Beyond Neighbor Influence: A Behavior-Driven Agent-Based Model of Silence

Ziqian Shi
Grupo de Sistemas Complejos,
Universidad Politécnica de Madrid,
Madrid, Spain
ziqian.shi@alumnos.upm.es

Rosa María Benito Zafrilla
Grupo de Sistemas Complejos,
Universidad Politécnica de Madrid,
Madrid, Spain
rosamaria.benito@upm.es

ABSTRACT

Silence is a complex social behavior, yet existing agent-based models (ABMs) predominantly treat it as neighbor-driven and rely on manual parameter tuning, leaving the agent's internal decision-making chain under-explored. Our study proposes a behavior-driven ABM in which each agent undergoes a three-stage decision-making chain: Dual Perception, Consistency Judgment, and Expression Decision. This process ultimately leads to a behavioral choice among silence, expression, and swing. To operationalize this chain while avoiding manual tuning, we first construct a personality-environment configuration, defined along the mental and material dimensions, which consists of agent's value orientations and the reward-punishment mechanism projected from the system outcome. Each agent is initially assigned a random configuration, then applies Dual Perception to calculate a consistency score J, which reflects the alignment between its own opinion and the system outcome. We finally introduce an expression function $N(J)$ as a probabilistic generator of behavioral choices, where each agent's probability distribution is shaped by its specific configuration.

Our model extends silence mechanisms by incorporating satisfaction-based silence alongside fear-based silence and embeds nonlinear features that trigger abrupt behavioral reversals or irreversible silence under severe conditions. Across extensive simulations (each run with 10^8 agents) under both randomized and controlled settings, the model demonstrates statistical and structural stability, showing that silence is mainly driven by agent behavior rather than neighbor influence. In summary, we introduce a behavior-driven baseline model that provides a new perspective for studying silence, and captures key patterns including the spiral of silence, acquiescent silence, and satisfaction-induced silence. It also offers potential applications in opinion dynamics and other complex social settings.

KEYWORDS

Agent-Based Model, Social Simulation, Silence, Opinion Dynamics, Decision-Making Chain

ACM Reference Format:

Ziqian Shi and Rosa María Benito Zafrilla. 2026. Beyond Neighbor Influence: A Behavior-Driven Agent-Based Model of Silence. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 8 pages. https://doi.org/10.65109/UPPI5986

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/UPPI5986

1 INTRODUCTION

Silence has attracted extensive attention across multiple fields, including political communication, social psychology, and behavioral science, as a common yet complex form of social behavior [10, 15, 19, 23]. From everyday interactions to contemporary online platforms, people do not always speak out; instead, choosing silence is common.

Research on silence has largely been grounded in the spiral of silence (SoS) theory, focusing on theoretical and empirical analyses of individual motivations (e.g., fear, conformity) or societal consequences (e.g., biased perception, opinion imbalance) [1, 5, 6, 16, 17, 21]. However, mainstream perspectives tend to overemphasize neighbor influence while paying little attention to the behavioral chain that individuals follow from cognition to decision-making. As a result, existing agent-based models (ABMs) face two main limitations: (1) silence is primarily modeled as a neighbor-driven outcome, with little attempt to systematically model the full agent behavioral chain as the dominant mechanism, and (2) models often rely heavily on manually tuned parameters [2, 3, 20, 22, 24, 25].

To address these problems, we design an ABM simulation free of manual parameter tuning, in which the agent decision-making chain serves as the primary driver and neighbor influence plays only a secondary role, thereby overcoming these modeling limitations. Specifically, the model assigns each agent a three-step decision-making chain: Value Perception, Consistency Judgment and Expression Decision. In this process, each agent evaluates the degree of consistency between its own opinion and the aggregated system outcome, and ultimately chooses among silence, expression, or swing. The system outcome is defined as an abstract result of the collective aggregation of agents' opinions, external influences and mutual interactions. It is projected back into the environment as a unified reward-punishment mechanism, enabling agents to perceive it within their decision-making chain. This design ensures that neighbor influence enters only indirectly and remains non-dominant.

Using the constructed model for large-scale simulations, our main goal is to demonstrate that a formally specified ABM, free of manual parameter tuning and driven by the agent's decision-making chain, can produce a pattern consistent with the stylized SoS phenomenon. Achieving this goal not only provides technical validation but also supports a broader claim, that agent decision-making may serve as the primary driver of silence.

The contribution of this study is threefold. First, it addresses the limitations of existing silence ABMs that depend on neighbor influence and manual parameter tuning, by providing a behavior-driven baseline model grounded in the full decision-making chain of agents. Second, it introduces a consistency score that links agent value orientations with the reward-punishment mechanism projected from the system outcome, providing a new lens to interpret silence. In addition, an expression function incorporates satisfaction-based silence alongside fear-based silence, enabling a single model to capture multiple key silence patterns. Third, it embeds a fixed-point sampling method into the behavioral chain, offering a potential tool for assessing behavioral probabilities under given societal conditions.

The remainder of this paper is organized as follows. In Section 2, we describe the model and simulation setup. In Section 3, we present the simulation results. In Section 4, we provide discussion of theoretical and applied implications. In Section 5, we present the main conclusions of this paper.

2 METHODOLOGY

2.1 Conceptual Foundations and Projection Rule

To construct a behavior-driven ABM, we treat silence as an agent's behavioral decision outcome, together with expression and swing, grounded in the agent's subjective judgment of the consistency between their opinion and the system outcome. The key concepts are defined as follows:

- **Agent:** The smallest behavioral unit situated within a specific social environment. Each agent maintains a fixed opinion toward a given event, but the specific behavioral choice (silence, expression, or swing) remains uncertain.
- **System Outcome:** An abstract collective state that emerges from the aggregation of all participating agents who maintain personal opinions toward a given event within a specific social environment, together with external influences and mutual interactions. This state may take two typical forms:
  - **Type I (Opinion outcome):** the convergence of all agents' opinions combined with environmental and neighbor influences.
  - **Type II (Expressive outcome):** the outcome shaped by opinions of expressing agents during the process, along with the corresponding external and neighbor influences present during the process.

These two types of system outcomes typically occur in sequence in real-world scenarios. For instance, in the U.S. political environment, the election event can be observed in two stages: during the campaigning period, the support configuration reflected only by expressive voters corresponds to type II; once the election results are announced, the final collective outcome corresponds to type I.

In our model, the system outcome is defined as an abstract collective state that cannot be directly observed by agents. To handle this, we project this outcome into the social environment as a unified reward-punishment mechanism, enabling agents to perceive it. Specifically, inspired by Maslow's hierarchy of needs [14], we consider individual motives as rooted in an inner pursuit orientation and represented by two dimensions: mental and material. Therefore, we introduce these two dimensions, which jointly represent each agent's value orientation and align with the reward-punishment mechanism. Neighbor influence is embedded in this mechanism and manifested as part of the mental or material rewards or punishments perceived by agents.

These structural choices are designed to support the agent's decision-making chain, which drives the model's dynamics (Figure 1). In particular, the chain consists of three stages: Dual Perception, Consistency Judgment and Expression Decision. The next section introduces this process in detail.

2.2 Description of the Agent Decision-Making Chain

In the decision-making chain, the first step for each agent is Dual Perception. At this stage, the agent receives the reward-punishment signals from the social environment along two dimensions, mental and material, but interprets them through its own value orientation. As a result, even under identical external influences, agents may perceive the situation very differently. This individualized perception process highlights the dominant nature of agent behavior.

The second stage, Consistency Judgment, transforms the agent's Dual Perception into a subjective consistency score of whether the current system outcome aligns with its opinion. If an agent perceives the score as negative, according to its own value orientation, it is likely to experience dissatisfaction or even fear, signaling a perceived misalignment with the system outcome. Conversely, when the score is positive, the agent feels satisfied, a psychological response that arises when its personal values are affirmed by the social environment. In such cases, the agent interprets the system outcome as consistent with its values.

The final stage is Expression Decision, where the agent chooses a behavior among silence, expression or swing. This decision is primarily guided by Consistency Judgment and is further modulated by the tolerance level of the social environment, which enhances the realism of the model. More precisely, when an agent perceives a very high level of consistency, it tends to remain silent because expression is no longer necessary [9, 18]. When it perceives a very high level of inconsistency, it also tends to remain silent, as expression is unlikely to change the outcome [17]. In contrast, when the perceived consistency score falls into a middle range, the agent considers expression potentially influential in shifting the system outcome and thus becomes more inclined to speak.

Through these designs, the model establishes a complete three-stage decision-making chain, whereby each agent undergoes Dual Perception, Consistency Judgment, and Expression Decision for a given event within a specific social environment. Based on this structure, our model simulates the behavioral regularities of agent silence under different levels of consistency between personal opinions and the system outcome.

2.3 Model Implementation

This section provides the description of the mathematical model implementation corresponding to the decision-making chain. The model is structured around three core components: (1) the personality-environment configuration, which defines each agent's unique

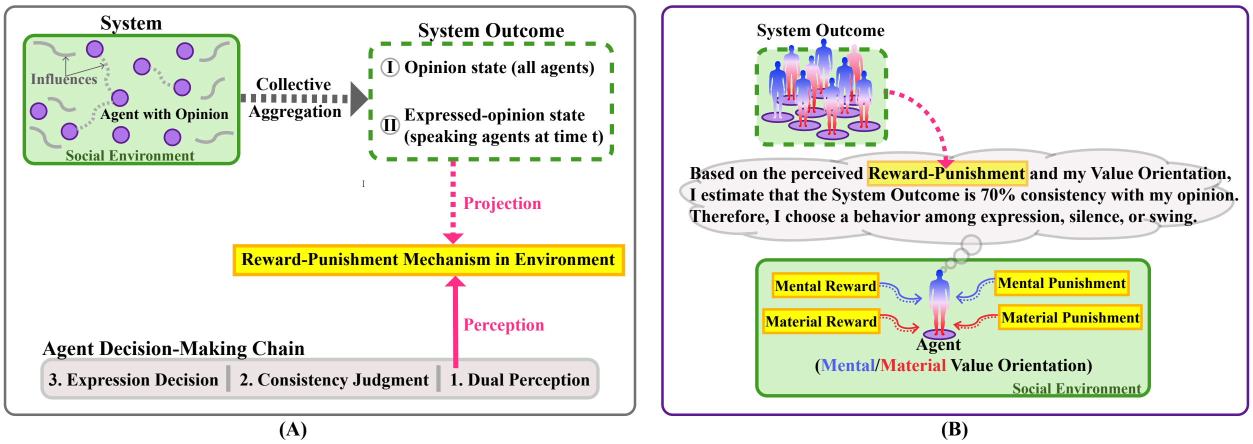

Figure 1: Panel A illustrates the overall modeling logic: in the process of agents deciding whether to express their opinions, all relevant influences, including each agent's opinion, interactions among agents (dashed links), and external factors from the environment (solid links), are collectively aggregated. This aggregation produces two types of system outcomes, projected back into the social environment for perception in the three-stage decision-making chain. Panel B illustrates how a single agent, guided by its value orientation, perceives rewards and punishments to evaluate the consistency between its own opinion and the system outcome, which then informs its decision.

value orientation and the surrounding social conditions; (2) the consistency score, which evaluates the alignment between agents and system outcomes; and (3) the expression function, which translates these evaluations into behavioral probabilities. In addition, Section 2.3.1 introduces embedded nonlinear mechanisms, namely collapse and silent points, which capture emergent dynamics beyond the core formulation. Together, these elements provide a complete specification of the model.

(1) **Personality-Environment Configuration**
Each agent is characterized by its own personality-environment configuration determined by six parameters (a, b, c, d, e, f). The following two components specify the structure of this configuration in detail:

- **Agent Value Orientation and Sensitivity Coefficient**
The value orientation parameters a and b are defined in the following way. Let $a \in [0, 1]$ denote the mental orientation of the agent and let $b \in [0, 1]$ denote the material orientation, subject to $a + b = 1$.
Therefore, the agent's sensitivity coefficients are $\alpha_1 = a$ (mental), and $\alpha_2 = b$ (material). These parameters act as fixed sensitivity coefficients and determine how strongly the agent perceives signals in each dimension.
- **Reward-Punishment Mechanism and Societal Tolerance**
Let $c, d, e, f \in [0, 1]$ represent the reward-punishment parameters generated by projecting the system outcome into the social environment, denoting respectively the mental reward, material reward, mental punishment, and material punishment.
Based on these parameters, we define the societal tolerance as:

$$S = c + d - e - f, \quad S \in [-2, 2] \tag{1}$$

The value of S determines the type of social environment: $S > 0$ indicates a tolerant society in which the reward is predominant over the punishment, $S < 0$ indicates a harsh society in which the punishment is prioritized over the reward, and $S = 0$ indicates a neutral society.
To ensure bounded output in simulation, we apply a logistic transformation: $\sigma(S) = \frac{1}{1+e^{-S}} \in [0.119, 0.881]$. Notice that $\sigma(S)$ serves only as a width-control, not as the variance of a distribution.

(2) **Consistency Score**
We define the agent's consistency score as:

$$\begin{aligned} J &= \Delta_{\text{mental}} + \Delta_{\text{material}} \\ &= \alpha_1\big((c - e) - a\big) + \alpha_2\big((d - f) - b\big) \end{aligned} \tag{2}$$

This formula is intended to provide agents with a more human-like cognitive evaluation, rather than mechanically calculating the reward-punishment difference. Inspired by theories of reference-dependent preferences in psychology and behavioral economics [12, 13], we assume that whether a given outcome is perceived as a reward depends not only on its reward-punishment difference but also on whether it surpasses the individual's reference points (value orientation). Accordingly, the model subtracts each agent's own value orientations from the corresponding net rewards, thereby simulating the cognitive process through which an agent judges whether the outcome is a reward or a punishment for itself.
Then, following theories emphasizing individual variability in reward sensitivity [4, 11], we design the model to scale the net rewards by the sensitivity coefficients before aggregation, thereby simulating the extent to which each agent perceives

rewards or punishments according to its own sensitivity configuration.

Through these designs, the feasible range of J is endogenously bounded between -2 and 0.5. Positive scores indicate alignment (stronger when larger), negative scores indicate misalignment (stronger when smaller), and $J = 0$ reflects neutrality. The broader negative domain arises naturally because in each dimension the value orientations can be as large as 1, while the reward-punishment differentials ($c - e$ or $d - f$) can be as small as -1, yielding a maximum negative gap of -2 when combined across both dimensions. Symmetrically, the positive domain is capped at 0.5 because the constraint $a + b = 1$ ensures that, together with the design of the sensitivity coefficients, it jointly restricts the maximum attainable value of J. This asymmetry is therefore a structural property of the formulation and corresponds to empirical findings that, in social contexts, dissatisfaction is more commonly observed than strong alignment [7, 8].

(3) **Expression Function**

To operationalize the Expression Decision stage, we define the Gaussian-shaped function:

$$N(J) = \exp\left(-\frac{J^2}{2\cdot(\sigma(S))^2}\right), \quad N(J) \in (0,1] \tag{3}$$

$\sigma(S)$ serves as the width-control parameter, and the function is centered at $J = 0$, representing the most neutral state. This function acts as a dynamic probability generator. In particular, each agent is associated with its own values of S and J: S determines the specific shape of $N(J)$, while the agent's consistency score J locates a point on $N(J)$. The vertical coordinate of this point then specifies the agent's baseline probability of expression (before swing adjustment). This mechanism constitutes the fixed-point sampling scheme (Figure 2). Accordingly, the probability of expression is given by:

$$P_e(J;S) = \begin{cases} N(J), & \text{if } J \le 0, \\ N(J) - N(0.5), & \text{if } J > 0. \end{cases} \tag{4}$$

with the boundary condition $P_e(0.5; S) = 0$.

Baseline silence probability is $P_s(J;S) = 1 - P_e(J;S)$.

To capture hesitation without extra parameters, we introduce swing behavior by splitting the two probabilities symmetrically. Specifically, $P_e^* = \frac{1}{2}P_e$, $P_s^* = \frac{1}{2}P_s$, $P_{\text{swing}}^* = \frac{1}{2}P_e + \frac{1}{2}P_s = \frac{1}{2}$, ensuring that $P_e^* + P_s^* + P_{\text{swing}}^* = 1$ and keeping expression-silence probabilities structurally stable across agents.

2.3.1 *Embedded Nonlinear Mechanisms: Collapse and Silent Points.*

To enhance the model's ability to represent real-world behavioral dynamics, two special points are embedded as intrinsic nonlinear mechanisms of the decision-making process: the collapse point and the silent point. These points are not case-specific outcomes but general features of the model, triggered under specific personality-environment configurations.

- **Collapse Point** models counter-intentional behavior arising under severe value conflict. At this point, an agent, who would otherwise be expected to express or remain silent, may instead make the opposite behavioral choice, driven by extreme misalignment between personal values and perceived system outcome.
- **Silent Point** represents a state of irreversible silence triggered by extreme punishment perception. At this point, the agent withdraws entirely from expression due to the overwhelming impact of negative external influences.

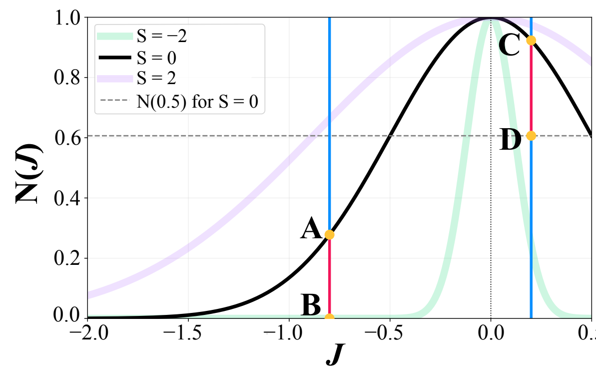

Figure 2: Expression function $N(J)$ versus agent's consistency score J for different values of the societal tolerance S. Using $S = 0$ as an example, the figure illustrates how the function serves as a probabilistic generator, with vertical lines at $J = -0.8$ and $J = 0.2$ indicating the two cases: $J \le 0$ evaluated against the horizontal axis and $J > 0$ against $N(0.5)$. Red line denotes expression probability and blue line denotes silence probability (baseline without swing behavior).

2.4 Simulation Procedure and Settings

The following procedure provides the general implementation of our model, designed to examine how silence behavior statistically emerges from the agent's three-stage decision-making chain under fully randomized personality-environment configurations:

- **Simulation Input Settings**
  Each simulation represents an agent executing its full three-stage decision-making chain for a given event within a specific social environment. The agent's behavior is generated through the probabilistic mechanism of $N(J)$. Six core parameters are sampled as follows: $a \in [0,1]$ and $b = 1 - a$; c, d, e, f are independently sampled from $[0,1]$. These designs ensure that all behavioral decisions emerge solely from randomized personality orientations and environmental conditions, thereby enabling the simulation process without manual parameter tuning.
- **Constraints for Nonlinear Responses**
  We introduce constraint conditions into the model to capture agents whose sampled configurations fall into extreme cases, thereby triggering nonlinear responses. Collapse and silent points are not manually assigned values but emerge whenever these constraints are satisfied. They are triggered only in agents with strong value orientation, defined as a > 0.9 or b > 0.9, representing extreme preference for mental or material orientation.

  - **Collapse Point** is activated when the perceived reward direction is in direct conflict with the agent's dominant value orientation. Specifically, collapse is triggered under the following two configurations: ($a > 0.9$, $c < 0.1$, $d > 0.9$) or ($b > 0.9$, $d < 0.1$, $c > 0.9$).
  - **Silent Point** is triggered when agents perceive extreme punishment precisely in the domain aligned with their core values. This occurs under either ($a > 0.9$, $e > 0.9$) or ($b > 0.9$, $f > 0.9$).
- **Simulation Output Settings**
  The simulation employs the logistic-transformed value of J, partitioned into 75 equal-width intervals to capture the continuous distribution of perceived consistency. Each simulated agent is assigned to its corresponding interval, and the decision outcome is recorded. For each interval, we compute the proportion of agents that choose silence, excluding agents whose outcomes are swing or expression.

The simulation procedure has two key design features that distinguish it from conventional silence ABM. First, it eliminates the need for manual parameter tuning and ensures a fully random behavioral distribution across the entire input space, thereby covering all possible personality-environment configurations. Second, agents are aggregated solely by their J value, ensuring that the estimated silence ratio is derived directly from perceived consistency rather than predetermined attributes.

2.4.1 Simulation Configurations: Base and Controlled Cases. This section introduces the simulation configurations, including a fully randomized base case and five controlled cases.

The base case is executed with 10 independent runs, each generating 10^8 agents, yielding a total of 10^9 samples. The purpose of the base case is to establish the model's statistical baseline as a benchmark silence pattern.

The controlled cases are designed to assess structural robustness, defined as the shape invariance of the silence pattern, up to horizontal clipping or amplitude scaling, under controlled changes in value orientation and societal tolerance. Each controlled setting generates 10^8 agents. In particular:

- **R1.** All six parameters are fully randomized under the constraint $a + b = 1$ and $S \leq 0$. After repeated checks, the proportion of agents with $S = 0$ is consistently below 0.15% per run, so it is considered negligible and not analyzed as a distinct case.
- **R2.** Same as R1 but constrained to $S \geq 0$.
- **R3.** Fixed agent value orientation to material ($a = 0,\ b = 1$); reward-punishment parameters (c, d, e, f) are fully randomized.
- **R4.** Same as R3, but with fixed agent value orientation to mental ($a = 1,\ b = 0$).
- **R5.** Same as R3, but with fixed agent value orientation to balanced ($a = 0.5,\ b = 0.5$).

3 RESULTS

3.1 Base Case Results

Before examining silence dynamics, we first inspect the total population distribution of agents across J intervals (Figure 3). Only 2.2% of agents fall into the $J > 0$ region, indicating that most agents perceive misalignment with the system outcome, while only a small minority perceive strong alignment. This skewed distribution arises from the asymmetric range of J and mirrors real-world observations that dissatisfaction is far more common than strong alignment [7, 8].

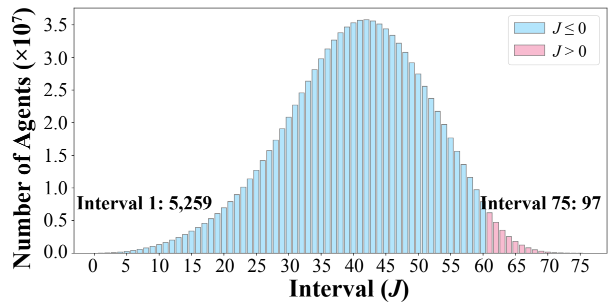

Figure 3: Agent population distribution across J-intervals, with J denoting the consistency score, aggregated over 10^9 samples. The value range of $J \in [-2, 0.5]$ is uniformly divided into 75 equal bins ($\Delta J \approx 0.0333$). The smallest population in the $J \leq 0$ category appears in Interval 1, and the smallest in the $J > 0$ category appears in Interval 75.

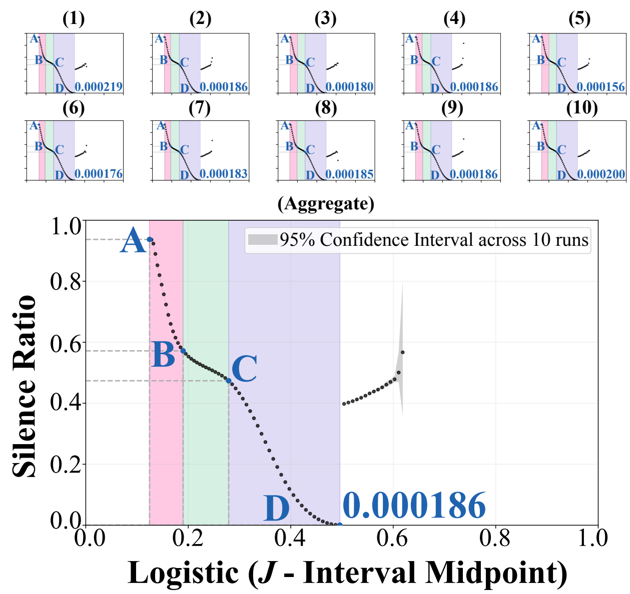

Figure 4: Silence pattern across 75 equal-width intervals of the consistency score J in the base case ($J \in [-2, 0.5]$). The silence ratio is plotted at the logistic-transformed midpoint of each interval. Aggregated results from 10 runs (10^8 agents each), with 95% confidence intervals depicted in gray. Four key points are marked: A denotes the peak, D the minimum, and B-C indicate an automatically detected relatively smooth descent segment. Top: detailed results for individual runs (1)-(10).

Table 1: Occurrence of silent and collapse points across 10 independent simulation runs in the base case, reported as counts and ratios relative to 10^8 agents per run.

Run	*# Silent*	*Ratio*	*# Collapse*	*Ratio*
1	1975804	0.01975804	199483	0.00199483
2	1976190	0.01976190	199801	0.00199801
3	1977334	0.01977334	200011	0.00200011
4	1977314	0.01977314	199633	0.00199633
5	1976778	0.01976778	200146	0.00200146
6	1975630	0.01975630	199707	0.00199707
7	1976425	0.01976425	199188	0.00199188
8	1976373	0.01976373	199852	0.00199852
9	1975770	0.01975770	199677	0.00199677
10	1975372	0.01975372	200450	0.00200450
Total	**19762990**	**0.01976299**	**1997948**	**0.00199780**

Building on this distribution, Figure 4 shows a consistent asymmetric U-shaped pattern across all runs, with narrow confidence intervals and divergence only at extreme J. The left segment of the pattern A-D shows a structured decline: A-B and C-D feature a rapid drop, while B-C remains relatively smoother. This segment roughly corresponds to intervals 15-35 of Figure 3, where the sample density remains sufficient. Therefore, B-C reflects a genuine behavioral buffer rather than sparse data. In this region, silence is no longer overwhelmingly dominant, but expression still entails a perceived risk for agents. At the right end, the proportion rises again, exhibiting a pattern where greater satisfaction corresponds to stronger silence. These findings support the statistical robustness of the model and align with the design of the decision-making chain.

Beyond the overall pattern, two nonlinear mechanisms, the silent point (about 2%) and the collapse point (about 0.2%), are consistently observed across simulations (Table 1). Although their aggregate impact is limited, they extend the explanatory scope by capturing boundary-condition behaviors.

Taken together, the left part of the model output shows that when agents experience fear of isolation, they choose silence. This supports our main goal: to demonstrate that a formally defined model based on the agent's decision-making chain can successfully reproduce the stylized SoS pattern. The right part shows that, given a state of satisfaction, silence becomes more pronounced as satisfaction intensifies. This interpretation aligns with the notion of acquiescent silence in organizational studies, where individuals remain silent out of acceptance of the status quo rather than fear [18]. It also aligns with emotion theory, which associates satisfaction with low action tendency [9].

In summary, the model simultaneously captures fear-based and satisfaction-based silence, reinforcing our broader claim that agent behavior, the decision-making chain, may serve as the dominant driver of silence.

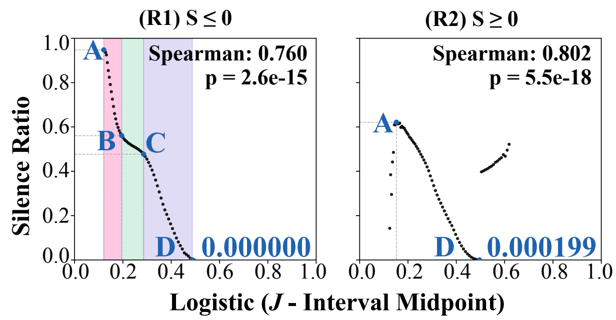

Figure 5: Subplots (R1)-(R2) show the silence pattern across the consistency score J under fully randomized sampling, subject to additional constraints on societal tolerance S, with key points A-D labeled as in Figure 4, and Spearman correlation reported relative to the aggregated base case.

3.2 Controlled Case Results and Comparative Analysis

In this subsection, we compare the controlled cases with the base case to evaluate the structural robustness of the model.

The population distribution of agents across societal tolerance consistently follows a bell-shaped form shown in Figure 6. This confirms that value orientation parameters and reward-punishment parameters are independent, and that agents preserve their autonomy rather than being mechanically determined by the environment.

For fixed value orientation parameters (a, b) of agents, the simulation results of cases R3-R5 demonstrate that value orientation parameters only affect the feasible range of the consistency score J, without altering the silence pattern. As shown in Figure 6, R3 and R4 correspond to material and mental agent orientations, both restricting J to $[-2, 0]$, while R5 imposes balanced agents, restricting the feasible domain to $[-1.5, 0.5]$. Despite these constraints, the silence pattern within the feasible domain remains almost identical to the base case, with Spearman correlations of $\rho = 1.000$ for R3 and R4 and $\rho = 0.955$ for R5, confirming the robustness of the structural form. A noteworthy feature appears in Figure 6, R5, where the initial part of the left segment looks unusually smooth. This effect arises because balanced value orientations and the corresponding balanced sensitivity coefficients jointly suppress extreme negative values of J, thereby reducing local fluctuations. However, the high Spearman correlation value already indicates that this smoothing does not represent a substantive deviation from the base case.

For constrained societal tolerance S, the simulation results of cases R1 and R2 show that reward-punishment parameters alter only the feasible ranges of J and S, while leaving the silence pattern structurally unchanged. In Figure 5, R1 constraining S limits the reward-punishment parameters and further restricts J to $[-2, 0]$. Within this domain, the silence pattern remains highly similar to the baseline ($\rho = 0.760$). In Figure 5, R2 representing a highly tolerant society, the overall shape remains similar to the baseline ($\rho = 0.802$). In this context, reduced concerns and stronger expression lead to a lower silence ratio at peak point and the disappearance of the three-stage descent on the left. The Spearman correlations of R1

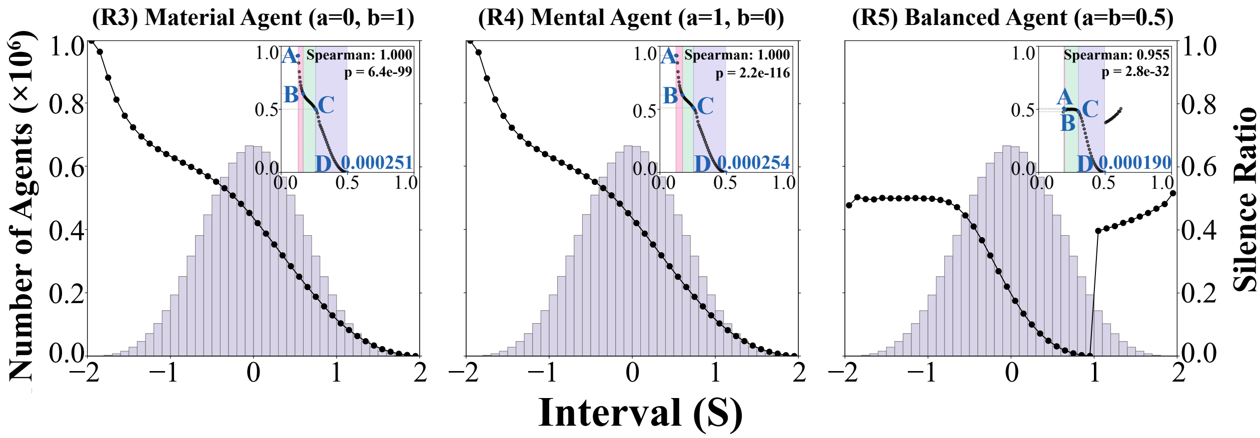

Figure 6: Silence outcomes under fixed value orientation configurations while reward-punishment parameters remain fully randomized. Main plots show agent counts and the silence ratio across equal-width intervals of societal tolerance S ($S \in [-2, 2]$). Insets show silence patterns across the consistency score J, evaluated at logistic-transformed interval midpoints, with key points A-D labeled as in Figure 4. Spearman correlations are reported against the aggregated base case.

and R2 are below 1 because the imposed constraints on S shift the pattern, which is consistent with the expression function design where S determines the peak width of $N(J)$.

Overall, these findings demonstrate that the silence pattern remains structurally stable regardless of personality-environment configuration parameters, and that variations in societal tolerance result in consistent adjustments in the expression function $N(J)$.

4 DISCUSSION

This study proposes a behavior-driven baseline model of silence, grounded in the full decision-making chain of agents rather than neighbor influence and avoiding manual parameter tuning. Through large-scale simulations of both the base case and controlled cases, we verify the statistical and structural robustness of the model. The results support the idea that agent decision-making may play a central role in driving silence.

Theoretically, the design of the consistency score provides a new lens to interpret silence as an outcome of subjective alignment between individual values and collective outcomes, highlighting the role of individual decision-making and cognitive autonomy. Moreover, the design of the expression function enables the model to incorporate satisfaction-based silence in addition to fear-based silence, thereby enriching theoretical accounts of silence.

Practically, the model introduces a fixed-point sampling scheme within the expression function. By fixing societal tolerance, different shapes of $N(J)$ can be generated. This provides a tool for roughly assessing behavioral probabilities under given societal conditions, with potential applications in opinion dynamics and other complex social scenarios.

5 CONCLUSION

This study presents a behavior-driven baseline model of silence, grounded in the full decision-making chain of agents rather than neighbor influence and designed to operate without manual parameter tuning. The core design lies in the implementation of the behavioral chain: we first link agent's value orientation with the reward-punishment mechanism projected from the system outcome to construct a consistency score, and then use the expression function to generate behavior choice probabilities based on agent's configuration, leading to silence, expression or swing. Our model provides a new perspective on silence as a form of subjective alignment between individual values and collective outcomes. Unlike conventional agent-based silence models that primarily reproduce silence under fear as in the SoS, it also incorporates satisfaction-based silence, thereby capturing multiple silence patterns. The design of special points further enhances its nonlinear boundaries. Large-scale simulations, including both the base case and controlled cases, verify the model's statistical and structural robustness, supporting the view that agent decision-making may be a central driver of silence. In practical terms, the fixed-point sampling scheme embedded in the model offers a potential tool for assessing behavioral probabilities under given societal conditions.

Future work includes extending the model with temporal dynamics and pursuing empirical validation, which would further consolidate its potential value for explaining silence in real-world contexts.

ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science and Innovation(Contract No. PID2021-122711NB-C21 and PID2024-157869NB-I00).

Additionally Ziqian SHI acknowledges the support of the China Scholarship Council program (Project ID: 202308410177).

REFERENCES

[1] S. E. Asch. 1951. Effects of group pressure upon the modification and distortion of judgments. In *Groups, leadership and men; research in human relations.* Carnegie Press, Oxford, England, 177–190.

[2] Benjamin Cabrera, Björn Ross, Daniel Röchert, Felix Brünker, and Stefan Stieglitz. 2021. The influence of community structure on opinion expression: an agent-based model. *Journal of Business Economics* 91, 9 (Nov. 2021), 1331–1355. https://doi.org/10.1007/s11573-021-01064-7

[3] Chun Cheng and Changbin Yu. 2019. Opinion dynamics with bounded confidence and group pressure. *Physica A: Statistical Mechanics and its Applications* 532 (Oct. 2019), 121900. https://doi.org/10.1016/j.physa.2019.121900

[4] Edward L. Deci and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior.* Springer US, Boston, MA. https://doi.org/10.1007/978-1-4899-2271-7

[5] Wolfgang Donsbach, Charles T. Salmon, and Yariv Tsfati (Eds.). 2014. *The Spiral of Silence: New Perspectives on Communication and Public Opinion.* Routledge, New York. https://doi.org/10.4324/9780203125007

[6] Mike Farjam and Karl Loxbo. 2024. Social conformity or attitude persistence? The bandwagon effect and the spiral of silence in a polarized context. *Journal of Elections, Public Opinion and Parties* 34, 3 (July 2024), 531–551. https://doi.org/10.1080/17457289.2023.2189730

[7] Andreas Flache and Michael W. Macy. 2011. Small Worlds and Cultural Polarization. *The Journal of Mathematical Sociology* 35, 1-3 (Jan. 2011), 146–176. https://doi.org/10.1080/0022250X.2010.532261

[8] R.S. Foa, A. Klassen, M. Slade, A. Rand, and R. Collins. 2020. *The Global Satisfaction with Democracy Report 2020.* Research Report. Centre for the Future of Democracy, University of Cambridge, Cambridge, United Kingdom.

[9] Nico H. Frijda. 1986. *The emotions.* Editions de la Maison des Sciences de l'Homme, Paris, France. Pages: xii, 544.

[10] Carroll J. Glynn, Andrew F. Hayes, and James Shanahan. 1997. Perceived Support for One's Opinions and Willingness to Speak Out: A Meta-Analysis of Survey Studies on the "Spiral of Silence". *Public Opinion Quarterly* 61, 3 (1997), 452. https://doi.org/10.1086/297808

[11] Jeffrey A. Gray. 1990. Brain Systems that Mediate both Emotion and Cognition. *Cognition & Emotion* 4, 3 (July 1990), 269–288. https://doi.org/10.1080/02699939008410799

[12] Daniel Kahneman and Amos Tversky. 1988. *Prospect theory: An analysis of decision under risk.* Cambridge University Press, New York, NY, US. https://doi.org/10.1017/CBO9780511609220.014 Pages: 214.

[13] Botond Köszegi and Matthew Rabin. 2006. A model of reference-dependent preferences. *Quarterly Journal of Economics* 121, 4 (Nov. 2006), 1133–1165. https://doi.org/10.1093/qje/121.4.1133

[14] A. H. Maslow. 1943. A theory of human motivation. *Psychological Review* 50, 4 (July 1943), 370–396. https://doi.org/10.1037/h0054346

[15] Elizabeth Wolfe Morrison and Frances J. Milliken. 2000. Organizational Silence: A Barrier to Change and Development in a Pluralistic World. *The Academy of Management Review* 25, 4 (Oct. 2000), 706. https://doi.org/10.2307/259200

[16] Kurt Neuwirth, Edward Frederick, and Charles Mayo. 2007. The Spiral of Silence and Fear of Isolation. *Journal of Communication* 57, 3 (Sept. 2007), 450–468. https://doi.org/10.1111/j.1460-2466.2007.00352.x

[17] Elisabeth Noelle-Neumann. 1974. The Spiral of Silence a Theory of Public Opinion. *Journal of Communication* 24, 2 (June 1974), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

[18] Craig C. Pinder and Karen P. Harlos. 2001. Employee silence: Quiescence and acquiescence as responses to perceived injustice. In *Research in Personnel and Human Resources Management.* Vol. 20. Emerald (MCB UP), Bingley, 331–369. https://doi.org/10.1016/S0742-7301(01)20007-3

[19] Deborah A. Prentice and Dale T. Miller. 1993. Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *Journal of Personality and Social Psychology* 64, 2 (Feb. 1993), 243–256. https://doi.org/10.1037/0022-3514.64.2.243

[20] Björn Ross, Laura Pilz, Benjamin Cabrera, Florian Brachten, German Neubaum, and Stefan Stieglitz. 2019. Are social bots a real threat? An agent-based model of the spiral of silence to analyse the impact of manipulative actors in social networks. *European Journal of Information Systems* 28, 4 (July 2019), 394–412. https://doi.org/10.1080/0960085X.2018.1560920

[21] Dietram A. Scheufle and Patricia Moy. 2000. Twenty-Five Years of The Spiral of Silence: A Conceptual Review and Empirical Outlook. *International Journal of Public Opinion Research* 12, 1 (March 2000), 3–28. https://doi.org/10.1093/ijpor/12.1.3

[22] Dongyoung Sohn and Nick Geidner. 2016. Collective Dynamics of the Spiral of Silence: The Role of Ego-Network Size. *International Journal of Public Opinion Research* 28, 1 (March 2016), 25–45. https://doi.org/10.1093/ijpor/edv005

[23] Subrahmaniam Tangirala and Rangaraj Ramanujam. 2008. Employee Silence on Critical Work Issues: The Cross-Level Effects of Procedural Justice Climate. *Personnel Psychology* 61, 1 (March 2008), 37–68. https://doi.org/10.1111/j.1744-6570.2008.00105.x

[24] Xicheng Yin, Hongwei Wang, Pei Yin, and Hengmin Zhu. 2019. Agent-based opinion formation modeling in social network: A perspective of social psychology. *Physica A: Statistical Mechanics and its Applications* 532 (Oct. 2019), 121786. https://doi.org/10.1016/j.physa.2019.121786

[25] Jiefan Zhu, Yiping Yao, Wenjie Tang, and Haoming Zhang. 2022. An agent-based model of opinion dynamics with attitude-hiding behaviors. *Physica A: Statistical Mechanics and its Applications* 603 (Oct. 2022), 127662. https://doi.org/10.1016/j.physa.2022.127662