A Theory of Cabinet-Making: The Politics of Inclusion, Exclusion, and Information

John W. Patty[*]

September 5, 2013

Abstract

I describe a model of strategic communication within groups in policy-making situations with decentralized policy-making authority. I show that, in a cheap-talk environment, inclusion and exclusion of agents can affect the credibility of signaling between agents and, accordingly, both the quality of individual policy decisions and social welfare. Somewhat surprisingly, the inclusion of agents can aid information aggregation and social welfare even when the added agents do not themselves communicate truthfully. Analogously, the results suggest an informational, social-welfare-based rationale for excluding agents not only from observing policy-relevant communication but also from observing the product of the communication precisely because the excluded agents possess decision-making authority. The results are applied to institutional questions regarding voluntary association among decision-makers and the propriety of allowing policy-makers to choose their own advisors.

[*] Department of Political Science and Center for New Institutional Social Sciences, Washington University in Saint Louis. Email: *jpatty@wustl.edu*. This research has benefited from the comments and suggestions of Justin Fox, Sean Gailmard, Francesco Squintani, and audience members at the NYU-LSE conference on political economy. In addition, I especially thank Maggie Penn for multiple incredibly helpful discussions and pointing out an error in an earlier draft. Of course, I am solely responsible for any and all remaining errors.

Information aggregation is inherently a collective activity. However, the goal of information pooling might itself be a collective decision (*e.g.*, a jury deciding the guilt or innocence of a defendant, a legislature amending and/or enacting a law) or it might consist of decentralized individual decisions. This article is about the second case: collective information sharing followed by individual decision-making. Furthermore, I focus on a high profile example of this: the formation of (informal or formal) information-sharing groups, or "cabinets," in a government. By "informal or formal," I mean to both avoid the question of formal assignment to positions of individuals who are then subsequently excluded from collective information sharing as well as the question of individuals who have no formal government position, but may nonetheless possess independent policy-making authority. I provide and discuss examples of each of these below.

Prior to that, it is important to note that, even in Presidential systems, some (and frequently many) subordinates have both *de jure* and *de facto* independent policy-making authority. Regardless of whether a direct corollary of statutory and/or constitutional design or the practical and logistical realities of the broad spans of control in modern governance institutions, the notion of an unambiguously "unitary executive" is simultaneously a useful and yet facile trope. Focusing on the United States federal government, for example, many statutes grant the *de jure* authority to promulgate binding regulations to officials other than the president.[1] Thus, even when one agent can be identified as "the principal" or "superior" among a group of individuals, this of course does not imply that the other actors in the group have no unilateral (even if partial) sway over policy decisions.

Taking both such unilateral authorities and possession of private, dispersed information as given, I consider the question of whether and when some of these agents can credibly share their information in "cheap-talk" settings. These are situations in which actors can communicate only through costless messages whose veracity can not be independently verified (at least not prior to the point at which the listeners must make tangible choices on the basis of the information gleaned from them). Cheap-talk communication represents the gold standard of credible or, more poetically, "trustworthy" communication. Intuitively, such communication is a collectively valuable phenomenon: I am interested in cases in which credible information aggregation is unambiguously and universally perceived as a socially- and individually-valued good, *per se*. Accordingly, the focus of the analysis presented in this article is when such communication can be bolstered through, for example, the exclusion of actors in the sense that they are not allowed to observe the messages other actors send to each other. The key finding that unites and relates various more fine-grained conclusions is that such exclusion can not only be (Pareto) socially optimal, but also that the circumstances that recommend such exclusion can be characterized and described in substantively interesting ways.

[1] Whether such statutes preclude presidential intervention is a contested matter (*e.g.*, Lessig and Sunstein (1994), Kagan (2001), Yoo, Calabresi and Colangelo (2005), Stack (2006), and Strauss (2006)). Recent empirical studies of the exertion, impact, and effectiveness of presidential attempts to control agency decision-making include Bressman and Vandenbergh (2006) & Mendelson (2009). Regardless of one's stance on the theoretical questions, however, practical realities imply that officials other than the president must from time-to-time make discretionary policy decisions.

While the model is very general, the point is particularly relevant for executive policymaking. In particular, high profile instances of high-ranking policy actors such as Cabinet-level officials being "out of the loop" with respect to policy decisions in their bailiwicks highlight the reality that such exclusion does occur even (or perhaps especially) when the stakes are high.[2] Conversely, concerns about the creation and utilization of informal advisors are a seemingly constant refrain in debates about the informational strategies relied upon in executive branch decision-making.[3]

Perhaps the principal contribution of the theory presented in this article to debates such as these is the provision and explication of an informational rationale for both exclusion and inclusion of actors in pre-decision policy making discussions. In other words, the theory provides an explanation for the design and construction of a set of selected actors (*e.g.*, who is "in the loop" and who is not) based on the credibility of advice and communication in pursuit of aggregating policy-relevant information. The explanation is not based on concerns about accountability—the motivations of every actor are (purposely) presumed to be known to all from the outset, so that (for example) signaling & career incentives based on any agent's worries about revealing his or her "type" to an external actor (such as a voter) play no role (*e.g.*, Banks (1990), Lohmann (1993, 1995), Duggan (2000), Canes-Wrone, Herron and Shotts (2001), Ashworth (2005, 2012)). Similarly, the rationale is not based on concerns about representation of various interests in the ultimate policy choice—in its representation of decision-making authority, the theory is purposely designed to set aside the coalition-formation and policy bargaining incentives that typically emerge in models of collective choice (*e.g.*, Romer and Rosenthal (1978), Baron and Ferejohn (1989), Krehbiel (1998)). Instead, the central dynamic that determines the optimal set of actors to include in "the room" in which policy information is exchanged prior to decision-making is a classic trade-off between including as many actors with meaningful decision-making authority and excluding those actors whose individual preferences are

[2] Recent examples include the famous exclusion of Secretary of State George Shultz from discussions and decisions that culminated in the Iran-Contra scandal during the Reagan Administration ("George Shultz was "cut out" of the information loop precisely because, with sound reasons, he disapproved of the operation. . . .His advice was unwanted by the President" (Harsch (1987))), the fragmented operations of Clinton's cabinet during the first two years of his administration ("It didn't take long for the Clinton cabinet to find itself out of the loop. The action is at the White House because that's where the president wants it. . . . Given the diversity of Clinton's assemblage and all the early personnel delays, government by group would have been a disaster." (Borger (1994))), and current debates about the composition of Obama's Cabinet and advisors ("Since his first days in office, Obama has been criticized for relying too heavily on a very small group of advisers, almost all in the White House, and for being largely disconnected from his Cabinet. . . .much of the real heavy lifting on foreign policy was done in the small circle that convened with the president for his morning security briefings. . .A similar pattern was apparent on the economic side, with Geithner, Lew and a few White House advisers enjoying great access and influence, and other members of the economic Cabinet often out of the loop." (Rothkopf (2013))).

[3] Famous examples of such debates include the role of Hillary Clinton in the development of the Clinton Administration's ultimately unsuccessful health care reform efforts (then-First Lady Clinton served as the chair of the Task Force on National Health Care Reform, which included the director of the Office of Management and Budget and the Secretaries of Health and Human Services, Treasury, Defense, Veterans Affairs, Commerce, and Labor), President George W. Bush's "energy task force" (formally the National Energy Policy Development Group), chaired by Vice President Cheney, and the appointment of "policy czars" outside of the normal advice and consent nomination process (Sholette (2010)). Both Hillary Clinton and Vice-President Cheney faced (and both ultimately overcame) court challenges to the composition of their task forces.

sufficiently different from the rest of the actors in the room so as to undermine the credibility of truthful, strategic cheap-talk communication within the room. The next section discusses a subset of the related literature in an attempt to properly place this article's framework and theoretical contribution.

1 Models of Information and Policy-making: An Overview

Of course, there have been many contributions to the theory of strategic communication, information aggregation, and policy-making. Space precludes anything approaching a complete summary of this broad literature. The framework utilized in this article is closely related to that used in models presented in Austen-Smith (1993), Wolinsky (2002), and Battaglini (2004). The theory presented in this article differs from those in that the information to be aggregated and potentially messaged is held by the same individuals who will make decisions. This distinction is particularly relevant when considering institutional design and welfare issues, as it renders impossible attempts to mitigate informational problems by simply picking a better advisor (typically one whose preferences are more consonant with those of the decision-maker). It is also a more realistic construction of the practical design problem faced when considering, for example, how to organize a self-regulating body within an industry.

The contributions of Hagenbach and Koessler (2010) and Galeotti, Ghiglino and Squintani (2013) are most closely related to the theory presented in this article. Each of these articles considers information transmission through networks with decentralized policy-making embedded in similar preference and information environments utilized in this article. Hagenbach and Koessler (2010) considers a different informational setting that is more general in that it allows for agents' signals to be of heterogeneous qualities (*i.e.*, some agents' signals are more informative about the underlying state of nature than others), but also imposes a constant marginal impact of truthful signals.[4] While their decision-making is very similar to that examined here, they focus on a common coordination incentive between agents that is qualitatively different from the incentives considered in this article. More importantly, Hagenbach and Koessler (2010) consider voluntary unilateral transmission of messages in the sense that each agent can decide to whom he or she wishes to send a message: thus, the communication network is endogenously generated in their framework.[5] The combination of these two features (coordination and endogenous network structures) implies that there is no general Pareto dominance relation between equilibria involving communication by different numbers of agents (Hagenbach and Koessler (2010), p. 1078).

Galeotti, Ghiglino and Squintani (2013) examine information aggregation through exogenously-

[4] Formally, the state of nature is equal to the sum of the agents' signals.

[5] Formally, a communication link from agent i to agent j if agent i plays a pure separating (*i.e.*, perfectly informative) strategy in terms of the messages agent i's sends to agent j.

specified network structures. By considering directed networks, their framework allows for the particularly interesting possibility of "one-way" communication, in which one agent i is able to send a message to agent j, but agent j is prohibited from sending a message to agent i. Galeotti, Ghiglino and Squintani (2013) presume (as do Hagenbach and Koessler (2010)) that each agent has equal decision-making authority in the sense that each agent's policy decision has the same impact on every other agents' payoff. By relaxing this assumption, the framework considered in this article allows for what I refer to as "purely advisory" agents, whose only impact on social welfare is through their private information as carried through the (equilibrium) impact of their messages on the policy choices of other agents with positive decision-making authority.

The model utilized in this article is also closely related to that used in Dewan and Squintani (2012), Patty and Penn (2012), and Gailmard and Patty (2013). Dewan and Squintani (2012) use the same informational environment to consider the creation and allocation of power within political factions and focus on the question of how decision-making authority might be transferred in equilibrium between agents prior to information aggregation in pursuit of more-informed (equilibrium) policy-making. Focusing on executive branch policy-making, Gailmard and Patty (2013) consider the potential impact of both endogenous power-sharing/delegation and transparency in a model of sequential decision-making. Taking a more abstract approach, Patty and Penn (2012) also consider sequential decision making and information aggregation, focusing in particular on the incentive and welfare impacts of different (small) network structures. The most important distinction between this paper and those is its focus on a specific feature of institutional design with a public messaging protocol: this focus obviously narrows the scope of application, but yields the benefit of results that are concomitantly stronger and more transparent.

With this article's theoretical focus situated relative to related work, I now proceed to the formal presentation of the model.

2 The Model

Let N denote a set of n individuals, $X = \mathbf{R}$ denote a policy space, and $\Theta = [0,1]$ denote a state space. Each individual $i \in N$ is initially (or formally) endowed with policy making authority $\alpha_i \geq 0$, which measures the degree of unilateral decision-making autonomy possessed by agent $i \in N$. The state of nature, $\theta \in \Theta$, is determined according to a distribution characterized by cumulative distribution function $F : [0,1] \to [0,1]$. Upon realization of θ according to F, each individual $i \in N$ receives a conditionally independent (and private) signal $s_i \in \{0,1\}$ according to the following probability mass function:

$$\Pr[s_i = x|\theta] = \begin{cases} 1 - \theta & \text{if } x = 0, \\ \theta & \text{if } x = 1. \end{cases}$$

Letting $g_i(\cdot|s_i)$ denote the probability density function of i's posterior probability distribution function of θ, given $s_i \in \{0,1\}$, this belief is given by

$$g_i(t|s=1) \quad = \quad \begin{cases} \frac{1-t}{1-E_F[\theta]} f(t) & \text{if } s_i = 0, \\ \frac{t}{E_F[\theta]} f(t) & \text{if } s_i = 1. \end{cases}$$

If F is the cumulative distribution function for the Uniform$[0,1]$ distribution, then

$$g_i(t|s_i) \quad = \quad \begin{cases} 2(1-t) & \text{if } s_i = 0, \\ 2(t) & \text{if } s_i = 1. \end{cases}$$

Note that the uniform distribution is a useful baseline, as it maximizes the *ex ante* informativeness of each agent's signal. In other words, this is the case in which information aggregation is most important to all agents from an *ex ante* perspective.[6]

Payoffs. Each player $i \in N$ chooses policy $y_i \in \mathbf{R}$, and denote the vector of these decisions by $y = (y_1, \ldots, y_n)$. Furthermore, each player $i \in N$ has a payoff function of the following form:

$$u_i(y, \theta; \beta) = -\sum_{j=1}^{n} \alpha_j (y_j - \theta - \beta_i)^2,$$

where $\beta_i \in \mathbf{R}$ denotes the *preference bias* of agent i and $\beta \equiv \{\beta_i\}_{i \in N}$ denotes the profiles of all preference biases. We assume throughout that these biases are common knowledge to all of the players. Note that the autonomy of each player j factors into the payoffs of every player (including j) by determining the importance of j's decision. Thus, setting $\alpha_j = 0$ is equivalent to eliminating j's decision-making authority.

Messaging. Throughout this article, I consider a binary messaging technology, where each message any individual sends must be either "0" or "1." I consider a classic and simple messaging environments in this article, which I refer to as "in the room" messaging, where all agents who are "in the room" communicate publicly with each other.[7] Thus, each agent $i \in N$ must choose only a single message, denoted by $m_i \in \{0,1\}$, to announce publicly to all other agents.

Policy-making. Following the messaging stage, each individual is presumed to make unilateral decisions that are private in the sense of not being observed by any other agent until after all agents'

[6] Accordingly, this baseline amplifies the importance of the results when they indicate that information is not aggregated in equilibrium or that optimal institutional design limits information aggregation.

[7] This type of messaging is also referred to as "public" messaging by many authors (*e.g.*, Goltsman and Pavlov (2011) and Farrell and Gibbons (1989)). I use the term "in the room" because of the article's focus on the incentive and welfare effects of inclusion and exclusion from the set of message-senders and listeners.

policy decisions have been made. Thus, policymaking in equilibrium will always be "truthful," because one's policy choice cannot affect the policy choices of any other agents.[8]

A player's posterior beliefs after m trials and k successes (*i.e.*, k occurrences of $s = 1$ and $m - k$ occurrences of $s = 0$) are characterized by a Beta$(k+1, m-k+1)$ distribution, so that

$$\begin{aligned} E(\theta|k,m) &= \frac{k+1}{m+2}, \text{ and} \\ V(\theta|k,m) &= \frac{(k+1)(m-k+1)}{(m+2)^2(m+3)}. \end{aligned}$$

Accordingly, the optimal policy choice for a policymaker, given (truthful) revelation of k successes and $m - k$ failures, is

$$y_i^*(k,m) = \frac{k+1}{m+2} + \beta_i. \tag{1}$$

Strategies and Equilibrium. I focus on pure strategy perfect Bayesian equilibria (referred to more simply as an equilibrium) in this article.[9] Accordingly, for each individual $i \in N$, i's strategy consists of a messaging strategy, $\mu_i : \{0,1\} \to \{0,1\}$, and a policy-making strategy, y_i. Sequential rationality in equilibrium pins down y_i^* as described in Equation (1). Accordingly, I characterize equilibria entirely by the vector of players' messaging strategies.

Letting $R = (N, \{\alpha_i\}_{i \in N}, \{\beta_i\}_{i \in N}) \equiv (N, \alpha, \beta)$ describe the strategic situation (or, more simply, "room"), the set of mixed (behavior) strategies for agent $i \in N$ is denoted by Σ_i and the set of $|N|$-dimensional strategy profiles is denoted by $\Sigma(R)$. Note that, given the assumed use of the "in the room" messaging protocol and the focus on pure strategy equilibria, an equilibrium can also be entirely characterized as a partition of the set of agents, N, into two (possibly empty) subsets, M and B, where M is the set of agents using a truthful messaging strategy and B is the set of agents who always choose the same message, regardless of the signal observed (*i.e.*, they are "babblers").

3 Equilibrium Analysis

I first derive the incentive compatibility conditions for any given agent to be truthful, under the presumption that every other agent will utilize his or her message and, indeed, that all agents will possess the same information (and, accordingly, beliefs about θ) following the messaging round and prior to their individual policy choices. The sequence of play can be thought of without generality as follows:

1. Each agent $i \in N$ simultaneously and privately observes s_i,

[8] This distinguishes this article's analysis from those in both Patty and Penn (2012) and Gailmard and Patty (2013).

[9] This approach is also used in Hagenbach and Koessler (2010), Galeotti, Ghiglino and Squintani (2013), Dewan and Squintani (2012), Gailmard and Patty (2013), and Patty and Penn (2012). Mixed strategy equilibria can exist in these settings, but characterization of such equilibria is very difficult due to the combinatorics of the underlying problem.

2. Each agent i simultaneously reveals $m_i \in \{0,1\}$,

3. Each agent i observes the set of all messages, $m = \{m_j\}_{j \in N}$.

4. Each agent i simultaneously chooses his or her policy, y_i.

5. Players receive payoffs and game concludes.

For any room $R = (N, \alpha, \beta)$ with $|N| = n + 1$, the incentive compatibilityx conditions for truthful messaging by any agent $j \in N$ (under the presumption that all other n agents are also being truthful) are:[10]

$$\sum_{i \in N-j} \alpha_i (\beta_j - \beta_i)^2 \leq \sum_{i \in N-j} \alpha_i \left(\beta_j - \beta_i - \frac{1}{n+3}\right)^2, \text{ and}$$
$$\sum_{i \in N-j} \alpha_i (\beta_j - \beta_i)^2 \leq \sum_{i \in N-j} \alpha_i \left(\beta_j - \beta_i + \frac{1}{n+3}\right)^2.$$

These are satisfied if and only if[11]

$$\left| \beta_j - \frac{\sum_{i \in N-j} \alpha_i \beta_i}{\sum_{i \in N-j} \alpha_i} \right| \leq \frac{1}{2(n+3)}. \tag{2}$$

Inequality (2) identifies, for each agent j who is "in the room," two factors as relevant for the incentive compatibility of truthfulness. I refer to these as the *weighted preference divergence* for agent j,

$$WD(j; \alpha, \beta) = \frac{1}{\sum_{i \in N-j} \alpha_i} \left| \sum_{i \in N-j} \alpha_i (\beta_j - \beta_i) \right|,$$

and the *manipulative impact*,

$$MI(n) = \frac{1}{2n+6}.$$

Weighted Preference Divergence, $WD(j; \alpha, \beta)$. The weighted preference divergence for any agent j measures the net weighted divergence of preferences between agent j and all other agents i who are in the room. Note that this measure is (intuitively) non-negative: weighted preference divergence is simply a weighted distance measure. More interestingly, this measure can be zero for an agent j even when there is preference heterogeneity in the room. For example, suppose that $n + 1 = 3$ and

$$\alpha_1 = \alpha_2 = \alpha_3 = 1, \qquad \text{and}$$
$$\beta_1 = -1, \quad \beta_2 = 0, \quad \beta_3 = 1.$$

[10] Below, I consider the possibility of incompletely truthful equilibria in which only some agents are truthful while others babble (and are accordingly ignored).

[11] Inequality (2) understandably mirrors, but does not duplicate, Inequality (4) in Hagenbach and Koessler (2010).

Then $WD(2;\alpha,\beta) = 0$: in an intuitive sense, agent 2's preference bias is exactly "in the middle" of the group's preference biases. Thus, the relative biases of agents 1 and 3 offset each other from agent 2's perspective. This implies that agent 2 can never gain from manipulating. Specifically, his or her incentive compatibility condition, according to (2), is:

$$\frac{1}{2}\left|(0-1)+(0+1)\right| \quad \leq \quad \frac{1}{2(2+3)} \Rightarrow 0 \leq \frac{1}{10},$$

which is obviously true. Furthermore, that this conclusion, while knife-edge, can be obtained with more agents and/or when the agents have unequal decision-making weights and imbalanced/asymmetric preference biases.

More generally, (2) establishes a general result that is interesting in its own right. Namely, incentive compatibility is more binding for agents with (relatively) "extreme" preference biases[12] than for agents with moderate preference biases. I now turn to discuss the second factor identified by Inequality (2), manipulative impact.

Manipulative impact, $MI(n)$. The manipulative impact measures the net change (in distance, not weighted by α) in policy choice by any agent based on the message of any other agent under the presumption that every agent believes that every other agent is being truthful. Recall that the framework presumes each agent's signal/information is of equal quality. Thus, when all agents are being truthful, the impact any agent j's message has on the sequentially rational policy choice by any other agent i is identical for all pairs of agents i, j. Because $MI(n)$ is strictly decreasing in n, Inequality 2 is increasingly difficult to satisfy for larger groups. Furthermore, this comparative static is *independent of the weighted preference divergence in the room.* This has an important implication. For any given room $R = (N,\alpha,\beta)$ with $\max_{i\in N}[WD(i;\alpha\beta)] > 0$ (*i.e.*, at least two agents have different preference biases), the addition of enough purely advisory agents—agents with no (or arbitrarily small) decision-making authority—to R will eventually cause Inequality 2 to be violated for some agent j. In other words, there is such a thing as "too much information," at least with respect to supporting truthful equilibria "in the room."

At first blush, this conclusion might seem to indicate a potential justification for excluding agents from the room, as in Meirowitz (2007). However, this intuition is only partly correct. As I return to later in the article (4.3), if the agents' policy choices are tied to their messages, then such a conclusion is potentially warranted, but not otherwise. This is because the informational contents of all agents' signals are identical. Accordingly, if one adds a purely advisory agent to the room and this causes an agent to "stop being truthful," the net informational impact of the addition is zero.[13]

[12] Here, "relatively" refers to the comparison of any agent's preference bias with those of the other agents in the room.

[13] Furthermore, as discussed in Section 3.1, this generalizes—after some attention is paid to equilibrium selection—to the cases in which the addition of an advisory agent leads to a violation of Inequality (2) for more than one agent.

3.1 Equilibrium Existence

Due to the cheap-talk nature of the messaging protocol, there is always at least one pure strategy equilibrium for any room $R = (N, \alpha, \beta)$. Specifically, a babbling equilibrium always exists in which (for example) every agent $i \in N$ always sends message $m_i = 0$ regardless of his or her signal and every agent i chooses $y_i = \frac{s_i+1}{3} + \beta_i$ regardless of the observed profile of messages. Thus, let $E(R) \subseteq \Sigma(R)$ denote the set of pure strategy equilibria for a room R. For any room $R = (N, \alpha, \beta)$ and any subset of agents $M \subseteq N$ an equilibrium is *M-truthful* if it satisfies

$$\forall i \in M \quad \mu_i(0) = 1 - \mu_i(1) \text{ and } \forall j \in N - M \quad \mu_j(0) = \mu_j(1). \tag{3}$$

An equilibrium $e \in E(R)$ is referred to as *completely truthful* if it is N-truthful. I now turn to the general question of existence of such an equilibrium.

Existence of a Truthful Equilibrium. The structure of the problem yields a simple necessary and sufficient condition for the existence of a completely truthful equilibrium for a given situation $R = (N, \alpha, \beta)$. This condition, which focuses attention on the agent with the maximal weighted preference divergence, is formally stated in Proposition 1.[14]

Proposition 1 *For any strategic situation $R = (N, \alpha, \beta)$, a completely truthful equilibrium exists for R if and only if*

$$\max_{j \in N} \left| \frac{\sum_{i \in N-j} \alpha_i(\beta_j - \beta_i)}{\sum_{i \in N-j} \alpha_i} \right| \leq \frac{1}{2(n+3)}. \tag{4}$$

The condition expressed in (4) neatly summarizes the relevance of the weighted preference divergence and manipulative impact characteristics of a group for the existence of completely truthful equilibria. In a nutshell, such an equilibrium depends on the values of these characteristics for the agent facing the most temptation to manipulate: this is the agent who faces the maximum level of weighted preference divergence.

Even if Inequality (4) does not hold, there can exist "incompletely truthful" equilibria in which only a subset of the agents are truthful (and, accordingly, listened to by the agents in the room). Letting $m = |M| < n + 1$ denote the number of agents who are playing a truthful strategy in an M-truthful strategy profile, the condition for such an equilibrium is

$$\forall j \in M, \left| \sum_{i \in M-j} \frac{\alpha_i(\beta_j - \beta_i)}{m+2} + \sum_{k \in N-M} \frac{\alpha_k(\beta_j - \beta_k)}{m+3} \right| \leq \sum_{i \in M-j} \frac{\alpha_i}{2(m+2)^2} + \sum_{k \in N-M} \frac{\alpha_k}{2(m+3)^2}. \tag{5}$$

The difference between the incentive compatibility constraint expressed in Inequality (2) and that expressed in Inequality (5) is due to the fact that any agent j who babbles (*i.e.*, $j \in N - M'$) will

[14] The proof of Proposition 1 is omitted, as it essentially follows directly from Inequality (2).

nonetheless use his or her own signal, s_j, in ultimately setting y_j and, furthermore, this fact is known by all those who signal truthfully in the equilibrium in question (*i.e.*, all agents $i \in M'$). Thus, the manipulative impact of a truthful agent's message varies across other agents, depending on whether those agents are babbling or not.

Before continuing, while one can always shrink the set of truthful message-senders in a truthful equilibrium and recover a (less informative) truthful equilibrium in which the new, smaller set of agents is exactly the set of truthful message-senders, it is not in general the case that one can shrink the set of *players* in a game possessing an M-truthful equilibrium and construct a truthful equilibrium of *any* size. Specifically, write $R = (N, \alpha, \beta) \subset R' = (N', \alpha', \beta')$ (*i.e.*, one room is a subset of another), if there is a mapping $f : N \to N'$ such that for all $i \neq j \in N$, $f(i) \neq f(j)$, $\alpha_i = \alpha_{f(i)}$, and $\beta_i = \beta_{f(i)}$. Then, the following proposition makes this point formally.

Proposition 2 *There exist rooms $R = (N, \alpha, \beta)$ and $R' = (N', \alpha', \beta')$ with $R \subset R'$ and R' possessing a N-truthful equilibrium, but R not possessing a M-truthful equilibrium for any $M \subseteq N$.*

Proof: Contained in the appendix. ∎

Intermediaries & Communication with Decentralized Decision-Making. The proof of Proposition 2 is constructive and illuminates an additional ancillary implication of the result: it is not difficult to construct examples in which the presence of an agent (or group of agents) with an intermediate bias can support truthful communication between agents with relatively extreme preferences.

In some ways, this result mirrors other results regarding the palliative effect of intermediaries on communication between agents with opposed policy preferences (*e.g.*, Kydd (2003), Ganguly and Ray (2006), Goltsman et al. (2009), Ivanov (2010)). However, the logic behind the conclusion in this context is different. In most models of mediation, the intermediary is (either strategically or sincerely) interested in obfuscating earlier messages "just enough" so as to make their content credible for the receiver: in effect and in the context of this model, a useful mediator removes enough information to reduce the manipulative impact faced by message-senders: manipulation is made unpalatable by virtue of the degree of policy impact meaningful manipulation after translation of the original message through the mediator's strategy (or, in sincere mediation settings, the "mediation protocol").

In this model, on the other hand, the informational content of any agent's message is beyond the reach of the new "mediating" agents (*i.e.*, the new agents with intermediate policy preferences).[15] Their presence in the room in this model supports truthful communication because of their independent decision-making authority: truthful signaling by the extreme agents becomes more attractive

[15] In particular, communication in this model is effectively simultaneous, whereas most mediation models are inherently sequential.

to those agents when the intermediary agents can also observe the extreme agents' messages. This point is made even more clearly by realizing that the important aspect of the agents' presence in terms of supporting the incentive compatibility of truthful communication by the extreme agents in these situations is actually bolstered by *barring these intermediary agents from signaling truthfully*. Thus, the example in the proof of the proposition is in some ways stronger than one might initially suppose, as it shows that one can introduce new agents so as to both establish credibility between existing agents *and* introduce additional information into all agents' decision calculuses.

3.2 Welfare Analysis

Equilibrium existence is interesting in its own right, but arguably the more relevant consideration for institutional design is the following formulation of *ex ante* expected social welfare from an equilibrium $e = \{\mu_i^*\}_{i \in N} \in E(R)$:[16]

$$SW(e; R) = -\sum_{i \in N} \alpha_i E_e[(y_i - \beta_i - \theta)^2], \tag{6}$$

where E_e denotes the expectation of $y_i(\{m_j\}_{j \in N})$ where m_j for each $j \in N$ is a random variable defined by $\mu_j^* : \{0, 1\} \to [0, 1]$.[17]

In a M-truthful equilibrium, Equation (6) reduces to[18]

$$SW(e, R) = -\frac{\sum_{i \in M} \alpha_i}{6|M| + 12} - \frac{\sum_{i \in N - M} \alpha_i}{6|M| + 18}. \tag{7}$$

Thus, the *ex ante* expected social welfare from an M-truthful equilibrium is higher than that from an M'-truthful equilibrium if and only if M contains more agents than M'.[19] This is stated formally in the following proposition.

Proposition 3 *For any room R and equilibria $e \in E(R)$ and $e' \in E(R)$, where e is M-truthful and e' is M'-truthful,*

$$|M| > |M'| \Rightarrow SW(e; R) > SW(e', R).$$

Proof: Contained in the appendix. ∎

[16] I refer to Equation 6 as a "formulation" because it is not the more traditional Benthamite sum of individual payoffs—it ignores the externalities experienced by each individual from the other agents' decisions. Equation 6 is more appropriate because it takes as given the individual agents' preferences and instead essentially focuses on the expected divergence between each agent's individual choice and his or her "target choice," $\beta_i + \theta$. In addition, the definition expressed in Equation 6 weights each individual's performance in this regard according to his or her decision-making authority.

[17] For reasons of space, I abuse notation a bit and omit the standard extension of μ_j to mixed (behavior) messaging strategies. Given the focus on and comparison of pure strategy equilibria in this article, such an extension is superfluous for the subsequent analysis.

[18] The two terms correspond to agents who are truthful and those who are babbling (but listening)—each agent in this second subset of agents also utilize his or her own signal when making his or her policy choice.

[19] Note that this is a simple cardinality comparison, it is not necessary that $M' \subset M$.

Social Ranking of Equilibria. For any room $R = (N, \alpha, \beta)$ and any strategy profile $s \in S(R)$, the *ex ante* expected payoff for agent $i \in N$ from s is denoted by $v_i(s, R)$. As is now well-known in cheap-talk games (Crawford and Sobel (1982)), *ex ante* expected equilibrium payoffs in this possess a great deal of useful structure across agents but this is unfortunately not true for interim expected payoffs (*i.e.*, agent i's conditional expected payoff from an equilibrium $e \in E(R)$ given observation of his or her signal, s_i). This reality, combined with the epistemological foundations of cheap-talk games,[20] motivates the use of *ex ante* expected payoffs for comparing equilibria.

Focusing on this notion of expected payoffs, the structure of the problem and equilibrium behavior implies that Proposition 3 can be strengthened: in particular, the pure strategy equilibria for any room R are also *Pareto*-ranked according to $SW(e, R)$.

Proposition 4 *For any room R and equilibria $e \in E(R)$ and $e' \in E(R)$, where e is M-truthful and e' is M'-truthful,*

$$|M| > |M'| \Rightarrow \{i \in N : v_i(e, R) > v_i(e', R)\} = N.$$

Proposition 4 is common in cheap-talk settings but is nonetheless still useful for my purposes in this article. Specifically, it obviates many questions about equilibrium selection that would hinder the analysis of "room design"—*i.e.*, questions about which agents to include and exclude from the room—to which I turn in the next section.

4 Optimal Rooms

The previous section considered existence and welfare properties of truthful equilibria. I now consider the question of "optimal rooms." In other words, letting

$$\mathcal{SW}(R) = \max_{e \in E(R)} [SW(e, R)].$$

denote the maximum equilibrium social welfare in a room R, what room maximizes $\mathcal{SW}(R)$? So simply stated, the problem does not represent an interesting question unless we take (say) the preference biases of the available agents as fixed prior to the room's "design." Accordingly, let $\mathcal{G} = (G, A, B)$ denote the latent *group* from which a room must be constructed: G is an index set of the agents in the group, A is a profile of $|G|$ (exogenous) individual decision-making authorities, and B is a profile of $|G|$ individual preference biases. We denote a special agent, whom I refer to as the *convener*, by $c \in G$. Without loss of generality, I normalize preferences by presuming that $\beta_c = 0$.

[20] Because individuals' conditional expected payoffs are obviously correlated with (and, in this case, would reveal) their private information, making institutional comparisons on the basis of interim expected payoffs places one in a realm where elicitation of individuals' rankings would require commitment to a mechanism (Myerson (1979)), the availability of which would obviate the need to engage in the cheap-talk game in the first place.

I consider a constrained optimization problem in which the convener must partition the set of agents into two sets, $G = N \cup O$, with $N \cap O = \varnothing$ and $c \in N$, such that N denotes the set of individuals inside the room, and O denoting the individuals left "outside." This constraint is meaningful in a couple of ways. First, the requirement that the convener be in the room is a binding constraint in some settings.[21] Second, it rules out "multiple room" designs. Even constraining each room to contain the convener (such that rooms might have overlap), such a multiple room design can dominate the best single room design. I return to this possibility in Section 4.2. The single room design problem is interesting in its own right, as (unmodeled) institutional constraints (such as open committee meeting, restrictions on *ex parte* contacts by the convener, and other "sunshine" laws) may preclude the creation of multiple rooms. In addition, the single room problem cleanly identifies the incentives faced by a "room designer."

In considering the single room design problem, I explore a simple optimization goal, which I refer to as *benevolent optimization*. Pursuing benevolent optimization, the convener seeks to maximize an analogue to maximum *ex ante* equilibrium social welfare, $\mathcal{SW}(R)$. Specifically, it involves maximization of the following function:

$$W_B(R, O) = \mathcal{SW}(R) - \frac{\sum_{i \in O} \alpha_i}{18}, \tag{8}$$

which implies that the optimization problem takes into account the cost of information lost not only *from* but also *by* those decision-makers excluded from the room.[22]

In general, maximization of (8) is potentially complicated when the agents have differing decision-making authorities (*i.e.*, $\alpha_i \neq \alpha_j$ for some agents i and j). Let $R_B(\mathcal{G})$ denote the room(s) that maximize (8) for a given group $|mathcalG$. Example 1, below, simultaneously illustrates the potential complications of constructing $R_B(\mathcal{G})$ and represents a proof of the following proposition, which states that benevolent optimization is not equivalent to choosing the room that supports an equilibrium that maximizes the number of truthful agents.

Proposition 5 *There exist groups $\mathcal{G}$ such that there are rooms $R' \subseteq \mathcal{G}$ with M-truthful equilibria such that M contains strictly more agents than are truthful in any truthful equilibrium supported by the optimal room under the benevolent optimization goal, $R_B(\mathcal{G})$.*

A second interesting aspect of benevolent optimization is that it might lead to the choice of a room in which one or more agents in the room are nonetheless uninformative. This is stated formally in the next proposition.

[21] Of course, this restriction might be justified as an exogenously imposed procedural requirement. I set aside consideration of this point for future research.

[22] Note that $\frac{\sum_{i \in O} \alpha_i}{18}$ is the variance of the state of nature and the policy decisions made by the excluded agents. Thus, I am assuming for simplicity that excluded agents do not form their own separate room(s). Such rooms would be beneficial to society and the convener. However, while this is an intriguing possibility, explicit consideration of this more general problem is intractable due to the combinatorics of the possible differ room configurations.

Proposition 6 *There exist groups $\mathcal{G}$ such that the equilibrium offering maximum* ex ante *expected social welfare in the optimal room under the benevolent optimization goal, $R_B(\mathcal{G}) = (N, \alpha, \beta) \subseteq \mathcal{G}$, is an M-truthful equilibrium for some $M \subset N$.*

The proof of Proposition 6 is constructive and consists of construction of a group $|mathcalG$ satisfying the statement of the proposition. The proof is simultaneously illustrative of a second characteristic of optimization. In particular, when comparing equilibria with equal numbers of truthful agents, social welfare will in general depend on the exact assignment of agents to truth-telling and babbling roles. This fact produces a succinct characterization of the socially optimal equilibrium in any given room. This characterization is summarized in the following proposition.

Proposition 7 *For any room R and M-truthful equilibrium $e \in E(R)$, if*

$$SW(e, R) = \mathcal{SW}(R)$$

then $i \in M$ and $j \in N - M$ implies that

$$\alpha_i \leq \alpha_j.$$

In words, Proposition 7 says that, in a socially (and, by virtue of Proposition 4, Pareto) optimal equilibrium, no truthful agent has strictly greater decision-authority than any non-truthful/babbling agent. The rationale behind this is identified by Equation (7): non-truthful agents in a M-truthful equilibrium can use their own information (*i.e.*, their own private signal) in addition to the $|M|$ truthful messages revealed by their colleagues.

Full exploration of this point is beyond the scope of this article. However, it is interesting to note that, if one thinks of the convener as having the greatest decision-making authority (an assumption I do not make in the general analysis), he or she will in some sense be the "least likely" to be offering informative opinions "in the room."[23]

Before moving on, the following example is intended to demonstrate a regularity that, though not uniformly dispositive, is undoubtedly present in most institutional design situations falling within the confines of the framework described in this article. Specifically, combining agents with decentralized information and decision-making authority immediately presents the convener with the challenge of "trading off" the value of an agent's private information with his or her decision-making authority. As described in Proposition 8 (Section 4.1, below), agents with no decision-making authority (*i.e.*, purely advisory agents) represent at best a technical complication for an institutional designer: adding such agents may, strictly speaking, not support a "more truthful" equilibrium, but they can never undermine the *ex ante* expected welfare of any Pareto optimal equilibria. Thus, Example 1 considers a situation in which one relatively large set of agents—who collectively-cum-

[23] This point provides an interesting angle on questions of "top-down transparency," as discussed in Gailmard and Patty (2013).

individually possess a large amount of policy-relevant information—holds a small amount of collective decision-making authority and another agent with an opposed preference bias (relative to the convener) possesses a commensurately small amount of private information but a disproportionately large degree of decision-making authority.

In the example, benevolent optimization calls for a room composed of only the convener and the single, powerful agent, excluding a large group of collectively-less-powerful but collective-more-informed agents. The framing of Example 1 is deliberately chosen so as to be both stark and counterintuitive. In particular, it controverts a common intuition that, because more information is, *ceteris paribus*, Pareto-superior to less (Proposition 4), social-welfare-maximization necessarily calls for the creation of the room with a greater amount of information aggregated by those making decisions (*i.e.*, more truthful messages).

Example 1 (Authority Trumps Information.) Suppose that the group G contains 10 agents, $G = \{c, 1, 2, \ldots, 9\}$, with preferences and authorities as follows:

$$\begin{aligned} \alpha_c &= 0.10, & \beta_c &= 0, \\ \alpha_1 &= 0.80, & \beta_1 &= -0.11, \\ \alpha_2 = \alpha_3 = \ldots \alpha_8 = \alpha_9 &= 0.0125, & \beta_2 = \beta_3 = \ldots = \beta_8 = \beta_9 &= 0.04. \end{aligned}$$

In this situation, one can verify that $R = N$ (excluding no agents from the room) is not compatible with a completely truthful equilibrium.[24] Furthermore, there is no M-truthful equilibrium for $R = N$ for any M containing any nonempty subset of the agents $\{1, 2, 3, \ldots, 8\}$. Indeed, the only M-truthful equilibrium with $R = N$ in this case is with $M = \{c\}$, in which the convener announces his or her signal truthfully, and every other agent babbles. According to the benevolent optimization goal, this equilibrium yields a payoff of

$$W_B(N, \varnothing) = -\frac{1}{18} \approx 0.056.$$

Thus, noting the structure of the problem and the incentive compatibility difficulties encountered in attempting to create a completely truthful equilibrium in $R = G$, calculate the values of the two obvious "next best" choices: a room with c and 1 and a room with c and all agents other than 1. V,

[24] Specifically, the incentive compatibility condition for agent 1 in a completely truthful equilibrium with $R = N$ requires

$$\frac{0.1(0.1) + 0.1(0.15)}{0.2} = 0.105 \leq \frac{1}{24} \approx 0.088,$$

which is not satisfied.

the values of these two arrangements are:

$$
\begin{aligned}
W_B(\{c,1\},\{2,\ldots,9\}) &= -\left(\frac{0.9}{24}+\frac{0.1}{18}\right) \approx -0.043, \\
W_B(\{c,2,\ldots,9\},\{1\}) &= -\left(\frac{0.20}{72}+\frac{0.8}{18}\right) \approx -0.047.
\end{aligned}
$$

Thus, while excluding only agent 1 from the room supports a truthful equilibrium with much more information being transmitted (9 signals are revealed truthfully versus only 2), the fact that

$$W_B(\{c,1\},\{2,\ldots,9\}) > W_B(\{c,2,\ldots,9\},\{1\})$$

implies that the excessive authority of agent 1 justifies excluding the other agents from the room so that agent 1 can make more informed policy based on the signal observed by c. △

4.1 What Kinds of Agents Are Problematic?

Following on the earlier discussion of the impact of adding purely advisory agents, note that adding a purely advisory agent affects only the manipulative impact, $MI(n)$. However, one can apply logic similar to that explicated above to establish that, if there is a truthful equilibrium for a given situation R, then after adding a set of $A > 0$ advisory agents (*i.e.*, agents having no decision-making authority: $\alpha_k = 0$ for each such agent k) there still exists a payoff-equivalent equilibrium in which each of these advisory agents simply babbles and the original $n+1$ agents in R behave truthfully.[25] This fact implies that the principal "risk" in terms of upsetting a informative equilibrium through the addition of an agent involves adding an agent whose preferences are extreme relative to the other agents in the room (as measured by the weighted preference divergence measures, $WD(j;\alpha,\beta)$).

Note that the addition of an agent to the room will potentially affect the weighted preference divergence of every existing agent. Thus, information aggregation in "in the room" messaging is most unambiguously hindered through the inclusion of a sufficiently extreme new agent *with positive decision-making authority*. This point is established formally in the next proposition.

Proposition 8 *Consider two rooms $R = (N,\alpha,\beta)$ and $R' = (N',\alpha',\beta')$ with $R \subset R'$. If $\mathcal{SW}(R') < \mathcal{SW}(R)$, then there exists $j \in N' - N$ such that $\alpha_j > 0$.*

Proposition 8 implies that reducing maximal equilibrium welfare through the introduction of new agents to a room requires that at least one of the new agents has independent decision-making authority. In line with the earlier discussion, Proposition 8 establishes that adding new agents can reduce social welfare in an unambiguous way only if the new agents include some "listeners" whose

[25] Other, strictly socially preferable equilibria may be supportable after the inclusion of the A advisory agents, but that is beside the point for my current purposes.

preferences are different from one or more of the existing agents and who also possess independent decision-making authority.

4.2 Multiple Rooms

As discussed earlier, a complete characterization of the "optimal room(s) problem" would allow the convener to assemble agents into multiple, possibly overlapping, rooms. Nonetheless, Example 1 concretely illustrates how such a construction might improve social welfare. The question of multiple rooms raises several interesting questions, including constraints on the inter-room structure, such as whether two or more rooms can have members in common. In addition, allowing for multiples rooms raises the question of the sequencing of both messages (if one agent is in multiple rooms, can he or she observe the communication in one room prior to sending a message in another?) and actions (can one room observe the policy decisions made by agents in another room prior to communicating and/or making their own policy choices?)—and, of course, this possibility greatly expands the space of potential room designs that the convener can consider.[26] This expansion of possibilities indicates why multiple rooms will generally be superior from a convener's standpoint to a single room.

4.3 Tying Messages to Actions

This article has considered pure cheap-talk communication. In many ways, this is a simplification that (somewhat ironically) hinders truthful communication. An immediate avenue for extending the analysis in this article would be to allow for agents to "put their money where their mouth is" by expressing their information through policy choice. Patty and Penn (2012) explicitly allow for this and explore its implications in various small (3-players) organizational forms in which sequential policy decisions can transmit information among agents. Their analysis is very different from a direct extension of the framework considered here in that they explicitly exclude the possibility of "in the room" messaging. Combining their focus with that of this article would immediately lead to the consideration of what might be called "staggered rooms" in policy-making, potentially involving multiple rounds of messaging among different subsets of agents, with a round of policy decisions by one or more other agents intervening between each round of messaging.[27]

While tying messages to actions suggests a number of interesting institutional choices that one could evaluate and, intuitively, make inter-agent communication more credible, it is also important to remember that such a change breaks the welfare ranking tidily summarized in Proposition 3. In

[26] Some of these aspects are considered within fixed institutional arrangements by Patty and Penn (2012).

[27] Given that each agent observes only a single signal, it might seem that each agent would need to message only once, but I conjecture that this might not be the case with multiple rooms, since moderate agents can credibly communicate with various audiences that might not be able to credibly communicate with each other, a point that is raised in a different (simultaneous-move) context by Galeotti, Ghiglino and Squintani (2013).

particular, one or more agents communicate in some way directly tied to their policy choices, then the social welfare, $SW(e, R)$, from an M-truthful equilibrium e with $M \notin \{\varnothing, N\}$ depends on the exact specification of which agents are being truthful (*i.e.*, which agents are in M). "Babbling" by an agent i whose message is tied in some way to his or her policy decision and who has positive decision authority $\alpha_i > 0$ is *per se* socially costly. While I do not pursue this line of inquiry farther in this article, it is useful to note that it provides an argument for possibly excluding *even purely advisory* agents from policy discussions.

5 Conclusion

Information and authority are each frequently dispersed in real-world policymaking organizations. Successfully eliciting private, dispersed information through public communication between strategic individuals depends on the definition of who "the public" is: the first principle of strategic communication, particularly cheap talk communication, is that it is easier for it to be credibly truthful between agents with similar preferences. This is particularly true when decision-making authority is also dispersed and individuals have preferences about each others' decisions, as any individual decision-maker may not be able to credibly commit to making decisions based on his or her colleague's message in a fashion that ameliorates the inherent differences in the agents' preferences.

This article has presented a theory of information-sharing an individual policy-making in such an environment and explored the incentives strategic (equilibrium) behavior within it induce for an actor tasked with choosing which agents will be allowed "in the room" (or, perhaps, brought "in the loop") in order to engage in public communication prior to policy choices being rendered. The results identify several regularities of these incentives, including that the inclusion of agents within the room—even if the new agents are not identical to any of the other agents already within the room and/or even when the added agents do not themselves communicate truthfully—can aid information aggregation and social welfare. Furthermore, the optimal room design need not maximize the level of information that can be aggregated in equilibrium and, for analogous reasons, the optimal room might purposely exclude one or more decision-makers precisely because they possess "too much" decision-making authority. Finally, of course, equilibrium information transmission is in general higher within groups with strongly similar preferences than among groups of individuals with more divergent policy goals.

More broadly, the theory clarifies that informational motivations (and hence social-welfare considerations) can in some cases justify excluding agents with exogenous and independent decision-making authority. Accordingly, the theory illuminates the coherence of an informational rationale for granting policy-makers discretion over their advisors. While the relevance or validity of such a rationale is an empirical question and must be weighed against other considerations such as worries about accountability, coordination, transparency, and representation, the theory presented here nonetheless present a (more than a knife-edge) cautionary tale regarding an immediate inference of unsavory motivations by decision-makers who "wall themselves off" and exclude decision-makers whose preferences differ from their own.

A Proofs

Proposition 2. *There exist rooms $R = (N, \alpha, \beta)$ and $R' = (N', \alpha', \beta')$ with $R \subset R'$ and R' possessing a N-truthful equilibrium, but R not possessing a M-truthful equilibrium for any $M \subseteq N$.*

Proof: The proof proceeds by constructing a pair of rooms R and R' satisfying the description of the proposition. Accordingly, let $N = \{1, 2, 3\}$ and R be defined as follows:

i	α_i	β_i
1	$1/3$	0
2	$1/3$	0.065
3	$1/3$	0.13

It can be verified that R possesses a completely truthful (N-truthful) equilibrium. However, $R' \subset R$ defined as follows

i	α_i	β_i
1	$1/3$	0
3	$1/3$	0.13

does not possess any truthful equilibria. ∎

Proposition 3. *For any room R and equilibria $e \in E(R)$ and $e' \in E(R)$, where e is M-truthful and e' is M'-truthful,*

$$|M| > |M'| \Rightarrow SW(e; R) > SW(e', R).$$

Proof: Fix any room R and pair of equilibria $e, e' \in E(R)$, where e is M-truthful and e' is M'-truthful. Define the following four subsets of agents (any of which may be empty):

$$\begin{aligned} T(0,0) &= (N - M) \cap (N - M'), \\ T(0,1) &= (N - M) \cap M', \\ T(1,0) &= M \cap (N - M'), \text{ and} \\ T(1,1) &= M \cap M'. \end{aligned}$$

These sets classify the agents according to their membership in M and M' (*i.e.*, whether they are truthful in neither, one, or both equilibria e and e'). Then, recalling Equation (7),

$$\begin{aligned} SW(e;R) - SW(e';R) &= \sum_{i \in N} \alpha_i E_{e'}[(y_i - \beta_i - \theta)^2] - \sum_{i \in N} \alpha_i E_e[(y_i - \beta_i - \theta)^2], \\ &= \sum_{i \in T(0,0)} \frac{\alpha_i(|M| - |M'|)}{6(|M'| + 3)(|M| + 3)} \\ &\quad + \sum_{i \in T(0,1)} \frac{\alpha_i(|M| + 1 - |M'|)}{6(|M'| + 2)(|M| + 3)} \\ &\quad + \sum_{i \in T(1,0)} \frac{\alpha_i(|M| - |M'| - 1)}{6(|M'| + 3)(|M| + 2)} \\ &\quad + \sum_{i \in T(1,1)} \frac{\alpha_i(|M| - |M'|)}{6(|M'| + 2)(|M| + 2)}, \end{aligned}$$

which, supposing that $|M| \neq |M'|$, is unambiguously positive so long as $|N| > 1$. However, if $|N| = 1$, $|M| = |M'| = 1$, as there is no untruthful equilibrium in this case, so that $SW(e;R) = SW(e';R)$ for all equilibria in the single agent case. Thus, if $|M| > |M'|$, it follows that $SW(e;R) > SW(e';R)$, as was to be shown. ∎

Proposition 4. *For any room R and equilibria $e \in E(R)$ and $e' \in E(R)$, where e is M-truthful and e' is M'-truthful,*

$$|M| > |M'| \Rightarrow \{i \in N : v_i(e,R) > v_i(e',R)\} = N.$$

Proof: Follows from straightforward calculations. ∎

Proposition 5. *There exist groups $\mathcal{G}$ such that there are rooms $R' \subseteq \mathcal{G}$ with M-truthful equilibria such that M contains strictly more agents than are truthful in any truthful equilibrium supported by the optimal room under the benevolent optimization goal, $R_B(\mathcal{G})$.*

Proof: Example 1 in the text. ∎

Proposition 6. *There exist groups $\mathcal{G}$ such that the equilibrium offering maximum* ex ante *expected social welfare in the optimal room under the benevolent optimization goal, $R_B(\mathcal{G}) = (N, \alpha, \beta) \subseteq \mathcal{G}$, is an M-truthful equilibrium for some $M \subset N$.*

Proof: The proof proceeds by constructing a group $\mathcal{G}$ satisfying the description of the proposition. Consider a group $\mathcal{G}$ consisting of a set of 3 agents, $N = \{1, 2, 3\}$, with authorities and preference biases, α and β, defined as follows:

i	α_i	β_i
1	$1/2$	0
2	$1/4$	0.11
3	$1/4$	0.11

Straightforward calculations confirm that there is no completely truthful (*i.e.*, N-truthful) equilibrium for R. Specifically, the incentive compatibility condition (Inequality (2)) for agent 1 in this case does not hold, because

$$0.11 > \frac{1}{10},$$

However, there are three M-truthful equilibria in which exactly two agents message truthfully (*i.e.*, one for each such designation of two distinct agents). Specifically, if $M = \{1,2\}$ (or, equivalently, $M = \{1,3\}$), the incentive compatibility conditions (Inequality (5)) reduce to

$$\begin{aligned} \left(\frac{1}{4}\right)\frac{0.11}{16} + \left(\frac{1}{4}\right)\frac{0.11}{5} &< \left(\frac{1}{4}\right)\frac{1}{32} + \left(\frac{1}{4}\right)\frac{1}{50}, \text{ and} \\ \left(\frac{1}{2}\right)\frac{0.11}{4} &< \left(\frac{1}{2}\right)\frac{1}{32} + \left(\frac{1}{4}\right)\frac{1}{50}, \end{aligned}$$

each of which holds, and if $M = \{2,3\}$, the incentive compatibility condition is

$$\left(\frac{1}{2}\right)\frac{0.11}{5} < \left(\tfrac{1}{4}\right)\tfrac{1}{32} + \left(\tfrac{1}{2}\right)\tfrac{1}{50},$$

which also holds, verifying that truthfulness is incentive compatible when all agents are in the room but exactly two of the agents are truthful and the remaining agent babbles. Finally, the optimal M-truthful equilibrium is that in which agents 2 and 3 signal truthfully and agent 1 babbles. Letting $\{2,3\}$ denote this equilibrium,

$$SW(\{2,3\}, R) = -\frac{0.5}{24} - \frac{0.5}{30} = -0.0375,$$

while the other two M-truthful equilibria with M containing two agents yield social welfare equal to

$$SW(\{1,2\}, R) = SW(\{1,3\}, R) = -\frac{0.75}{24} - \frac{0.25}{30} \approx -0.0396.$$

Thus, the optimal room with respect to benevolent optimization is $R_B(\mathcal{G}) = \{1,2,3\}$, but the optimal M-truthful equilibrium consists of a strict subset of agents truthfully signaling, as was to be shown. ∎

Proposition 7. *For any room R and M-truthful equilibrium $e \in E(R)$, if*

$$SW(e,R) = \mathcal{SW}(R)$$

then $i \in M$ and $j \in N - M$ implies that

$$\alpha_i \le \alpha_j.$$

Proof: Fix a room $R = (N, \alpha, \beta)$ and let $e = \{\mu_i^*\}_{i \in N} \in E(R)$ be an M-truthful equilibrium with

$$SW(e, R) = \mathcal{SW}(R).$$

Then, for the purpose of obtaining a contradiction, suppose contrary to the statement of the proposition that there exists $\hat{i}, \hat{j} \in N$ with $\hat{i} \in M$ and $\hat{j} \in N - M$ such that $\alpha_{\hat{i}} > \alpha_{\hat{j}}$. Let $e^{ij} = \{\mu_k'\}_{k \in N}$ denote the strategy profile defined as follows:

$$\mu_k' = \begin{cases} \mu_j^* & \text{if } k = i \\ \mu_i^* & \text{if } k = j \\ \mu_k^* & \text{if } k \notin \{i, j\} \end{cases},$$

and let $M^{ij} = M \cup \{j\} - \{i\}$ denote the set of agents sending truthful messages under e^{ij}. It is simple to verify that the incentive compatibility condition Inequality (5) is satisfied for each agent $k \in M^{ij}$, given that $e \in E(R)$ is an M-truthful equilibrium, so that e^{ij} is an M^{ij}-truthful equilibrium. Accordingly, note the following:

$$SW(e^{ij}, R) - SW(e, R) = \frac{\sum_{i \in M} \alpha_i}{6|M| + 12} + \frac{\sum_{j \in N-M} \alpha_j}{6|M| + 18} - \frac{\sum_{i \in M^{ij}} \alpha_i}{6|M^{ij}| + 12} - \frac{\sum_{j \in N - M^{ij}} \alpha_j}{6|M^{ij}| + 18},$$

which, noting that $|M| = |M^{ij}|$ by construction, reduces to

$$\begin{aligned} SW(e^{ij}, R) - SW(e, R) &= \frac{\alpha_{\hat{i}}}{6|M| + 12} - \frac{\alpha_{\hat{i}}}{6|M| + 18} + \frac{\alpha_{\hat{j}}}{6|M| + 18} - \frac{\alpha_{\hat{j}}}{6|M| + 12}, \\ &= \frac{6\left(\alpha_{\hat{i}} - \alpha_{\hat{j}}\right)}{(6|M| + 18)(6|M| + 12)}, \end{aligned}$$

so that, if $\alpha_{\hat{i}} > \alpha_{\hat{j}}$, $SW(e^{ij}, R) > SW(e, R)$, contradicting the supposition that $SW(e, R) = \mathcal{SW}(R)$. Thus, because supposing that $\alpha_{\hat{i}} > \alpha_{\hat{j}}$ for a truthful agent $\hat{i}$ and babbling agent $\hat{j}$ in an M-truthful equilibrium yielding maximum social welfare leads to a contradiction, it must be the case that, in a M-truthful equilibrium yielding maximum social welfare, $i \in M$ and $j \in N - M$ implies that $\alpha_i \leq \alpha_j$, as was to be shown. ∎

Proposition 8. *Consider two rooms $R = (N, \alpha, \beta)$ and $R' = (N', \alpha', \beta')$ with $R \subset R'$. If $\mathcal{SW}(R') < \mathcal{SW}(R)$, then there exists $j \in N' - N$ such that $\alpha_j > 0$.*

Proof: Take two rooms $R = (N, \alpha, \beta)$ and $R' = (N', \alpha', \beta')$ with $R \subset R'$ satisfying $\mathcal{SW}(R') < \mathcal{SW}(R)$. Then suppose, contrary to the proposition, that $\max_{j \in N' - N}[\alpha_j] = 0$. Let $e^* \in E(R)$ denote an M-truthful (perhaps completely truthful) equilibrium satisfying

$$SW(e^*, R) = \mathcal{SW}(R).$$

If $\max_{j \in N' - N}[\alpha_j] = 0$, then e^* can be extended to an M-truthful equilibrium for R', $e' \in E(R')$ simply by having each agent $j \in N' - N$ babble and each agent $i \in N$ use the same strategy as prescribed in e^*. Inequality (5) is satisfied for each agent $i \in M$ under e' because the left hand and right hand sides are identical under e' and e^* and e^* is an M-truthful equilibrium by construction. Thus,

$$\mathcal{SW}(R') \geq SW(e', R') = SW(e^*, R) = \mathcal{SW}(R),$$

contradicting the supposition that $\mathcal{SW}(R') < \mathcal{SW}(R)$. Accordingly, supposing that $\max_{j \in N' - N}[\alpha_j] = 0$ leads to a contradiction, implying that for any two rooms $R = (N, \alpha, \beta)$ and $R' = (N', \alpha', \beta')$ with $R \subset R'$ and $\mathcal{SW}(R') < \mathcal{SW}(R)$, there must exist $j \in N' - N$ such that $\alpha_j > 0$, as was to be shown. ∎

References

Ashworth, Scott. 2005. "Reputational Dynamics and Political Careers." *Journal of Law, Economics, and Organization* 21(2):441–466.

Ashworth, Scott. 2012. "Electoral Accountability: Recent Theoretical and Empirical Work." *Annual Review of Political Science* 15:183–201.

Austen-Smith, David. 1993. "Interested Experts and Policy Advice: Multiple Referrals under Open Rule." *Games and Economic Behavior* 5(1):3–43.

Banks, Jeffrey S. 1990. "Monopoly Agenda Control and Asymmetric Information." *The Quarterly Journal of Economics* 105(2):445–464.

Baron, David and John Ferejohn. 1989. "Bargaining in Legislatures." *American Political Science Review* 83:1181–1206.

Battaglini, Marco. 2004. "Policy advice with imperfectly informed experts." *Advances in theoretical Economics* 4(1).

Borger, Gloria. 1994. "Leon Panetta: Cabinet Maker." US News & World Report.

Bressman, Lisa Schultz and Michael P Vandenbergh. 2006. "Inside the Administrative State: A Critical Look at the Practice of Presidential Control." *Mich. L. Rev.* 105:47.

Canes-Wrone, Brandice, Michael C. Herron and Kenneth W. Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science* 45(3):532–550.

Crawford, Vincent P. and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50(6):1431–1451.

Dewan, Torun and Francesco Squintani. 2012. "The Role of Party Factions: An Information Aggregation Approach." Working Paper, University of Warwick.

Duggan, John. 2000. "Repeated Elections with Asymmetric Information." *Economics & Politics* 12(2):109–135.

Farrell, Joseph and Robert Gibbons. 1989. "Cheap talk with two audiences." *The American Economic Review* 79(5):1214–1223.

Gailmard, Sean and John W. Patty. 2013. "Giving Advice *vs.* Making Decisions: Transparency, Information, and Delegation." Working paper, Washington University in Saint Louis.

Galeotti, Andrea, Christian Ghiglino and Francesco Squintani. 2013. "Strategic Information Transmission in Networks." Forthcoming, *Journal of Economic Theory*.

Ganguly, Chirantan and Indrajit Ray. 2006. "On Mediated Equilibria of Cheap-Talk Games." V mimeo. University of Birmingham.

Goltsman, Maria and Gregory Pavlov. 2011. "How to talk to multiple audiences." *Games and Economic Behavior* 72(1):100–122.

Goltsman, Maria, Johannes Hörner, Gregory Pavlov and Francesco Squintani. 2009. "Mediation, Arbitration and Negotiation." *Journal of Economic Theory* 144(4):1397–1420.

Hagenbach, Jeanne and Frédéric Koessler. 2010. "Strategic communication networks." *Review of Economic Studies* 77(3):1072–1099.

Harsch, Joseph C. 1987. "Cut Out of the Loop." Christian Science Monitor.

Ivanov, Maxim. 2010. "Communication via a Strategic Mediator." *Journal of Economic Theory* 145(2):869–884.

Kagan, Elena. 2001. "Presidential Administration." *Harvard Law Review* pp. 2245–2385.

Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of U.S. Lawmaking*. Chicago, IL: University of Chicago Press.

Kydd, Andrew. 2003. "Which Side Are You On? Bias, Credibility, and Mediation." *American Journal of Political Science* 47(4):597–611.

Lessig, Lawrence and Cass Sunstein. 1994. "The President and the Administration." *Columbia Law Review* 94(1):5–129.

Lohmann, Susanne. 1993. "A Signaling Model of Informative and Manipulative Political Action." *American Political Science Review* 88:319–333.

Lohmann, Susanne. 1995. "Information, access, and contributions: A signaling model of lobbying." *Public Choice* 85(3):267–284.

Meirowitz, Adam. 2007. "In Defense of Exclusionary Deliberation: Communication and Voting with Private Beliefs and Values." *Journal of Theoretical Politics* 19(3):301.

Mendelson, Nina A. 2009. "Disclosing Political Oversight of Agency Decision Making." *Mich. L. Rev.* 108:1127.

Myerson, Roger B. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47(1):61–73.

Patty, John W. and Elizabeth Maggie Penn. 2012. "Sequential Decision-Making & Information Aggregation in Small Networks." Working Paper, Washington University in St. Louis.

Romer, Thomas and Howard Rosenthal. 1978. "Political Resource Allocation, Controlled Agendas, and the Status Quo." *Public Choice* 33:27–43.

Rothkopf, David J. 2013. "Obamas Cabinet: From a team of rivals to the usual suspects." Washington Post.

Sholette, Kevin. 2010. "American Czars, The." *Cornell Journal of Law & Public Policy.* 20:219.

Stack, Kevin. 2006. "The President's Statutory Powers to Administer the Laws." *Columbia Law Review* 106(2):263–323.

Strauss, Peter L. 2006. "Overseer, or the Decider—The President in Administrative Law." *George Washington Law Review* 75:696–760.

Wolinsky, Asher. 2002. "Eliciting information from multiple experts." *Games and Economic Behavior* 41(1):141–160.

Yoo, Christopher S., Steven G. Calabresi and Anthony J. Colangelo. 2005. "The Unitary Executive in the Modern Era, 1945-2004." *Iowa Law Review* 90(2):601–731.