Supporting Information

Fort 10.1073/pnas.1200662109

SI Text

Parameter Values and Observed Neolithic Front Speed Range. ***Initial growth rate a_N.*** Consider individuals of a farming population N in an area S. For low values of the population density $N(x,y,t)$, the first equation of Eq. **10** can be linearized, and the net reproduction function becomes

$$R_T[N(x,y,t)] \approx e^{a_N T} N(x,y,t). \qquad \textbf{[S1]}$$

Then, integration of Eq. **9** over the whole area S (so that there is no net migration, i.e., the first and second terms on the right-hand side cancel out) yields for a single population ($f = 0$),

$$P_N(t+T) \approx e^{a_N T} P_N(t), \qquad \textbf{[S2]}$$

and the total population number $P_N = \int_S N(x,y,t)\, dS$ grows exponentially, in agreement with observations (later on, the growth becomes gradually slower until the saturation density is reached, thereby yielding the typical S-shaped logistic curve). A continuous-time exponential dynamics $P_N(t) \approx e^{a_N t} P_N(t=0)$, consistent with the discrete-time Eq. **S2**, has been fitted to observed data for several human populations of farmers who settled in spaces originally empty of farming populations. Fits for the Pitcairn, Bass Strait, Tristan da Cunha Islands, and the United States yield the values $a_N = (0.02995 \pm 0.00119)\text{y}^{-1}$, $a_N = (0.02626 \pm 0.00052)\text{y}^{-1}$, $a_N = (0.02527 \pm 0.00032)\text{y}^{-1}$, and $a_N = (0.03135 \pm 0.00063)\text{y}^{-1}$, respectively, which imply the overall range $a_N = (0.028 \pm 0.005)\text{y}^{-1}$, with 80% confidence level (1); this is the range used in Fig. 1.

Besides the ethnographic estimations of a_N above, it is also very interesting that in recent years Guerrero et al. (2) have performed estimations directly from archaeological data based on the rise in fertility (detected as a rise in the proportions of immature skeletons in early Neolithic cemeteries) and a sample of 45 reference historic life tables. In this way they have estimated $a_N = 0.024\text{y}^{-1}$, which is within the range $a_N = (0.028 \pm 0.005)\text{y}^{-1}$ applied in our paper.

Generation time T. A theory of front propagation with distributed delays (3) (i.e., with several possible values of T and a probability for each possible value) showed that T is the mean age difference between a person and all of her/his children, not just with the oldest one (the latter is sometimes called the generation time in demography). A statistical analysis of the observed values of T for preindustrial farmers yields the range $(T = 32 \pm 3)$ y, with 80% confidence level (4); this is the range used in Fig. 1.

Dispersal kernel. A dispersal kernel is defined as the probability of motion as a function of the distance between the initial and final locations. Strictly, when analyzing population front spread, the dispersal kernel should be estimated as the probability as a function of distance between birthplaces of a parent and his/her children (3). Unfortunately, such kernels have not been measured for any preindustrial farming population. However, several authors have considered the following approximations to estimate the dispersal kernel of preindustrial farmers:

i) Mating distances are often used (5) and are defined as the distances between the birthplaces of spouses. For preindustrial farmers, such distances have been measured for the Issocongos in the Central African Republic (5). The corresponding probabilities and distances are $\{p_j\} = \{0.42; 0.23; 0.16; 0.08; 0.07; 0.02; 0.01; 0.01\}$, $\{\Delta_j\} = \{2.3; 7.3; 15; 25; 35; 45; 55; 100\}$ km, where the distances are the central points in the published histogram (1); this is the kernel used in Fig. 1 in the main paper.

ii) Another approximation is given by distances between the birthplace and the current place of residence. For preindustrial farmers, reliable data of this kind are only known for the Majangir in Ethiopia (1, 5, 6), but unfortunately only for age groups 10–19 y old and 20–29 y old. For our purposes, the ideal data would be for 29–35 y old (i.e., the range of T found above). The most reasonable of these data are thus those in the 20- to 29-y interval. They yield the following probabilities and distances (1, 6): $\{p_j\} = \{0.40; 0.17; 0.17; 0.26\}$, $\{r_j\} = \{2.4; 14.5; 36.2; 60.4\}$ km. Note that this kernel is less detailed than that in (*i*) above, so this is an additional reason to consider computations based on this kernel as less precise than those in the main text, but despite this, its predictions (Fig. S1) lead to the same conclusions as Figs. 1 and 2, as we now explain. In Fig. S1*A*, the observed range of C (hatched vertical rectangle), the observed speed range (hatched horizontal rectangle), and the predicted speed range (i.e., that between the full and dashed curves) have a consistency interval (black area, $1.0 \le C \le 1.3$). For this interval, in Fig. S1*B* the cultural effect is $27 \pm 3\%$, still lower than the effect for the Issocongos kernel used in the main text ($40 \pm 8\%$). Thus, the main conclusion (that the demic effect was larger than the cultural one) remains the same.

iii) There are also estimations of mating distances for nonagriculturalist populations, such as hunter-gatherers (7) (Agta, Aka, and !Kung) and horticulturalists (7) (Yanomano), but they are not necessary here.

Conversion intensity C. This parameter obviously depends on the cultural trait transmitted, agriculture in our case, and must be estimated from the observed cultural dynamics of hunter-gatherers becoming farmers. Quantitative observations of this kind exist for the Agta living in the Philippines (8), but they correspond to the case $P_N \gg P_P$, and in this limit the interaction term (8) becomes $I = fP_P$, so we cannot estimate the parameter $C = f/\gamma$, which appears in Eq. **5** for the front speed. However, if $P_N \ll P_P$, the interaction term (8) becomes $I = CP_N$, so we obtain from Eq. **1** that $P_N' = (1+C)P_N$, i.e., $C = \Delta P_N / P_N$, where $\Delta P_N = P_N' - P_N$ is the number of hunter-gatherers converted to farming per generation. Fortunately, some quantitative observations for this case ($P_N \ll P_P$) exist for the Ache living in Paraguay (9). Consider first the following example. In the 1970s a Protestant missionary and his family contacted with a band of 28 Ache hunter-gatherers and converted them to farming. If we take into account that about half of the hunter-gatherers died due to virgin soil epidemics in all contacts without medical attention (9), then $\Delta P_N = 14$ and, assuming that the number of individuals of the missionary family P_N was in the range 3–6, we find the range 2.3–4.7 for C. Similar estimations yield the C ranges 5.3–9.7 for a Catholic mission ($\Delta P_N = 26$), 1.0–2.0 for an Evangelical mission ($\Delta P_N = 6$), and 5.5–10.9 for a farm that became a reservation ($\Delta P_N = 109$, $10 \le P_N \le 20$) (9). Therefore, we have used the overall range 1.0–10.9 for C in our paper. Note that the upper limit ($C = 10.9$) would be larger if the deaths due to virgin soil epidemics were neglected, but the observed and predicted speeds would imply the same consistency range in Fig. 1 (namely, $1.0 \le C \le 2.5$), so Fig. 2 and all of the results in our paper would not be affected. Note also that if some other case study yielded lower values of C, then the cultural effect would be still smaller (Fig. 2 and Fig. S1*B*), thus the main conclusion of the paper (that the demic contribution was more important than the cultural one) would not change.

Speed range. The speed of the Neolithic transition has been estimated previously from an analysis of 735 Neolithic sites in the Near East and Europe (4). Linear fits were performed by computing great-circle

and shortest-path distances. Great-circle distances are defined as shortest paths between two geographic points on the Earth, considered a sphere. Shortest-path distances take into account the Mediterranean but allow for short sea trips, as implied by the existence of Neolithic sites on islands such as Cyprus (for details, see supplemental text 2 in ref. 4). Several Neolithic origins were used to compute distances, and the origin yielding the highest correlation coefficient ($R > 0.8$ in all cases) was used to estimate the speed range using dates vs. distances regressions (they should be preferred to distances vs. dates, because radiocarbon dates have some error, whereas the great circle-to-shortest path comparison is devised to take care of distance uncertainties). Great-circle distances yielded the speed range 0.9–1.1 km/y (4), whereas shortest-path distances yielded 1.1–1.3 km/y (95% confidence intervals). The overall range is thus 0.9–1.3 km/y, which is the range used in Fig. 1.

Frequency-Dependent Cultural Transmission. Here we will show that frequency-dependent models of cultural transmission do not change the conclusions of our paper.

In the beginning of *Methods*, we introduced a model (10) that leads to the result that the probability that a P individual becomes N is fu, i.e., proportional to the N frequency $u = \frac{P_N}{P_N + P_P}$. However, in many examples of cultural transmission, f is not a constant but depends on u. Usually, f is replaced by $f + h[2u - 1]$ (11–13), where the terms within the square brackets give the effect of so-called frequency-dependent biases, and the parameter f includes (if they exist) the effect of other biases (11, 12) (namely, direct biases, i.e., those due to the intrinsic merit of the cultural trait, and indirect biases, e.g., those due to the prestige of individuals) (14). The parameter h measures the strength of frequency-dependent transmission, which is also called conformist transmission, and its minimum possible value is $h = 0$ (then our model in the main text is recovered) (11–13). Eq. **1** is thus generalized into

$$\begin{cases} P_N' = P_N + \dfrac{P_N P_P}{P_N + \gamma P_P}\left(f + h\left[2\dfrac{P_N}{P_N + P_P} - 1\right]\right) \\ P_P' = P_P - \dfrac{P_N P_P}{P_N + \gamma P_P}\left(f + h\left[2\dfrac{P_N}{P_N + P_P} - 1\right]\right) \end{cases}. \qquad \textbf{[S3]}$$

It is easy to see that using Eq. **S3** instead of Eq. **1**, the final result for the front speed (Eq. **5**) is the same with C replaced by $C - h$. However, the speed is an increasing function of C (Fig. 1 and Fig. S1*A*). Therefore, for any value of C, frequency-dependent transmission ($h \neq 0$) will lead to a slower speed than the model considered in the main text ($h = 0$). Thus, the cultural effect (Fig. 2 and Fig. S1*B*) will be surely smaller. Moreover, this result would not change for other frequency-dependent generalizations because frequency-dependent effects are well known to lead to slower cultural transmission for $u \approx 0$ (11, 12) than the non–frequency-dependent case ($h = 0$), and this is precisely the reason why the front speed (and thus the cultural effect) is smaller than in the main text ($h = 0$). Thus, we can safely conclude that demic diffusion will be more important than cultural diffusion.

The spatial spread of frequency-dependent cultural traits can be also formulated in terms of payoffs depending on frequency (15, 16), although this does not seem necessary for the case of agriculture. Moreover, in such more complicated models it may be extremely difficult (or even impossible) to estimate all of the necessary parameter values from empirical data to compare the results of the models to archaeological observations.

Nonlocal Cultural Transmission. The model in Eqs. **3** and **4** is such that without cultural transmission ($f = 0$) a wave of advance propagates, but without demic diffusion [$\phi_N(\Delta_x, \Delta_y) = 0$] a wave cannot propagate, due to the assumption that cultural transmission takes place locally, i.e., kernels are used for the dispersal of individuals but not of culture. The transmission of culture is modeled by the following term in Eq. **4**:

$$f\frac{R_T[N(x,y,t)]R_T[P(x,y,t)]}{R_T[N(x,y,t)] + \gamma R_T[P(x,y,t)]}, \qquad \textbf{[S4]}$$

where all population densities are evaluated at the same spatial point, (x,y); this is the reason why the transmission of culture is local in this model. Let us now include an additional, nonlocal mechanism of culture transmission. For this purpose we assume that hunter-gatherers visit farmers and copy their culture [this is more reasonable according to ethnographical observations (9, 17, 18), but a model in which farmers visit hunter-gatherers would yield the same final result for the front speed]. Under this assumption, we have to evaluate the hunter-gatherer population density P at the point (x,y) where conversion takes place (i.e., where the number of farmers increases), and the farmer population density N at other points $(x + \Delta_x, y + \Delta_y)$, which are visited by hunter-gatherers with probability given by a kernel of visits, say $\phi_P'(\Delta_x, \Delta_y)$. Then Eq. 3 is generalized into

$$\begin{aligned} N(x,y,t+T) = &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{N}(x+\Delta_x, y+\Delta_y, t)\phi_N(\Delta_x,\Delta_y)d\Delta_x d\Delta_y \\ &+ \int_{-\infty}^{\infty} \phi_P'(\Delta_x,\Delta_y)d\Delta_x d\Delta_y f\frac{R_T[N(x+\Delta_x, y+\Delta_y, t)]R_T[P(x,y,t)]}{R_T[N(x+\Delta_x, y+\Delta_y, t)] + \gamma R_T[P(x,y,t)]}, \end{aligned} \qquad \textbf{[S5]}$$

where in the right-hand side, the new additional term corresponds to nonlocal cultural transmission. In the first term, $\tilde{N}(x,y,t)$ is given by Eq. 4, i.e.,

$$\begin{aligned} \tilde{N}(x,y,t) &\equiv R_T[N(x,y,t)] + f\frac{R_T[N(x,y,t)]R_T[P(x,y,t)]}{R_T[N(x,y,t)] + \gamma R_T[P(x,y,t)]} \\ &\approx e^{a_N t}(1 + C)N(x,y,t), \end{aligned} \qquad \textbf{[S6]}$$

where we have linearized the population densities a the leading edge of the wave of advance (*Methods*) and applied that $C = f/\gamma$. The demic model is recovered if there is no local cultural transmission ($f = 0$ and $C = 0$) neither nonlocal cultural transmission [$\phi_P'(\Delta_x, \Delta_y) = 0$]. By combining the former two equations and linearizing, we arrive at

$$\begin{aligned} N(x,y,t+T) \approx &\, e^{a_N T}(1 + C)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N(x+\Delta_x, y+\Delta_y, t)\phi_N(\Delta_x,\Delta_y)d\Delta_x d\Delta_y \\ &+ e^{a_N T} C\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N(x+\Delta_x, y+\Delta_y, t)\phi_P'(\Delta_x,\Delta_y)d\Delta_x d\Delta_y, \end{aligned} \qquad \textbf{[S7]}$$

which generalizes Eq. 17. Recall that for the demic kernel, N individuals have probabilities p_j to disperse at distances r_j (j = 1, 2, . . . , M). Analogously, assume that for the cultural kernel, P individuals have probabilities p_j' to disperse at distances r_j' (j = 1, 2, . . . , N). Then, by following the same approach as in *Methods*, we obtain for the front speed

$$s = \min_{\lambda > 0}\frac{a_N T + \ln\left[(1 + C)\left(\sum_{j=1}^{M} p_j I_0(\lambda r_j)\right)\right] + C\left(\sum_{j=1}^{N} p_j' I_0(\lambda r_j')\right)}{T\lambda}, \qquad \textbf{[S8]}$$

which generalizes Eq. **5** with an additional term due to the cultural kernel. Without cultural transmission ($C = 0$), we recover the speed in Fort et al. (19) (purely demic model).

Humans often visit the places where their relatives live, and the dispersal kernels of hunter-gatherers and farmers are rather similar (ref. 20 and references therein), so it is reasonable to

assume that the migration (demic) and the visit (cultural) kernels are approximately the same. Then the previous equation is simplified into

$$s \approx \min_{\lambda>0} \frac{a_N T + \ln\left[(1+2C)\left(\sum_{j=1}^{M} p_j I_0(\lambda r_j)\right)\right]}{T\lambda}. \qquad \textbf{[S9]}$$

The predictions of this equation are shown in Fig. S2*A*. Interestingly, the asymptotic speeds s^* for $C \to \infty$ are the same as without nonlocal transmission (Fig. 1), which is very reasonable, because s^* is simply the maximum kernel distance divided by the generation time (as explained in *Results*). However, for $C \to 0$, the local and nonlocal models also yield the same speeds (because then there is no cultural transmission and the wave propagation is purely demic). For these reasons, and the obvious fact that the speed increases with increasing values of C (i.e., with more hunter-gatherers becoming farmers per generation), the predictions of the local and nonlocal model (Fig. S2*A*) are not substantially different. Indeed, the cultural effect for the nonlocal model (Fig. S2*B*) is $44 \pm 4\%$, rather similar to that for the local model in the main text ($40 \pm 8\%$). In other words, assuming that nonlocal cultural transmission took place in the spread of such a complex trait as farming, it would have led to an additional contribution to those of the migration of people (demic spread) and local cultural transmission, but this does not change the conclusion that the cultural effect was significant, albeit less important, than the demic one.

Finally, note that in Eqs. **S7**–**S9** we have assumed the same value C for the conversion intensities of the local and nonlocal components of cultural transmission. Certainly, it is reasonable to expect than for the nonlocal component, the efficiency of cultural transmission (and thus the value of C) could be lower than for the local component. However, obviously this would lead to still-smaller differences relative to the local model and, therefore, to the same conclusion as that reached in the previous paragraph. Similarly, a cultural kernel with smaller distances than the demic one would yield the same conclusion.

Reaction-Diffusion Model. In the main text we have added the effect of cultural transmission to a previously known model of demic spread (19). The differences between that demic model and Fisher's wave of advance model (which is also demic, but based on regular diffusion) are that the former takes into account a dispersal kernel as well as the cohabitation time between newborn children and their parents. When including cultural transmission, the question arises whether simpler models, based on regular diffusion, lead to similar conclusions. One such model is Fisher's wave of advance model extended to include cultural transmission, namely (Eq. **2**):

$$\begin{cases} \dfrac{\partial N}{\partial t} = D_N \nabla^2 N + F(N) + \dfrac{f}{T}\dfrac{NP}{N+\gamma P} \\ \dfrac{\partial P}{\partial t} = D_P \nabla^2 P + F(P) - \dfrac{f}{T}\dfrac{NP}{N+\gamma P} \end{cases}. \qquad \textbf{[S10]}$$

This model assumes the validity of several Taylor expansions, and this approximation leads to the operator ∇^2, which is typical in diffusion theory. Let us thus refer to Eq. **S10** as our reaction-diffusion model; it is less precise than that in the main text because of the Taylor expansions mentioned, and also because it neglects the cohabitation effect (see the text above and below Eq. **11**). Using the procedure explained in *Methods*, it is easy to show that the speed of the waves of advance of farmers for the reaction-diffusion model described by Eq. **S10** is

$$s_{reaction-diffusion} = 2\sqrt{\left(a_N + \frac{C}{T}\right) D_N}, \qquad \textbf{[S11]}$$

where $D_N = \frac{1}{4T}\sum_{j=1}^{M} p_j r_j^2$ (21). Without cultural transmission ($C = 0$), Eq. **S11** reduces to Fisher's speed $2\sqrt{a_N D_N}$, as it should (22). Fisher's speed was applied by Ammerman and Cavalli-Sforza in their demic diffusion model (5). The maximum and minimum speeds, obtained from Eq. **S11**, are shown in Fig. S3*A*. Note that these speeds do not reach a finite bound for $C \to \infty$. In contrast, the kernel model used in the main text (also shown for comparison in Fig. S3*A*) has a maximum speed (s^* in Fig. S3*A*, which is the maximum kernel distance divided by the generation time; *Results*). We think that this also shows the limitations of using reaction-diffusion models compared with dispersal-kernel equations in realistic models of range expansions. In fact, such limitations are well known in dispersal ecology (23).

If we used the reaction-diffusion model predictions (dash-dotted and dotted curves in Fig. S3*A*), then the observed ranges from *Parameter Values and Observed Neolithic Front Speed Range* (0.9–1.3 km/y for the speed and 1.0–10.9 for C) imply that $1.0 \le C \le 5.0$, and the cultural effect is $45 \pm 15\%$, a little higher than for the kernel model ($40 \pm 8\%$, from Fig. 2).

However, it is important to recall that (as explained above) conclusions based on regular diffusion (reaction-diffusion model) are surely less precise that conclusions based on the kernel model (indeed, this is why we have used the kernel model in the main text). Thus, the most precise range for the cultural conversion intensity C is clearly that obtained from the kernel model, namely $1.0 \le C \le 2.5$ (see main text). For this range of C, the cultural effect predicted by the reaction-diffusion model (Fig. S3*B*) is $40 \pm 10\%$, almost the same as that predicted by the kernel model in the main text ($40 \pm 8\%$).

It may seem surprising that the cultural effect predicted by the reaction-diffusion model is similar to that predicted by the kernel model, because intuitively we expect a long-distance kernel to yield faster speeds than a reaction-diffusion model (for low-enough C, see Fig. S3*A*), and at first sight, it could seem that this might underlie the primacy of the demic effect over the cultural one. However, consider a typical value of C (e.g., $C = 2$). Certainly, the reaction-diffusion model does lead to a slower speed than the kernel model (Fig. S3*A*), but their predictions for the cultural effect are similar (Fig. S3*B*) because this effect is computed as the difference between the speed for the value of C considered (e.g., $C = 2$) minus the speed for $C = 0$ (demic model), divided by the former. However, the speed for $C = 0$ is also slower for the reaction-diffusion than for the kernel model (Fig. S3*A* for $C = 0$), which is why the cultural effect predicted by the kernel model and by the reaction-diffusion model are rather similar for realistic values of C (Fig. S3*B*).

Besides the model given by Eq. **S10**, other reaction-diffusion models have been proposed in the literature. Some such models (5, 24, 25) consider two populations with a Lotka–Volterra interaction (*Methods*); one of them was successfully applied to test the ability of principal component analysis to separate the effects of several migrations on the genetic composition of populations (25), thereby lending support to the application of principal components to real genetic data (26). A three-population model (again with Lotka–Volterra cultural transmission) was used by Aoki et al. (27) to determine the theoretical conditions under which genetic clines form, and this model has been generalized in two directions. On one hand, Patterson et al. (28) have extended it to nonhomogeneous environments, and compared its predictions to the archaeological observations of the Neolithic transition in the Indian subcontinent. On the other hand, Ackland et al. (29) have introduced additional competition terms to explain the formation of cultural and linguistic boundaries in

Europe, India, and Southern Africa. Besides those Lotka–Volterra reaction-diffusion models, an integrodifference model (also with Lotka–Volterra interaction) has been applied to the computation of the neolithization time (30). In terms of archaeology, the main difference between the present work and those previous papers is that here we have estimated the importance of cultural transmission on the front speed.

Numerical Simulations. To confirm the validity of Eq. **5**, numerical simulations of Eqs. **3** and **4** were performed as follows. A square grid with 3,000·3,000 nodes was considered. Initially there were no farmers in any node [$N(x,y) = 0$] except at the central one, where $N(x,y) = K_N$, and all nodes were saturated with hunter-gatherers [$P(x,y) = K_P$], but the front speed was the same if using $P(x,y) = 0$ at the central node and/or using the initial conditions in the central node for a finite region. At each node, the population densities $N(x,y,t)$ and $P(x,y,t)$ were updated according to the following three-step cycle (each cycle corresponding to one generation).

Step 1. Logistic growth was applied by computing the new population densities as given by Eq. **10** with the observed values $K_N =$ 1.28 km^{-2} (30), $K_P = 0.064$ km^{-2} (30), and (20) $a_P = 0.59$ gen^{-1} (however, none of these values had any effect on the front speed), and the values of a_N and T considered in each case (Fig. 1 legend).

Step 2. Cultural transmission was applied by using Eq. **4** and several values of f and γ, with ratios $C = f/\gamma$ corresponding to the symbols in Fig. 1 (but only the ratio $C = f/\gamma$ had an effect on the front speed); if at some nodes this yielded a negative population density [$\tilde{P}(x,y,t) < 0$], which makes no biological sense, we limited cultural transmission so that $\tilde{P}(x,y,t) = 0$ at those nodes. Similarly, if at some nodes this step yielded a value above the saturation density [$\tilde{N}(x,y,t) > K_N$], which makes no biological sense either, the effect cultural transmission was limited so that $\tilde{N}(x,y,t) = K_N$ at those nodes.

Step 3. Population dispersal was applied according to Eq. **3** with a single observed kernel for computational simplicity: $\{p_j\} =$ {0.42; 0.23; 0.16; 0.08; 0.07; 0.02; 0.01; 0.01} for $\{r_j\} =$ {2.3; 7.3; 15; 25; 35; 45; 55; 100} km $\approx \{1; 3; 6; 10; 14; 18; 22; 40\}2.5$ km (as stated in the main text and explained in *Parameter Values and Observed Neolithic Front Speed Range*). This dispersal step consisted of distributing each population density at each "original" node into the nodes located on the edges of a set of squares centered at the original node: 8 nodes on a square with side $2d$ and a total probability of 0.42; 8·3 nodes on a square with side $2{\cdot}3d$ and total probability 0.23; . . . ; 8·40 nodes on a square with side $2{\cdot}40d$ and total probability 0.01. The value of d was computed so that the mean dispersal distance of all those jumps is equal to that of the observed kernel ($\sum p_i r_i = 12.075$ km, yielding $d = 2.166$ km).

This three-step cycle was repeated at all nodes for both population densities, until the Neolithic front speed was constant (this happened after ~30 generations or cycles, and took less than 3.5 h of computing time). The speed was computed by a linear fit, over the last 10 generations, of the front position [defined as the position at which $N(x,y,t) = 0.1K_N$, but varying the factor 0.1 did not change the results]. The speed was computed both along a vertical (or horizontal) direction and along a diagonal direction, and both results were averaged. In this way, the simulation speeds (rhombus in Fig. 1 in the main text) were obtained, and the differences between them and the analytical results from Eq. **5** (curves in Fig. 1) are below 5%, mainly due to the fact that it is necessary to discretize space in the simulations. Indeed, according to Eqs. **3** and **5**, individuals can jump along all directions, but simulations can be computed only on a finite number of points (the grid nodes), so dispersal can take place only along a finite number of directions. We can check this effect simply by recalling that the maximum front speed is $\frac{\Delta_{max}}{T}$ and is obtained for $C \to \infty$ (*Results*). In discrete space, for the kernel above and $T = 29$ y, this implies a maxim speed of $\frac{40d}{29} = 2.99$ km/y along the vertical/horizontal direction, and of $\frac{40\sqrt{2}}{29} = 4.22$ km/y along the diagonal direction. Both speeds agree exactly with the corresponding simulation results for sufficiently large values of C ($C = 10,000$ in Fig. 1). The average of those two speeds is 3.61 km/y, and corresponds exactly to the rhombus at Fig. 1 *Upper Right*.

Finally, the front profile provided by the simulations can be easily used to estimate the front width [i.e., the distance along which the value of $N(x,y,t)$ changes from approximately zero to K_N]. For the realistic range of C used in the main text ($1.0 \le C \le 10.9$), the front width obtained from the simulations is 320–430 km. Unfortunately, to the best of our knowledge, the front width has not been measured directly by archaeologists. However, an indirect estimate seems possible thanks to the fact that Shennan and Edinborough (31) succeeded in estimating the population number as a function of time in three countries. In this way, a population rise to a ceiling was detected, corresponding to the Neolithic transition in Germany (from 6,550 cal y B.C.), Poland (from 6,400 cal y B.C.), and Denmark (from 4,000 cal y B.C.). In all three cases, the time required for the population to saturate was ~300 y (figure 3 in ref. 31). Multiplying this time interval by the mean observed speed, namely 1.1 km/y (Fig. 1), yields an estimation of 330 km for the front width, which is in reasonable agreement with our estimated interval 320–430 km from the numerical simulations. We stress, however, that a direct comparison would be more accurate, and that this could in principle become possible if in the future the population number as a function of time is estimated in smaller areas.

1. Isern N, Fort J, Pérez-Losada J (2008) Realistic dispersion kernels applied to cohabitation reaction-dispersion equations. *J Stat Mech* 2008(10):P10012.
2. Guerrero E, Naji S, Bocquet-Appel JP (2008) The signal of the Neolithic demographic transition in the Levant. *The Neolithic Demographic Transition and Its Consequences*, eds Bocquet-Appel JP, Bar-Yosef O (Springer, Berlin).
3. Fort J, Jana D, Humet JM (2004) Multidelayed random walks: Theory and application to the neolithic transition in Europe. *Phys Rev E Stat Nonlin Soft Matter Phys* 70(3 Pt 1):031913.
4. Pinhasi R, Fort J, Ammerman AJ (2005) Tracing the origin and spread of agriculture in Europe. *PLoS Biol* 3(12):e410.
5. Ammerman AJ, Cavalli-Sforza LL (1984) *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ Press, Princeton, NJ).
6. Stauder J (1971) *The Majangir: Ecology and Society of a Southwest Ethiopian People* (Cambridge Univ Press, Cambridge).
7. MacDonald DH, Hewlett BS (1999) Reproductive interests and forager mobility. *Curr Anthropol* 40(4):501–523.
8. Early JD, Headland TN (1998) *Population Dynamics of a Philippine Rain Forest People. The San Ildefonso Agta* (Univ Press Florida, Gainesville, FL).
9. Hill K, Hurtado AM (1996) *Ache Life History* (de Gruyter, New York).
10. Cavalli-Sforza LL, Feldman MW (1981) *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton Univ Press, Princeton, NJ).
11. Boyd R, Richerson PJ (1985) *Culture and the Evolutionary Process* (Univ Chicago Press, Chicago).
12. Henrich J (2001) Cultural transmission and the diffusion of innovations: Adoption dynamics indicate that biased cultural transmission is the predominant force in behavioral change. *Am Anthropol* 103(4):992–1013.
13. Kandler A, Steele J (2009) Innovation diffusion in time and space: Effects of social information and income inequality. *Diff Fund* 11(3):1–17.
14. Reyes-García V, et al. (2008) Do the aged and knowledgeable men enjoy more prestige? A test of the predictions from the prestige-bias model of cultural transmission. *Evol Hum Behav* 29(4):275–281.
15. Boyd R, Richerson PJ (2002) Group beneficial norms can spread rapidly in a structured population. *J Theor Biol* 215(3):287–296.
16. Lehmann L, Feldman MW, Foster KR (2008) Cultural transmission can inhibit the evolution of altruistic helping. *Am Nat* 172(1):12–24.
17. Newman JL (1970) *The Ecological Basis for Subsistence Change Among the Sandawe of Tanzania* (Natl Acad Sci, Washington, DC).
18. Lee RB (1979) *The! Kung San* (Cambridge Univ Press, Cambridge).
19. Fort J, Pérez-Losada J, Isern N (2007) Fronts from integrodifference equations and persistence effects on the Neolithic transition. *Phys Rev E Stat Nonlin Soft Matter Phys* 76(3 Pt 1):031913.

20. Fort J, Pujol T, Cavalli-Sforza LL (2004) Palaeolithic populations and waves of advance. *Camb Archaeol J* 14(1):53–61.
21. Fort J, Méndez V (1999) Time-delayed theory of the Neolithic transition in Europe. *Phys Rev Lett* 82(4):867–870.
22. Fisher RA (1937) The wave of advance of advantageous genes. *Ann Eugen* 7: 355–369.
23. Clark JS (1998) Why trees migrate so fast: Confronting theory with dispersal biology and the paleorecord. *Am Nat* 152(2):204–224.
24. Fort J, Méndez V (2002) Wavefronts in time-delayed systems. Theory and comparison to experiment. *Rep Prog Phys* 65(6):895–954.
25. Rendine S, Piazza A, Cavalli-Sforza LL (1986) Simulation and separation by principal components of multiple demic expansions in Europe. *Am Nat* 128(5):681–706.
26. Menozzi P, Piazza A, Cavalli-Sforza LL (1978) Synthetic maps of human gene frequencies in Europeans. *Science* 201(4358):786–792.
27. Aoki K, Shida M, Shigesada N (1996) Travelling wave solutions for the spread of farmers into a region occupied by hunter-gatherers. *Theor Popul Biol* 50(1):1–17.
28. Patterson MA, Sarson GR, Sarson HC, Shukurov A (2010) Modeling the Neolithic transition in a heterogeneous environment. *J Arch Sci* 37(11):2929–2937.
29. Ackland GJ, Signitzer M, Stratford K, Cohen MH (2007) Cultural hitchhiking on the wave of advance of beneficial technologies. *Proc Natl Acad Sci USA* 104(21):8714–8719.
30. Fort J, Pérez-Losada J, Suñol JJ, Escoda L, Massaneda JM (2008) Integro-difference equations for interacting species and the Neolithic transition. *New J Phys* 10(4):043045.
31. Shennan S, Edinborough K (2007) Prehistoric population history: From the late glacial to the late Neolithic in central and northern Europe. *J Arch Sci* 34(8):1339–1345.

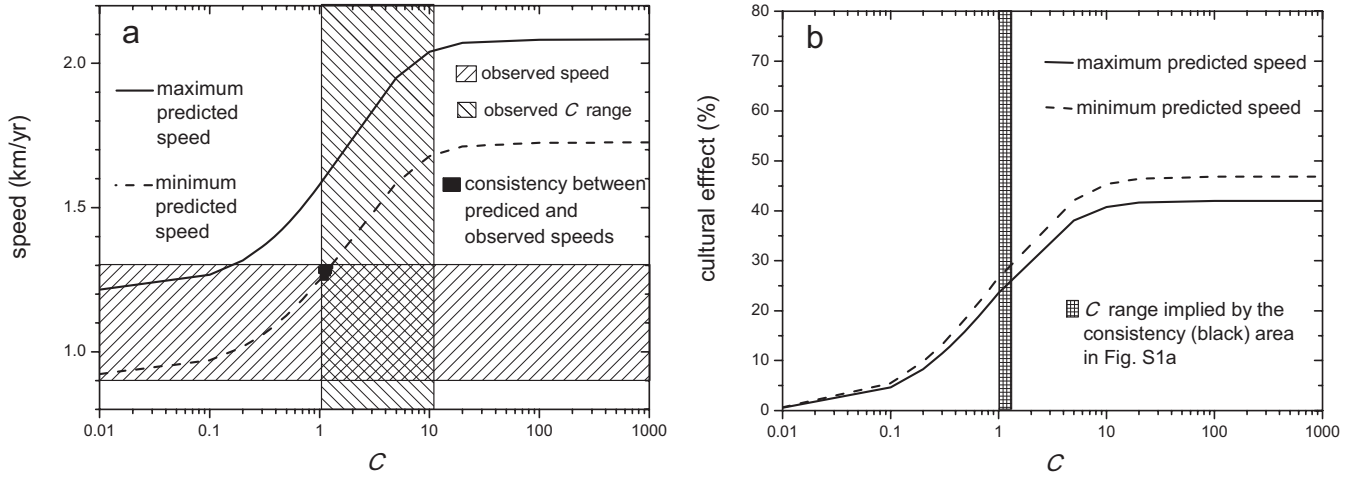

Fig. S1. Predicted speed (*A*) and cultural effect on the Neolithic front speed (*B*) using the Majangir dispersal kernel discussed in *Dispersal kernel*, point *ii*.

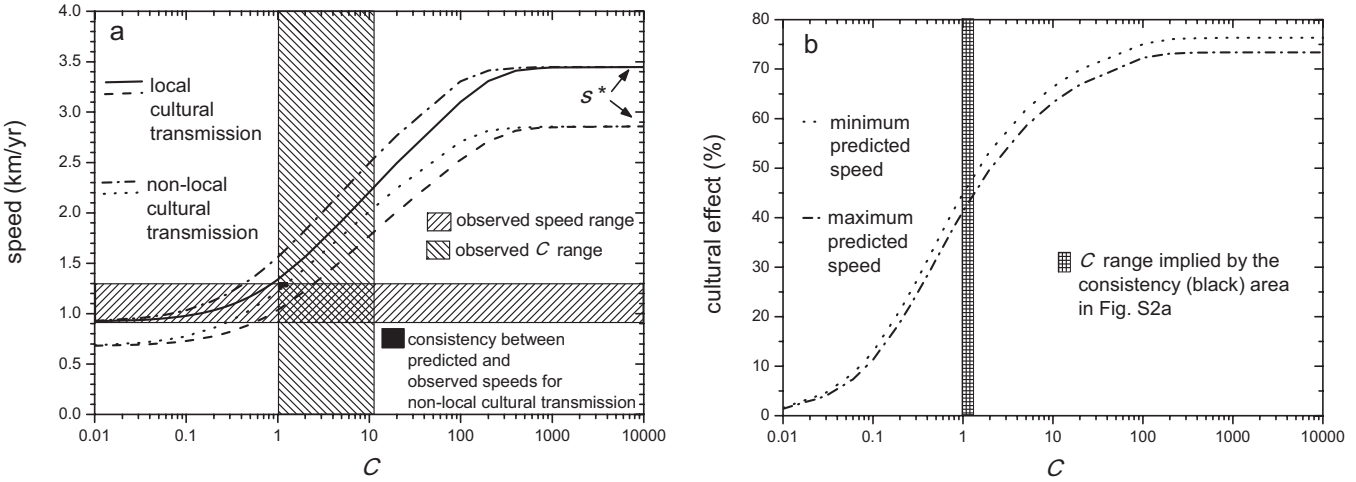

Fig. S2. Predicted maximum and minimum speeds (*A*) and cultural effect on the Neolithic front speed (*B*), including nonlocal cultural transmission. In *A*, the maximum and minimum speeds under local transmission are also shown for comparison (those two curves are the same as in Fig. 1).

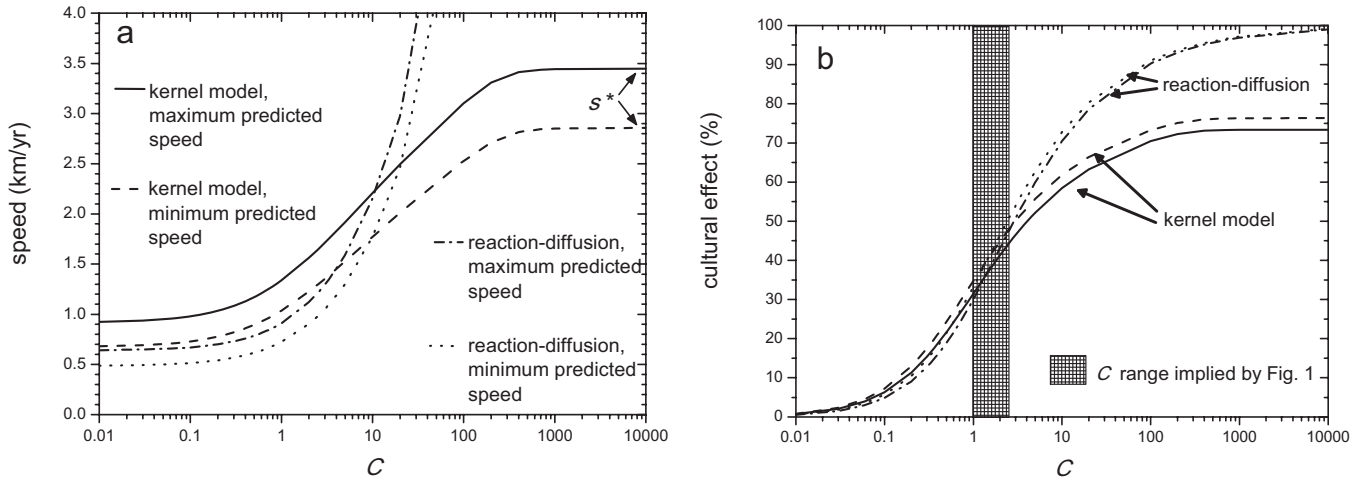

Fig. S3. Predicted maximum and minimum speeds (*A*) and cultural effect on the Neolithic front speed (*B*) according to our approximate, reaction-diffusion model (Eqs. **S10** and **S11**). The results from our more-precise kernel-cohabitation model (main text) are also shown for comparison (the corresponding curves, labeled as kernel model, are the same as in Figs. 1 and 2).