Appendix

Appendix 1: Identical definition for level 1 aggregation

As noted in the main text, the level 1 aggregation distinguishes disaggregated states by how many users select a choice but not who select it. To characterise this property, we first define an operator called *swapper* that swaps choices of users in the same group. When choices of two users i, j in the same group are swapped, this operation is denoted by $\mathbf{c}^{\mathrm{swap}}(\mathbf{c}, [(i,j)])$. The iterative operation of an arbitrary number of swaps is denoted by $\mathbf{c}^{\mathrm{swap}}(\mathbf{c}, \sigma)$, where σ denotes how the choices of users are swapped (e.g. if $\sigma = [(i,j),(k,l)]$, choices of users i, j are swapped first, and then choices of users k, l are swapped). Note that users in the same parenthesis must belong to the same group, while those not in the same parenthesis need not belong to the same group (e.g. users k, l need not belong to the same group as users i, j in the example). We call σ the *sequence of swaps* and let Σ_{Swp} be the set of all available sequences of swaps. Note that Σ_{Swp} constitutes a symmetric group, whose elements are permutations that swaps choices of users belonging to the same group.

The following proposition can be stated using the sequence of swaps:

PROPOSITION 1. *A is the level 1 aggregation if and only if*

$$\begin{aligned}
&\mathbf{c}' = \mathbf{c}^{\mathrm{swap}}(\mathbf{c}, \sigma) \\
\Leftrightarrow\ \ & a(\mathbf{c}; A) = a(\mathbf{c}'; A) \ \ \forall \mathbf{c}, \mathbf{c}' \in C, \ \ \exists \sigma \in \Sigma_{\mathrm{Swp}}.
\end{aligned} \tag{A1}$$

and Equation (12) holds.

Proof. Because any swapper does not change $x(\mathbf{c})$, we have

$$\mathbf{c}' = \mathbf{c}^{\mathrm{swap}}(\mathbf{c}, \sigma) \Leftrightarrow \mathbf{x}(\mathbf{c}) = \mathbf{x}(\mathbf{c}') \ \ \exists \sigma \in \Sigma_{\mathrm{Swp}}. \tag{A2}$$

This implies that Equations (11) and (A1) are identical. □

Appendix 2: Proof of Theorem 1

Combining Equations (5) and (6), we have

$$\begin{aligned}
&p(a, t+1; \mathbf{c}^0) = \sum_{\mathbf{c} \in a} p(\mathbf{c}, t+1; \mathbf{c}^0) \\
&= \sum_{a' \in A} \sum_{\mathbf{c}' \in a'} \sum_{\mathbf{c} \in a} \prod_{i \in I} \\
&\qquad \{ r K_{c_i}(a') + (1-r)\delta(c_i, c_i') \} \, p(\mathbf{c}', t; \mathbf{c}^0).
\end{aligned} \tag{A3}$$

$K_{c_i}(\mathbf{c}')$ is now replaced with $K_{c_i}(a')$ owing to Equation (12). In addition, Equation (48) implies that the innermost summation of the right-hand side of Equation (A3) takes the sum for all combinations of swaps. Therefore, it does not depend on $\mathbf{c}'$ but a' and can be denoted by

$$q_A(a',a) = \sum_{\mathbf{c}\in a}\prod_{i\in I}\{rK_{c_i}(a') + (1-r)\delta(c_i,c_i')\}. \tag{A4}$$

Substituting $q_A(a',a)$ into Equation (A3), we have

$$\begin{aligned} p(a,t+1;\mathbf{c}^0) &= \sum_{a'\in A} q_A(a',a)\sum_{\mathbf{c}'\in a'} p(\mathbf{c}',t;\mathbf{c}^0) \\ &= \sum_{a'\in A} q_A(a',a)p(a',t;\mathbf{c}^0). \end{aligned} \tag{A5}$$

Letting $p(a,t;a^0) = \sum_{\mathbf{c}^0\in a} p(a,t;\mathbf{c}^0)$, we finally obtain

$$p(a,t+1;a^0) = \sum_{a'\in A} q_A(a',a)p(a',t;a^0), \tag{A6}$$

which derives Equation (16). $\square$

Appendix 3: Proof of Theorem 2 and Corollary 2.1

Consider any pair of $\mathbf{c}^1,\mathbf{c}^2 \in a \in A$, where A is a level 1 aggregation. Let $p(\mathbf{c}^1,t;\mathbf{c}^0) = p(\mathbf{c}^2,t;\mathbf{c}^0)$ and denote them by $p(a,t;\mathbf{c}^0)$. Then,

$$\begin{aligned} &p(\mathbf{c}^1,t+1;\mathbf{c}^0) - p(\mathbf{c}^2,t+1;\mathbf{c}^0) = \\ &\sum_{a'\in A} p(a,t;\mathbf{c}^0)\sum_{\mathbf{c}'\in a'} \\ &\qquad\left[\begin{array}{l} \prod_{i\in I}\left\{rK_{c_i^1}(a') + (1-r)\delta(c_i^1,c_i')\right\} \\ -\prod_{i\in I}\left\{rK_{c_i^2}(a') + (1-r)\delta(c_i^2,c_i')\right\} \end{array}\right]. \end{aligned}$$

Because the right-hand side of Equation (A7) takes the sum for all combinations of swaps in a', the term in the square brackets is zero.

To prove the corollary, let the initial state of a disaggregated Markov chain be a corner state. It lets the flat probability condition hold when $t = 0$, which leads to the corollary by letting $t \to \infty$ in Theorem (2). $\square$

Appendix 4: Mixing time with different thresholds of the TV distance

A smaller threshold than 0.25 yields greater t_C^*. How t_C^* increases by decreasing the threshold depends on the property of the TV distance, but according to Levin and Peres (2017) (Section 4), it is upper bounded by a value that decays exponentially with respect to t. Thus, if this upper bound is a good approximation of the TV distance, using 0.25^k as a threshold (where $k \geq 1$) will increase the MCMT by $\alpha(k-1)$, where α is a positive constant.

Appendix 5: Proof of Theorem 4

The following equations/inequalities, which is derived by Equation (20), prove the theorem:

$$\begin{aligned}
&\|\mathbf{p}_A^1 - \mathbf{p}_A^2\|_{\mathrm{TV}} = \frac{1}{2}\sum_{a\in A}|p_1(a) - p_2(a)| \\
&= \frac{1}{2}\sum_{a^+\in A^+}\sum_{a\in A(a^+)}|p_1(a) - p_2(a)| \\
&\geq \frac{1}{2}\sum_{a^+\in A^+}\left|p_1(a^+) - p_2(a^+)\right| \|\mathbf{p}_A^1 - \mathbf{p}_A^2\|_{\mathrm{TV}} \\
&= \|\mathbf{p}_{A^+}^1 - \mathbf{p}_{A^+}^2\|_{\mathrm{TV}}
\end{aligned} \tag{A7}$$

This equation also holds when A^+ and A are replaced by A and C, respectively. $\square$

Appendix 6: Proof of Theorem 5

The flat probability condition implies

$$p(a) = |C(a)|p(\mathbf{c}) \;\; \forall \mathbf{c} \in C(a), \tag{A8}$$

where $C(a)$ is a set of disaggregated states combined to aggregate set a and $|C(a)|$ the number of disaggregated sets in a. Therefore, we have

$$\sum_{\mathbf{c}\in C(a)} |p^1(\mathbf{c}) - p^2(\mathbf{c})| = |p^1(a) - p^2(a)|, \tag{A9}$$

which constitutes the theorem. $\square$

Appendix 7: Proof of Theorem 6

$\delta_{a^+}(\mathbf{p}_A^1, \mathbf{p}_A^2) \geq 0$ owing to the triangle inequality. In addition,

$$\begin{aligned}
&\sum_{a \in A(a^+)} |p^1(a) - p^2(a)| \\
&= \left| \sum_{a \in A(a^+)} p^1(a) - \sum_{a \in A(a^+)} p^2(a) \right| \\
&= |p^1(a^+) - p^2(a^+)|
\end{aligned} \tag{A10}$$

holds for all $a^+ \in \tilde{A}^+$ as either $p^1(a) \geq p^2(a)$ or $p^1(a) \leq p^2(a)$ holds for all $a \in A(a^+)$ by the definition of $\tilde{A}^+$, which leads to $\delta_{a^+}(\mathbf{p}_A^1, \mathbf{p}_A^2) = 0$. $\square$

Appendix 8: Matrix calculation

The equations applicable for the level 1 aggregation denoted by A is shown below. All the equations can also be applied to disaggregated states C by replacing A with C.

While we can use the standard diagonalisation technique to find the power of the transition matrix, the following algorithm is easier to implement and faster when the number of multiplication to find the stationary distribution is not many.

1. Let $k = 1$ and $Q_{A1} = Q_A$.
2. Let $Q_{Ak+1} \leftarrow (Q_{Ak})^2$
3. Check if every pair of columns in Q_{Ak+1} are sufficiently close each other. If so, finish the algorithm. Any column of Q_{Ak+1} includes the stationary distribution.
4. Let $k \leftarrow k + 1$ and go back to Step 2.

We can use the bisection method to find an MCMT because the TV distances of disaggregated and level 1 aggregated states are monotonically decreasing with respect to t (see Exercise 4.2 of Levin and Peres (2017)). We can reduce the calculation cost of the bisection search by keeping Q_{Ak} in the memory when finding the stationary distribution. For example, when the TV distance becomes less than 0.25 between $k = 3$ (i.e. $Q_{A3} = Q_A^4$) and $k = 4$ (i.e. $Q_{A4} = Q_A^8$), we can find the TV distance at $t = 6$ by $Q_A^6 = Q_A^2 Q_A^4 = Q_{A2} Q_{A3}$.

Appendix 9: Accuracy estimation of the Monte Carlo method

To roughly estimate how the error is correlated with the number of states, consider two probability distributions $p_1(a)$ and $p_2(a)$ and the frequencies of occurrence by these probabilities, denoted by $f_1(a)$ and

$f_2(a)$, respectively. Approximating the distributions of $f_1(a)$ and $f_2(a)$ by the binomial distribution, i.e. $f_1(a) \sim \mathrm{Bi}(p_1(a), N_{\mathrm{sim}})$ and $f_2(a) \sim \mathrm{Bi}(p_2(a), N_{\mathrm{sim}})$, we have

$$\frac{f_1(a) - f_2(a)}{N_{\mathrm{sim}}} \sim \mathrm{N}\left(p_1(a) - p_2(a), \frac{p_1(a) + p_2(a)}{N_{\mathrm{sim}}}\right) \tag{A11}$$

as an estimated difference between $p_1(a)$ and $p_2(a)$, when N_{sim} is large and $p_1(a)$ and $p_2(a)$ are small so that the binomial distributions can be approximated by normal distributions. The relative standard deviation of this value, denoted by

$$\sigma_{\mathrm{rd}}(a) = \frac{1}{|p_1(a) - p_2(a)|}\sqrt{\frac{p_1(a) + p_2(a)}{N_{\mathrm{sim}}}} \tag{A12}$$

should be sufficiently less than one to maintain accuracy in estimating TV distances and MCMTs. This σ_{rd} can be lower bounded as follows.

$$\sigma_{\mathrm{rd}} \geq \frac{1}{\sqrt{0.5 N_{\mathrm{sim}} p_{\mathrm{m}}(a)}}. \tag{A13}$$

This result implies that we need to employ a number that is sufficiently greater than p_{m}^{-1} as N_{sim}.

References

Levin DA, Peres Y (2017) *Markov chains and mixing times*, volume 107 (Providence, RI, USA: American Mathematical Society).