

Supplemental Material A

Materials and Methods

Vector Representation of Names

We use a dataset from the Social Security Administration (SSA) reporting names from Social Security card applications for births in the United States in each year from 1880 to 2020, excluding names occurring less than 5 times in a given year to maintain individual privacy. The original dataset contained 100,364 unique names across all years. In our analysis, we exclude names that appear in less than 25% of the years; that is, we retain names that occur in at least 35 of the 141 years in the dataset. This leaves us with 17,180 unique names.

To embed the names in a vector space, we use a pre-trained model produced using the fastText algorithm from Mikolov et al. (2018). The model uses 2 million word vectors trained with subword information on Common Crawl data in the English language, consisting of 600 billion tokens. Since fastText is based on co-occurrences of strings of characters, the model can produce a vector for every name in the SSA dataset, including those that do not explicitly appear in the Common Crawl data. The fastText model represents each name as a 300-dimensional vector.

To calculate the mean name vector for each year, we average the name vectors weighted by the number of babies receiving each name that year. We use the vector representation for each name to calculate the Euclidean distance between that vector and the mean vector for each year that the name appears in the dataset. We present results using Euclidean distance for consistency with the general model proposed in this paper; we get similar results using cosine distance in analyses presented in Figures SM1 and SM2. To calculate popularity, we divide the number of babies given each name by the total number of babies born in the year and included in the dataset. (That is, the denominator is missing babies with names excluded from our dataset.) We convert proportions to percentages.

The Social Networks

We borrow Jin, Girvan, and Newman’s Model II algorithm for growing undirected social networks (Jin et al., 2001) and modify it to generate directed social networks with $N = 100$ people, each of whom can observe up to a maximum of z_{\max} neighbors. The network is initialized with all 100 people and no connections. The following three steps are then repeated 100 times:

1. Choose 3 pairs of individuals uniformly at random. For each pair i and j , if i observes less than z_{\max} people and does not already observe j , then i begins to observe j ; else, if j observes less than z_{\max} people and does not already observe i , then j begins to observe i .
2. Randomly select a fraction r of the triads i , j , and k such that i observes k and k observes j or that i and j both observe k . If i observes less than z_{\max} people and does not already observe j , then i begins to observe j .
3. Randomly select and break 0.5% of connections (rounded up).

All 25 social networks, measures of their structural properties, and the Python source code used to create them are made available on OSF and can be accessed at osf.io/s4t6r.

The Better-Reply Dynamics

Our computational model adopts a specification of the better-reply dynamics in which at each time step, one randomly selected individual searches for (and upon discovery, adopts) a better reply to the current population profile. Initial strategies are randomly (uniformly) distributed. We check for convergence every 1000 time steps by checking whether any individual can find a better reply. The Python source code and complete output data are available on OSF and can be accessed at osf.io/s4t6r.

Co-evolving Social Networks and Identities

We again assume there are $N = 100$ people. We consider the space of identities with $m = 1$, $d = 1$, and $\{a..b\} = \{0..9\}$. We set the maximum number of neighbors that an individual can handle (i.e., maximum out-degree) to be $z_{\max} = 5$. In this model, in contrast to the earlier model, each time step corresponds to a single individual considering a single change (either to his identity or his network), rather than searching for (i.e., repeatedly considering) such a change. We allow the dynamics to run for up to 2,000,000 time steps before cutting them off and classifying them as non-convergent, and we check for convergence every 1000 time steps.

Initially people have no network connections and strategies are randomly distributed. At each time step, there is an equal 50% chance of considering a change in identity or a change in the network. In the former case, a randomly selected individual considers switching to a randomly selected new identity and does so only if the switch increases his utility. In the latter case, the probability of considering a new connection from person i to person j is proportional to $1 + 2000(\tau_{\text{in}} + \tau_{\text{out}})$, where τ_{in} is the number of triads in which i and j both observe some other individual k , and τ_{out} is the number of triads in which i observes some other individual k , who then observes j . If person i already has z_{\max} connections to other people, then the potential connection to j is considered jointly with breaking one of i 's existing connections. Person i goes through with the change only if it would increase his utility or if he previously had no connections (i.e, undefined utility). The Python source code and output data are made available on OSF and can be accessed at osf.io/s4t6r.

Supplemental Material B

Mathematical Appendix

Formal Definitions

We can express person i 's neighbors' average identity as

$$\bar{x}_{\eta(i)} = \frac{1}{|\eta(i)|} \sum_{j \in \eta(i)} x_j.$$

We can express the number of i 's neighbors who adopt the same expression of identity trait μ as person i as

$$\tilde{n}_{i,\mu}(X; \eta(i)) = \sum_{j \in \eta(i)} \delta(x_{i,\mu}, x_{j,\mu}),$$

where δ is the Kronecker delta function. Then

$$\tilde{n}_i(X; \eta(i)) = \frac{1}{m} \sum_{\mu} \tilde{n}_{i,\mu}(X; \eta(i))$$

is the average number of neighbors sharing one's traits (across all the aspects of identity).

In a well-mixed population, we set $\eta(i) = \{j : j \neq i\}$ to recover $n_{i,\mu}(X)$ and $n_i(X)$ for all i .

Supplementary Results and Proofs

Lemma 1. *In a well-mixed population with utility functions given in Equation (1), the game has an exact potential function:*

$$\Phi(X) = - \sum_{i=1}^N \frac{N-1}{N} \|x_i - \bar{x}\|^2 + \frac{1}{2} \lambda n_i(X).$$

Proof. Consider a change in the profile of identities $X \rightarrow X'$ resulting from person i alone changing his identity $x_i \rightarrow x'_i$, i.e., such that $x'_j = x_j$ for all $j \neq i$. We need only show that the change in the potential function equals the change in i 's utility:

$$\Phi(X') - \Phi(X) = u_i(X') - u_i(X).$$

We express the change in the potential function as a sum of the changes in each term:

$$\begin{aligned} \Phi(X') - \Phi(X) = \\ \sum_{j=1}^N \frac{N-1}{N} \left(\|x_j - \bar{x}\|^2 - \|x'_j - \bar{x}'\|^2 \right) + \sum_{j=1}^N \frac{1}{2} \lambda \left(n_j(X) - n_j(X') \right). \end{aligned}$$

We consider each of the two summations separately.

We expand the first sum:

$$\begin{aligned} \sum_{j=1}^N \frac{N-1}{N} \left(\|x_j - \bar{x}\|^2 - \|x'_j - \bar{x}'\|^2 \right) = \\ \frac{N-1}{N} \left(\|x_i - \bar{x}\|^2 - \|x'_i - \bar{x}'\|^2 \right) + \\ \sum_{j \neq i} \frac{N-1}{N} \left(\|x_j - \bar{x}\|^2 - \|x'_j - \bar{x}'\|^2 \right). \quad (3) \end{aligned}$$

We find it useful to express the average identity as $\bar{x} = \frac{N-1}{N} \bar{x}_{-i} + \frac{1}{N} x_i$. Plugging in to the first term in Equation (3), we have:

$$\|x_i - \bar{x}\|^2 - \|x'_i - \bar{x}'\|^2 = \left(\frac{N-1}{N} \right)^2 \left(\|x_i - \bar{x}_{-i}\|^2 - \|x'_i - \bar{x}_{-i}\|^2 \right).$$

Plugging in to the second term in Equation (3), expanding and canceling off common terms, we have for any $j \neq i$:

$$\begin{aligned} \|x_j - \bar{x}\|^2 - \|x'_j - \bar{x}'\|^2 = \\ \frac{1}{N^2} \left(\|x_i - \bar{x}_{-i}\|^2 - \|x'_i - \bar{x}_{-i}\|^2 \right) + \frac{2}{N} (x_j - \bar{x}_{-i}) \cdot (x_i - x'_i). \end{aligned}$$

Observe that the last term here drops out when we sum over all $j \neq i$ because

$\sum_{j \neq i} (x_j - \bar{x}_{-i}) = 0$. The first term does not depend on j , so summing over all $j \neq i$ just

multiplies this term by a factor of $(N - 1)$. Putting it all together, we find that Equation (3) simplifies to:

$$\begin{aligned}
\sum_{j=1}^N \frac{N-1}{N} \left(\|x_j - \bar{x}\|^2 - \|x'_j - \bar{x}'\|^2 \right) \\
&= \left(\frac{(N-1)^3}{N^3} + \frac{(N-1)^2}{N^3} \right) \left(\|x_i - \bar{x}_{-i}\|^2 - \|x'_i - \bar{x}_{-i}\|^2 \right) \\
&= \left(\frac{N-1}{N} \right)^2 \left(\|x_i - \bar{x}_{-i}\|^2 - \|x'_i - \bar{x}_{-i}\|^2 \right) \\
&= \|x_i - \bar{x}\|^2 - \|x'_i - \bar{x}'\|^2. \quad (4)
\end{aligned}$$

Now, returning to the second part of the change in the potential function, we can use the formal definition of $n_j(X)$ to write:

$$\sum_{j=1}^N \frac{1}{2} \lambda (n_j(X) - n_j(X')) = \frac{1}{2} \lambda \frac{1}{m} \sum_{\mu} \sum_{j=1}^N \sum_{k \neq j} \left(\delta(x_{j,\mu}, x_{k,\mu}) - \delta(x'_{j,\mu}, x'_{k,\mu}) \right).$$

The terms cancel whenever $j \neq i$ and $k \neq i$, so we are left with:

$$\begin{aligned}
\sum_{j=1}^N \frac{1}{2} \lambda (n_j(X) - n_j(X')) &= \\
&\frac{1}{2} \lambda \frac{1}{m} \sum_{\mu} \left(\sum_{j \neq i} \left(\delta(x_{j,\mu}, x_{i,\mu}) - \delta(x'_{j,\mu}, x'_{i,\mu}) \right) + \sum_{k \neq i} \left(\delta(x_{i,\mu}, x_{k,\mu}) - \delta(x'_{i,\mu}, x'_{k,\mu}) \right) \right) \\
&= \lambda \frac{1}{m} \sum_{\mu} \sum_{j \neq i} \left(\delta(x_{j,\mu}, x_{i,\mu}) - \delta(x'_{j,\mu}, x'_{i,\mu}) \right) \\
&= \lambda \frac{1}{m} \sum_{\mu} (n_{i,\mu}(X) - n_{i,\mu}(X')) = \lambda (n_i(X) - n_i(X')). \quad (5)
\end{aligned}$$

Putting Equations (4) and (5) together, we have now shown that

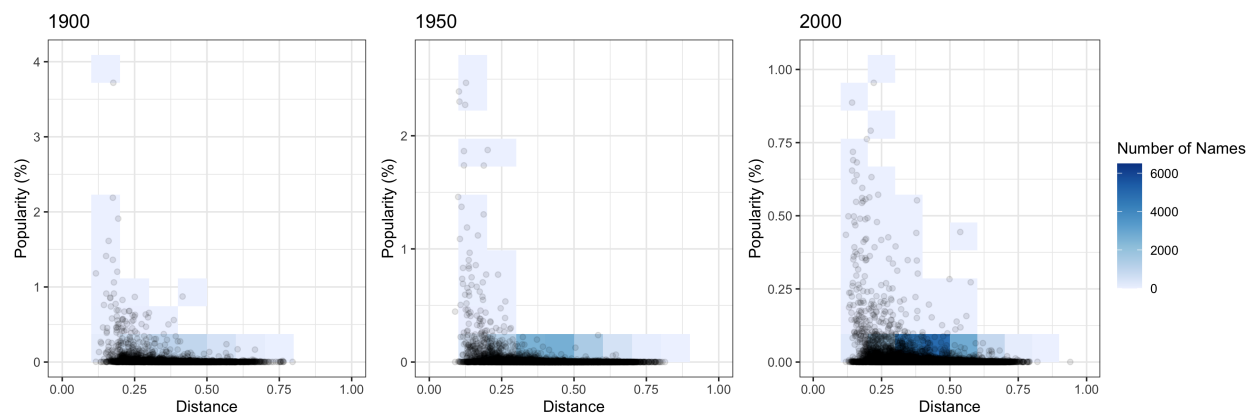
$$\Phi(X') - \Phi(X) = u_i(X') - u_i(X). \quad \square$$

Proof of Theorem 1

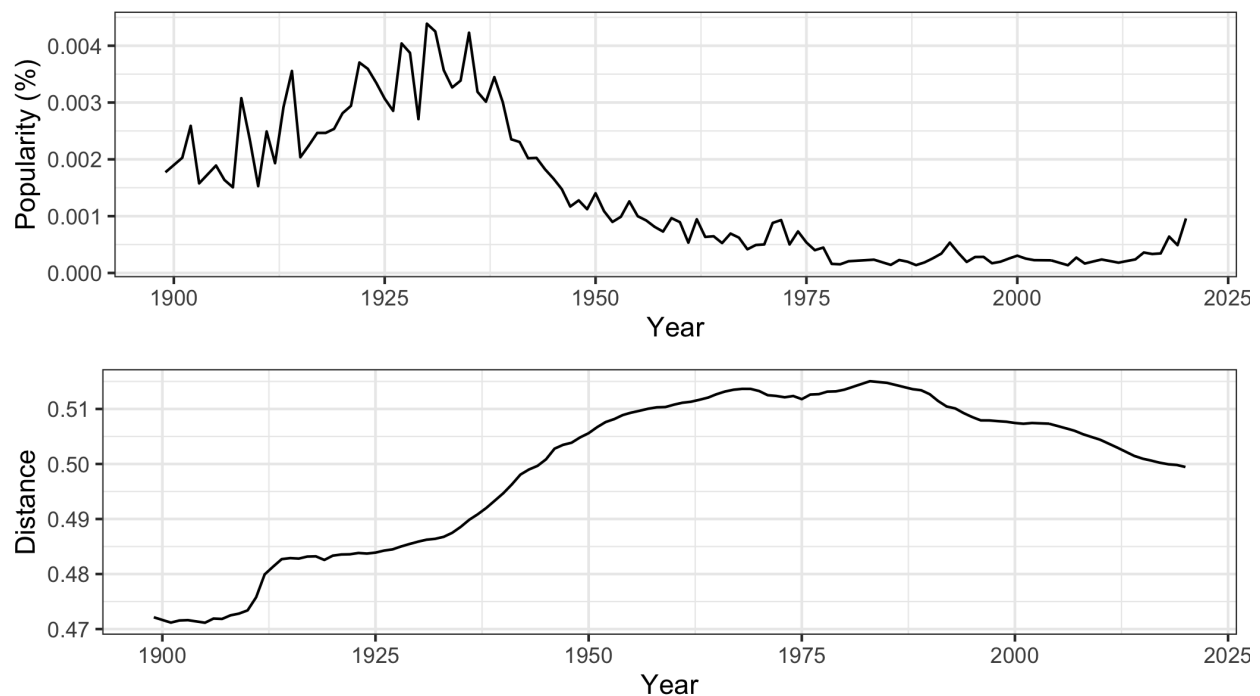
Theorem 1 follows from Lemma 1 by Monderer and Shapley's 1996b argument. \square

Supplemental Material C

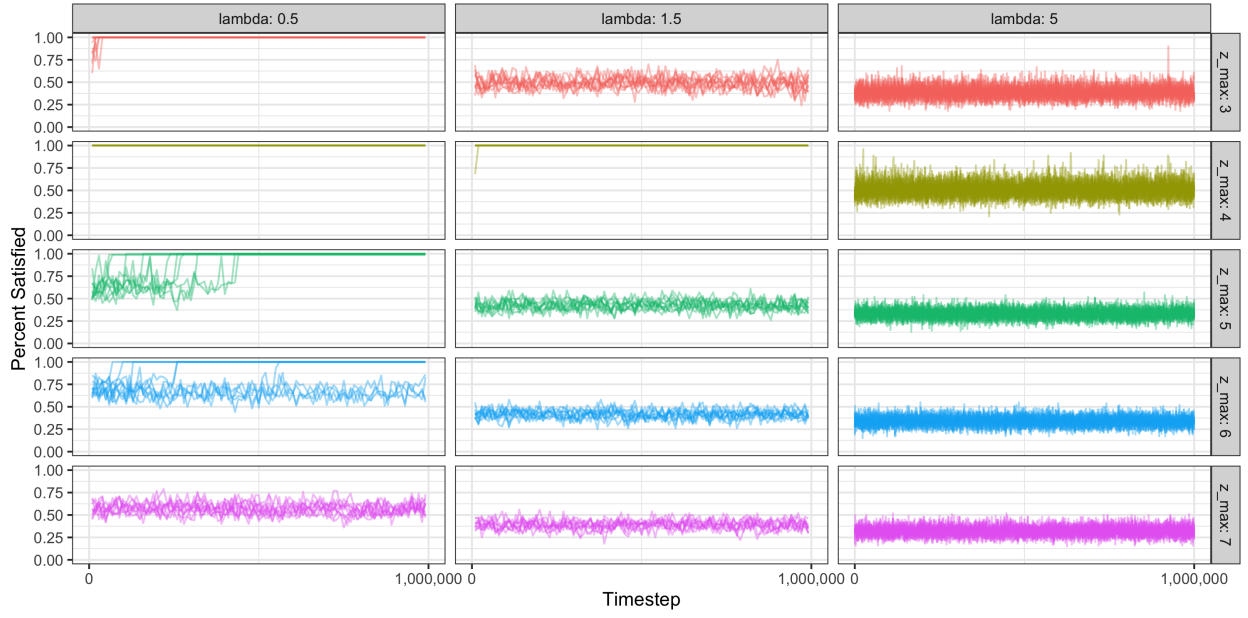
Supplemental Figures

**Figure SM1**

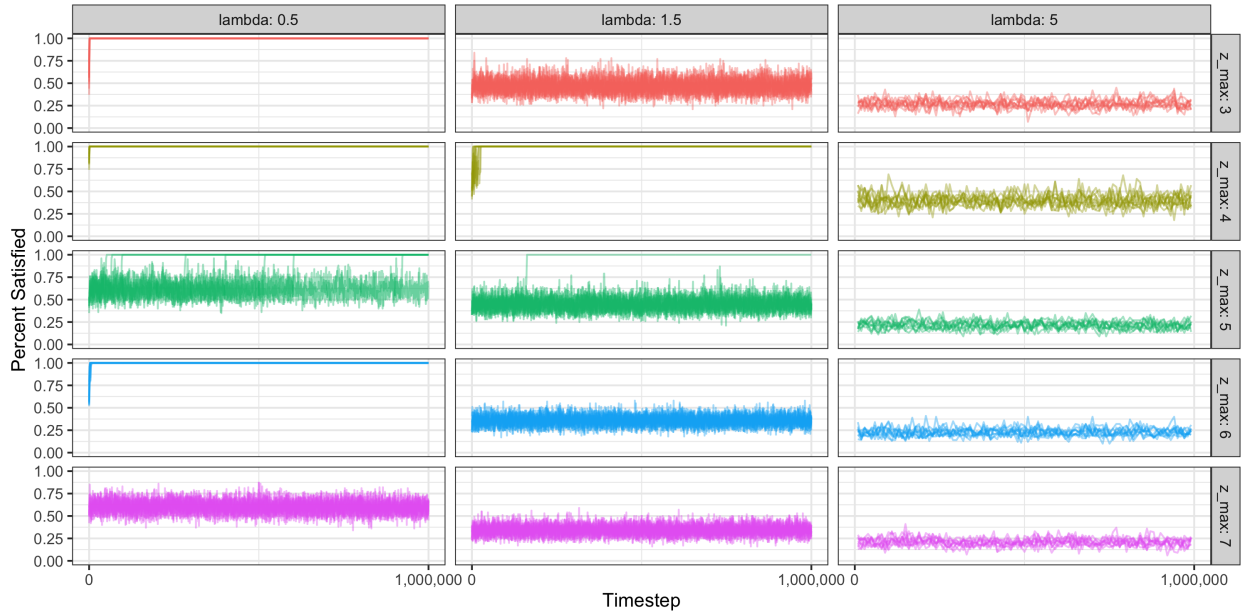
Cosine distance between each name's vector representation and the mean name vector, along with the popularity of these names (i.e., the percentage of babies receiving each name) in the years 1900, 1950, and 2000. Background shading indicates the number of names in each bin.

**Figure SM2**

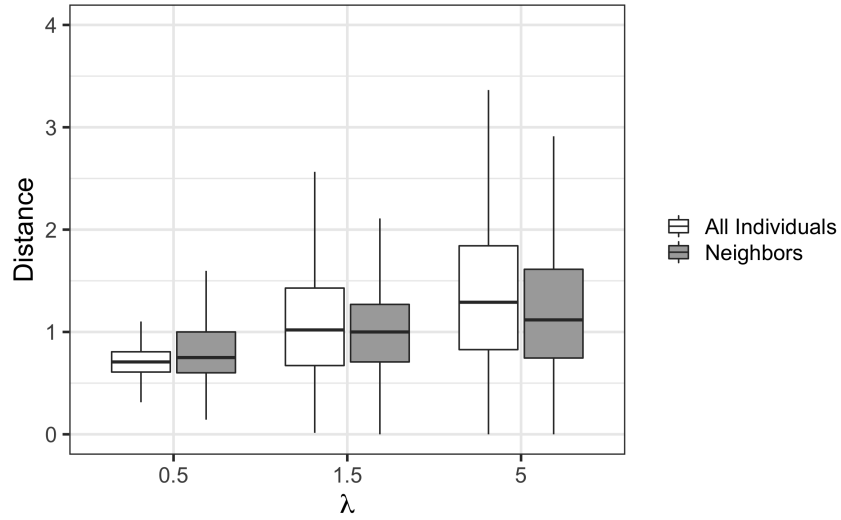
Cosine distance between the vector representation of "Wynona" and the mean name vector, as well as the popularity of the name "Wynona" for each year the name was given.

**Figure SM3**

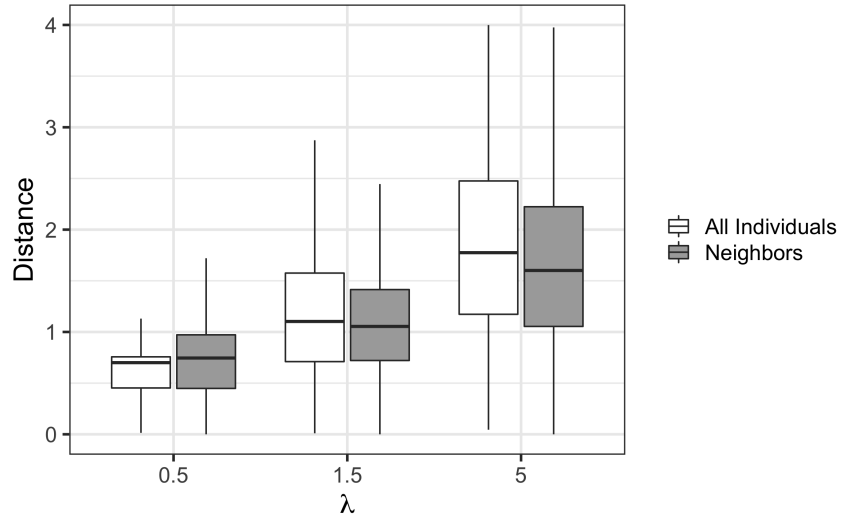
Percentage of individuals satisfied over 1,000,000 time steps for each trial with $m = 1$, $d = 2$, and varying λ , for networks with $r = 1$ and varying z_{\max} .

**Figure SM4**

Percentage of individuals satisfied over 1,000,000 time steps for each trial with $m = 2$, $d = 1$, and varying λ , for networks with $r = 1$ and varying z_{\max} .

**Figure SM5**

Box plots showing distances from individuals' identities to the average identity of all individuals in the population and to the average identity of their neighbors in the network, measured at the 10,000th time step, for $m = 1$, $d = 2$, and varying λ , aggregating trials across the different networks. The differences between the average distance to the population mean identity and the average distance to the mean of one's neighbors' identities are all significant with $p < .001$ in paired t-tests.

**Figure SM6**

Box plots showing distances from individuals' identities to the average identity of all individuals in the population and to the average identity of their neighbors in the network, measured at the 10,000th time step, for $m = 2$, $d = 1$, and varying λ , aggregating trials across the different networks. The differences between the average distance to the population mean identity and the average distance to the mean of one's neighbors' identities are all significant with $p < .001$ in paired t-tests.