$ denotes the belief held after updating the information α starting from the belief μ. Given the form of indirect preferences and the fact that the speaker's belief when he speaks is $B(\mu^{s(t),t},\sigma_t)$, the speaker chooses m_t to minimize $\|M * (B(\mu^{s(t),t},\sigma_t)) - M * (B(\mu^{r(t),t},m_t))\|$: he sends the message that minimizes the distance between his new most preferred decision and the future dictator's.

We assume that although agents act, whenever they can, in order to maximize the expected utility derived from the collective decision, they do not observe their own utility. Therefore, agents learn nothing on their own, and their beliefs are affected only by the messages they hear. This strong assumption is needed if we are to focus on limits on learning imposed by communication. If agents could learn from their utility level, then they would learn the true state of the world at once.[1]

III.3. The Signal Structure

To make the problem interesting, signals must be neither too uninformative, nor too informative. On the one hand, we are interested in the extent to which strategic communication may leave room for some disagreement. Therefore, a desirable feature of the model should be that making the infinite sequence of signals public would cause all agents to learn the true state of the world: this will imply that any failure to learn the truth results from a communication failure. On the other hand, to make strategic communication nontrivial, signals should not be completely informative: if they were, then given identical preferences, the agent who observed the signal would simply report it, and would be believed.

We make some very specific assumptions about the signal structure. It is assumed to allow not only for pure signals (providing information about x_s only or y_s only) but also for "mixed" signals providing information simultaneously along both dimensions, with various "ratios of informativeness" along both directions. We assume that this relative informativeness (the direction of the signal) along the two dimensions is drawn at random and becomes known to all agents, but that this piece of information alone tells nothing about the state of the world. More precisely, the signal can be decomposed in two steps.

First, a direction $\theta \in [0,\pi)$ and a "signal intensity" $a \in [0,A]$ (with $A < 1/2\sqrt{2}$) are drawn at random, from a probability distribution independent of the state of the world, with strictly positive density everywhere. θ and a become common knowledge, but they do not provide any information about the state of the world.

1. A similar assumption is made by Hart [1985].

Then, the agent randomly selected to be the speaker observes a signal that can take two values $S_{a,\theta}$ or $S_{a,\theta+\pi}$. We write $f_{ij}(S_{a,\theta})$ for the probability of observing $S_{a,\theta}$ if the state of the world is (i,j).

We assume that the functions f_{ij} are given by

$$f_{ij}(S_{a,\theta}) = \tfrac{1}{2} + a(\beta_i \cos\theta + \beta_j \sin\theta), \text{ with } \beta_0 = -1, \beta_1 = 1.$$

This implies that $f_{ij}(S_{a,\theta}) + f_{ij}(S_{a,\theta+\pi}) = 1$. The assumption $A < 1/2\sqrt{2}$ implies that $f_{ij}(S_{a,\theta}) > 0$ for all i, j, a, and θ. This functional form has the following simple property: if μ is the belief assigning equal weights to all four states of the world (so that $M * (\mu) = (\tfrac{1}{2},\tfrac{1}{2})$), then $M * (B(\mu,S_{a,\theta})) = M * (\mu) + a(\cos\theta,\sin\theta)$, so that θ is the direction of the change of the agent's most preferred decision, and a is the magnitude of this change.

This example allows us to make the assumption of common knowledge of the signal direction more precise: the direction θ (mod π) is some commonly known information about "what the signal is about." For example, $\theta = 0$ mod π (respectively, $\pi/2$ mod π) means that the signal provides information only about x_s and not at all about y_s (respectively, the opposite). All other directions are mixed, and are more and more informative about y_s (relative to x_s) the closer θ is to the vertical.

The signal structure assumed above is one of many that would yield the same results: what is needed is only that informativeness is bounded and that in infinitely many periods, each direction θ occurs and is common knowledge with a strictly positive probability.

III.4. Role of the Various Assumptions

Continuum of agents. This assumption is necessary: with finitely many agents, each of them would observe infinitely many signals directly with probability one, and would therefore learn the truth even without any communication.

Public communication. If communication were not bound to be public, then the speaker would communicate the signal at least to his own group, since there is no prior divergence within a group. Therefore, each group would learn infinitely many signals, and converge to the truth with probability one. The necessity of this assumption, as well as that of the continuum assumption, highlights the fact that this paper may be more relevant to think about the political debate (involving many agents and characterized by a "public" character) than about other communication situations.

Infinite impatience. This assumption is made for simplicity. Given the complexity of modeling forward-looking behavior in this stochastic context, we do not really know to what extent it is necessary for the results. At the very least, the results should carry over, by continuity, to the case of very impatient agents.

Knowledge of the signal direction. This assumption is made for simplicity. Together with the fact that there are only two signals along each direction (for a given intensity), it implies that the pure equilibria of the communication games are either completely uninformative or completely informative, which facilitates the analysis.

IV. Steady States

As we stressed in subsection III.2 above, the existence of a continuum of agents implies that in the beginning of any period t, almost all agents in group i have the same beliefs $\mu^{i,t}$, resulting from updating all the messages sent between periods 0 and t starting from $\mu^{i,0}$. This, and the assumption that the random determination of speakers and receivers is independent across periods, ensures that in the beginning of period $t + 1$, the speaker's and the receiver's respective beliefs are $\mu^{s(t),t}$ and $\mu^{r(t),t}$ with probability one. This allows us to keep track of the entire dynamics by using $(\mu^{1,t},\mu^{2,t})$ as an exhaustive state variable.

As always in pure communication games, the question of equilibrium selection arises, since there always exists, alongside any other equilibrium, a "babbling" equilibrium where no communication takes place. However, since there are only two signals (once the uncertainty about a and θ has been resolved), there exists for each belief distribution and signal direction either the babbling equilibrium only, or the babbling equilibrium and the "communicative" equilibrium where the speaker reveals the signal.[2] Whenever such a communicative equilibrium exists, we are going to select it. This defines a stochastic process $(\mu^{1,t},\mu^{2,t})_{t\geq 0}$ starting from some arbitrary initial condition $(\mu^{1,0},\mu^{2,0})$. We first characterize its steady states, which will allow us to describe its dynamics in Section V.

A steady state is a pair of beliefs (μ^1,μ^2) such that no further

2. We omit mixed equilibrium, since one can easily show that for any belief distribution, they occur for a zero-measure set of signal directions and intensities.

communication is possible, i.e., such that if $(\mu^{1,t},\mu^{2,t}) = (\mu^1,\mu^2)$ then $(\mu^{1,t+1},\mu^{2,t+1}) = (\mu^1,\mu^2)$ with probability 1.

Notation. For i,j in $\{0,1\}^2$ let δ_{ij} denote the vertex of Δ corresponding to the probability distribution assigning probability 1 to the event $\{s = (i,j)\}$, and 0 to the other three events. We are going to call the segments $[\delta_{ij},\delta_{i'j'}]$ (where $(i,i') \neq (j,j')$) the edges of Δ. They correspond to beliefs assigning a positive probability only to the two states of the world (i,j) and (i',j'). There are six such edges, since there are four vertices.

We first show why there are many steady states characterized by a one-dimensional conflict of beliefs, that is, such that all beliefs belong to the same edge of Δ. Proposition 1 states that there exist steady-state belief distributions with support in the interior of any given edge.

PROPOSITION 1. Consider an edge L of Δ. There exists a set of steady states $(\mu^1,\mu^2) \in L^2$ which has a nonempty interior in L^2, and therefore, a strictly positive measure.

The proof is in the Appendix. Proposition 1 is very intuitive: if all beliefs are on the same edge of Δ, the disagreement is one-dimensional.[3] Therefore, any new information (signals or messages) changes beliefs in a way that makes new most preferred policies still belong to the same one-dimensional set. If beliefs are initially far apart, then, since the informativeness of signals is limited, extreme beliefs remain extreme whatever the signals (and, therefore, the messages) they observe: posterior beliefs are determined mostly by prior beliefs rather than by new information. This implies that if the speaker is a left-winger and the receiver a right-winger, the speaker would always prefer the receiver to believe he observed a left-wing signal, irrespective of the signal he truly observed. This means he cannot be credible, and agents are stuck in this polarized belief distribution where no further communication can take place.[4] The distance between

3. These steady-state beliefs distributions may or may not display correlation between x and y: if the edge containing both beliefs is $[\delta_{0i},\delta_{1i}]$ (respectively, $[\delta_{i0},\delta_{i1}]$) for $i = 0$ or 1, then both groups agree that $y_s = i$ (respectively, $x_s = i$), so that the disagreement is only about the dimension x (respectively, y) and beliefs are not correlated. But if the edge is $[\delta_{00},\delta_{11}]$ (respectively, $[\delta_{01},\delta_{10}]$), there is a positive (respectively, negative) correlation between x_s and y_s.

4. Notice that the truth is even more "extreme" than any belief in a one-dimensional steady state. This is an unfortunate consequence of the modeling choice characterized by only two states of the world along each dimension. In the

priors in our model is the equivalent of a distance between preferences, and Proposition 1 is the equivalent, in our context, of results stating that the greater the distance of preferences, the less communication takes place.

What are the other steady states? Proposition 2 provides a partial answer by showing that a belief distribution such that no belief belongs to an edge of Δ is not a steady state. It is the most important result of the paper, and it relies on the orthogonal argument mentioned in the introduction: as long as agents assign a positive probability to at least three states of the world, there exists a signal direction moving the receiver's induced most preferred decision "orthogonally" to the disagreement, and information can be credibly transmitted along this direction. Given the infinite horizon, communication will take place in the future with probability one, and such a belief distribution is not a steady state. In other words, Proposition 2 establishes a partial converse of the result in Proposition 1: the steady states of Proposition 1 are almost the only ones (there also exist steady states such that all beliefs are located on different edges of Δ, but we do not mention them because, as will appear in Section V below, the probability of converging toward them is zero if the initial beliefs are not degenerate).

PROPOSITION 2. Assume (μ^1,μ^2) is such that μ^1 and μ^2 assign a strictly positive probability to more than three states of the world (that is, none of these beliefs belongs to an edge of Δ). Then (μ^1,μ^2) is not a steady state.

Sketch of the proof of Proposition 2.[5] The argument summarized here is illustrated in Figure II. Let us write M_i for $M*(\mu^i)$ (the most preferred decision of group i agents before any communication takes place). One can show that, writing M_i^+ for $M*(B(\mu^i,S_{a,\theta}))$ and M_i^- for $M*(B(\mu^i,S_{a,\theta+\pi}))$, there exists a signal direction θ such that

- $M_2^- M_2^+$ (for example) is orthogonal to $M_1 M_2$. (If a belief is not one-dimensional, then there exists a signal direction moving the corresponding most preferred decision along any given direction, here the one orthogonal to $M_1 M_2$.)

real world we should think of a continuum of possible states of the world, so that some beliefs may be, on average, to the left of the truth, while some others are to the right of the truth.

5. The full proof is available from the author upon request.

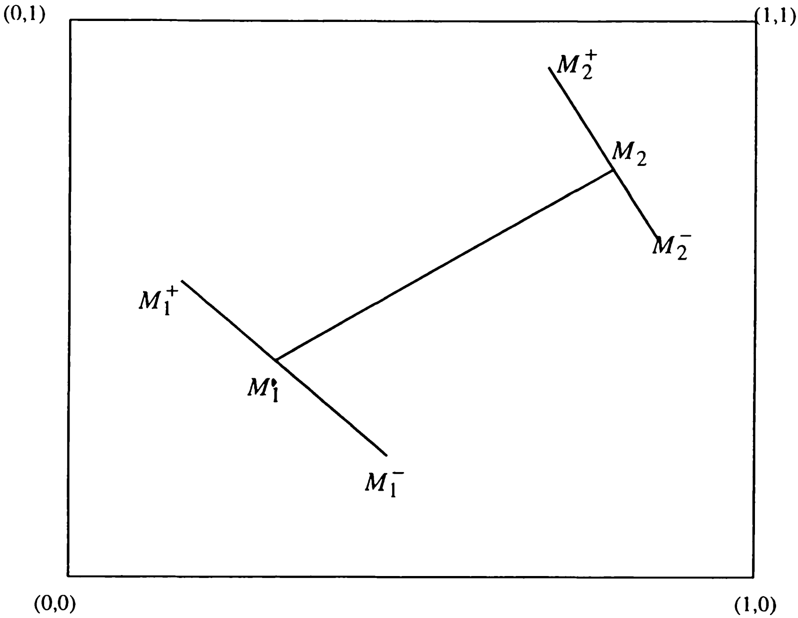

FIGURE II

- M_1^+ and M_2^+ are located on the same side of the M_1M_2. (This amounts to saying that although the two beliefs do not move along the same direction, the directions along which they move form an acute angle.[6])
- The distances $M_iM_i^+$ and $M_iM_i^-$ are roughly equal if a is small enough (if the signal is little informative, then M_i^+ and M_i^- are approximately equally likely).

This implies that $M_1^+M_2^+ < M_1^+M_2^-$ and $M_1^-M_2^- < M_1^-M_2^+$. Therefore, a group 1 agent wants to truthfully report $S_{a,\theta}$ if he observed $S_{a,\theta}$ and expects to be believed: his most preferred decision is M_1^+ and the receiver in group 2 is going to choose M_2^+ if he reports $S_{a,\theta}$, M_2^- otherwise. The form of the indirect utility function implies that the sender wants to minimize the distance between his most preferred decision and the receiver's. The

6. This is a difficult point: it is not true that in general the directions of two belief changes induced by the same signal form an acute angle. The result we prove is that for any given pair of beliefs and direction θ, there exists a signal moving one of the most preferred decisions along the direction θ, and the other one along a direction forming an acute angle with θ.

inequality $M_1^+ M_2^+ < M_1^+ M_2^-$ implies that this is done by reporting $S_{a,\theta}$. Similarly, the second inequality implies that a group 1 agent wants to truthfully report $S_{a,\theta+\pi}$ if he observed $S_{a,\theta+\pi}$.

Therefore, there exists a communicative equilibrium when the signal direction is θ, the intensity is small, and the sender is in group 1. By continuity there exist many in a neighborhood around θ, so they occur with a strictly positive probability. This means that with probability one, some communication will take place in some future period, and beliefs are going to change. Therefore, (μ^1,μ^2) is not a steady state.

V. Dynamics, Stability, and Convergence

For any initial condition $(\mu^{1,0},\mu^{2,0})$, the stochastic process $(\mu^{1,t},\mu^{2,t})$ is a bounded martingale, just as any Bayesian learning process. That is, conditional on a member of group i's belief $\mu^{i,t}$ at time t, $E(\mu^{i,t+1}|\mu^{i,t}) = \mu^{i,t}$. Therefore, the martingale convergence theorem applies (see Neveu [1975]), and as t goes to infinity, $(\mu^{1,t}\mu^{2,t})$ converges toward some steady state (μ^1,μ^2) with probability 1.[7] Having characterized the set of steady states in Section IV, we examine in this section how the probability distribution of limit beliefs on that set varies as a function of initial beliefs. The description of the belief dynamics starts with the following restriction about possible limits.

PROPOSITION 3. Assume that $\mu^{1,0}$ and $\mu^{2,0}$ have full support. Then the limit belief distribution (μ^1,μ^2) is such that

(i) either μ^1 and μ^2 assign probability 1 to the true state of the world s, i.e., $\mu^1 = \mu^2 = \delta_s$;
(ii) or μ^1 and μ^2 both belong to some edge $[\delta_s,\delta_w]$ of Δ containing the true belief.

Proof. See the Appendix.

The second case corresponds to convergence toward the steady states described in Proposition 1, characterized by a one-dimensional conflict. Although these are not the only steady states, Proposition 3 rules out convergence to any other (for example, belief distributions such that beliefs belong to different

7. Strictly speaking, the martingale convergence theorem only implies that the system will converge somewhere with probability one. The limit has to be a steady state (in the sense defined in Section III above) only if the transition correspondence has adequate continuity properties, which is the case if we select the most informative equilibrium at each stage (selecting the truth-telling equilibrium whenever it exists defines a continuous transition correspondence).

edges). It remains to prove that one-dimensional conflict indeed happens at the limit with a strictly positive probability.

Although we do not fully describe the dynamics, we focus below on the following two cases: the one where the initial belief distribution is close to a one-dimensional steady state (Proposition 4), and the one where initial beliefs are close to each other (Proposition 5).

Proposition 4 shows that if the initial belief distribution is close to a one-dimensional steady state, then there is positive probability of converging toward a nearby one, and that this probability converges to one as the initial belief distribution converges to such a steady state. This result is important: it means that the steady states of Proposition 1 are locally stable (as long as they assign a positive probability to the truth): if such a steady state is locally perturbed, the system converges back to a similar one with positive probability. This is a weak notion of stability, since we do not prove that if a steady state is locally perturbed, the system goes back to it with probability one—only that with a large probability it does not go very far. This proposition implies that the steady states of Proposition 1 are meaningful: convergence toward a one-dimensional conflict occurs with a positive probability for a positive-measure set of initial beliefs.

PROPOSITION 4. Consider an edge L of Δ containing the true belief δ_s, a steady state (μ^{*1},μ^{*2}) in the interior of L^2, and a neighborhood W of (μ^{*1},μ^{*2}) in L^2. There exists a neighborhood V of (μ^{*1},μ^{*2}) in Δ^2 such that if $(\mu^{1,0},\mu^{2,0}) \in V$ then with a strictly positive probability $(\mu^{1,t},\mu^{2,t})_{t\geq 0}$ converges toward a limit belonging to W. This probability converges to 1 as $(\mu^{1,0},\mu^{2,0})$ converge toward (μ^{*1},μ^{*2}).

The main idea of the proof (available from the author) is as follows: assume that steady-state beliefs assign a zero probability to the event "$y_s = 1$," so that they are located on the horizontal axis defined by the belief "$y_s = 0$," assumed to be true. Since (μ^{*1},μ^{*2}) is a steady state, "horizontal" communication is impossible. By continuity, communication at nearby belief distributions is almost only about the "vertical" variable y_s, so that beliefs move orthogonally to the horizontal axis. An agent's expectation of his own belief change is of course zero. But since the truth is that "$y_s = 0$," agents' expectations of their vertical movement, based on their own beliefs (assigning a positive probability to the event "$y_s = 1$")

are biased upward: the true vertical movement is downward on average, meaning that agents tend to move back toward the horizontal axis "$y_s = 0$."

The next result qualifies Proposition 4: if initial beliefs are close enough to each other, then convergence to the truth—and total agreement—occurs with probability one.

PROPOSITION 5. If the ratios $\mu_p^{1,0}\mu_q^{2,0}/\mu_p^{2,0}\mu_q^{1,0}$ (for all pairs (p,q) of states of the world) are close enough to 1, then all beliefs converge toward the truth with probability 1.

Proof of Proposition 5. See the Appendix.

This result is stronger than a continuity property around identical initial beliefs: one could have expected the probability of convergence to the truth to converge to one, instead of being equal to one around a given initial distribution with identical beliefs. The proof relies on the following idea: since in period t almost all agents in both groups observe the same information (the sequence of messages up to period t), Bayes' rule implies that $\mu_p^{1,t}\mu_q^{2,t}/\mu_p^{2,t}\mu_q^{1,t}$ is independent of t, so that if it is close to 1 initially, it must also be close to 1 at any steady state, meaning that the limit beliefs are close to each other. However, by an argument close to the proof of Proposition 1, one-dimensional conflict is impossible when beliefs are close. So the limit beliefs must be the same, and hence both be true.

To summarize, we have shown in this section that convergence toward one-dimensional conflict occurs with a positive probability for a positive measure-set (though not the entire set) of initial belief distributions.

VI. Concluding Remarks

One may wonder how sensitive the results are to the various assumptions. As explained in subsection III.4, they would fail for obvious reasons if there were a finite number of agents, or if an agent could speak to his own group only. We have a less clear idea about the possibility of relaxing the assumption of infinite impatience, as the forward-looking behavior this would induce makes the analysis of the communication game difficult (at the very least, a continuity argument should allow our results to carry over to the case of very impatient agents, with the set of possible steady states converging toward the one found here as the rate of time preference goes to infinity). Similarly, we do not know to which class of utility functions our results can be extended. The reason for this uncertainty is the very counterintuitive nature of the geometry of Bayesian updating in several dimensions, exemplified by the fact that the changes of the most preferred decisions of two agents having observed the same signal can go in very different directions (Section IV). These problems may weaken the orthogonality argument. Extending the results to more than two groups would be particularly welcome: it would make the link with the evidence from political science more meaningful. It should be possible, at least, to extend our results, by continuity, to the case where the initial belief distribution is close to two main groups—for example, if all groups except two are small enough, or if all initial beliefs can be partitioned into two sets of neighboring beliefs.

Generalizing the model to more than two dimensions seems relatively straightforward: after extending the signal structure in a natural way, it should be sufficient to consider a plane containing both beliefs in the corresponding simplex, and to apply Proposition 2 to that plane by restricting the attention to signals leaving beliefs in that plane. This would show that steady-state beliefs must belong to a common edge of the simplex.

Further developments in Bayesian theory might allow us to assess how general our argument is, and how relevant it may be in terms of political economy. This should be the subject of future research.

Appendix

Proof of Proposition 1. If $L = [\delta_{ij},\delta_{i',j'}]$ (with $(i,i') \neq (j,j')$), then each belief belonging to L can be summarized by a single number p in $[0,1]$ denoting the probability assigned to the state of the world (i,j), the probability assigned to (i',j') being then equal to $1 - p$.

We consider a belief distribution such that the respective beliefs of groups 1 and 2 are p and p'. We assume that there exists a communicative equilibrium when the signal direction is θ, the intensity is a, and the speaker belongs to group 1. Given the signal structure, for all signal directions θ'

$$\frac{1}{\Gamma} < \frac{f_{ij}(S_{a,\theta'})}{f_{i'j'}(S_{a,\theta'})} < \Gamma$$

with $\Gamma = (1 + 2A\sqrt{2})/(1 - 2A\sqrt{2})$. Therefore, if in equilibrium the speaker reports the signal, the posterior beliefs of members of groups 1 and 2 belong, respectively, to the intervals,

$$\left[\frac{p}{p + (1-p)\Gamma}, \frac{p\Gamma}{p\Gamma + (1-p)}\right]$$

$$\text{and} \quad \left[\frac{p'}{p' + (1-p')\Gamma}, \frac{p'\Gamma}{p'\Gamma + 1 - p'}\right].$$

If $p\Gamma/(p\Gamma + 1 - p) < p'/(p' + (1 - p')\Gamma)$, then whatever the signal was, the speaker's posterior belief assigns a lower probability to (i,j) than the receiver would after observing any signal. Therefore, if the receiver expects the speaker to report the signal truthfully, then the best response for the speaker is to report the signal increasing the probability assigned by the receiver to (i,j), that is, $S_{a,\theta}$ if $f_{ij}(S_{a,\theta})/f_{i'j'}(S_{a,\theta}) > f_{ij}(S_{a,\theta+\pi})/f_{i'j'}(S_{a,\theta+\pi})$, $S_{a,\theta+\pi}$ otherwise.

Since this best response is independent of which signal the speaker observed, truthful reporting of the signal is not an equilibrium, and there is no communicative equilibrium. By the same argument, there exists no communicative equilibrium with the speaker in group 2 if $p\Gamma/(p\Gamma + 1 - p) < p'/(p' + (1 - p')\Gamma)$. This inequality is satisfied by a positive measure subset of L^2, which yields the result.

QED

Proof of Proposition 3. A general result about Bayesian learning is that if a belief assigns initially a strictly positive probability to the true state of the world, then with probability one its limit does as well (see Aghion et al. [1991]). If case (i) does not hold, there exists a state of the world $w \neq s$ such that $\mu_w^1 > 0$. Since almost all agents in both groups (all except those who directly observed a signal) received the same information between periods 0 and ∞, Bayes' rule implies that

$$\frac{\mu_w^1 \mu_s^{1,0}}{\mu_w^{1,0} \mu_s^1} = \frac{\mu_w^2 \mu_s^{2,0}}{\mu_w^{2,0} \mu_s^2} = \lim_{t \to \infty} \frac{\Pr(\text{messages up to } t|w)}{\Pr(\text{messages up to } t|s)}.$$

Therefore $\mu_w^2 > 0$. We know from Proposition 2 that neither μ^1 nor μ^2 can be in the interior of Δ, for (μ^1, μ^2) would not be a steady state then. Therefore, both μ^1 and μ^2 belong to $[\delta_s, \delta_w]$.

Proof of Proposition 5. We consider a pair of initial beliefs $(\mu^{1,0},\mu^{2,0})$ such that for all pairs of states of the world (p,q), $|\mu_p^{1,0}\mu_q^{2,0}/\mu_p^{2,0}\mu_q^{1,0} - 1| < \epsilon$. We want to show that if ϵ is small enough, then with probability 1, $\lim_{t\infty} (\mu^{1,t},\mu^{2,t}) = (\delta_s,\delta_s)$.

Let us assume that with a positive probability, $\lim_{t\infty} (\mu^{1,t},\mu^{2,t}) = (\mu^{1\infty},\mu^{2\infty}) \neq (\delta_s,\delta_s)$. By Proposition 2 we know that $\mu^{1\infty}$ and $\mu^{2\infty}$ belong to the same edge of Δ. Let us assume, for example, that they both assign weight only to the states of the world (0,0) and (1,0). Let us define for $i = 1,2$, $p_i = \mu_{10}^{i\infty}$, and let us assume that $p_1 < p_2$ (group 1 is "to the left" of group 2).

Step 1. Assume that the signal direction is horizontal ($\theta = 0$) and the signal intensity is maximal ($a = A$). The signal structure implies that for $i = 1,2$,

$$B(\mu^{i\infty},S_{A,0})(1,0) = \frac{Cp_i}{Cp_i + 1 - p_i}$$

$$\text{and} \qquad B(\mu^{i\infty},S_{A,\pi}(1,0) = \frac{p_i}{p_i + C(1 - p_i)},$$

with $C = (1 + 2A)/(1 - 2A) > 1$. Since $(\mu^{1\infty},\mu^{2\infty})$ is a steady state, it must be the case that whenever the speaker is in group 1, he wants to report the "left-wing signal" $S_{A,\pi}$ even after observing the "right-wing signal" $S_{A,0}$, which is true if

$$\left|\frac{p_1C}{p_1C + 1 - p_1} - \frac{p_2}{p_2 + C(1 - p_2)}\right| < \left|\frac{p_1C}{p_1C + 1 - p_1} - \frac{p_2C}{p_2C + 1 - p_2}\right|$$

or

$$|(p_1 - p_2)\,C| > \frac{p_2C + 1 - p_2}{p_2 + C(1 - p_2)}\,|p_1(1 - p_2)\,C^2 - p_2\,(1 - p_1)|.$$

$C > 1$ and $p_1 < p_2$ imply then the inequality

$$C^2\left|1 - \frac{p_2\,(1 - p_1)}{p_1\,(1 - p_2)}\right| > \left|C^2 - \frac{p_2\,(1 - p_2)}{p_1(1 - p_2)}\right|.$$

Step 2. For every period t, Bayes' rule implies that

$$\frac{\mu_{10}^{1,t}\mu_{00}^{1,0}}{\mu_{00}^{1,t}\mu_{10}^{1,0}} = \frac{\mu_{10}^{2,t}\mu_{00}^{2,0}}{\mu_{00}^{1,t}\mu_{10}^{1,0}} = \frac{\Pr(\text{messages up to } t|(1,0))}{\Pr(\text{messages up to } t|(0,0))},$$

so that

$$\frac{\mu_{10}^{1,t}\mu_{00}^{2,t}}{\mu_{00}^{1,t}\mu_{10}^{2,t}} = \frac{\mu_{10}^{1,0}\mu_{00}^{2,0}}{\mu_{00}^{1,0}\mu_{10}^{2,0}}.$$

Taking the limit of this identity as t tends to infinity yields $p_2(1-p_1)/p_1(1-p_2) = \mu_{10}^{20}\mu_{00}^{10}/\mu_{00}^{20}\mu_{10}^{10}$.

Step 3. Steps 1 and 2 imply that if for all pairs (p,q) of states of the world,

$$\left|\frac{\mu_p^{10}\mu_q^{20}}{\mu_p^{20}\mu_q^{10}} - 1\right| < \epsilon, \quad \text{then} \quad C^2\left|1 - \frac{p_2(1-p_1)}{p_1(1-p_2)}\right|$$

$$> \left|C^2 - \frac{p_2(1-p_1)}{p_1(1-p_2)}\right| \quad \text{and} \quad \left|1 - \frac{p_2(1-p_1)}{p_1(1-p_2)}\right| < \epsilon$$

so that

$$\epsilon C^2 > C^2\left|1 - \frac{p_2(1-p_1)}{p_1(1-p_2)}\right| > \left|C^2 - \frac{p_2(1-p_1)}{p_1(1-p_2)}\right| > C^2 - 1 - \epsilon,$$

which cannot be true if ϵ is close enough to zero, given that $C > 1$. This proves that if ϵ is close enough to zero, then $(\mu^{1\infty},\mu^{2\infty})$ cannot be a steady state characterized by a one-dimensional conflict, so that (by Proposition 2), $(\mu^{1\infty},\mu^{2\infty}) = (\delta_s,\delta_s)$.

QED

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

Aghion, Philippe, Patrick Bolton, Christopher Harris, and Bruno Jullien, "Optimal Learning by Experimentation," *Review of Economic Studies,* LVIII (1991), 621–654.

Arrow, Kenneth, *Social Choice and Individual Values* (New York: John Wiley and Sons, 1963).

Banerjee, Abhijit, and Rohini Somanathan, "A Simple Model of Voice," mimeo, MIT, 1998.

Crawford, Vincent, and Joseph Sobel, "Strategic Information Transmission," *Econometrica,* L (1982), 1431–1451.

Cukierman, Alex, and Mariano Tommasi, "When Does It Take a Nixon to Go to China?" *American Economic Review,* LXXXVIII (1998), 180–197.

Hart, Sergiu, "Nonzero-Sum Two-Person Repeated Games with Incomplete Information," *Mathematics of Operations Research,* X (1985), 117–153.

Neveu, Jacques, *Discrete-Parameter Martingales* (Oxford: North-Holland, 1975).

Piketty, Thomas, "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics,* CX (1995), 551–584.

Poole, Keith, and Howard Rosenthal, "Patterns of Congressional Voting," *American Journal of Political Science,* XXXV (1991), 228–278.

Snyder, James, "Constituency Preferences: California Ballot Propositions, 1974–1990," *Legislative Studies Quarterly,* XXI (1996), 463–487.