Supplementary Material to "The origin of the central dogma through conflicting multilevel selection" in Proceedings of the Royal Society B http://dx.doi.org/10.1098/rspb.2019.1359

Nobuto Takeuchi^{*1,2} and Kunihiko Kaneko^{1,3}

¹Research Center for Complex Systems Biology, Graduate School of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan
²School of Biological Sciences, Faculty of Science, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand
³Department of Basic Science, Graduate School of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153-8902, Japan

1 Supporting Texts

1.1 An alternative agent-based model in which coexistence between P and Q is selectively neutral

In this section, we describe an alternative agent-based model in which coexistence between P and Q is neutral with respect to cellular-level selection. In the agent-based model described in the main text, coexistence between P and Q is favoured by cellular-level selection. This is due to a specific rule about complex formation, which implies that replicators multiply fastest if both P and Q provide and receive catalysis (see Methods for details). To ascertain that this specific rule about complex formation does not critically affect results, we additionally examined an alternative model in which replicators multiply fastest even if only either P or Q provides and receives catalysis. In this model, cellular-level selection does not favour coexistence between P and Q while it still tends to maximise the multiplication rate of replicators within protocells.

In the alternative model, the reaction rate constants of complex formation are defined as a function of the k_{pt}^c values of a replicator serving as a catalyst as follows:

$$\max(k_{\mathrm{P}t}^c, k_{\mathrm{Q}t}^c) \frac{k_{pt}^c}{k_{\mathrm{P}t}^c + k_{\mathrm{Q}t}^c}.$$

^{*}Corresponding author. E-mail: nobuto.takeuchi@auckland.ac.nz

Under this definition, two replicators, denoted by X and Y, form a complex at a rate proportional to $\max(k_{Py}^x, k_{Qy}^x) + \max(k_{Px}^y, k_{Qx}^y) \leq 2k_{\max}$ if all possible complexes are considered, where x and y are the replicator types of X and Y, respectively (in the original model, this rate is proportional to $\sum_p k_{py}^x + k_{px}^y \leq 4k_{\max}$). Accordingly, replicators multiply fastest not only if $k_{pt}^c = k_{\max}$ for all combinations of c, p, and t, but also if $k_{cc}^c = k_{\max}$ for either c = P or c = Q and $k_{pt}^c = 0$ for all the other combinations of c, p, and t. In other words, replicators multiply fastest even if only either P or Q provides and receives catalysis (this is in contrast to the model described in the main text). While cellular-level selection always tends to maximise the multiplication rate of replicators within protocells, it is indifferent to how this maximisation is achieved. Therefore, cellular-level selection does not necessarily tend to maximise k_{pt}^c values for all combinations of c, p, and t; i.e., it does not necessarily favour coexistence between P and Q.

To examine the effect of coexistence between P and Q on symmetry breaking, we simulated the alternative model described above with two initial conditions, symmetric and asymmetric. In the symmetric initial condition, both P and Q were present—this is the same initial condition as used for the original agent-based model. In the asymmetric initial condition, only Q was present (see Fig. S2 for details)—this condition might be closer to what is typically imagined in the RNA world hypothesis. For both initial conditions, the model displays the same three-fold symmetry breaking as displayed by the original model (Fig. S2), indicating that the results do not depend on whether coexistence between P and Q is favoured by cellular-level selection.

1.2 Alternative agent-based models in which the mutation of k_{pt}^c is modelled differently

In this section, we describe alternative models for the mutation of k_{pt}^c . In the agentbased model described in the main text, the mutation of k_{pt}^c is modelled as unbiased random walks in a half-open interval $(-\infty, k_{\max})$ with a reflecting boundary at $k_{pt}^c = k_{\max}$. To ascertain that this specific model of mutation does not critically affect results, we additionally examined two alternative models of mutation. The first alternative model is nearly the same as the above, except that the reflecting boundary condition is set at $k_{pt}^c = 0$. In the second alternative model, each k_{pt}^c value is mutated by multiplying $\exp(\epsilon)$, where ϵ is a number randomly drawn from a uniform distribution on the interval $(-\delta_{\max}, \delta_{\max})$, with a reflecting boundary at $k_{pt}^c = k_{\max}$. Both models of mutation produce essentially the same result as described in the main text (Figs. S3 and S4), indicating that the results do not depend on the specific models of mutation.

1.3 The derivation of equation (1)

In this section, we describe the derivation of equations (1) that is outlined in Methods.

To derive equations (1), we simplified the agent-based model in two ways. First, we assumed that k_{pt}^c is independent of p and t. Under this assumption, a catalyst does not distinguish the replicator types of templates (i.e., $k_{pt}^c = k_{pt'}^c$ for $t \neq t'$) and products (i.e., $k_{pt}^c = k_{pt'}^c$ for $p \neq p'$). This assumption excludes the possibility of numerical symmetry breaking, but still allows catalytic and informatic symmetry breaking as described in the main text (see Results).

Second, we abstracted away chemical reactions by defining ω_{ij}^t as the probability that replicator j of type t in protocell i is replicated or transcribed per unit time. Let $n_{ij}^t(\tau)$

be the population size of this replicator at time τ . Then, the dynamics of $n_{ij}^t(\tau)$ can be mathematically described as

$$\begin{bmatrix} n_{ij}^{\mathrm{P}}(\tau+1)\\ n_{ij}^{\mathrm{Q}}(\tau+1) \end{bmatrix} = \begin{bmatrix} \omega_{ij}^{\mathrm{P}} & \omega_{ij}^{\mathrm{Q}}\\ \omega_{ij}^{\mathrm{P}} & \omega_{ij}^{\mathrm{Q}} \end{bmatrix} \begin{bmatrix} n_{ij}^{\mathrm{P}}(\tau)\\ n_{ij}^{\mathrm{Q}}(\tau) \end{bmatrix}.$$
 (S1)

The fitness of the replicator can be defined as the dominant eigenvalue λ_{ij} of the 2 × 2 matrix on the right-hand side of equation (S1). The equilibrium frequencies of P and Q are given by the right eigenvector v_{ij} associated with λ_{ij} . Fisher's reproductive values of P and Q are given by the corresponding left eigenvector u_{ij} . These eigenvalue and eigenvectors are calculated as follows:

$$\lambda_{ij} = \omega_{ij}^{\mathrm{P}} + \omega_{ij}^{\mathrm{Q}}, \qquad \boldsymbol{v}_{ij} = \begin{bmatrix} 1\\1 \end{bmatrix}, \qquad \boldsymbol{u}_{ij} = \begin{bmatrix} \omega_{ij}^{\mathrm{P}} & \omega_{ij}^{\mathrm{Q}} \end{bmatrix}.$$
(S2)

Based on the above simplification, we now derive equations (1). For concreteness, we focus on the evolution of the average catalytic activity of P (denoted by \bar{k}^{P} in the main text). However, the same method of derivation is applicable to that of Q if P and Q are swapped.

Let $\kappa_{ij}^{\rm P}$ be the catalytic activity of replicator j of type P in protocell i (we use κ instead of k to distinguish $\kappa_{ij}^{\rm P}$ from $k_{pt}^{\rm P}$). Price's equation [1, 2] states that

$$\langle \lambda_{\tilde{i}\tilde{j}} \rangle \Delta \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \rangle = \sigma_{\tilde{i}}^{2} [\langle \lambda_{i\tilde{j}} \rangle, \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle] + \mathbb{E}_{\tilde{i}} [\sigma_{i\tilde{j}}^{2} [\lambda_{ij}, \kappa_{ij}^{\mathrm{P}}]]$$
(S3)

where $\langle x_{i\tilde{j}} \rangle$, $\langle x_{\tilde{i}\tilde{j}} \rangle$, and $\mathbb{E}_{\tilde{i}}[x]$ are x averaged over the indices marked with tildes, $\sigma_{\tilde{i}}^2[x, y]$ is the covariance between x and y over protocells, and $\sigma_{\tilde{i}\tilde{j}}^2[x, y]$ is the covariance between x and y over the replicators in protocell i (one replicator is always counted as one sample in calculating all moments). Below, we show that equation (S3) is approximated by equations (1) up to the second moments of $\langle \kappa_{i\tilde{j}}^{\rm P} \rangle$ and $\kappa_{ij}^{\rm P}$, namely, $\sigma_{\tilde{i}}^2[\langle \kappa_{i\tilde{j}}^{\rm P} \rangle, \langle \kappa_{i\tilde{j}}^{\rm P} \rangle]$ and $\mathbb{E}_{\tilde{i}}[\sigma_{i\tilde{i}}^2[\kappa_{ij}^{\rm P}, \kappa_{ij}^{\rm P}]]$.

To approximate the first term on the right-hand side of equation (S3), we assume that $\langle \lambda_{i\tilde{j}} \rangle$ is a function of $\langle \kappa_{i\tilde{j}}^{\rm P} \rangle$ and $\langle \kappa_{i\tilde{j}}^{\rm Q} \rangle$ that can be expanded as a Taylor series around $\langle \kappa_{i\tilde{j}}^{\rm P} \rangle$ and $\langle \kappa_{i\tilde{j}}^{\rm Q} \rangle$. Substituting this series into $\sigma_{\tilde{i}}^2 [\langle \lambda_{i\tilde{j}} \rangle, \langle \kappa_{i\tilde{j}}^{\rm P} \rangle]$, we obtain

$$\sigma_{\tilde{i}}^{2}[\langle\lambda_{i\tilde{j}}\rangle,\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle] = \sum_{c\in\{\mathrm{P},\mathrm{Q}\}} \frac{\partial\langle\lambda_{i\tilde{j}}\rangle}{\partial\langle\kappa_{i\tilde{j}}^{c}\rangle} \sigma_{\tilde{i}}^{2}[\langle\kappa_{i\tilde{j}}^{\mathrm{P}}\rangle,\langle\kappa_{i\tilde{j}}^{c}\rangle] + O(\sigma_{\tilde{i}}^{3}), \tag{S4}$$

where $O(\sigma_{\tilde{i}}^3)$ consists of terms involving the third or higher (mixed) central moments of $\langle \kappa_{\tilde{i}\tilde{j}}^{\rm P} \rangle$ and $\langle \kappa_{\tilde{i}\tilde{j}}^{\rm Q} \rangle$ over protocells [3].

To approximate the second term on the right-hand side of equation (S3), we likewise assume that λ_{ij} is a function of $\kappa_{ij}^{\rm P}$ and $\kappa_{ij}^{\rm Q}$ that can be expanded as a Taylor series around $\langle \kappa_{ij}^{\rm P} \rangle$ and $\langle \kappa_{ij}^{\rm Q} \rangle$. Substituting this series into $\sigma_{ij}^2 [\lambda_{ij}, \kappa_{ij}^{\rm P}]$, we obtain

$$\sigma_{i\tilde{j}}^{2}[\lambda_{ij},\kappa_{ij}^{\mathrm{P}}] = \sum_{c \in \{\mathrm{P},\mathrm{Q}\}} \frac{\partial \lambda_{ij}}{\partial \kappa_{ij}^{c}} \sigma_{i\tilde{j}}^{2}[\kappa_{ij}^{\mathrm{P}},\kappa_{ij}^{c}] + O(\sigma_{i\tilde{j}}^{3}),$$

where $O(\sigma_{i\bar{j}}^3)$ consists of terms involving the third or higher (mixed) central moments of $\kappa_{ij}^{\rm P}$ and $\kappa_{ij}^{\rm Q}$ over the replicators in protocell *i* [3]. Applying $\mathbb{E}_{\tilde{i}}$ to both sides of the above

equation and assuming that $\partial \lambda_{ij} / \partial \kappa_{ij}^c$ is independent of $\sigma_{ij}^2[\kappa_{ij}^{\rm P}, \kappa_{ij}^c]$, we obtain

$$\mathbb{E}_{\tilde{i}}\left[\sigma_{i\tilde{j}}^{2}[\lambda_{ij},\kappa_{ij}^{\mathrm{P}}]\right] = \sum_{c \in \{\mathrm{P},\mathrm{Q}\}} \mathbb{E}_{\tilde{i}}\left[\frac{\partial \lambda_{ij}}{\partial \kappa_{ij}^{c}}\right] \mathbb{E}_{\tilde{i}}\left[\sigma_{i\tilde{j}}^{2}[\kappa_{ij}^{\mathrm{P}},\kappa_{ij}^{c}]\right] + \mathbb{E}_{\tilde{i}}\left[O(\sigma_{i\tilde{j}}^{3})\right].$$
(S5)

Substituting equations (S4) and (S5) into equation (S3), we obtain

$$\Delta \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \rangle = \frac{1}{\langle \lambda_{\tilde{i}\tilde{j}} \rangle} \sum_{c \in \{\mathrm{P},\mathrm{Q}\}} \left(\frac{\partial \langle \lambda_{i\tilde{j}} \rangle}{\partial \langle \kappa_{i\tilde{j}}^{c} \rangle} \sigma_{\tilde{i}}^{2} [\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle, \langle \kappa_{i\tilde{j}}^{c} \rangle] + \mathbb{E}_{\tilde{i}} \left[\frac{\partial \lambda_{ij}}{\partial \kappa_{ij}^{c}} \right] \mathbb{E}_{\tilde{i}} \left[\sigma_{i\tilde{j}}^{2} [\kappa_{ij}^{\mathrm{P}}, \kappa_{ij}^{c}] \right] \right) + O', \quad (\mathrm{S6})$$

where $O' = O(\sigma_{\tilde{i}}^3) + E_{\tilde{i}}[O(\sigma_{\tilde{i}}^3)].$

Next, we assume that covariances $\sigma_{\tilde{i}}^2[\langle \kappa_{i\tilde{j}}^{\rm P} \rangle, \langle \kappa_{i\tilde{j}}^{\rm Q} \rangle]$ and $\mathbb{E}_{\tilde{i}}[\sigma_{i\tilde{j}}^2[\kappa_{ij}^{\rm P}, \kappa_{ij}^{\rm Q}]]$ are negligible because the mutation of $\kappa_{ij}^{\rm P}$ and that of $\kappa_{ij}^{\rm Q}$ are uncorrelated in the simulation model (this assumption is alternatively justified in Supplementary Material Text 1.6). Under this assumption, equation (S6) is transformed into

$$\Delta \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \rangle = \frac{1}{\langle \lambda_{\tilde{i}\tilde{j}} \rangle} \left(\frac{\partial \langle \lambda_{i\tilde{j}} \rangle}{\partial \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle} \sigma_{\tilde{i}}^{2} [\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle, \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle] + \mathbb{E}_{\tilde{i}} \left[\frac{\partial \lambda_{ij}}{\partial \kappa_{ij}^{\mathrm{P}}} \right] \mathbb{E}_{\tilde{i}} \left[\sigma_{i\tilde{j}}^{2} [\kappa_{ij}^{\mathrm{P}}, \kappa_{ij}^{\mathrm{P}}] \right] \right) + O'.$$
(S7)

Using equation (S2) (i.e., $\lambda_{ij} = \omega_{ij}^{\rm P} + \omega_{ij}^{\rm Q}$), we can transform equation (S7) into

$$\Delta \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \rangle = \frac{1}{\langle \lambda_{\tilde{i}\tilde{j}} \rangle} \sum_{t \in \{\mathrm{P},\mathrm{Q}\}} \left(\frac{\partial \langle \omega_{i\tilde{j}}^{t} \rangle}{\partial \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle} \sigma_{\tilde{i}}^{2} [\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle, \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle] + \mathbb{E}_{\tilde{i}} \left[\frac{\partial \omega_{ij}^{t}}{\partial \kappa_{ij}^{\mathrm{P}}} \right] \mathbb{E}_{\tilde{i}} \left[\sigma_{i\tilde{j}}^{2} [\kappa_{ij}^{\mathrm{P}}, \kappa_{ij}^{\mathrm{P}}] \right] \right) + O'.$$
(S8)

Moreover, it can be shown that

$$\begin{split} \mathbb{E}_{\tilde{i}} \begin{bmatrix} \frac{\partial \omega_{ij}^{t}}{\partial \kappa_{ij}^{c}} \Big|_{\substack{\kappa_{ij}^{\mathrm{P}} = \langle \kappa_{ij}^{\mathrm{P}} \rangle \\ \kappa_{ij}^{\mathrm{Q}} = \langle \kappa_{ij}^{\mathrm{Q}} \rangle} \end{bmatrix} &= \mathbb{E}_{\tilde{i}} \begin{bmatrix} \omega_{ij}^{t} (\langle \kappa_{ij}^{\mathrm{P}} \rangle, \langle \kappa_{ij}^{\mathrm{Q}} \rangle) \frac{\partial \ln \omega_{ij}^{t}}{\partial \kappa_{ij}^{c}} \Big|_{\substack{\kappa_{ij}^{\mathrm{P}} = \langle \kappa_{ij}^{\mathrm{Q}} \rangle \\ \kappa_{ij}^{\mathrm{Q}} = \langle \kappa_{ij}^{\mathrm{Q}} \rangle} \end{bmatrix} \\ &= \mathbb{E}_{\tilde{i}} \begin{bmatrix} \omega_{ij}^{t} (\langle \kappa_{ij}^{\mathrm{P}} \rangle, \langle \kappa_{ij}^{\mathrm{Q}} \rangle) \end{bmatrix} \mathbb{E}_{\tilde{i}} \begin{bmatrix} \frac{\partial \ln \omega_{ij}^{t}}{\partial \kappa_{ij}^{c}} \Big|_{\substack{\kappa_{ij}^{\mathrm{P}} = \langle \kappa_{ij}^{\mathrm{P}} \rangle \\ \kappa_{ij}^{\mathrm{Q}} = \langle \kappa_{ij}^{\mathrm{Q}} \rangle} \end{bmatrix} + O(\sigma_{i}^{2}) \\ &= \langle \omega_{\tilde{i}j}^{t} \rangle \mathbb{E}_{\tilde{i}} \begin{bmatrix} \frac{\partial \ln \omega_{ij}^{t}}{\partial \kappa_{ij}^{c}} \Big|_{\substack{\kappa_{ij}^{\mathrm{P}} = \langle \kappa_{ij}^{\mathrm{P}} \rangle \\ \kappa_{ij}^{\mathrm{Q}} = \langle \kappa_{ij}^{\mathrm{Q}} \rangle} \end{bmatrix} + \mathbb{E}_{\tilde{i}} \begin{bmatrix} O(\sigma_{ij}^{2}) \end{bmatrix} + O(\sigma_{i}^{2}). \end{split}$$

Using the above equation, we can transform equation (S8) into

$$\Delta \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \rangle = \sum_{t \in \{\mathrm{P},\mathrm{Q}\}} \frac{\langle \omega_{\tilde{i}\tilde{j}}^{t} \rangle}{\langle \lambda_{\tilde{i}\tilde{j}} \rangle} \left(\frac{\partial \ln \langle \omega_{i\tilde{j}}^{t} \rangle}{\partial \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle} \sigma_{\tilde{i}}^{2} [\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle, \langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle] + \mathbb{E}_{\tilde{i}} \left[\frac{\partial \ln \omega_{ij}^{t}}{\partial \kappa_{ij}^{\mathrm{P}}} \right] \mathbb{E}_{\tilde{i}} \left[\sigma_{i\tilde{j}}^{2} [\kappa_{ij}^{\mathrm{P}}, \kappa_{ij}^{\mathrm{P}}] \right] \right) + O'', \quad (S9)$$

where $O'' = O' + O(\sigma_{\tilde{i}}^2) \mathbb{E}_{\tilde{i}}[O(\sigma_{\tilde{i}}^2)] + \mathbb{E}_{\tilde{i}}[O(\sigma_{\tilde{i}}^2)]\mathbb{E}_{\tilde{i}}[O(\sigma_{\tilde{i}}^2)]$. We adopt the following notation:

$$\begin{split} \bar{\omega}^t &= \frac{\langle \omega_{\tilde{i}\tilde{j}}^t \rangle}{\langle \lambda_{\tilde{i}\tilde{j}} \rangle}, \qquad \qquad \sigma_{\rm cel}^2 = \sigma_{\tilde{i}}^2 [\langle \kappa_{i\tilde{j}}^{\rm P} \rangle, \langle \kappa_{i\tilde{j}}^{\rm P} \rangle], \qquad \qquad \sigma_{\rm mol}^2 = \mathbb{E}_{\tilde{i}} [\sigma_{i\tilde{j}}^2 [\kappa_{ij}^{\rm P}, \kappa_{ij}^{\rm P}]], \\ \bar{k}^{\rm P} &= \langle \kappa_{\tilde{i}\tilde{j}}^{\rm P} \rangle, \qquad \qquad \gamma_{\rm P}^{\rm P} = -\mathbb{E}_{\tilde{i}} \left[\frac{\partial \ln \omega_{ij}^{\rm P}}{\partial \kappa_{ij}^{\rm P}} \right], \qquad \qquad \beta_{\rm P}^t = \frac{\partial \ln \langle \omega_{i\tilde{j}}^t \rangle}{\partial \langle \kappa_{i\tilde{j}}^{\rm P} \rangle}, \end{split}$$

where $\bar{\omega}^t$ is the normalised average reproductive value of type-*t* replicators, σ_{cel}^2 , σ_{mol}^2 , and \bar{k}^P are the simplification of the notation, γ_P^P is an average decrease in the replication rate of a type-P replicator due to an increase in its own catalytic activity, and β_P^t is an increase in the average replication rate of type-*t* replicators in a protocell due to an increase in the average catalytic activity of type-P replicators in that protocell.

We assume that V is so large that $\langle \kappa_{ij}^{\rm P} \rangle$ and $\kappa_{ij}^{\rm P}$ can be regarded as mathematically independent of each other, provided *i* and *j* are fixed (if *i* and *j* are varied, $\langle \kappa_{ij}^{\rm P} \rangle$ and $\kappa_{ij}^{\rm P}$ may be statistically correlated). Under this assumption, increasing $\kappa_{ij}^{\rm P}$ does not increase $\langle \kappa_{ij}^{\rm P} \rangle$, so that $\gamma_{\rm P}^{\rm P}$ reflects only the cost of providing catalysis at the molecular level. Likewise, increasing $\langle \kappa_{ij}^{\rm P} \rangle$ does not increase $\kappa_{ij}^{\rm P}$, so that $\beta_{\rm P}^{\rm t}$ reflects only the benefit of receiving catalysis at the cellular level. Moreover, the independence of $\langle \kappa_{ij}^{\rm P} \rangle$ from $\kappa_{ij}^{\rm P}$ implies that $\partial \omega_{ij}^{\rm Q} / \partial \kappa_{ij}^{\rm P} = 0$, which permits the following interpretation: if a replicator of type P provides more catalysis, its transcripts, which is of type Q, pay no extra cost (i.e., $\gamma_{\rm P}^{\rm Q} = 0$).

Using the above notation and the fact that $\partial \omega_{ij}^{\rm Q} / \partial \kappa_{ij}^{\rm P} = 0$, we can transform equation (S9) into

$$\Delta \bar{k}^{\rm P} \approx \bar{\omega}^{\rm P} (b_{\rm P}^{\rm P} \sigma_{\rm cel}^2 - \gamma_{\rm P}^{\rm P} \sigma_{\rm mol}^2) + \bar{\omega}^{\rm Q} b_{\rm P}^{\rm Q} \sigma_{\rm cel}^2, \tag{S10}$$

where O'' is omitted. Equation (S10) is identical to equations (1).

Finally, to derive the equation for $\Delta \bar{k}^{Q}$ (i.e., $\Delta \langle \kappa_{ij}^{Q} \rangle$), we swap P and Q in the above derivation. Moreover, we assume that $\sigma_{\tilde{i}}^{2}[\langle \kappa_{ij}^{Q} \rangle, \langle \kappa_{ij}^{Q} \rangle] = \sigma_{\tilde{i}}^{2}[\langle \kappa_{ij}^{P} \rangle, \langle \kappa_{ij}^{P} \rangle]$ and $\mathbb{E}_{\tilde{i}}[\sigma_{ij}^{2}[\kappa_{ij}^{Q}, \kappa_{ij}^{Q}]] = \mathbb{E}_{\tilde{i}}[\sigma_{ij}^{2}[\kappa_{ij}^{P}, \kappa_{ij}^{P}]]$ because no difference is a priori assumed between P and Q.

1.4 The mathematical analysis of numerical symmetry breaking

In this section, we show that numerical symmetry breaking occurs because while it is neither favoured nor disfavoured by molecular-level selection, it is favoured by cellularlevel selection if catalytic and informatic symmetry breaking has occurred. To this end, we will again simplify the agent-based model into mathematical equations in a mannar analogous to that used to derive equations (1).

Before describing the mathematical analysis, we first need to note that the proximate as opposed to ultimate—cause of numerical symmetry breaking is the self-replication of catalysts (i.e., $k_{cc}^c > 0$, where c is the replicator type of catalysts) in the absence of the reverse transcription of catalysts (i.e., $k_{tc}^c = 0$, where t is the replicator type of templates). This fact can be inferred from the following two results. First, when catalytic, informatic, and numerical symmetry breaking occurs, the replication and transcription of templates are catalysed at about the same rate, i.e., $k_{tt}^c \approx k_{ct}^c$ (Fig. 2b). Therefore, the replication and transcription of templates cannot cause numerical asymmetry. Second, when catalytic and informatic symmetry breaking occurs without numerical symmetry breaking, the selfreplication of catalysts is absent (Fig. S5). Taken together, these results indicate that the proximate cause of numerical symmetry breaking is the self-replication of catalysts in the absence of the reverse transcription of catalysts. Therefore, to understand why numerical symmetry breaking occurs, we need to understand why the self-replication of catalysts evolves if catalytic and informatic symmetry breaking has occurred.

To address the above question, we assume that replicators have already undergone catalytic and informatic symmetry breaking and consider how the fitness of those replicators depends on the self-replication of catalysts. The population dynamics of replicators with catalytic and informatic asymmetry can be described as follows. Let $n_{ij}^t(\tau)$ be the population size of replicator j of type t in protocell i at time τ . Let catalysts and templates be P and Q, respectively. Then, the dynamics of $n_{ij}^t(\tau)$ is mathematically described as follows:

$$\begin{bmatrix} n_{ij}^{\rm P}(\tau+1)\\ n_{ij}^{\rm Q}(\tau+1) \end{bmatrix} = \begin{bmatrix} w_{ij}^{\rm PP} & \omega_{ij}^{\rm Q}\\ 0 & \omega_{ij}^{\rm Q} \end{bmatrix} \begin{bmatrix} n_{ij}^{\rm P}(\tau)\\ n_{ij}^{\rm Q}(\tau) \end{bmatrix},$$
(S11)

where w_{ij}^{PP} is the self-replication probability of catalysts, and ω_{ij}^{Q} is the replication and transcription probabilities of templates, which are assumed to be identical to each other. The fitness of replicators can be defined as the dominant eigenvalue (denoted by λ_{ij}) of the 2 × 2 matrix on the right-hand side of equation (S11):

$$\lambda_{ij} = \begin{cases} \omega_{ij}^{\mathbf{Q}} & \text{if } \omega_{ij}^{\mathbf{Q}} > w_{ij}^{\mathbf{PP}} \\ w_{ij}^{\mathbf{PP}} & \text{otherwise.} \end{cases}$$
(S12)

The associated right eigenvector, which determines the stationary frequencies of P and Q, is

$$\boldsymbol{v}_{ij} = \begin{cases} \frac{1}{2 - w_{ij}^{\text{PP}}/\omega_{ij}^{\text{Q}}} \begin{bmatrix} 1\\ 1 - w_{ij}^{\text{PP}}/\omega_{ij}^{\text{Q}} \end{bmatrix} & \text{if } \omega_{ij}^{\text{Q}} > w_{ij}^{\text{PP}} \\ & \\ & \\ \begin{bmatrix} 1\\ 0 \end{bmatrix} & \text{otherwise.} \end{cases}$$
(S13)

Equation (S13) shows that we must assume $\omega_{ij}^{\text{Q}} > w_{ij}^{\text{PP}}$ in order for P and Q to coexist. Equation (S13) also shows that the frequency of catalysts (i.e., $1/(2 - w_{ij}^{\text{PP}}/\omega_{ij}^{\text{Q}})$) increases with the self-replication of catalysts (i.e., w_{ij}^{PP}), as stated in the beginning of this section.

We first examine whether the self-replication of catalysts is favoured by molecularlevel selection. To this end, we consider how the fitness of replicators (i.e., λ_{ij}) depends on the self-replication of catalysts (i.e., w_{ij}^{PP}). According to equation (S12), λ_{ij} does not directly depend on w_{ij}^{PP} . However, λ_{ij} can indirectly depend on w_{ij}^{PP} because λ_{ij} increases with the frequency of catalysts in a protocell (i.e., $\mathbb{E}_{i\tilde{j}}[1/(2 - w_{i\tilde{j}}^{\text{PP}}/\omega_{i\tilde{j}}^{\text{Q}})])$. This frequency increases with w_{ij}^{PP} if V is so small that a particular replicator can influence the frequency of catalysts in the protocell. However, if λ_{ij} increases with w_{ij}^{PP} , the average fitness of replicators in the protocell (i.e., $\langle \lambda_{i\tilde{j}} \rangle$) must also increase. Therefore, we need to consider the relative fitness (i.e., $\lambda_{ij}/\langle \lambda_{i\tilde{j}} \rangle$). The relative fitness is independent of w_{ij}^{PP} because catalysis is equally shared among templates within a protocell. Therefore, the self-replication of catalysts is neither favoured not disfavoured by molecular-level selection.

We next examine whether the self-replication of catalysts is favoured by cellular-level selection. To this end, we consider how the fitness of a protocell depends on the average self-replication of catalysts in that protocell (i.e., $\langle w_{ij}^{\rm PP} \rangle$). The fitness of a protocell can be defined as the average fitness of the replicators in that protocell (i.e., $\langle \lambda_{ij} \rangle$). According to equation (S12), $\langle \lambda_{ij} \rangle$ does not directly depend $\langle w_{ij}^{\rm PP} \rangle$. However, $\langle \lambda_{ij} \rangle$ indirectly depends on $\langle w_{ij}^{\rm PP} \rangle$ because $\langle \lambda_{ij} \rangle$ increases with the frequency of catalysts in a protocell (i.e., $\mathbb{E}_{ij}[1/(2 - w_{ij}^{\rm PP}/\omega_{ij}^{\rm Q})])$). This frequency increases with $\langle w_{ij}^{\rm PP} \rangle$, so that $\langle \lambda_{ij} \rangle$ must also increase with $\langle w_{ij}^{\rm PP} \rangle$. Therefore, the self-replication of catalysts is favoured by cellular-level selection.

Taken together, the above considerations indicate that the self-replication of catalysts is neutral with respect to molecular-level selection, but advantageous with respect to cellular-level selection. Therefore, numerical symmetry breaking results from the maximisation of fitness at the cellular level in the presence of catalytic and informatic asymmetry.

Finally, we mention an important consequence of numerical symmetry breaking. Numerical symmetry breaking causes a bottleneck effect on the population of replicators within a protocell. This bottleneck effect increases among-cell variance relative to within-cell variance (i.e., $\sigma_{cel}^2/\sigma_{mol}^2$); therefore, it has a stabilising effect on protocells [4, 5]. In this regard, numerical symmetry breaking can be compared to life-cycle bottlenecks displayed by multicellular organisms and eusocial colonies (i.e., an organism or colony develops from only one or a few propagules), which are considered to reduce within-group conflict [6–8].

1.5 The hierarchical Wright-Fisher model

In this section, we describe a model that stochastically simulates the population dynamics described by equations (1), in which σ_{mol}^2 and σ_{cel}^2 are treated as dynamic variables dependent on m and V.

The simplifications involved in the derivation of equations (1), while illuminating, make the comparison between equations (1) and the agent-based model indirect. Specifically, equations (1) cannot be compared with the agent-based model in terms of the same parameters, because the equations treat $\sigma_{\rm mol}^2$ and $\sigma_{\rm cel}^2$ as parameters, which are actually dynamic variables dependent on m and V in the agent-based model. To fill this gap, we constructed a model that stochastically simulates the population dynamics described by equations (1) and treats $\sigma_{\rm mol}^2$ and $\sigma_{\rm cel}^2$ as dynamic variables dependent on m and V. This model is formulated as a hierarchical Wright-Fisher process. Replicators are

This model is formulated as a hierarchical Wright-Fisher process. Replicators are partitioned into a number of groups (hereafter, protocells). Each replicator is individually assigned replicator type $c \in \{P, Q\}$ and two k^c values. The fitness of a replicator is calculated according to equation (S14). In each generation, replicators are replicated or transcribed with probabilities proportional to ω_{ij}^c , so that the population dynamics matches equation (S1) on average. After the replication-transcription step, the protocells containing greater than V replicators are divided with their replicators randomly distributed between the two daughter cells. The protocells containing no replicators are discarded.

The mutation of k^c is modelled as unbiased random walks with reflecting boundaries. That is, each k^c value of a replicator is mutated with a probability m per replication or transcription by adding a number randomly drawn from a uniform distribution on the interval $(-\delta_{\text{mut}}, \delta_{\text{mut}})$ ($\delta_{\text{mut}} = 0.1$). The values of k^c are bounded in [0, 1] with reflecting boundaries at both bounds.

To determine the condition for symmetry breaking, we simulated the above Wright-Fisher model for various values of V and m. The simulations show that symmetry breaking occurs only if V and m are sufficiently large (Fig. S8), a result that is consistent with the outcomes of the original agent-based model (Fig. 2). Given that the Wright-Fisher model involves many of the simplifications involved in equations (1), the above consistency supports the validity of the symmetry breaking mechanism described by equations (1).

1.6 The phase-plane analysis

In this section, we describe the phase-plane analysis outlined in Methods.

To perform the phase-plane analysis depicted in Fig. 3, we adapted equations (1) by defining ω_{ij}^t as a specific function of κ_{ij}^t (see the previous section for the meaning of ω_{ij}^t

and κ_{ij}^t). The following definition was employed:

$$\omega_{ij}^{t} = e^{\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle + \langle \kappa_{i\tilde{j}}^{\mathrm{Q}} \rangle} \frac{e^{-s\kappa_{ij}^{\mathrm{P}}}}{\langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{P}}} \rangle + \langle e^{-s\kappa_{i\tilde{j}}^{\mathrm{Q}}} \rangle}.$$
(S14)

where the factor $e^{\langle \kappa_{ij}^{\rm P} \rangle + \langle \kappa_{ij}^{\rm Q} \rangle}$ represents the cellular-level benefit of catalysis provided by the replicators in protocell *i*, the numerator $e^{-s\kappa_{ij}^t}$ represents the molecular-level cost of catalysis provided by the focal replicator, the denominator $1/(\langle e^{-s\kappa_{ij}^{\rm P}} \rangle + \langle e^{-s\kappa_{ij}^{\rm Q}} \rangle)$ normalises the cost, and *s* is the cost-benefit ratio. The above definition of ω_{ij}^t was chosen to satisfy the requirement that a replicator faces the trade-off between providing catalysis and serving as a template, so that γ_t^t and β_c^t are positive; for example, if the cost γ_t^t were negative, it would actually be a benefit, so that there would be no trade-off. This requirement is satisfied if $\partial \omega_{ij}^t / \partial \kappa_{ij}^t < 0$ and $\partial \langle \omega_{ij}^t \rangle / \partial \langle \kappa_{ij}^c \rangle > 0$ for c = t and $c \neq t$. Apart from this requirement, the definition was arbitrarily chosen for simplicity.

Under the definition of ω_{ij}^t in equation (S14), we obtain equations describing the evolution of $\langle \kappa_{ij}^c \rangle$ (denoted as k^c in the main text) as follows. Since the evolution of $\langle \kappa_{ij}^c \rangle$ is described by equation (S6), we substitute equation (S14) into equation (S6). For this substitution, we need to calculate the derivatives of fitness. According to equation (S2), the fitness of a replicator is $\lambda_{ij} = \omega_{ij}^{\rm P} + \omega_{ij}^{\rm Q}$. Therefore,

$$\begin{split} \mathbb{E}_{\tilde{i}} \bigg[\frac{\partial \lambda_{ij}}{\partial \kappa_{ij}^{c}} \bigg|_{\substack{\kappa_{ij}^{\mathrm{P}} = \langle \kappa_{ij}^{\mathrm{P}} \rangle \\ \kappa_{ij}^{\mathrm{Q}} = \langle \kappa_{ij}^{\mathrm{Q}} \rangle}}{\sum_{\substack{\alpha \in \mathcal{I} \\ \kappa_{ij}^{\mathrm{Q}} = \langle \kappa_{ij}^{\mathrm{Q}} \rangle \\ \kappa_{ij}^{\mathrm{Q}} = \langle \kappa_{ij}^{\mathrm{Q}} \rangle}} = -ce^{\langle \kappa_{ij}^{\mathrm{P}} \rangle + \langle \kappa_{ij}^{\mathrm{Q}} \rangle} \frac{e^{-s \langle \kappa_{ij}^{\mathrm{P}} \rangle}}{e^{-s \langle \kappa_{ij}^{\mathrm{Q}} \rangle}} + \langle e^{-s\kappa_{ij}^{\mathrm{Q}}} \rangle} \bigg] \\ = -ce^{\langle \kappa_{ij}^{\mathrm{P}} \rangle + \langle \kappa_{ij}^{\mathrm{Q}} \rangle} \frac{e^{-s \langle \kappa_{ij}^{\mathrm{P}} \rangle}}{e^{-s \langle \kappa_{ij}^{\mathrm{P}} \rangle} + e^{-s \langle \kappa_{ij}^{\mathrm{Q}} \rangle}} + \mathbb{E}_{\tilde{i}} \big[O(\sigma_{ij}^{2}) \big] + O(\sigma_{\tilde{i}}^{2}). \end{split}$$

Moreover, the average fitness of replicators in a protocell is $\langle \lambda_{i\tilde{j}} \rangle = e^{\langle \kappa_{i\tilde{j}}^{\mathrm{P}} \rangle + \langle \kappa_{i\tilde{j}}^{\mathrm{Q}} \rangle}$, so

$$\left. \frac{\partial \langle \lambda_{i\tilde{j}} \rangle}{\partial \langle \kappa^c_{i\tilde{j}} \rangle} \right|_{\substack{\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle = \langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle \\ \langle \kappa^Q_{i\tilde{j}} \rangle = \langle \kappa^Q_{i\tilde{j}} \rangle}} = e^{\langle \kappa^{\mathrm{P}}_{i\tilde{j}} \rangle + \langle \kappa^{\mathrm{Q}}_{i\tilde{j}} \rangle}.$$

We substitute these derivatives into equation (S6) and use the fact that

$$\left\langle \lambda_{\tilde{i}\tilde{j}} \right\rangle = e^{\left\langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \right\rangle + \left\langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}} \right\rangle} + O(\sigma_{\tilde{i}}^{2})$$

to obtain

$$\Delta \langle \kappa_{\tilde{i}\tilde{j}}^c \rangle = (1 + \rho_{\rm cel})\sigma_{\rm cel}^2 - s \frac{e^{-s\langle \kappa_{\tilde{i}\tilde{j}}^c \rangle} + \rho_{\rm mol} e^{-s\langle \kappa_{\tilde{i}\tilde{j}}^c \rangle}}{e^{-s\langle \kappa_{\tilde{i}\tilde{j}}^{\rm P} \rangle} + e^{-s\langle \kappa_{\tilde{i}\tilde{j}}^{\rm Q} \rangle}} \sigma_{\rm mol}^2 + O'', \tag{S15}$$

where $c' \neq c$, ρ_{cel} is the correlation coefficient between $\langle \kappa_{ij}^{P} \rangle$ and $\langle \kappa_{ij}^{Q} \rangle$ (i.e., $\rho_{cel} = \sigma_{\tilde{i}}^{2} [\langle \kappa_{ij}^{P} \rangle, \langle \kappa_{ij}^{Q} \rangle] / \sigma_{cel}^{2}$), and ρ_{mol} is the average correlation coefficient between κ_{ij}^{P} and κ_{ij}^{Q} (i.e., $\rho_{mol} = \mathbb{E}_{\tilde{i}} [\sigma_{ij}^{2} [\kappa_{ij}^{P}, \kappa_{ij}^{Q}]] / \sigma_{mol}^{2}$). To derive equation (S15), we have assumed that the variances of $\langle \kappa_{ij}^{c} \rangle$ and κ_{ij}^{c} are independent of c; i.e., $\sigma_{cel}^{2} = \sigma_{\tilde{i}}^{2} [\langle \kappa_{ij}^{c} \rangle, \langle \kappa_{ij}^{c} \rangle]$ and $\sigma_{mol}^{2} = \mathbb{E}_{\tilde{i}} [\sigma_{ij}^{2} [\kappa_{ij}^{c}, \kappa_{ij}^{c}]]$ for c = P and c = Q.

Equation (S15) can be expressed in a compact form as follows:

$$\begin{bmatrix} \Delta \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \rangle \\ \Delta \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}} \rangle \end{bmatrix} = \sigma_{\mathrm{tot}}^2 \nabla [RB - (1-R)C] + O'',$$

where ∇ is a nabla operator (i.e., $\nabla = [\partial/\partial \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} \rangle, \partial/\partial \langle \kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}} \rangle]^{\mathrm{T}}$, where ^T denotes transpose), $\sigma_{\mathrm{tot}}^{2} = \sigma_{\mathrm{mol}}^{2} + \sigma_{\mathrm{cel}}^{2}, R = \sigma_{\mathrm{cel}}^{2}/(\sigma_{\mathrm{cel}}^{2} + \sigma_{\mathrm{mol}}^{2}), B = (1 + \rho_{\mathrm{cel}})(\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} + \kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}), \text{ and } C = (\rho_{\mathrm{mol}} - 1)\ln(e^{-s\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}}} + e^{-s\kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}}) + \rho_{\mathrm{mol}}s(\kappa_{\tilde{i}\tilde{j}}^{\mathrm{P}} + \kappa_{\tilde{i}\tilde{j}}^{\mathrm{Q}}). R \text{ can be interpreted as the regression coefficient of <math>\langle \kappa_{\tilde{i}\tilde{j}}^{c} \rangle$ on κ_{ij}^{c} [9] and, therefore, the coefficient of genetic relatedness [10]. The potential function RB - (1 - R)C can then be interpreted as inclusive fitness.

Next, we set $\rho_{\text{mol}} = 0$ and $\rho_{\text{cel}} = 0$ in equations (S15) and let $\langle \kappa_{\tilde{i}\tilde{j}}^c \rangle$ be denoted by \bar{k}^c , obtaining

$$\Delta \bar{k}^{c} = \sigma_{cel}^{2} - s \frac{e^{-sk^{c}}}{e^{-s\bar{k}^{P}} + e^{-s\bar{k}^{Q}}} \sigma_{mol}^{2} + O''$$

$$= \frac{e^{-s\bar{k}^{c}}}{e^{-s\bar{k}^{P}} + e^{-s\bar{k}^{Q}}} (\sigma_{cel}^{2} - s\sigma_{mol}^{2}) + \frac{e^{-s\bar{k}^{c'}}}{e^{-s\bar{k}^{P}} + e^{-s\bar{k}^{Q}}} \sigma_{cel}^{2} + O'',$$
(S16)

where $c' \neq c$. Comparing equations (S16) and (S10), we infer that

$$\begin{split} \bar{\omega}^c &= \frac{e^{-s\bar{k}^c}}{e^{-s\bar{k}^{\mathrm{P}}} + e^{-s\bar{k}^{\mathrm{Q}}}}, \\ \gamma^c_c &= s, \\ \beta^t_c &= 1, \end{split}$$

which are identical to equations (5).

Next, we omit O'' in equation (S16) and replace Δ with time derivative $d/d\tau$, obtaining

$$\frac{d}{d\tau}\bar{k}^c = \sigma_{\rm cel}^2 - s\frac{e^{-sk^c}}{e^{-s\bar{k}^{\rm P}} + e^{-s\bar{k}^{\rm Q}}}\sigma_{\rm mol}^2.$$
(S17)

Finally, to allow for the restriction on the range of \bar{k}^c (i.e., $\bar{k}^c \in [0, k_{\max}]$), we multiply the right-hand side of equation (S17) with a function, denoted by $\Theta(\bar{k}^c)$, that is 1 if $0 < \bar{k}^c < k_{\max}$ and 0 if $\bar{k}^c = 0$ or $\bar{k}^c = k_{\max}$. Multiplying $\Theta(\bar{k}^c)$ with the right-hand side of equation (S17), we obtain

$$\frac{d}{d\tau}\bar{k}^c = \Theta(\bar{k}^c) \left[\sigma_{\rm cel}^2 - s \frac{e^{-s\bar{k}^c}}{e^{-s\bar{k}^{\rm P}} + e^{-s\bar{k}^{\rm Q}}} \sigma_{\rm mol}^2 \right].$$

The above equation was numerically integrated for s = 1 to obtain the phase-plane portrait depicted in Fig. 3.

Equation (S15) allows for statistical correlations between $\kappa_{ij}^{\rm P}$ and $\kappa_{ij}^{\rm Q}$ at the molecular and cellular levels, i.e., $\rho_{\rm mol}$ and $\rho_{\rm cel}$. Therefore, it can be used to examine the consequence of ignoring these correlations, which is one of the simplifications made in the derivation of equations (1) described in Supplementary Material Text 1.3. For this sake, we calculate the nullcline of $\Delta \langle \kappa_{\tilde{i}\tilde{j}}^c \rangle$. Setting $\Delta \langle \kappa_{\tilde{i}\tilde{j}}^c \rangle = 0$ in equation (S15) and omitting O'', we obtain

$$\left\langle \kappa_{\tilde{i}\tilde{j}}^{c'} \right\rangle \approx \left\langle \kappa_{\tilde{i}\tilde{j}}^{c} \right\rangle + s^{-1} \ln \frac{\rho_{\rm mol} s \sigma_{\rm mol}^2 - (1 + \rho_{\rm cel}) \sigma_{\rm cel}^2}{(1 + \rho_{\rm cel}) \sigma_{\rm cel}^2 - s \sigma_{\rm mol}^2}$$

This equation shows that all parameters only appear in the intercept of the nullcline with the $\langle \kappa_{\tilde{i}\tilde{j}}^{c'} \rangle$ -axis. Let us denote this intercept as $s^{-1} \ln I$. The way I qualitatively depends on $\sigma_{\rm cel}^2$ and $s\sigma_{\rm mol}^2$ is independent of $\rho_{\rm cel}$ because $-1 < \rho_{\rm cel} < 1$. Therefore, we

can assume that $\rho_{\rm cel} = 0$ without loss of generality. Next, to see how $\rho_{\rm mol}$ influences I, we focus on the singularity of I by setting $(1 + \rho_{\rm cel})\sigma_{\rm cel}^2 = s\sigma_{\rm mol}^2 + \epsilon$, where $\epsilon > 0$. Then, $I = (1 - \rho_{\rm mol})s\sigma_{\rm mol}^2/\epsilon - \rho_{\rm mol}$. The way I qualitatively depends on $s\sigma_{\rm mol}^2/\epsilon$ is independent of $\rho_{\rm mol}$ because $-1 < \rho_{\rm mol} < 1$. Therefore, we can assume that $\rho_{\rm mol} = 0$ without loss of generality. Taken together, these calculations show that ignoring correlations between $\kappa_{ij}^{\rm P}$ and $\kappa_{ij}^{\rm Q}$ does not qualitatively affect the results, supporting the validity of equations (1).

2 Supporting Figures



Figure S1: The evolutionary dynamics of the agent-based model. **a**, The dynamics of k_{pt}^c averaged over all replicators for parameters corresponding to 'no symmetry breaking' in Fig. 2a: V = 178 and m = 0.01. **b**, Catalytic activities evolved in a. **c**, **d**, Parameters corresponding to 'uncategorised' in Fig. 2a: V = 178 and m = 0.1. **e**, **f**, Parameters corresponding to 'incomplete symmetry breaking' in Fig. 2a: V = 562 and m = 0.01. **g**, **h**, Parameters corresponding to 'incomplete symmetry breaking' in Fig. 2a: V = 178 and m = 0.01. **g**, **h**, Parameters corresponding to 'incomplete symmetry breaking' in Fig. 2a: V = 1778 and m = 0.01.



Figure S2: Symmetry breaking with an alternative definition of complex formation rates (see Supplementary Material Text 1.1). The rate constants of complex formation were defined in such a way that coexistence between P and Q is neither favoured nor disfavoured by cellular-level selection. **a**, Phase diagram with a symmetric initial condition: $k_{pt}^c = 1$ for all combinations of c, p, and t, with both P and Q present at the beginning of each simulation. The symbols are the same as in Fig. 2a, except that the circles include cases in which one replicator type goes extinct. **b**, Dynamics of k_{pt}^c averaged over all replicators for m = 0.01 and V = 10000 in a. **c**, Phase diagram with an asymmetric initial condition: $k_{QQ}^Q = 1$ and $k_{pt}^c = 0$ for all the other combinations of c, p, and t, with only Q present at the beginning of each simulation. The symbols are the same as in a, except that stars indicate the extinction of replicators. **d** Dynamics of k_{pt}^c averaged over all replicators for m = 0.01 and V = 10000 in b.



Figure S3: Symmetry breaking with reflecting mutation (see Supplementary Material Text 1.2). The mutation of k_{pt}^c is modelled as unbiased random walk with reflecting boundaries at 0 and 1. **a**, Phase diagram. The symbols are the same as in Fig. 2a $(t_{\min} > 3.9 \times 10^7 \text{ for } m = 0.1 \text{ and } V = 10000)$. **b** Dynamics of k_{pt}^c averaged over all replicators. m = 0.01 and V = 10000. Three-fold symmetry breaking occurs. **c**, m = 0.0562 and V = 10000. Numerical symmetry breaking is slight. **d**, m = 0.00178 and V = 10000. Numerical symmetry breaking is slight. **e**, **f**, **g**, Catalytic activities evolved in b, c, d, respectively.



Figure S4: Symmetry breaking with log-space mutation (see Supplementary Material Text 1.2). The mutation of k_{pt}^c is modelled as unbiased random walks on a logarithmic scale. **a**, Phase diagram. The symbols are the same as in Fig. 2a ($t_{\min} > 3.9 \times 10^7$ only for m = 0.1 and V = 10000). **b**, Dynamics of k_{pt}^c averaged over all replicators. m = 0.01 and V = 10000. Three-fold symmetry breaking occurs. **c**, m = 0.1 and V = 10000. No numerical symmetry breaking occurs. **d**, m = 0.00178 and V = 10000. No numerical symmetry breaking occurs. **e**, **f**, **g**, Catalytic activities evolved in b, c, d, respectively.



Figure S5: The absence of numerical symmetry breaking for small m and large V (see Supplementary Material Text 1.4). **a**, **b**, The dynamics of k_{pt}^c averaged over all replicators is shown for V = 10000 and m = 0.001 with two different initial conditions: a symmetric initial condition, where $k_{pt}^c = 1$ (a); an asymmetric initial condition, where $k_{PP}^P = 0.95$, $k_{PQ}^P = 0.1$, $k_{QP}^P = 1$, $k_{QQ}^P = 1$, and $k_{pt}^Q = 0.1$ (b). The self-replication of catalysts does not evolve for the symmetric initial condition, whereas it is maintained for the asymmetric initial condition ($t_{\min} > 1.2 \times 10^7$). The dependence of the results on the initial conditions suggests the presence of bistability for V = 10000 and m = 0.001. **c**, **d**, The frequencies of P (catalysts) and Q (templates) are plotted as the functions of time. Numerical symmetry breaking does not occur for the symmetric initial condition, whereas it occurs for the asymmetric initial condition. The results indicate that numerical asymmetry depends on the self-replication of catalysts. **e**, **f**, Catalytic activities evolved for the symmetric initial condition (e) and for the asymmetric initial condition (f).



Figure S6: The effect of symmetry breaking on catalytic activities. The fraction of replicators $1 - N_{\rm S}/N_{\rm tot}$, which is a proxy for the overall catalytic activity of replicators, is shown as a function of m and V, where $N_{\rm S}$ is the total number of S molecules in the system, and $N_{\rm tot} = N_{\rm P} + N_{\rm Q} + N_{\rm S}$. **a**, The original model, which allows symmetry breaking (i.e., Fig. 1). **b**, The model that excludes the possibility of symmetry breaking; specifically, it allows only one type of replicator (either P or Q). Black squares indicate extinction (i.e. $N_{\rm tot} = N_{\rm S}$). $t_{\rm min} > 1.5 \times 10^7$.



Figure S7: Result for large m and V values. The dynamics of the agent-based model is shown for m = 0.1 and $V = 10^5$, parameters outside the range examined in Fig. 2a and Fig. S6a. **a**, The dynamics of k_{pt}^c averaged over all replicators. **b**, The dynamics of the fraction of replicators $1 - N_{\rm S}/N_{\rm tot}$, where $N_{\rm tot}$ and $N_{\rm S}$ are the total numbers of particles and S molecules in the system, respectively. $t_{\rm min} > 1.8 \times 10^6$.



Figure S8: Symmetry breaking in a hierarchical Wright-Fisher model (see Supplementary Material Text 1.5). The model stochastically simulates the population dynamics described by equations (1), treating σ_{mol}^2 and σ_{cel}^2 as variables dependent on m and V (see Supplementary Material Text 1.5). **a**, Phase diagram. Circles indicate no symmetry breaking (i.e., $\bar{k}^{\rm P} \approx \bar{k}^{\rm Q} \approx 1$); diamonds, symmetry breaking (i.e., $\bar{k}^c \approx 0$ and $\bar{k}^{c'} \approx 1$ for $c \neq c'$; stars, extinction (i.e., $\bar{k}^{\rm P} \approx \bar{k}^{\rm Q} \approx 0$). s = 1 (cost-benefit ratio). The total number of replicators was 50V (approximately 130 protocells throughout simulations). The initial condition was $k^{\rm P} = k^{\rm Q} = 1$ for all replicators. Each simulation was run for 4×10^5 generations. The extinction (i.e., $\bar{k}^{\rm P} \approx \bar{k}^{\rm Q} \approx 0$) for large m and V is consistent with the phase-plane analysis of equations (1), which also shows extinction (i.e., $\bar{k}^{\rm P} \approx \bar{k}^{\rm Q} \approx 0$ for sufficiently large $\sigma_{\rm mol}^2 / \sigma_{\rm cel}^2$ (parameters outside the range examined in Fig. 3). The discrepancy between Fig. S8a and Fig. 2a is due the simplifying assumption made in equations (1) that k_{pt}^c is independent of p and t. If k_{pt}^c is allowed to depend on p and t, the flow of information from templates to catalysts can become completely unidirectional. Such unidirectional flow of information can resolve the dilemma between catalysing and templating and leads to the maintenance of high catalytic activities as described in Results. **b**, The dynamics of \bar{k}^c for m = 0.001 and V = 1000 (no symmetry breaking). c, m = 0.01 and V = 1000 (symmetry breaking). d, m = 0.1 and V = 1000(extinction).

References

- [1] Price GR. Extension of covariance selection mathematics. Annals of Human Genetics. 1972;35(4):485–490.
- [2] Hamilton WD. Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox R, editor. Biosocial Anthoroplogy. London: Malaby Press; 1975. p. 133–153.
- [3] Iwasa Y, Pomiankowski A, Nee S. The evolution of costly mate preferences II. the 'handicap' principle. Evolution. 1991;45(6):1431–1442.
- [4] Kaneko K, Yomo T. On a kinetic origin of heredity: minority control in a replicating system with mutually catalytic molecules. Journal of Theoretical Biology. 2002;214(4):563–576.
- [5] Takeuchi N, Hogeweg P, Kaneko K. The origin of a primordial genome through spontaneous symmetry breaking. Nature Communications. 2017;8(1):250.
- [6] Maynard Smith J, Szathmáry E. The Major Transitions in Evolution. Oxford: W. H. Freeman/Spektrum; 1995.
- [7] Michod RE. Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality. Princeton, NJ: Princeton University Press; 1999.
- [8] Bourke AFG. Principles of Social Evolution. Oxford, UK: Oxford University Press; 2011.
- [9] Rice SH. Evolutionary Theory: Mathematical and Conceptual Foundations. Sunderland, MA, USA: Sinauer Associates; 2004.
- [10] Hamilton WD. Selfish and spiteful behaviour in an evolutionary model. Nature. 1970;228(5277):1218–1220.