Supplementary Appendix

Collaborative Research Process

Following emerging best practices in systems science modeling, our research team not only collaborated frequently with one another, but also convened multiple "milestone" check-in meetings (either in-person or virtual as circumstances dictated). During these meetings, we were able to supplement our team's content and modeling expertise with input from a group of individuals with substantial relevant practical experience. This Expert Advisory Group (n = 5 at the outset, 4 after the unfortunate passing of one of its members) consisted of a small group of carefully selected consultants who were current or former practitioners from state health departments and were capable of substantial participation and enthusiastic about our goal of creating and using an agent-based model (ABM) to provide insights into causes of and potential solutions to mis-implementation. The Group provided invaluable feedback throughout each phase of the work described in this document: agent-based model design, parameterization, testing, and experimentation. These milestone meetings were a source of key input into our research efforts; the timing and content of each was as follows:

Model design (September 2019)

For the initial model design workshop, the goal was to identify the important system-level drivers of mis-implementation. The guiding question that our group focused on was 'What are key actors and dynamics we want to represent'? To answer this question, we engaged in activities intended to generate a preliminary model design that we could use as a starting point for iterative design, development, and testing. The process was informed by group model-building methodology.[1–3] First, we collaboratively determined a set of relevant barriers and facilitators that might affect mis-implementation levels experienced in public health department settings, as well as a list of key mechanistic pathways involved in the outcomes of interest. We then ranked this list based on importance of factors, eliminating those deemed to have minimal impact. Next, we worked as a group to develop causal loop diagrams representing the hypothesized mechanisms and describing how factors were dynamically related to one another. Based on these materials, individual attendees with agent-based modeling expertise created multiple candidate design summaries that were presented to the group. Each design summary contained a preliminary description of key model elements (agents and their properties, actions and rules, along with environmental structure), a list of required data elements, and a list of potential outputs that could be generated. Through extensive discussion and refinement of these options, consensus was reached that gave us the intended preliminary model design.

Model parameterization (January 2020)

After further refining the model derived from the last Expert Advisory Group meeting, we had an initial set of data requirements to satisfactorily parameterize the model so that it meaningfully represented salient, real-world individuals, organizations, behaviors, and processes. The guiding question for this meeting was 'How should actors in our model and dynamics look based on extant sources of evidence'? By answering this, we had a preliminary strategy to ground the ABM in extant data, literature, theory, and expert guidance. This included parameter values established during the meaning through a consensus-building process. At this meeting the Group also identified the need for additional data to support the agent-based model's parameterization, leading to the design and deployment of the Supplemental stakeholder interviews described below.

Model testing (December 2020)

An initial "baseline" parameterization of the model was established based on the strategy established in the last meeting. During this meeting, initial findings from the model were presented to the Group highlighting: a) mis-implementation patterns, b) intervention decision patterns, and c) organizational structures generated by the model, and c) model dynamics including distributions of evidence-based decision-making (EBDM) ability, information sharing propensity over time, and frequencies of intervention effectiveness reporting and responses. For this session, the guiding question was 'Is model behavior sensible'? The goal was to establish whether experts with deep experience with and intuition about the operation of state health departments felt that the model had "generative sufficiency."[4] That is, using their own knowledge as a qualitative reference mode, did the model exhibit behavior patterns similar to those in the real world? The clear consensus was that it did so.

Model experimentation (March 2021)

During this meeting, we collectively determined a set of experimental conditions that we would use for model runs that reflected "counterfactual scenarios." The guiding question for this session was 'How can we use our model to productively explore ways in which organizational changes affect mis-implementation'? We wanted to identify a defined and finite (but potentially large) set of scenarios that could both be reflected with our existing model (i.e., could be characterized by specific changes in model parameters) and would be of interest to the field. We were able to successfully do so, resulting in the experiment strategy described below.

Model results (August 2021)

The goal of this final meeting was to present results of model experiments to the experts and research team. A guiding question was 'What are the implications for organizational intervention?' The discussion during this session is reflected in the narrative of our main text.

Literature Review

Based on input from the Expert Advisory Group, a literature review was performed to inform the design of model elements that characterize leadership's intervention continuation decision-making processes and employees' information sharing dynamics.

Literature review questions

Specifically, the literature review aimed at answering the following questions:

- How likely are employees to share information with their supervisors? What are the factors related to this?
- How often do supervisors make changes based on this information? What factors affect this?

Literature review protocol

The following databases were used for this literature review: EBSCO Databases, Academic Search Complete, Business Source Complete, MEDLINE, PsychINFO, and SocINDEX. To perform the search, we employed the following terms: "organizations" AND "information sharing" were used with either "supervisor," “employee,” or “employee voice.” The search was limited to academic journal articles from 2009 to 2019. The search resulted in 624 articles, of which 173 were reviewed in further depth. Of the studies reviewed, 16 were found to be of particular interest to this literature review. From these 16 articles, several additional useful studies were identified from the article citations, leading to a total of 38 articles.[5–42]

Literature review summary findings

The data extracted from the selected articles were summarized into five different topics: (1) Employee Voice; (2) Prevalence of Employee Voice; (3) Factors Related to Employee Voice; (4) Supervisor Response to Employee Voice; (5) Effect of Voicing Dissenting Viewpoint on Team Decisions Making.

This summary document informed our selection of functional forms representing our model’s core dynamics: information sharing and intervention continuation decisions.

Model Design

The Mis-Implementation (“MI”) model is an agent-based model (“ABM”) of intra-organizational communication and decision-making. Agents represent individuals within an organization (i.e., state health department), each of whom operate at a particular organizational level and have attributes that affect whether and how they share pertinent information about active interventions. Leadership within the organization uses available information to make decisions about intervention continuation that result in levels of mis-implementation (i.e., continuing ineffective interventions or terminating effective interventions) over time.

Each simulation run stochastically generates a single organization and all its constituent agents. The model uses a time step (i.e., granularity) of one month, and each simulation run is comprised of 36 steps, thus representing three years. The values for all parameters mentioned and described in this section are given in the parameterization section below.

Model Environment

The model begins by generating an organization, which is comprised of a list A of agents and a list T of interventions. At initialization, the number of levels m in the organization is drawn from a truncated normal distribution $N_{[1.5,\mu_m+2.5]}(\mu_m, \sigma_m)$ and rounded to the nearest integer.

The organization's agents are created iteratively starting with a single top-level agent (the "top agent" or "leader"). The remaining levels are filled in by iterating over all agents in each subsequent level and creating for each several subordinate agents in the level below. Each supervisor agent's subordinate count is drawn independently from a truncated normal distribution $N_{[0,50]}(\mu_o, \sigma_o)$. The process is completed when the m^{th} level has been generated, and the agents' supervisor-subordinate relationships constitute the full structure of the organization.

The list T is always initialized with 50 active interventions. Each intervention T_k has the following properties:

- Evidence support, $V_k \in [-1,1]$
- Age, $G_k \in [0, \infty]$
- External stakeholder support, $S_k \in [0,1]$
- Funder support, $F_k \in [0,1]$

Each intervention's properties are drawn from four appropriately truncated normal distributions with a prescribed correlation matrix. Intervention age (time in years since its initiation), the length of time an intervention has been implemented and maintained, is in units of years and increases by 1/12 each time step.

Agent Properties
Each agent A_i has the following properties:

- Level $L_i \in [1, m]$ in organizational hierarchy
- Supervisor $\text{super}(A_i)$, the agent to which they directly report
- List of subordinates $\text{subs}(A_i)$, agents directly reporting to this agent. This list is empty for agents in the m^{th} level.
- Ability to engage in evidence-based decision making ("EBDM"), $B_i \in [0,1]$
- Information sharing propensity ("sharing"), $H_i \in [0,1]$
- Assessment of current interventions, $R_{i,k} \in [-1,1]$ for *k* in 1, …, 50.

At each agent's initialization, EBDM and sharing values are drawn from truncated normal distributions on [0,1] with separate means and variances.

Each agent's intervention assessments can be thought of as their estimates of each intervention's evidence support V_k. At initialization, these assessments are generated for each agent by taking a weighted average of each intervention's true effectiveness and a uniform noise term, $u_{i,k} \sim U(-1,1)$: $R_{i,k} = B_i V_k + (1 - B_i) u_{i,k}$.

Update Rules (Model Behavior)

Agent Actions
Each model time step, the agents are activated in random order. Upon activation, agent A_i performs the following actions in order:

- Report: for each intervention, agent may send an assessment signal to their supervisor attempting to update their supervisor's assessment the intervention. For intervention T_k, the signal is sent with probability $p(\text{report}) = H_i|R_{i,k}|(1 - C_1 S_k)(1 - C_1 F_k)$ where C_1 is the reporting probability external influence coefficient.

- Update EBDM: Each non-leader agent ($L_i > 1$) moves their EBDM towards the EBDM value of their supervisor. This drift is asymmetrical: if $B_j > B_i, B_i^{new} = B_i^{old} + C_1$; if $B_j < B_i, B_i^{new} = B_i^{old} - C_2$, where $A_j = \text{super}(A_i)$ and C_1 and C_2 are the EBDM upward and downward steps, with $C_2 < C_1$.

If a report is sent from agent A_i to agent A_j on intervention T_k, the assessment signal is received with a probability that begins as the agent's information sharing propensity value and decreases as the assessment difference between agents grows larger: $p(\text{receive signal}) = H_j - |R_{j,k} - R_{i,k}| + C_1$, where C_1 is the assessment update probability constant term.

If the assessment signal is received, A_j updates their intervention assessment by $R_{j,k}^{new} = R_{j,k}^{old} + C_2(R_{i,k} - R_{j,k}^{old})$ where C_2 is the assessment update scale; agent A_i will then update their sharing upward: $H_i^{new} = H_i^{old} + C_3$, where C_3 is the sharing update size. If a report was sent but the signal was not received, A_i shifts their sharing down: $H_i^{new} = H_i^{old} - C_3$.

Administrative Actions

Throughout the model, interventions are evaluated and potentially discontinued.

- Intervention evaluation: An intervention evaluation consists of a complete round of information sharing by every agent in the organization, followed by a continuation decision by the top agent.
  - During the evaluation, information sharing reports are sent and received with 100% probability; they cause assessment updates as described above but do not trigger updates to agent sharing scores.
  - Every intervention receives a guaranteed evaluation every simulated year, in month 6. Additionally, each intervention may be evaluated in any other month with probability equal to the monthly intervention evaluation probability.
- Continuation decision: At the end of an evaluation, the intervention may be discontinued. This occurs for intervention T_k with probability $p(\text{discontinuation}_k) = C_1 + C_2 R_{1,k} + C_3 G_k + C_4 S_k + C_5 F_k$, where A_1 is the top agent, C_1 is the continuation function constant term, C_2 is the perception coefficient, C_3 is the age coefficient, C_4 is the stakeholder coefficient, and C_5 is the funder coefficient.
- Intervention replacement: If an intervention is discontinued, it is replaced with a new intervention with some constant probability. If it is not replaced, the number of active interventions will decrease and is never higher than its initial value of 50. A replacement intervention will have age zero, and effectiveness, stakeholder support, and funder support values drawn from a conditional multivariate distribution that preserves the correlations with age and other attributes.

Model Development

The model was written in Python,[43] using elements of the Mesa agent-based modelling framework,[44] as well as the NumPy,[45] SciPy,[46] and pandas packages.[47] Iterative testing was conducted throughout development, primarily consisting of examination of model output from sweeps of plausible as well as extreme parameter values and verifying expected behavior.

Model Parameterization ("baseline")

In Table S1, we present the specific parameter values that collectively comprise a simulation run under what we refer to as our "baseline" condition (i.e. the condition that produces organizational behavior, including mis-implementation, that corresponds to available data). With the exception of input from the Expert Advisory Group discussed above, data sources used to obtain each of these parameter values are described in greater detail below.

Description	Model Element	Parameter Value	Source Description
Number of organization levels (mean)	Organizational Structure	5	Supplemental Stakeholder Interviews
Number of organization levels (SD)	Organizational Structure	0.8	
Number of subordinates assigned to supervisors (mean)	Organizational Structure	5.1	Survey Data
Number of subordinates assigned to supervisors (SD)	Organizational Structure	1.7	
Intervention count	Active Interventions	50	Supplemental Stakeholder Interviews
Intervention age at run start (mean)	Active Interventions	5	Initial Stakeholder Interviews
Intervention age at run start (SD)	Active Interventions	3	
Intervention evidence support (mean)	Active Interventions	0.3	Supplemental Stakeholder Interviews
Intervention evidence support (SD)	Active Interventions	0.5	
Intervention stakeholder support (mean)	Active Interventions	0.7	
Intervention stakeholder support (SD)	Active Interventions	0.2	
Intervention funder support (mean)	Active Interventions	0.4	
Intervention funder support (SD)	Active Interventions	0.2	
Intervention age/evidence support (correlation)	Active Interventions	0.2	

Intervention age/stakeholder support (correlation)	Active Interventions	0.25	
Intervention age/funder support (correlation)	Active Interventions	0.15	
Intervention evidence support/ stakeholder support (correlation)	Active Interventions	0.35	
Intervention evidence support/funder support (correlation)	Active Interventions	0.3	
Intervention stakeholder support/funder support (correlation)	Active Interventions	0	
Initial Agent EBDM (mean)	EBDM Ability	0.75	Survey Data
Initial Agent EBDM (SD)	EBDM Ability	0.14	
Initial Agent information sharing propensity (mean)	Information Sharing Propensity	0.72	
Initial Agent information sharing propensity (SD)	Information Sharing Propensity	0.23	
Report probability function, external influence term	Intervention Assessment Reporting	0.9	Model Calibration
Assessment update probability function, constant term	Intervention Assessment Reporting	0.8	
Assessment update magnitude (scale)	Intervention Assessment Reporting	0.4	
Information sharing propensity update magnitude (step)	Information Sharing Propensity	0.1	
Continuation decision function, perception term	Continuation Decisions	-0.025	
Continuation decision function, age term	Continuation Decisions	-0.025	
Continuation decision function, stakeholder support term	Continuation Decisions	-0.1	
Continuation decision function, funder support term	Continuation Decisions	-0.1	
Continuation decision function, constant term	Continuation Decisions	0.75	
Probability of new age 0 intervention replacement	Continuation Decisions	0.33	Expert Advisory Group
EBDM update magnitude upwards (step)	EBDM Ability	0.083	

EBDM update magnitude downwards (step)	EBDM Ability	0.028	
Monthly intervention evaluation probability	Leadership Review	0.008	

Table S1: List of parameter values used in the model "baseline" condition. Model element and summary-level information about data sources for each values are also presented here.

Data Sources for Baseline Parameters

Survey Data

We utilized data from a national survey of state health department employees the research team conducted in 2018. [48] We also used data from a separate survey of a different sample of state health department employees conducted in 2016.[49] Parameter values were derived as follows:

Agent EBDM: For parameters used to stochastically initialize agents' EBDM ability values, we combined (with equal weighting) four variables from the survey: (1) I am knowledgeable about evidence-based public health processes, (2) I have the skills I need to modify evidence-based interventions from one priority population to another, (3) I have the ability to lead efforts in evidence-based public health in my work unit, and (4) I have the skills to manage program and policy change within my work unit.

Each of the variables was on a 1-5 Likert scale with 1 meaning strongly disagree with having the skill to 5 strongly agree with having the skill. The mean and standard deviation of the resulting combined variable were translated into the 0-1 scale used in the model.

Agent Information Sharing: For parameters used to stochastically initialize agents' information sharing values, we used one variable from each of the survey response datasets at our disposal: (1) I have the skills to effectively communicate the value of evidence-based interventions to leaders in my agency. (responses on a 1-5 Likert scale)[48] and (2) ...My work unit encourages communication and collaboration. (responses on a 1-7 Likert scale)[49]. Means and standard deviations for each variable were translated into the 0-1 scale used into the model and then combined with equal weighting.

Initial Stakeholder Interviews

Based on survey data collected from respondents in 2018, eight state health departments were purposively selected based on reported levels of mis-implementation.[50] Employees (n=45) participated in semi-structured interviews in 2019. Solicited descriptions of active interventions were coded and used to parameterize intervention age parameters.

Supplemental Stakeholder Interviews

Based on input from the core research team and the Expert Advisory Group in meetings held in January 2020, it was determined that additional data were needed to support the agent-based model's parameterization.

The research team members developed an interview guide to obtain these supplemental data. The interview guide's development, testing, and IRB (institutional review board) approval occurred between March and June 2020. The instrument was designed to collect responses representing the perceived evidence political, funders, and stakeholder support for active interventions as well as perceived relationships between age, support, and effectiveness of programs and interventions. In addition, general questions about organizational structure (i.e., numbers of hierarchical levels and employees per supervisor) were included. The data collection happened in July and August 2020 and was conducted by two research team members and one public health consultant. In total, 13 interviews were conducted with chronic disease (former or current) directors/branch managers, one program manager, and one supervisor from state health departments of 13 states in the US. Because interview protocols and questions were designed specifically for parameterization of the ABM, translation into model parameter values was straightforward (e.g., distributions of responses on a 1-10 scale were translated onto a 0-1 scale for use in the model).

Model Calibration Exercise

We used pooled responses to four questions from two surveys: the first, conducted in 2014 (n=1,237) asking (1) In your opinion, how often do programs continue that should have ended? and (2) In your opinion, how often do programs end that should not have ended?, with responses given on a 1-5 Likert scale[51], and the second, conducted in 2018 (n=643) asking (3) How often do ineffective programs, overseen by your work unit, continue when they should have ended? and (4) How often do effective programs overseen by your work unit end when they should have continued?, with responses given on a 1-5 Likert scale[48]. We translated all responses naively from categorical to continuous values (i.e., 1=0, 2=.25, 3=.5, 4=.75, 5=1) for comparison to model output.

We conducted a large sweep of the parameter space of all remaining free parameters. We calculated the rates of continuation and discontinuation mis-implementation for each parameter combination over 50 simulation repetitions, and compared these rates to survey response data.

The calibrated parameter values were those that minimized the distance between the mean model output and survey data in the phase space of the two mis-implementation rates:

$$\sqrt{(\text{model continuation MI rate} - \text{survey continuation MI rate})^2 + (\text{model discontinuation MI rate} - \text{survey discontinuation MI rate})^2}$$

Although we do not present these values in the paper to avoid implying to our audience a more rigorous assessment of model behavior than we were able to conduct, we subjected output to two statistical analyses. The first was an observation of the values that we minimized (i.e., straightforward comparison of means, averaged across repeated model runs). The best fit condition, which we selected as our "baseline," was a mean difference of 0.033. The second used the mean Kolmogorov-Smirnov test statistic comparing the empirical CDFs of the survey and model data (averaged over all runs, then over both MI types); for the baseline condition, this resulted in a KS of 0.34.

Model Experimentation

Experiment Mechanics (main)

We implement several different "experiments," applied alone or in combination to the baseline model described above, that represent broad changes in policies and practices modifying the structure of organizations and behavior of agents and leadership. Effects are described in general here, with specific values explored given in Table S2:

- EBDM boost: multiply the mean of the organization EBDM distribution by a fixed coefficient, with a maximum EBDM mean value of 1. Agent EBDM values are drawn from the distribution and will change on average by a factor of the EBDM multiplier.
- Sharing boost: after agents are generated, multiply the sharing score of every agent below the max level by a fixed coefficient, with a maximum sharing score of 1.
- Altered organization structure distributions: organizations will have taller or wider average structures for differing values of the mean level count and subordinates/supervisor, μ_m and μ_o.
- Continuation function coefficient changes: alter the dependence of the discontinuation probability on top agent perception, intervention age, and stakeholder and funder support.
    - For each of our baseline conditions (original and covid-adjustment), the discontinuation probability of an average intervention (with mean value for all four intervention properties and top agent perception equal to intervention evidence support) is used as an additional internal parameter. We constrain the probability of discontinuation of an average intervention equal to this baseline value. If the age, stakeholder, or funder coefficients change, the perception coefficient is adjusted to maintain this constraint.

Parameter	Baseline value	Policy values			
agent EBDM boost	1	1.1	1.3	1.5	
agent sharing boost	1	1.3	1.5	1.3, top 3 levels	1.5, top 5 levels
organization structure: (mean levels, mean subordinates/supervisor)	(5,5.1)	(6,3.8)	(4,8)	(3,14)	
continuation function age coefficient	-0.025	0			
continuation function stakeholder coefficient	-0.1	0			
continuation function funder coefficient	-0.1	0			

Table S2: The model policy sweep included the cartesian product over all lists of policy values listed in this table.

Additional Results

Main Results (full)

Run Condition	MI, continuation (mean)	MI, continuation (SD)	MI, discontinuation (mean)	MI, discontinuation (SD)
Baseline	0.463	0.108	0.481	0.060
10% EBDM boost	0.455	0.085	0.484	0.060
30% EBDM boost	0.451	0.077	0.480	0.063
50% EBDM boost	0.442	0.097	0.478	0.062
30% sharing boost top 3 levels	0.445	0.088	0.471	0.068
50% sharing boost top 3 levels	0.447	0.109	0.468	0.066
30% sharing boost all agents	0.432	0.096	0.486	0.069
50% sharing boost all agents	0.441	0.098	0.484	0.058
Tall orgs: ~6 levels ~3.8 emp/sup	0.453	0.112	0.472	0.062
Wide orgs 1: ~4 levels ~8 emp/sup	0.443	0.085	0.502	0.076
Wide orgs 2: ~3 levels ~14 emp/sup	0.479	0.107	0.501	0.080
Continuation: no age term	0.253	0.081	0.468	0.054
Continuation: no age or stakeholder terms; perception & funder only	0.143	0.064	0.458	0.058
Continuation: perception term only	0.114	0.061	0.436	0.064

Table S3: Distributions of mis-implementation frequencies (continuation and discontinuation) from run conditions displayed in main text Figure 3.

Synergistic Intervention Effects

Incorporated within our experimental application of our ABM was an assessment of synergistic impacts of multiple approaches. We identified a small set of combined approaches that displayed synergy (i.e., their impact on either type of mis-implementation was greater than the sum of both approaches applied individually). These are summarized in Table S3. The largest synergistic impacts were obtained by applying a change in decision-making processes alongside another approach, especially enhancing organizational EBDM capacity.

MI Type	Change 1	Change 2	MI rate change	Synergy contribution
continuation	Wider org structure 2	Remove Intervention Age from Continuation Decisions	-0.229	-0.035
continuation	Wider org structure 2	50% Information sharing increase (organization-wide)	-0.036	-0.031
continuation	Taller org structure	30% EBDM Increase	-0.033	-0.011
continuation	Taller org structure	10% EBDM multiplier 1.1	-0.025	-0.007
continuation	Wider org structure 1	Management sharing multiplier 1.3	-0.042	-0.004
discontinuation	EBDM multiplier 1.5	Remove Intervention Age and External Stakeholder Support from Continuation Decisions	-0.079	-0.054
discontinuation	EBDM multiplier 1.3	Only Use Perceived Effectiveness in Continuation Decisions	-0.089	-0.044
discontinuation	EBDM multiplier 1.5	Only Use Perceived Effectiveness in Continuation Decisions	-0.088	-0.041
discontinuation	EBDM multiplier 1.3	Remove Intervention Age and External Stakeholder Support from Continuation Decisions	-0.063	-0.039
discontinuation	EBDM multiplier 1.1	Only Use Perceived Effectiveness in Continuation Decisions	-0.058	-0.017
discontinuation	EBDM multiplier 1.1	Remove Intervention Age and External Stakeholder Support from Continuation Decisions	-0.036	-0.016
discontinuation	Taller org structure	All org sharing multiplier 1.3	-0.014	-0.010

discontinuation	EBDM multiplier 1.3	Remove Intervention Age from Continuation Decisions	-0.024	-0.010
discontinuation	EBDM multiplier 1.1	Remove Intervention Age from Continuation Decisions	-0.017	-0.007

Table S4: combinations of experiments that displayed synergy. The type of mis-implementation for which synergy was observed is given in the leftmost column, with total impact on that type of mis-implementation and synergetic effect (relative to estimates of the changes applied alone) given in the rightmost columns, respectively.

Sensitivity Analysis

We performed sensitivity analysis by computing the model on a reduced sweep of the policy space for several additional "sensitivity conditions". We explore different regimes for intervention replacement, fixing agent properties, and an alternate free parameter calibration that produces a "second best" fit with available mis-implementation frequency data. The conditions that fix agent properties, (e.g., "Fixed EBDM") disable the mechanics that allow these values to update during model run; policies that alter these values for the entire model run operate normally.

Sensitivity Condition	Perception update prob., constant term	Intervention replacement prob.	EBDM update steps	Sharing update size
Baseline	0.8	0.33	+ 0.083, - 0.028	0.1
Alternate calibration	**0.5**	0.33	+ 0.083, - 0.028	0.1
Always replace interventions	0.8	**1**	+ 0.083, - 0.028	0.1
Never replace interventions	0.8	**0**	+ 0.083, - 0.028	0.1
Fixed EBDM	0.8	0.33	**+ 0, - 0**	0.1
Fixed info sharing	0.8	0.33	+ 0.083, - 0.028	**0**

Table S5: parameter values differing from baseline (shown in bold) for each of the five conditions included in our sensitivity analysis.

We visualize experiments across sensitivity conditions using a "policy effect" scale in the figure below. This scale is units of fractional change the no-policy MI rate for each condition.

Significantly, the effect of changes to the coefficients of the continuation function remains strong across all sensitivity conditions. Other policies do see deflections in one direction or the other for some conditions.

The "alternate calibration" condition represents only a small perturbation to quantitative model dynamics, and here we see only a modest improvement in the effect of sharing and EBDM multipliers on

MI rate. With that being said, the minor sensitivity of results to variation in a calibrated free parameter suggests that future data collection around these parameters might provide improved—and potentially useful from a policy and practice perspective—model precision. The other sensitivity conditions represent "extreme condition" testing, with key model dynamics and organizational processes set to their maximal or minimal values. Although this fortunately did not result in the model behaving unreasonably (e.g. all interventions being discontinued when up for consideration by leadership), these runs displayed larger variation in policy effects. Taken together, this indicates that our model dynamics are ones that matter quite a bit for our outcome of interest. Thus, further data collection and experimentation around them may be warranted.

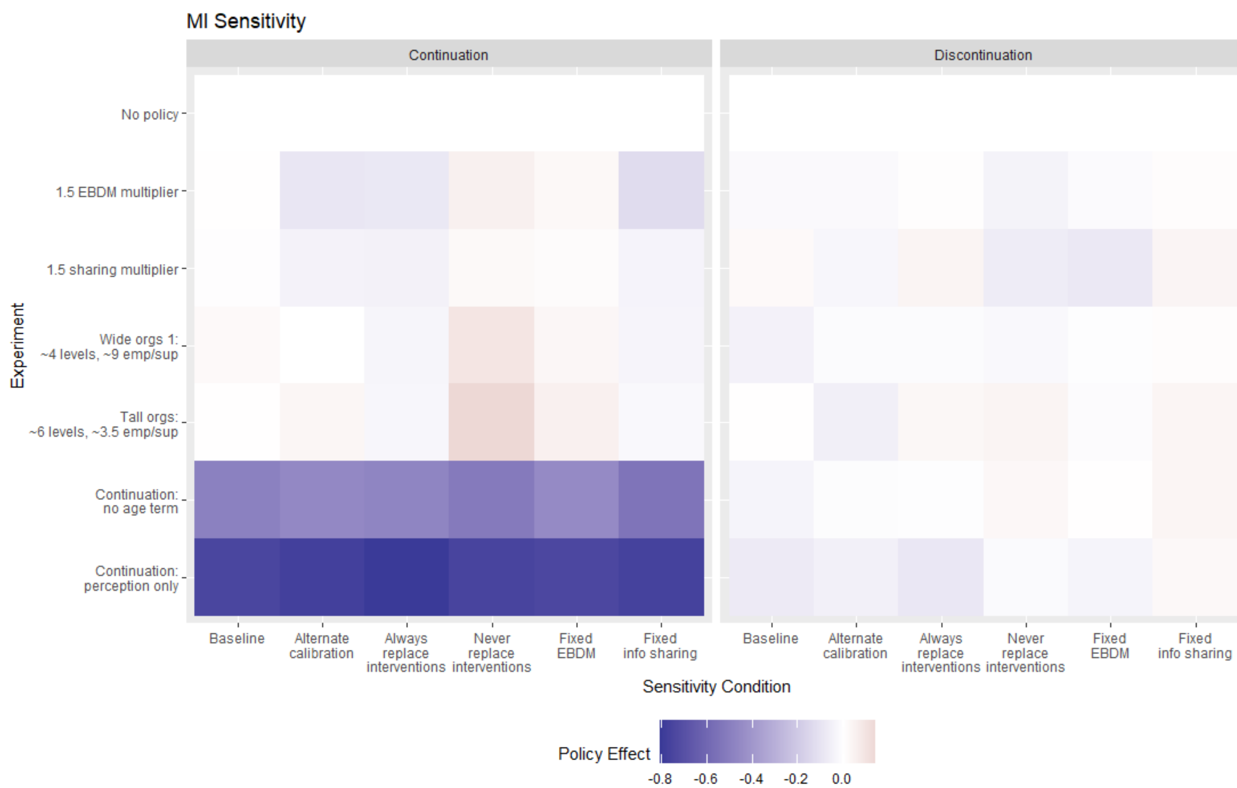

Figure S1: effect of "policy" changes in each experiment relative to baseline (no policy change) under each sensitivity analysis condition.

Additional Model Testing

We performed additional tests of our model to explore boundary adequacy (i.e., selectively removing dynamic pathways) and extreme conditions (i.e., initialization with uniformly maximum and minimum agent attributes. These tests suggested that our model was satisfactorily robust to manipulation. We reproduced main text Figure 3 below (Figure S2) for the sake of reader convenience, and then showed comparable results (Figures S3-S9) for these test conditions.

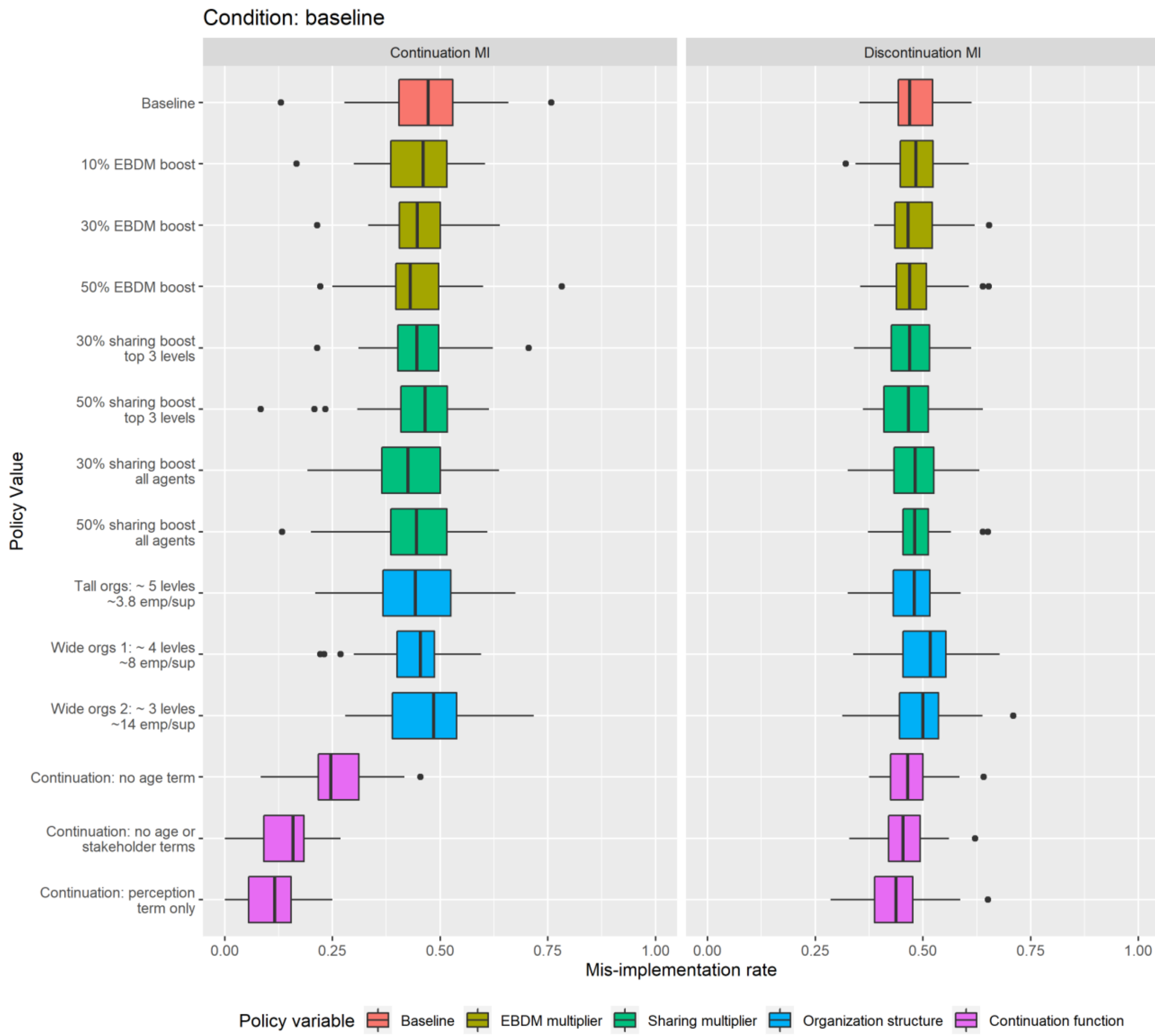

Figure S2: Results from runs depicted in main manuscript Figure 3. Box-plot distributions of mis-implementation frequency (i.e., continuation of ineffective interventions or discontinuation of effective ones, respectively shown in the left and right panels) under the baseline as well as all "single intervention" policy value conditions. This includes "EBDM boost" conditions in which agents are initialized with larger EBDM values, "sharing boost" conditions in which agents are initialized with larger information sharing values, alternative organizational structures in which the organization is initialized such that it is either "wider" or "taller" than in the baseline condition, and conditions in which leadership utilizes different strategies for making continuation decisions. Median values are shown as vertical lines, the 25th and 75th percentile values as left and right box edges, 95% confidence intervals as horizontal lines, and outlier values as dots. Frequency values are shown on the x-axis and the sole deviation from the baseline condition noted on y-axis (other than the baseline itself). For ease of comparison, the baseline median is presented as a dashed line.

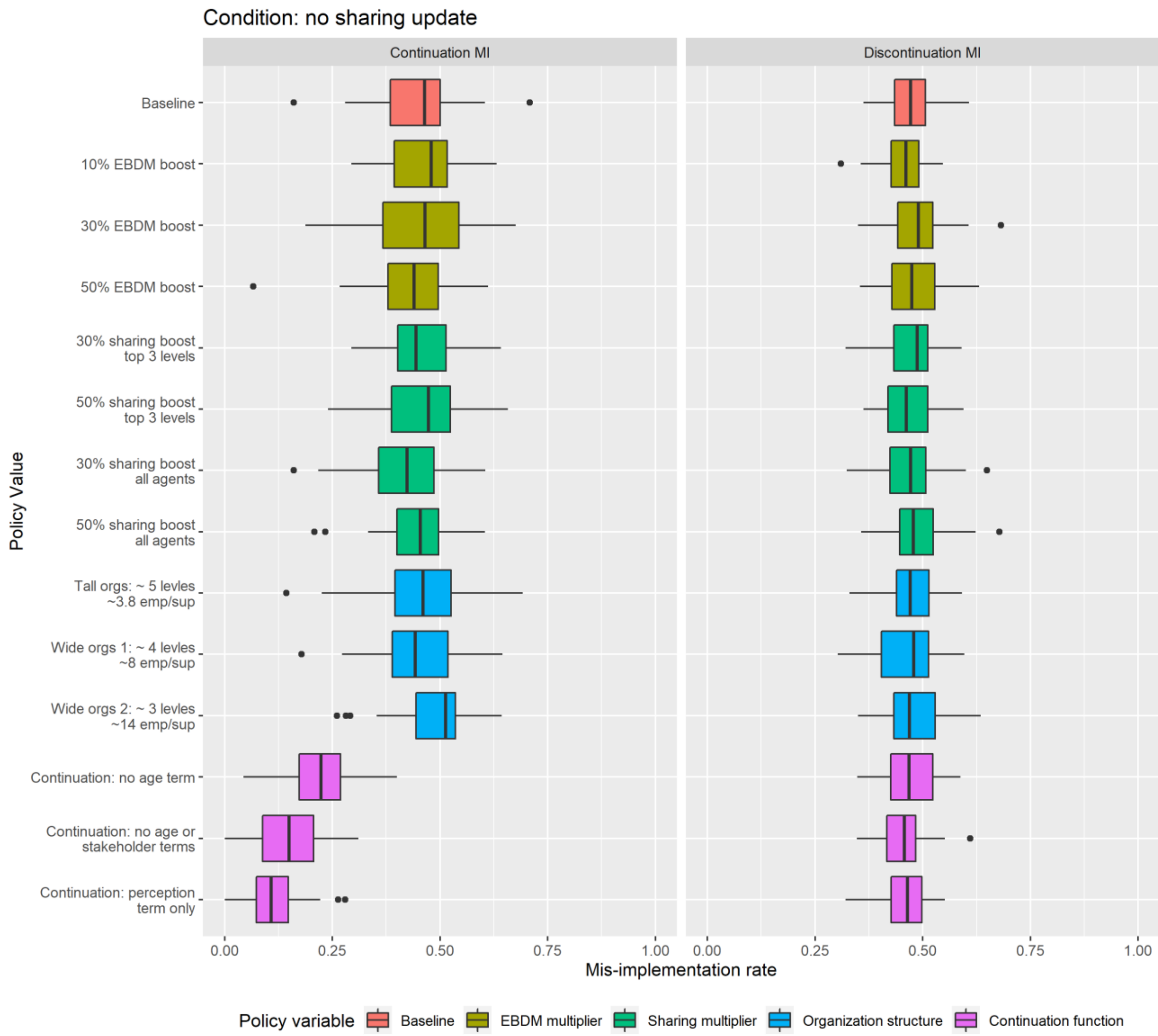

Figure S3: Results from runs with information sharing propensity updating dynamics turned off. Box-plot distributions of mis-implementation frequency (i.e., continuation of ineffective interventions or discontinuation of effective ones, respectively shown in the left and right panels) under the baseline as well as all “single intervention” policy value conditions. This includes “EBDM boost” conditions in which agents are initialized with larger EBDM values, “sharing boost” conditions in which agents are initialized with larger information sharing values, alternative organizational structures in which the organization is initialized such that it is either “wider” or “taller” than in the baseline condition, and conditions in which leadership utilizes different strategies for making continuation decisions. Median values are shown as vertical lines, the 25th and 75th percentile values as left and right box edges, 95% confidence intervals as horizontal lines, and outlier values as dots. Frequency values are shown on the x-axis and the sole deviation from the baseline condition noted on y-axis (other than the baseline itself). For ease of comparison, the baseline median is presented as a dashed line.

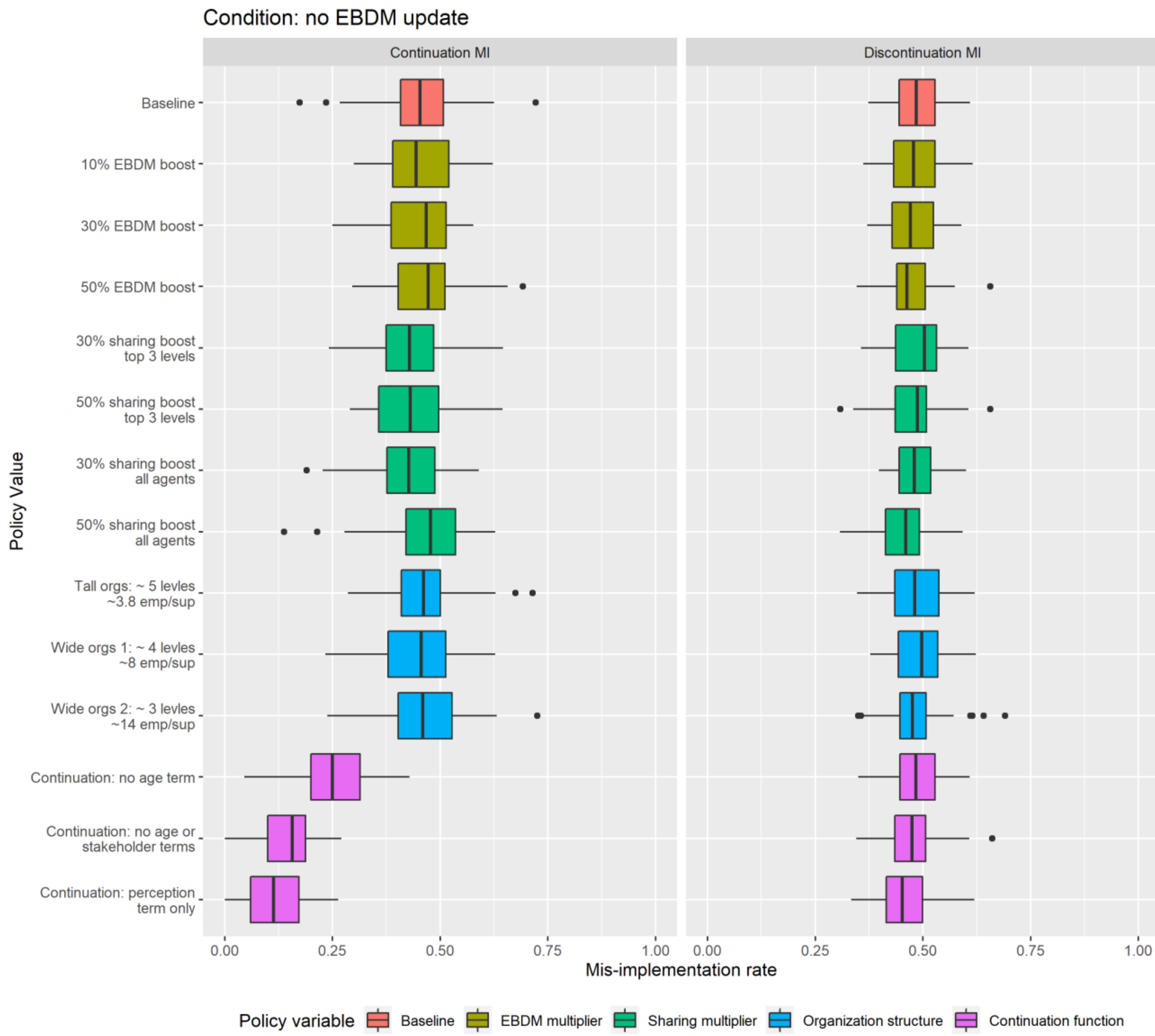

Figure S4: Results from runs with EBDM updating dynamics turned off. Box-plot distributions of mis-implementation frequency (i.e., continuation of ineffective interventions or discontinuation of effective ones, respectively shown in the left and right panels) under the baseline as well as all “single intervention” policy value conditions. This includes “EBDM boost” conditions in which agents are initialized with larger EBDM values, “sharing boost” conditions in which agents are initialized with larger information sharing values, alternative organizational structures in which the organization is initialized such that it is either “wider” or “taller” than in the baseline condition, and conditions in which leadership utilizes different strategies for making continuation decisions. Median values are shown as vertical lines, the 25th and 75th percentile values as left and right box edges, 95% confidence intervals as horizontal lines, and outlier values as dots. Frequency values are shown on the x-axis and the sole deviation from the baseline condition noted on y-axis (other than the baseline itself). For ease of comparison, the baseline median is presented as a dashed line.

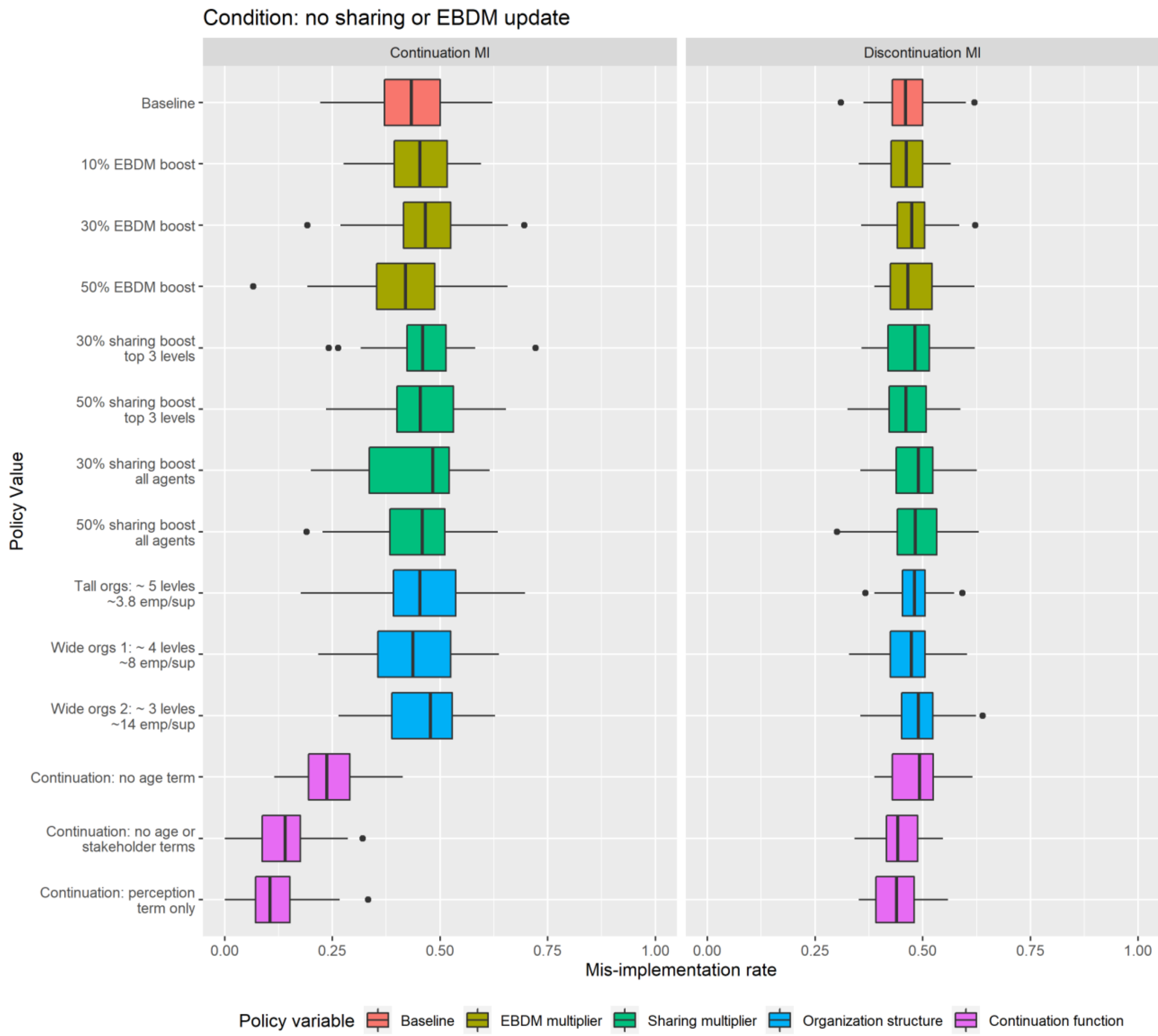

Figure S5: Results from runs with information sharing propensity and EBDM update dynamics turned off. Box-plot distributions of mis-implementation frequency (i.e., continuation of ineffective interventions or discontinuation of effective ones, respectively shown in the left and right panels) under the baseline as well as all “single intervention” policy value conditions. This includes “EBDM boost” conditions in which agents are initialized with larger EBDM values, “sharing boost” conditions in which agents are initialized with larger information sharing values, alternative organizational structures in which the organization is initialized such that it is either “wider” or “taller” than in the baseline condition, and conditions in which leadership utilizes different strategies for making continuation decisions. Median values are shown as vertical lines, the 25th and 75th percentile values as left and right box edges, 95% confidence intervals as horizontal lines, and outlier values as dots. Frequency values are shown on the x-axis and the sole deviation from the baseline condition noted on y-axis (other than the baseline itself). For ease of comparison, the baseline median is presented as a dashed line.

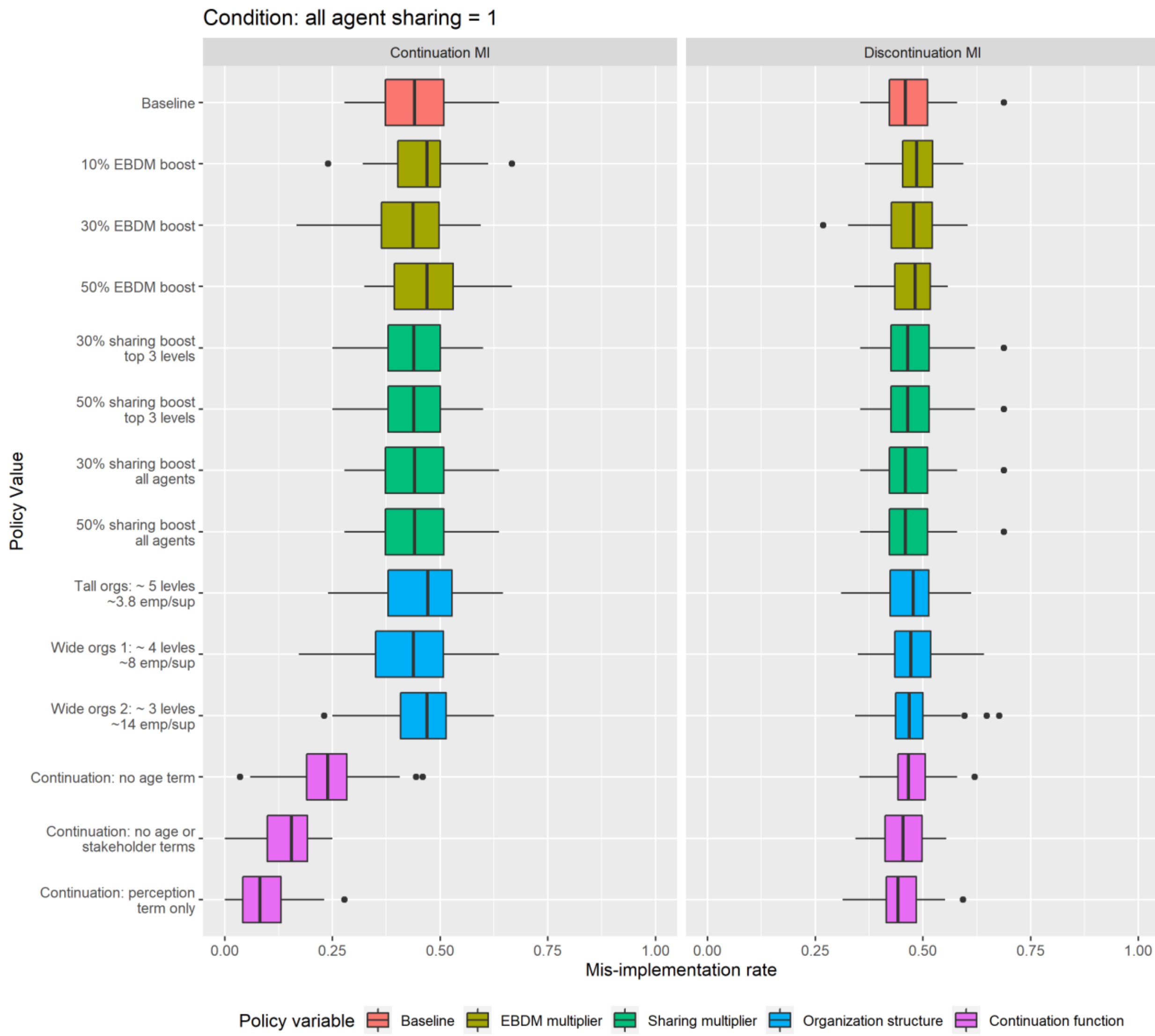

Figure S6: Results from runs with maximum information-sharing propensity at initialization. Box-plot distributions of mis-implementation frequency (i.e., continuation of ineffective interventions or discontinuation of effective ones, respectively shown in the left and right panels) under the baseline as well as all “single intervention” policy value conditions. This includes “EBDM boost” conditions in which agents are initialized with larger EBDM values, “sharing boost” conditions in which agents are initialized with larger information sharing values, alternative organizational structures in which the organization is initialized such that it is either “wider” or “taller” than in the baseline condition, and conditions in which leadership utilizes different strategies for making continuation decisions. Median values are shown as vertical lines, the 25th and 75th percentile values as left and right box edges, 95% confidence intervals as horizontal lines, and outlier values as dots. Frequency values are shown on the x-axis and the sole deviation from the baseline condition noted on y-axis (other than the baseline itself). For ease of comparison, the baseline median is presented as a dashed line.

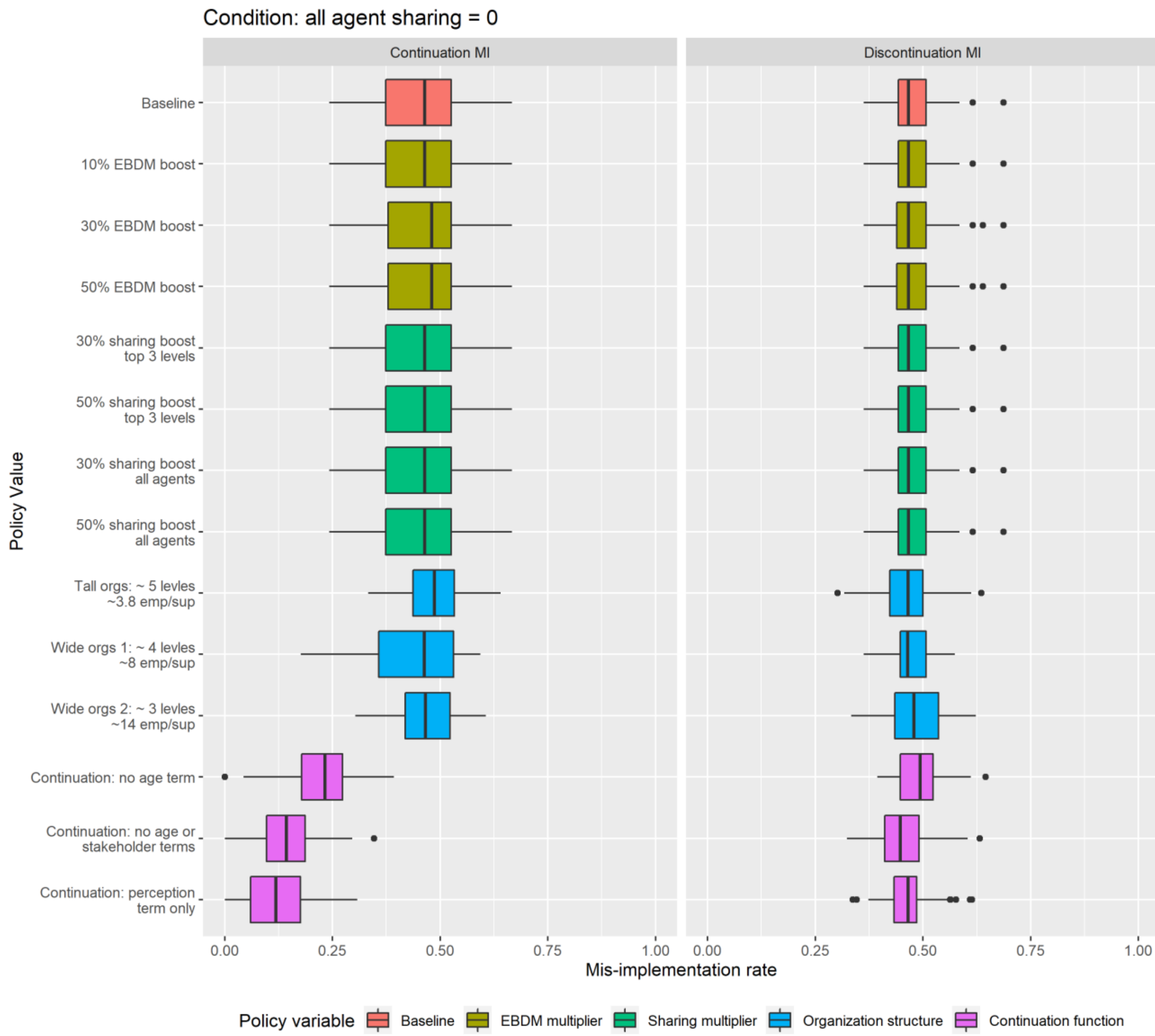

Figure S7: Results from runs with minimum information-sharing propensity at initialization. Box-plot distributions of mis-implementation frequency (i.e., continuation of ineffective interventions or discontinuation of effective ones, respectively shown in the left and right panels) under the baseline as well as all “single intervention” policy value conditions. This includes “EBDM boost” conditions in which agents are initialized with larger EBDM values, “sharing boost” conditions in which agents are initialized with larger information sharing values, alternative organizational structures in which the organization is initialized such that it is either “wider” or “taller” than in the baseline condition, and conditions in which leadership utilizes different strategies for making continuation decisions. Median values are shown as vertical lines, the 25th and 75th percentile values as left and right box edges, 95% confidence intervals as horizontal lines, and outlier values as dots. Frequency values are shown on the x-axis and the sole deviation from the baseline condition noted on y-axis (other than the baseline itself). For ease of comparison, the baseline median is presented as a dashed line.

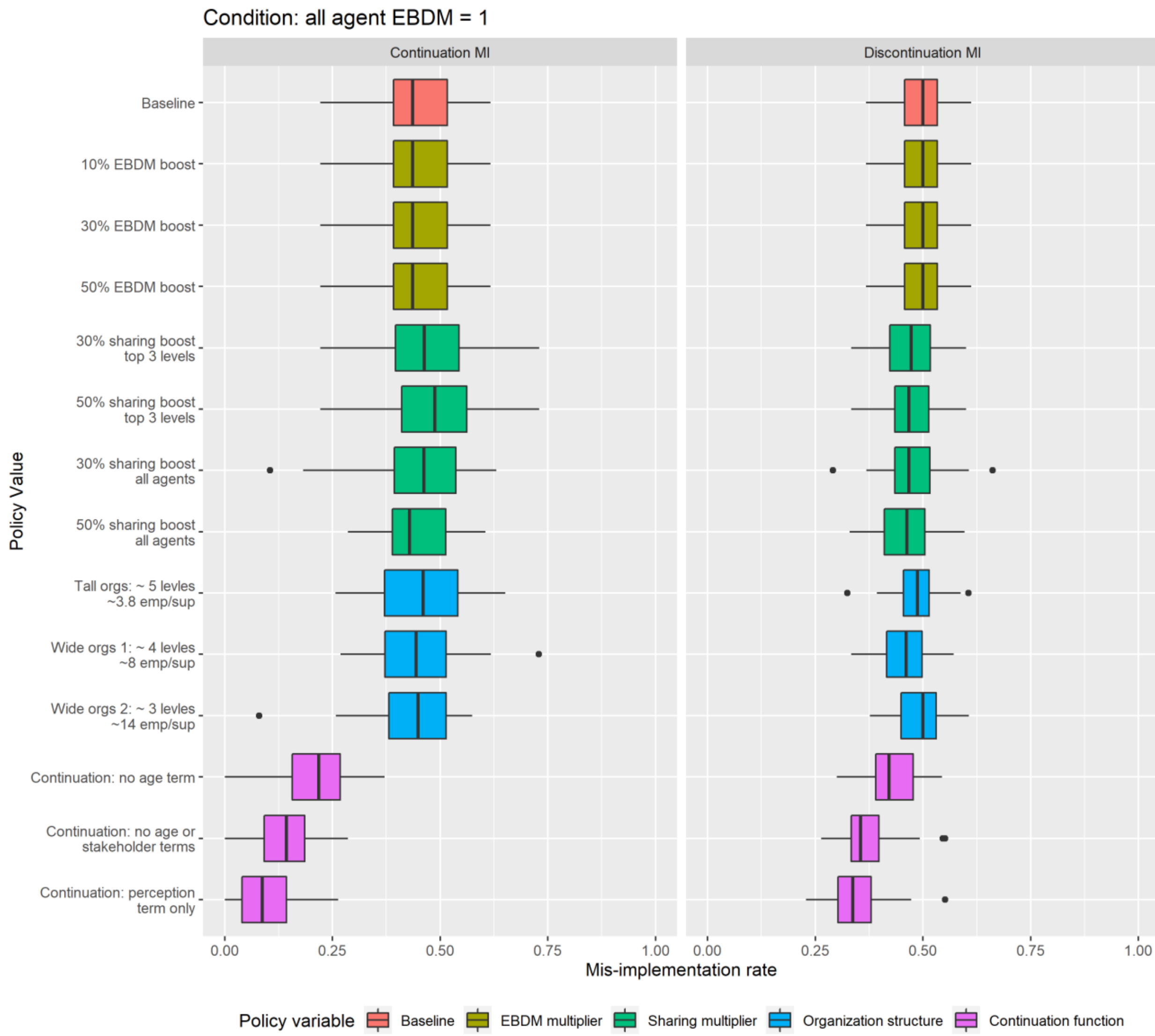

Figure S8: Results from runs with maximum information-sharing propensity at initialization. Box-plot distributions of mis-implementation frequency (i.e., continuation of ineffective interventions or discontinuation of effective ones, respectively shown in the left and right panels) under the baseline as well as all “single intervention” policy value conditions. This includes “EBDM boost” conditions in which agents are initialized with larger EBDM values, “sharing boost” conditions in which agents are initialized with larger information sharing values, alternative organizational structures in which the organization is initialized such that it is either “wider” or “taller” than in the baseline condition, and conditions in which leadership utilizes different strategies for making continuation decisions. Median values are shown as vertical lines, the 25th and 75th percentile values as left and right box edges, 95% confidence intervals as horizontal lines, and outlier values as dots. Frequency values are shown on the x-axis and the sole deviation from the baseline condition noted on y-axis (other than the baseline itself). For ease of comparison, the baseline median is presented as a dashed line.

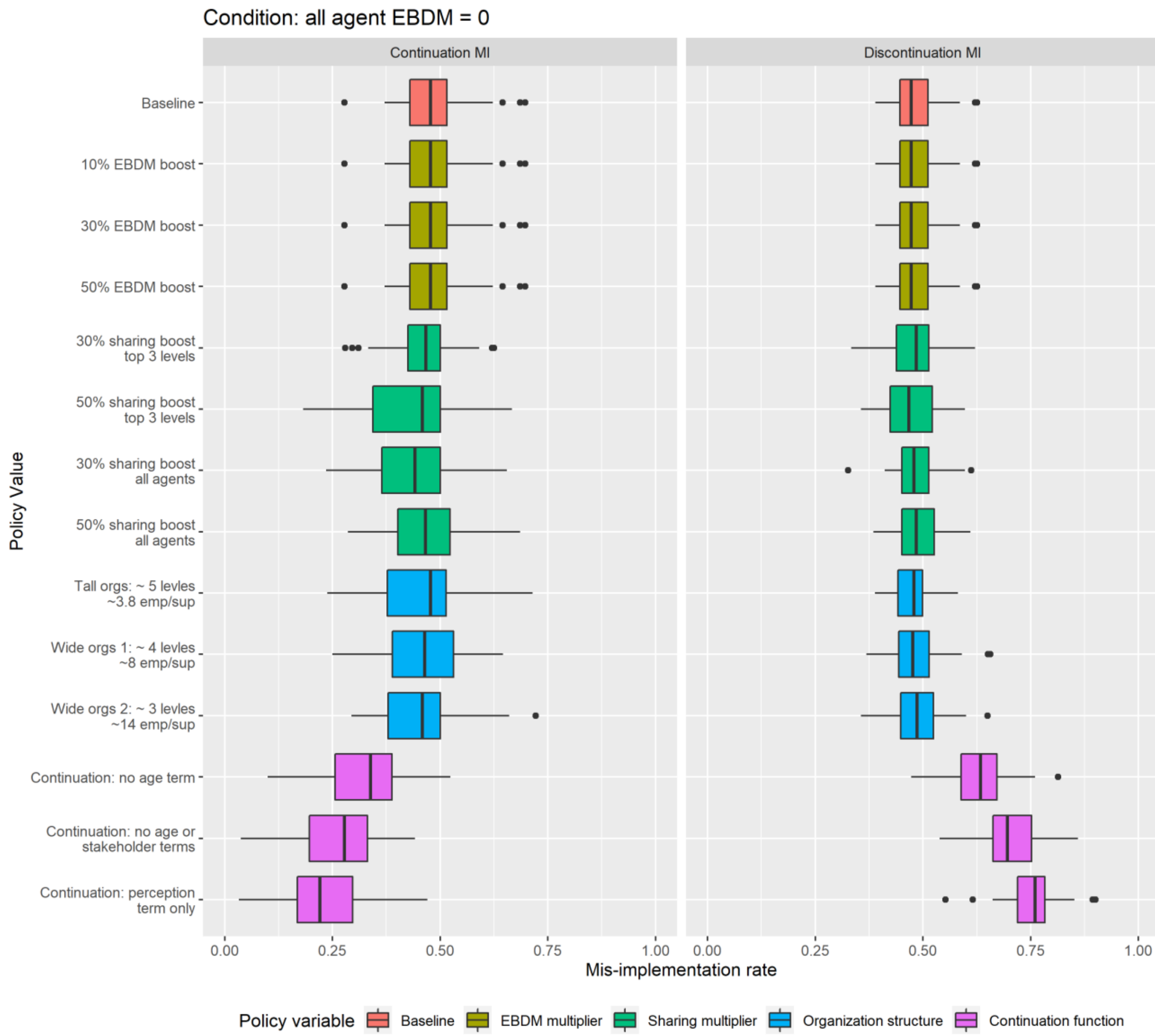

Figure S9: Results from runs with minimum EBDM capacity at initialization. Box-plot distributions of mis-implementation frequency (i.e., continuation of ineffective interventions or discontinuation of effective ones, respectively shown in the left and right panels) under the baseline as well as all “single intervention” policy value conditions. This includes “EBDM boost” conditions in which agents are initialized with larger EBDM values, “sharing boost” conditions in which agents are initialized with larger information sharing values, alternative organizational structures in which the organization is initialized such that it is either “wider” or “taller” than in the baseline condition, and conditions in which leadership utilizes different strategies for making continuation decisions. Median values are shown as vertical lines, the 25th and 75th percentile values as left and right box edges, 95% confidence intervals as horizontal lines, and outlier values as dots. Frequency values are shown on the x-axis and the sole deviation from the baseline condition noted on y-axis (other than the baseline itself). For ease of comparison, the baseline median is presented as a dashed line.

References

1. Thomas IM, Reilly SR. Group Model Building. *J Public Health Manag Pract*. 2015;21:S79-S83. doi:10.1097/PHH.0000000000000209

2. Kasman M, Owen B, Hayward J. A community-based complex systems approach to high school completion. *Syst Res Behav Sci*. 2017;34(3):267-276. doi:10.1002/sres.2409

3. Hovmand PS, Andersen DF, Rouwette E, Richardson GP, Rux K, Calhoun A. Group model-building 'scripts' as a collaborative planning tool. *Syst Res Behav Sci*. 2012;29(2):179-193. doi:10.1002/sres.2105

4. Epstein JM. *Generative Social Science*. Princeton University Press; 2012.

5. Grant AM, Gino F, Hofmann DA. Reversing the extraverted leadership advantage: The role of employee proactivity. *Acad Manage J*. 2011;54(3):528-550. doi:10.5465/amj.2011.61968043

6. See KE, Morrison EW, Rothman NB, Soll JB. The detrimental effects of power on confidence, advice taking, and accuracy. *Organ Behav Hum Decis Process*. 2011;116(2):272-285. doi:10.1016/j.obhdp.2011.07.006

7. Grant AM. Rocking the boat but keeping it steady: The role of emotion regulation in employee voice. *Acad Manage J*. 2013;56(6):1703-1723. doi:10.5465/amj.2011.0035

8. Burris ER, Rockmann KW, Kimmons YS. The value of voice to managers: Employee identification and the content of voice. *Acad Manage J*. 2017;60(6):2099-2125. doi:10.5465/amj.2014.0320

9. Garner JT. Open doors and iron cages: Supervisors' responses to employee dissent. *Int J Bus Commun*. 2016;53(1):27-54. doi:10.1177/2329488414525466

10. Garner JT. Making waves at work: Perceived effectiveness and appropriateness of organizational dissent messages. *Manag Commun Q*. 2012;26(2):224-240. doi:10.1177/0893318911431803

11. Kish-Gephart JJ, Detert JR, Treviño LK, Edmondson AC. Silenced by fear: The nature, sources, and consequences of fear at work. *Res Organ Behav*. 2009;29:163-193. doi:10.1016/j.riob.2009.07.002

12. Harvey P, Martinko MJ, Douglas SC. Causal perceptions and the decision to speak up or pipe down. In: *Voice and Silence in Organizations*. Emerald Bingley; 2009:63-82.

13. Edwards MS, Ashkanasy NM, Gardner J. Deciding to speak up or to remain silent following observed wrongdoing: The role of discrete emotions and climate of silence. In: *Voice and Silence in Organizations*. Emerald Group Publishing Bingley, UK; 2009:83-109.

14. Detert JR, Edmondson AC. Implicit voice theories: Taken-for-granted rules of self-censorship at work. *Acad Manage J*. 2011;54(3):461-488. doi:10.5465/AMJ.2011.61967925

15. Fast NJ, Burris ER, Bartel CA. Managing to stay in the dark: Managerial self-efficacy, ego defensiveness, and the aversion to employee voice. *Acad Manage J*. 2014;57(4):1013-1034. doi:10.5465/amj.2012.0393

16. Botero IC, Van Dyne L. Employee voice behavior: Interactive effects of LMX and power distance in the United States and Colombia. *Manag Commun Q*. 2009;23(1):84-104. doi:10.1177/0893318909335415

17. Takeuchi R, Chen Z, Cheung SY. Applying uncertainty management theory to employee voice behavior: An integrative investigation. *Pers Psychol*. 2012;65(2):283-323. doi:10.1111/j.1744-6570.2012.01247.x

18. Edmondson AC. Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *J Manag Stud*. 2003;40(6):1419-1452. doi:10.1111/1467-6486.00386

19. Stamper CL, Dyne LV. Work status and organizational citizenship behavior: A field study of restaurant employees. *J Organ Behav Int J Ind Occup Organ Psychol Behav*. 2001;22(5):517-536. doi:10.1002/job.100

20. Ashford SJ, Rothbard NP, Piderit SK, Dutton JE. Out on a limb: The role of context and impression management in selling gender-equity issues. *Adm Sci Q*. 1998;43(1):23-57. doi:10.2307/2393590

21. Detert JR, Burris ER. Leadership behavior and employee voice: Is the door really open? *Acad Manage J*. 2007;50(4):869-884. doi:10.5465/amj.2007.26279183

22. Naus F, Van Iterson A, Roe R. Organizational cynicism: Extending the exit, voice, loyalty, and neglect model of employees' responses to adverse conditions in the workplace. *Hum Relat*. 2007;60(5):683-718. doi:10.1177/0018726707079198

23. LePine JA, Van Dyne L. Voice and cooperative behavior as contrasting forms of contextual performance: evidence of differential relationships with big five personality characteristics and cognitive ability. *J Appl Psychol*. 2001;86(2):326. doi:10.1037/0021-9010.86.2.326

24. Crant JM, Kim TY, Wang J. Dispositional antecedents of demonstration and usefulness of voice behavior. *J Bus Psychol*. 2011;26(3):285-297. doi:10.1007/s10869-010-9197-y

25. Venkataramani V, Tangirala S. When and why do central employees speak up? An examination of mediating and moderating variables. *J Appl Psychol*. 2010;95(3):582. doi:10.1037/a0018315

26. Tangirala S, Ramanujam R. Ask and you shall hear (but not always): Examining the relationship between manager consultation and employee voice. *Pers Psychol*. 2012;65(2):251-282. doi:10.1111/j.1744-6570.2012.01248.x

27. Kassing JW. Speaking up: Identifying employees' upward dissent strategies. *Manag Commun Q*. 2002;16(2):187-209. doi:10.1177/089331802237234

28. Frazier ML, Fainshmidt S. Voice climate, work outcomes, and the mediating role of psychological empowerment: A multilevel examination. *Group Organ Manag*. 2012;37(6):691-715. doi:10.1177/1059601112463960

29. Morrison EW, Wheeler-Smith SL, Kamdar D. Speaking up in groups: a cross-level study of group voice climate and voice. *J Appl Psychol*. 2011;96(1):183. doi:10.1037/a0020744

30. Miceli MP, Near JP. *Blowing the Whistle: The Organizational and Legal Implications for Companies and Employees*. Lexington Books; 1992.

31. Tangirala S, Kamdar D, Venkataramani V, Parke MR. Doing right versus getting ahead: The effects of duty and achievement orientations on employees' voice. *J Appl Psychol*. 2013;98(6):1040. doi:10.1037/a0033855

32. Nikolaou I, Vakola M, Bourantas D. Who speaks up at work? Dispositional influences on employees' voice behavior. *Pers Rev*. 2008. doi:10.1108/00483480810906892

33. Tangirala S, Ramanujam R. Exploring nonlinearity in employee voice: The effects of personal control and organizational identification. *Acad Manage J*. 2008;51(6):1189-1203. doi:10.5465/amj.2008.35732719

34. Tangirala S, Ramanujam R. Employee silence on critical work issues: The cross level effects of procedural justice climate. *Pers Psychol*. 2008;61(1):37-68. doi:10.1111/j.1744-6570.2008.00105.x

35. Liu W, Zhu R, Yang Y. I warn you because I like you: Voice behavior, employee identifications, and transformational leadership. *Leadersh Q*. 2010;21(1):189-202. doi:10.1016/j.leaqua.2009.10.014

36. Liang J, Farh CI, Farh JL. Psychological antecedents of promotive and prohibitive voice: A two-wave examination. *Acad Manage J*. 2012;55(1):71-92. doi:10.5465/amj.2010.0176

37. Fuller JB, Marler LE, Hester K. Promoting felt responsibility for constructive change and proactive behavior: Exploring aspects of an elaborated model of work design. *J Organ Behav Int J Ind Occup Organ Psychol Behav*. 2006;27(8):1089-1120. doi:10.1002/job.408

38. Souba W, Way D, Lucey C, Sedmak D, Notestine M. Elephants in academic medicine. *Acad Med*. 2011;86(12):1492-1499. doi:10.1097/ACM.0b013e3182356559

39. Detert JR, Treviño LK. Speaking up to higher-ups: How supervisors and skip-level leaders influence employee voice. *Organ Sci*. 2010;21(1):249-270. doi:10.1287/orsc.1080.0405

40. Milliken FJ, Morrison EW, Hewlin PF. An exploratory study of employee silence: Issues that employees don't communicate upward and why. *J Manag Stud*. 2003;40(6):1453-1476. doi:10.1111/1467-6486.00387

41. Morrison EW. Employee voice and silence. *Annu Rev Organ Psychol Organ Behav*. 2014;1(1):173-197. doi:10.1146/annurev-orgpsych-031413-091328

42. Schulz-Hardt S, Brodbeck F, Mojzisch A, Kerschreiter R, Frey D. Group Decision Making in Hidden Profile Situations: Dissent as a Facilitator for Decision Quality. *J Pers Soc Psychol*. 2007;91:1080-1093. doi:10.1037/0022-3514.91.6.1080

43. McKinney W. Data Structures for Statistical Computing in Python. In: Walt S van der, Millman J, eds. *Proceedings of the 9th Python in Science Conference*. ; 2010:56-61. doi:10.25080/Majora-92bf1922-00a

44. Kazil J, Masad D, Crooks A. Utilizing Python for Agent-Based Modeling: The Mesa Framework. In: Thomson R, Bisgin H, Dancy C, Hyder A, Hussain M, eds. *Social, Cultural, and Behavioral Modeling*. Cham: Springer International Publishing; 2020:308-317.

45. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357-362. doi:10.1038/s41586-020-2649-2

46. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261-272. doi:10.1038/s41592-019-0686-2

47. The pandas development team. *Pandas-Dev/Pandas: Pandas*. Zenodo; 2020. doi:10.5281/zenodo.3509134

48. Padek MM, Mazzucca S, Allen P, et al. Patterns and correlates of mis-implementation in state chronic disease public health practice in the United States. *BMC Public Health*. 2021;21(1):1-11. doi:10.1186/s12889-020-10101-z

49. Eyler AA, Valko C, Ramadas R, Macchi M, Fershteyn Z, Brownson RC. Administrative Evidence-based Practices in State Chronic Disease Practitioners. *Am J Prev Med*. 2018;54(2):275-283. doi:10.1016/j.amepre.2017.09.006

50. Padek M, Allen P, Erwin PC, et al. Toward optimal implementation of cancer prevention and control programs in public health: a study protocol on mis-implementation. *Implement Sci*. 2018;13(1):1-13. doi:10.1186/s13012-018-0742-9

51. Brownson RC, Allen P, Jacob RR, et al. Understanding mis-implementation in public health practice. *Am J Prev Med*. 2015;48(5):543-551. doi:10.1016/j.amepre.2014.11.015