PNAS
www.pnas.org

Supplementary Information for

Individual variations lead to universal and cross-species patterns of social behavior

Sang Hyun Choi, Vikyath D. Rao, Tim Gernat, Adam Hamilton, Gene E. Robinson, Nigel Goldenfeld

Corresponding Author: Nigel Goldenfeld
E-mail: nigel@uiuc.edu

This PDF file includes:

Supplementary text
Figs. S1 to S5
References for SI reference citations

Supporting Information Text

1. Quantile-quantile plots of the pair interaction time distribution

Our model predicts the pair interaction time distribution to be exponential. We have generated the quantile-quantile (QQ) plots of pair interaction times of honeybees and humans to verify the prediction. If the data followed the proposed probability density, a QQ plot would show a linear line. QQ plot was meant to be a preliminary test to examine the probability density of the pair interaction times, so only a few pairs from each dataset have been tested. The pair with the most number of interactions from each dataset has been chosen to be demonstrated here for clearer examination of the distribution.

As shown in Fig. S1, although some pair interaction times appear to be exponential, there are data that deviate for long times. The deviation is always upward, which implies that it is not due to statistical fluctuations but is instead a systematic effect. The upward deviation indicates that the actual probability density decays more slowly than the proposed one. However, even the deviating data align well with the reference line for short times. The deviating region at long times also looks linear although it does not lie on the reference line. The linearity suggests that the deviating region is exponential as well, but with a slower decaying rate than the short time region. A weighted sum of two exponential distributions, sometimes known as a hyperexponential distribution, exhibits the same behavior. Its limiting behavior at small values of the argument is the faster decaying of the two summed exponentials whilst its limiting behavior at large values is the other one. Therefore, we deduce that the pair interaction time distribution is hyperexponential, and is well approximated by just two terms.

2. Calculation of the interaction time distribution.

As discussed in the main text, the probability density of the rates is $p(\omega) \sim \omega^{\alpha}$. Integrating the pair interaction time distribution over the rates gives

$$f(t) = \int_0^{\epsilon} d\omega p(\omega) f(\omega, t) = \int_0^{\epsilon} d\omega \omega^{\alpha+1} e^{-\omega t} = t^{-(2+\alpha)} \gamma(2+\alpha, \epsilon t) \qquad [S1]$$

where $\gamma(a, z) \equiv \int_0^z t^{a-1} e^{-t} dt$ is the incomplete gamma function. Since $p(\omega) \sim \omega^{\alpha}$ is for small ω limit, the integration is performed up to a finite value ϵ. $\gamma(a, z)$ approaches a constant $\Gamma(a) \equiv \int_0^{\infty} t^{a-1} e^{-t} dt$ as $z \to \infty$ (a is fixed in our case), which means that equation S1 would become $f(t) \sim t^{-(2+\alpha)}$ for large ϵt. $\gamma(a, z)$ starts to saturate to a constant value at z of order 1. Therefore, the large ϵt condition translates into $t > 1/\epsilon$. We assume ϵ to be small, but it is the largest value of ω in the population (equation S1). So the condition $t > 1/\epsilon$ is satisfied for most t, and we observe the heavy-tailed distribution $f(t) \sim t^{-(2+\alpha)}$.

The pair interaction time distribution turns out to be better represented by a hyperexponential distribution than an exponential distribution. However, this modification does not affect the final result $f(t) \sim t^{-(2+\alpha)}$. Now the pair interaction time distribution is $f_{pair}(t) = g\omega_1 e^{-\omega_1 t} + (1-g)\omega_2 e^{-\omega_2 t}$ where g is a weight on one of the two exponential terms. Since we now have two rates ω_1 and ω_2 with relation $\omega_1 > \omega_2$, the integration becomes

$$f(t) = \int_0^{\epsilon} d\omega_1 p(\omega_1) \int_0^{\omega_1} d\omega_2 p(\omega_2) \left\{ g\omega_1 e^{-\omega_1 t} + (1-g)\omega_2 e^{-\omega_2 t} \right\}. \qquad [S2]$$

The first term is calculated as

$$\int_0^{\epsilon} d\omega_1 \omega_1^{\alpha} \frac{g}{\alpha+1} \omega_1^{\alpha+2} e^{-\omega_1 t} = \frac{g}{\alpha+1} t^{-(3+2\alpha)} \gamma(3+2\alpha, \epsilon t) \sim t^{-(3+2\alpha)}. \qquad [S3]$$

The last asymptotic relation is for large t. Here "large t" means $t > 1/\epsilon$ as stated in the previous paragraph. The first integral of the second term gives us $(1-g)t^{-(2+\alpha)}\gamma(2+\alpha, \omega_1 t)$. Then the second term is calculated as

$$(1-g)t^{-(2+\alpha)} \int_0^{\epsilon} d\omega_1 \omega_1^{\alpha} \gamma(2+\alpha, \omega_1 t) \sim t^{-(2+\alpha)} \qquad [S4]$$

for large t because $\gamma(2+\alpha, \omega_1 t)$ is essentially a constant for that limit. Therefore, the interaction time distribution for the whole population becomes

$$f(t) \sim At^{-(3+2\alpha)} + Bt^{-(2+\alpha)} \sim t^{-(2+\alpha)} \qquad [S5]$$

where A and B are constants. As shown in equation S5, the scaling of $f(t)$ does not change even after the modification of the pair interaction time distribution.

Although we observe only two energy barriers, hence a hyperexponential pair interaction time distribution, the theory in principle could allow the pair interaction time distribution to be an arbitrary sum of exponential distributions $\sum_i g_i \omega_i e^{-\omega_i t}$. The same calculation as above gives $f(t) \sim t^{-(2+\alpha)}$ in this case as well.

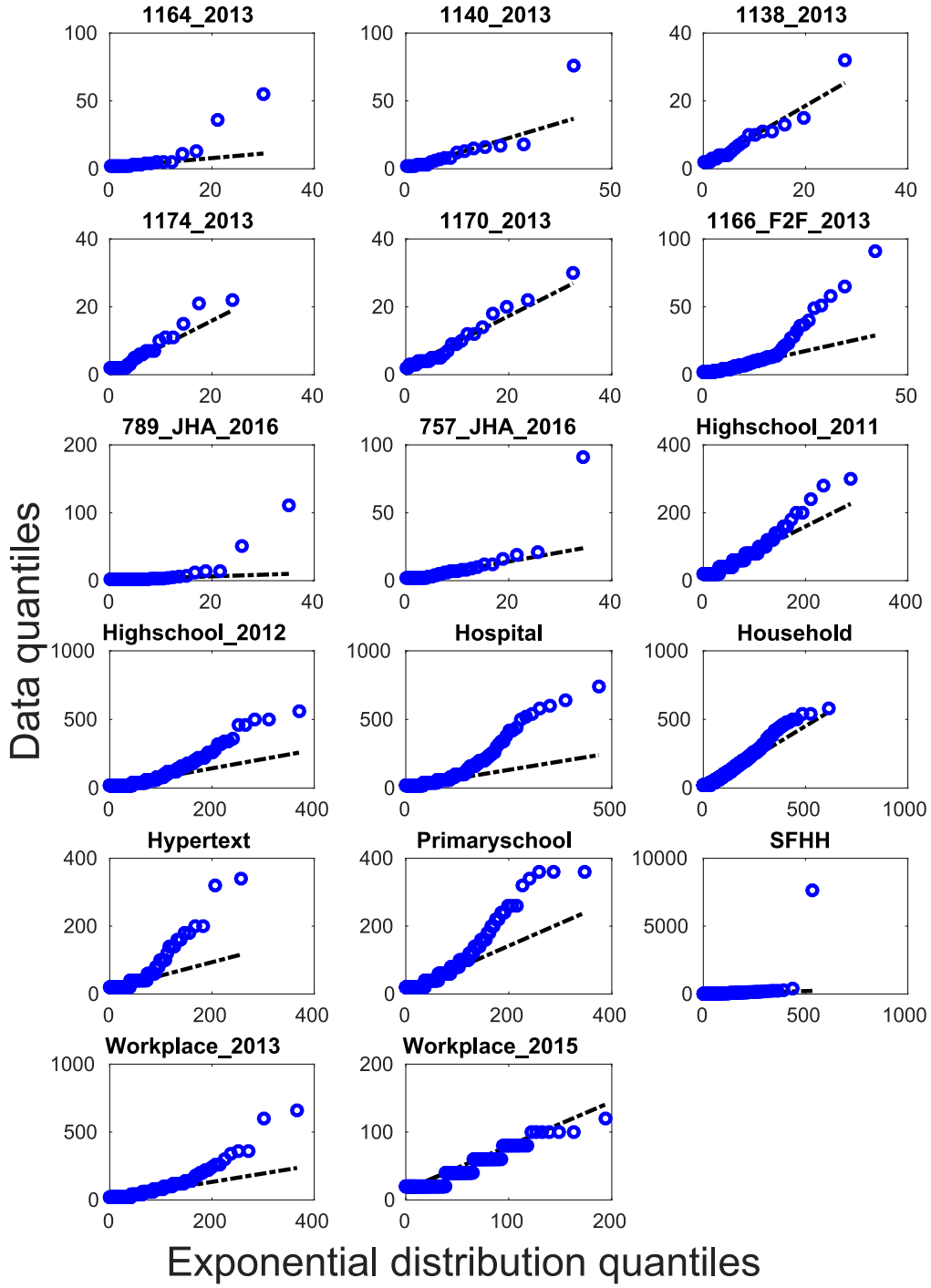

Fig. S1. Quantile-quantile plots of the most frequently-interacting pair from each honeybee and human dataset. If the data followed the proposed distribution, the data points would lie on the dashed reference line.

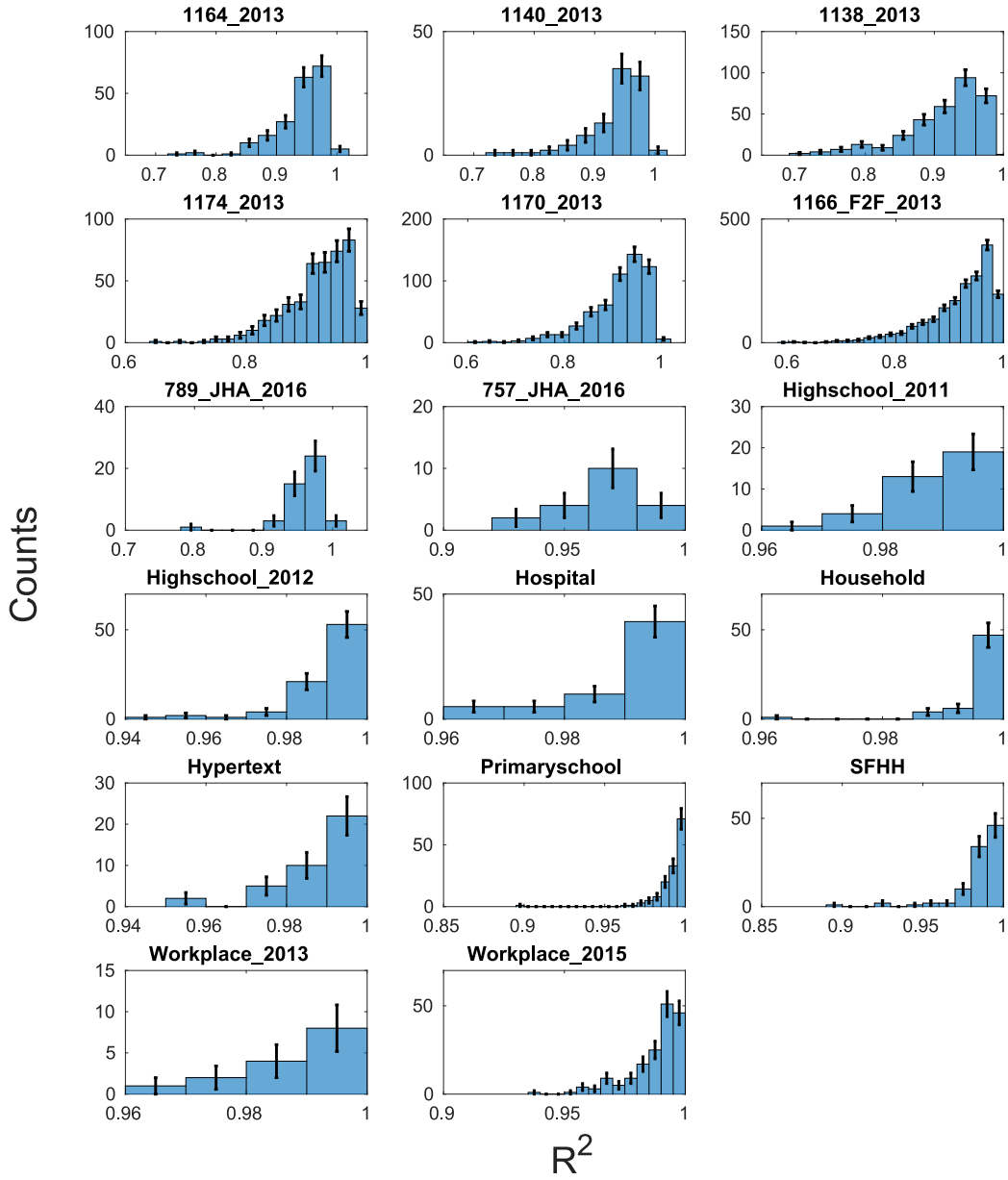

Fig. S2. Histogram of R^2 of fitting of the pair interaction time distribution to hyperexponential distribution for all honeybee and human datasets. For all figures, error bars indicate the standard error.

3. Fitting of the pair interaction time distribution to the hyperexponential distribution.

We have fitted the empirical cumulative distribution function (ECDF) of the pair interaction times to the cumulative distribution function (CDF) of a hyperexponential distribution $F_{pair}(t) = 1 - ge^{-\omega_1 t} - (1-g)e^{-\omega_2 t}$. Fig. S2 shows the histogram of R^2 values of the fitting. It is not a single value but a distribution of values because fitting has been done to each pair and R^2 has been collected from each pair.

As shown in Fig. S2, most values of R^2 are close to 1, which means that the pair interaction time distributions are well represented by a hyperexponential distribution.

Fig. S3 shows the histogram of weight g on the first exponential term in the hyperexponential distribution. Again it shows not a single value but a distribution of values because the fitting has been done for each pair.

The peak near 1 in Fig. S3a-f shows that one of the two exponential terms has a much higher weight than the other in these datasets. On the other hand, other datasets show more uniform distributions, signifying that both exponential terms, or both energy barriers, have non-negligible effect on many pairs. The broad distributions in Fig. S3i-q suggest that the multi-dimensionality of the potential landscape of social interaction is more prominent in humans than in honeybees. Yet, as shown in Fig. S3g and h, honeybee colonies with partial JHA treatment exhibit similar distributions of weights as humans. Despite the robust heavy tail in the interaction time distribution, Fig. S3 shows that the presence of honeybees with JHA treatment does influence the social interaction of honeybees. The precise way that it affects the honeybee social interactions needs further study and is beyond the scope of this work.

4. Exponential barrier height distribution.

Our model predicts that the barrier height distribution is the extreme value distribution for maxima. However, since the heavy tail at large time is dominated by high energy barriers, we take the large barrier height limit and obtain the exponential distribution $p(E) \sim e^{-(\alpha+1)E}$ where E is the energy barrier height in units of an effective energy scale that depends on details of social interactions that are outside the level of description of the present theory and α is some parameter. Using Kramers' rate formula $\omega = \omega_0 e^{-E}$ where ω_0 is some constant, we express the mean pair interaction time distribution as $p(\tau) \sim \tau^{-(2+\alpha)}$. It has the same scaling as the interaction time distribution $f(t) \sim t^{-(2+\alpha)}$. So by comparing the exponent of $f(t)$ and $p(\tau)$, we verify the exponential energy barrier height distribution.

Fig. S4 shows the scaling of the probability densities of interaction time $f(t)$, mean pair interaction time from fitting $p(\tau_{ij})$ and mean pair interaction time from averaging $p(\tau_i)$ for all honeybee and human datasets. Here the index i labels a pair, and the index j labels an energy barrier. τ_{ij} is the mean escape time for each energy barrier and is obtained from fitting of the pair interaction time distribution. With this index notation, CDF of a hyperexponential distribution to which we have fitted ECDF of the pair interaction time distribution is written as $F_i(t) = 1 - g_{i1}e^{-\omega_{i1}t} - g_{i2}e^{-\omega_{i2}t}$ where g_{ij} is the weight with $g_{i2} = 1 - g_{i1}$ and ω_{ij} is the rate of each exponential term. These g_{i1}, ω_{i1} and ω_{i2} are fitting parameters. We obtain two mean pair interaction times τ_{i1} and τ_{i2} for each pair i by taking the reciprocal of ω_{i1} and ω_{i2} respectively. On the other hand, τ_i is average of all interaction times for a pair. So $\tau_i = \frac{1}{n_{i1}+n_{i2}}\left(\sum_{k=1}^{n_{i1}} t_{i1k} + \sum_{k=1}^{n_{i2}} t_{i2k}\right)$ where t_{ijk} is the interaction time, the index k labels each interaction event, and n_{ij} is the number of interactions associated with each energy barrier j. In this notation, the mean pair interaction time from fitting is $\tau_{ij} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} t_{ijk}$. So the relation between τ_{ij} and τ_i is written as $\tau_i = \sum_j g_{ij}\tau_{ij}$ where the weight is $g_{ij} \equiv n_{ij}/\sum_j n_{ij}$.

In Fig. S4, data points that extend to very large values (such as of order of 10^{14} seconds) were omitted because they are an artifact from fitting. These outliers correspond to one or two pairs out of all analyzed pairs in the colony. The artifact arises because the pair interaction time distribution for some pairs is better described by an exponential distribution than a hyperexponential distribution. In such a case, ω_{i2} takes a very small value close to 0 because $1 - g_{i1}e^{-\omega_{i1}t} - (1-g_{i1})e^{-\omega_{i2}t} \to g_{i1}(1 - e^{-\omega_{i1}t})$ for $\omega_{i2} \to 0$. Then taking the value of g_{i1} close to 1 gives us $1 - e^{-\omega_{i1}t}$ which is CDF of an exponential distribution. $\omega_{i2} \to 0$ corresponds to $\tau_{i2} \to \infty$, which explains the large τ_{ij}'s in some datasets. The pair interaction time distributions that yield such large τ_{ij} indeed have g_{ij} values close to 1 and are fitted well to exponential distributions.

For all datasets, $p(\tau_{ij})$ exhibits the same scaling as $f(t)$, verifying the exponential distribution of energy barriers. On the other hand, as shown in Fig. S4l and n, $p(\tau_i)$ doesn't show the same scaling for some datasets, and this feature is discussed in the next section.

5. The mean pair interaction time distribution and the stable law.

In this section, we discuss the difference between the two kinds of mean pair interaction time distributions $p(\tau_{ij})$ and $p(\tau_i)$ to emphasize that $p(\tau_{ij})$ is the right distribution to use. τ_{ij} represents the mean time needed to jump over an energy barrier j, while τ_i is the average of the whole pair interaction times. So the quantity associated with the energy barrier E through $\omega = \omega_0 e^{-E}$ is τ_{ij}. It would be more accurate to write the Kramers rate formula as $\omega_{ij} = \omega_0 e^{-E_{ij}}$ according to this notation. τ_i is the normalized linear sum of τ_{ij}'s.

It is very tempting to use τ_i instead of τ_{ij} because simply averaging the pair interaction times is much easier than fitting the pair interaction time distributions. Fig. S4 shows whether it matters which mean pair interaction time we use; while $p(\tau_i)$ has the same scaling as $p(\tau_{ij})$ and thus as the population interaction time distribution $f(t)$ for most datasets, it is not the case for Household (Fig. S4l) and Primaryschool (Fig. S4n). The deviation of $p(\tau_i)$ is clearer in Primaryschool (Fig. S4n) because of better statistics associated with the larger dataset. These two particular datasets differ from the rest by having $f(t)$ that

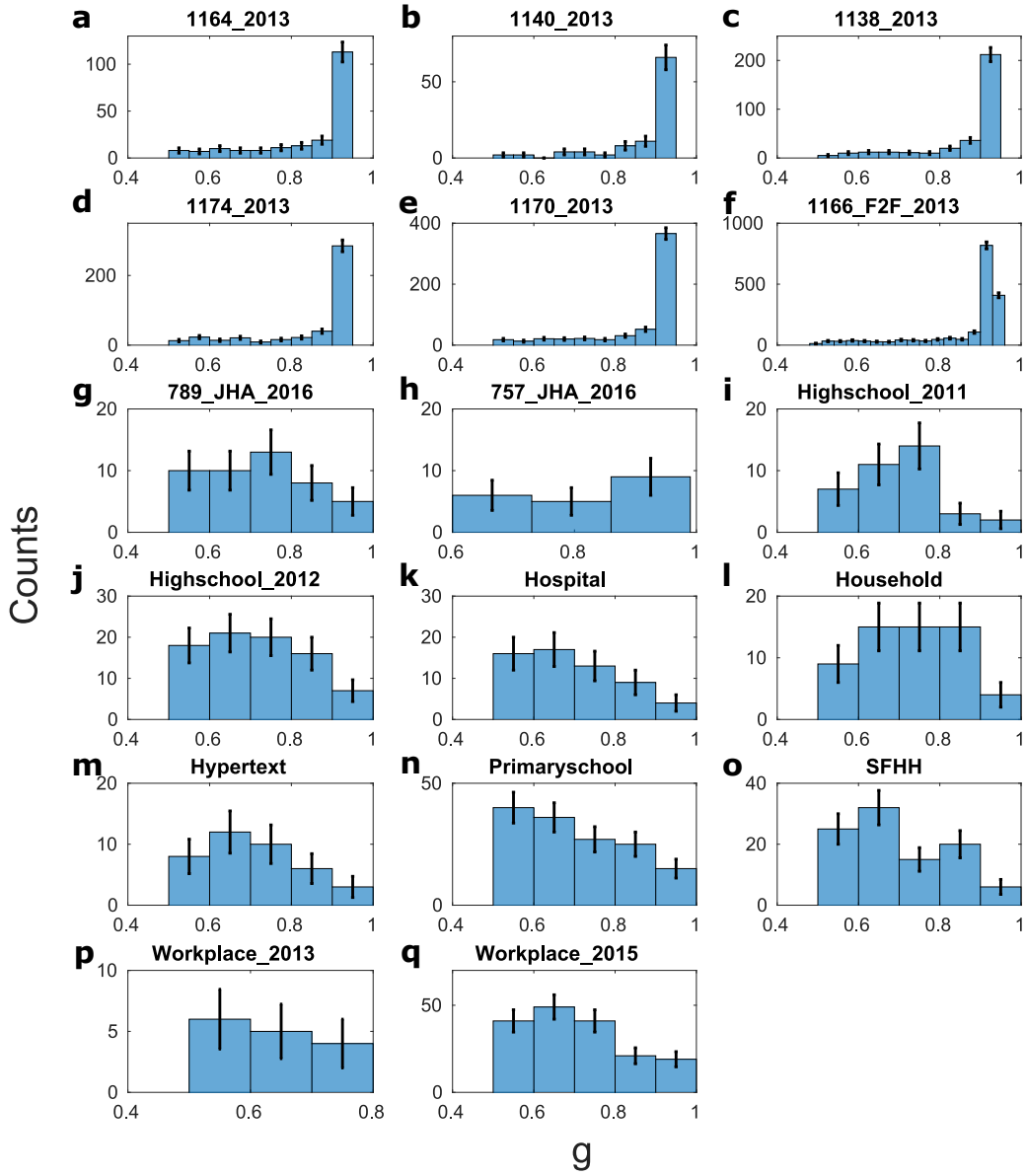

Fig. S3. Histogram of weight g in hyperexponential fitting of the pair interaction time distribution for all honeybee and human datasets. g can only be between 0.5 and 1. For all figures, error bars indicate the standard error.

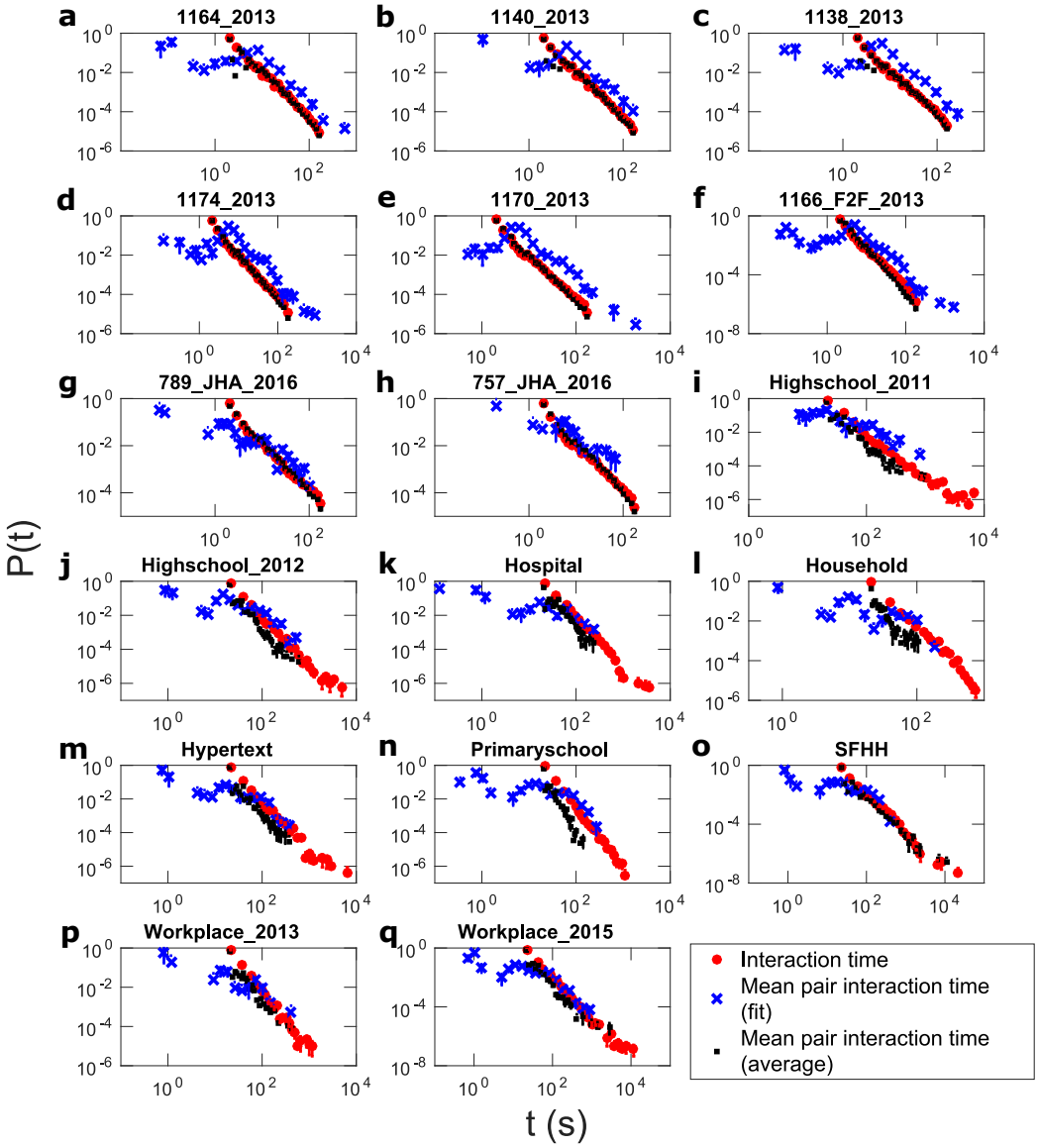

Fig. S4. Comparison of scaling between the interaction time distribution and the mean pair interaction time distributions for all honeybee and human datasets. The probability densities of mean pair interaction times obtained by different methods are separately plotted. From **a** to **q**, the number of mean pair interaction times from averaging used to generate the plot is 200723, 143571, 129653, 174317, 212685, 472914, 88441, 76810, 1710, 2220, 1139, 891, 2196, 8316, 9889, 754, 4273, which is the same as the number of detected pairs. The number of mean pair interaction times from fitting used is 197, 99, 328, 443, 561, 1806, 46, 20, 37, 82, 59, 58, 39, 143, 98, 15, 171, which is the same as the number of pairs used for fitting. For all figures, error bars indicate the standard error. Some error bars are not visible because they are smaller than the marker size. Lower error bars for bins of count 1 could not be drawn on a logarithmic scale.

decays faster than t^{-3}; the exponent of $f(t)$ of Household is -3.4 and that of Primaryschool is -3.6. The other datasets have $f(t)$ with exponents from -2 to -3, with specific values are provided in the main text.

Now we explain why the value of the exponent of $f(t)$ is relevant to whether or not $p(\tau_i)$ has the same scaling as $p(\tau_{ij})$ by using the so-called stable law (S1), sometimes referred to as the generalized central limit theorem. A probability density is said to be stable if a linear combination of independent and identically distributed (i.i.d.) random variables drawn from this probability density is also a random variable from the same probability density up to location and scale parameters. The normal distribution is a well-known example of a stable distribution. The stable law states that the probability density of normalized sums of i.i.d. random variables converges to a stable distribution (S1), and specifies the conditions that determine the appropriate stable distribution. In our case, the random variables are simply the pair interaction times.

We are interested in a set of random variables, drawn from an empirical probability distribution that we will call the parent or source distribution. We wish to know what will be the probability distribution for a linear combination of these variables. If the parent distribution is stable, then the stable law will imply that the linear combination is also distributed according to the same distribution. If not then it will be distributed according to a different distribution.

The distribution of the linear combination of random variables depends on the asymptotic behavior of the parent distribution. If the parent distribution $f(x)$ has the asymptotic behavior $f(x) \sim x^{-(\alpha+1)}$ with $\alpha \geq 2$ for large x, the distribution of the sums converges to a normal distribution. This result is the central limit theorem. On the other hand, if $0 < \alpha < 2$, the parent distribution itself is stable, so the limit distribution of the sums converges to the parent distribution $x^{-(\alpha+1)}$.

In our case, the parent distribution is $p(\tau_{ij})$, and the distribution of the normalized sums is $p(\tau_i)$. The datasets with the same scaling of $p(\tau_{ij})$, $p(\tau_i)$ and $f(t)$ all have $f(t)$ with exponents from -2 to -3. According to the theory in the main text, $f(t)$ and $p(\tau_{ij})$ always have the same exponents, which is verified by all the datasets in Fig. S4; thus $p(\tau_{ij})$ has exponents in the range from -2 to -3. It can be re-written as $p(\tau_{ij}) \sim \tau_{ij}^{-(\alpha+1)}$ with $1 < \alpha < 2$. Indeed, we see that α is in the range where $p(\tau_{ij})$ is stable, and so we expect the same scaling between $p(\tau_i)$ and $p(\tau_{ij})$, assuming that the variables are indeed i.i.d. In conclusion, $f(t)$ has the same scaling as $p(\tau_{ij})$ due to our theory, while $p(\tau_{ij})$ and $p(\tau_i)$ have the same scaling because $p(\tau_{ij})$ is the parent distribution of $p(\tau_i)$. Thus all three variables have the same scaling when $\alpha + 1$ is in the range from -2 to -3.

On the other hand, Household (Fig. S4l) and Primaryschool (Fig. S4n) have $f(t)$, and thus $p(\tau_{ij})$, with exponents -3.4 and -3.6 respectively. These exponents correspond to $\alpha + 1$. For these datasets, $\alpha > 2$, showing that $p(\tau_{ij})$ is not a stable distribution. Therefore, the stable law predicts that $p(\tau_i)$ does not have the same scaling as $p(\tau_{ij})$ and $f(t)$. The deviation of $p(\tau_i)$ from the other two distributions in Fig. S4l and n is consistent with this prediction.

Although the stable law predicts that the probability density of the normalized sum of i.i.d. random variables that are not stable converges to a Gaussian, $p(\tau_i)$ is not Gaussian as shown in Fig. S4l and n. It is not Gaussian because the number of summed τ_{ij} is only two, as there are two energy barriers for each pair. The convergence to a Gaussian only occurs if the number of summands is large (e.g. more than 10). In addition, even if the number of summands were large, burstiness of temporal social networks of honeybees and humans suggests that τ_{ij} could be not i.i.d. Then $p(\tau_i)$ would be expected to exhibit an exponential tail rather than a Gaussian tail, because the distribution of the normalized sum of correlated random variables converges to the Fisher-Tippett-Gumbel distribution, which has an exponential tail (S2).

In summary, $f(t)$ and $p(\tau_{ij})$ always have the same scaling, as predicted by our theory. However, as demonstrated in Fig. S4l and n, $p(\tau_i)$ may or may not have the same scaling as $f(t)$ depending on the exponent that characterizes the asymptotic behavior of $f(t)$, in accord with the stable law. This observation shows that we cannot ignore the fact that there are two energy barriers per pair, and we indeed need to obtain the mean pair interaction times by fitting the pair interaction time distribution to the hyperexponential function.

Note that the stable law does not hold between $f(t)$ and $p(\tau_{ij})$ although they have the same exponent. τ_{ij} is a normalized sum of pair interaction times associated with one energy barrier whose probability density is exponential. So the parent distribution for τ_{ij} is not $f(t)$. Furthermore, each τ_{ij} has its own parent distribution labeled by indices i and j, which is an exponential distribution with different parameter values. The stable law concerns the normalized sum of random variables from the same probability density. Therefore, one cannot predict the form of $p(\tau_{ij})$ from $f(t)$.

Conversely, our theory is only able to predict the form of $f(t)$ from $p(\tau_{ij})$ precisely because $f(t)$ is not the parent distribution for τ_{ij} and so the stable law does not hold between them. As a result, the scaling of $p(\tau_{ij})$ is purely determined by $p(E_{ij})$.

6. Lorenz plots.

Heterogeneity or variability in the population is represented by the barrier height distribution in our model. However, the model does not dictate whether the barrier height is determined by a specific pair or a particular individual, which poses the question about the individuality in social interaction. To explore the effect of individuality in a way that is independent of our theory, we have calculated the Gini coefficients (S3) for the total number of interactions and the total number of partners in addition to the total interaction time shown in the main text. We have chosen the Gini coefficient to quantify the different degree of dominance in social interactions by each individual because we are analyzing interaction times, which are shared between a pair, and thus it is nontrivial to decouple individual contribution.

Fig. S5 shows the Lorenz plots (S4) for the total number of interactions and the total number of partners of honeybees and humans. The larger deviation from the reference line in Fig. S5b and d than in Fig. S5a and c respectively indicates that individual differences are larger in human communities than honeybee communities. However, Fig. S5a and c still show a deviation from the straight line, signifying that honeybee individuals are different. This result still does not tell us how much individuality contributes to social interaction. Given that the energy barrier is a pairwise property, we can only tell that the

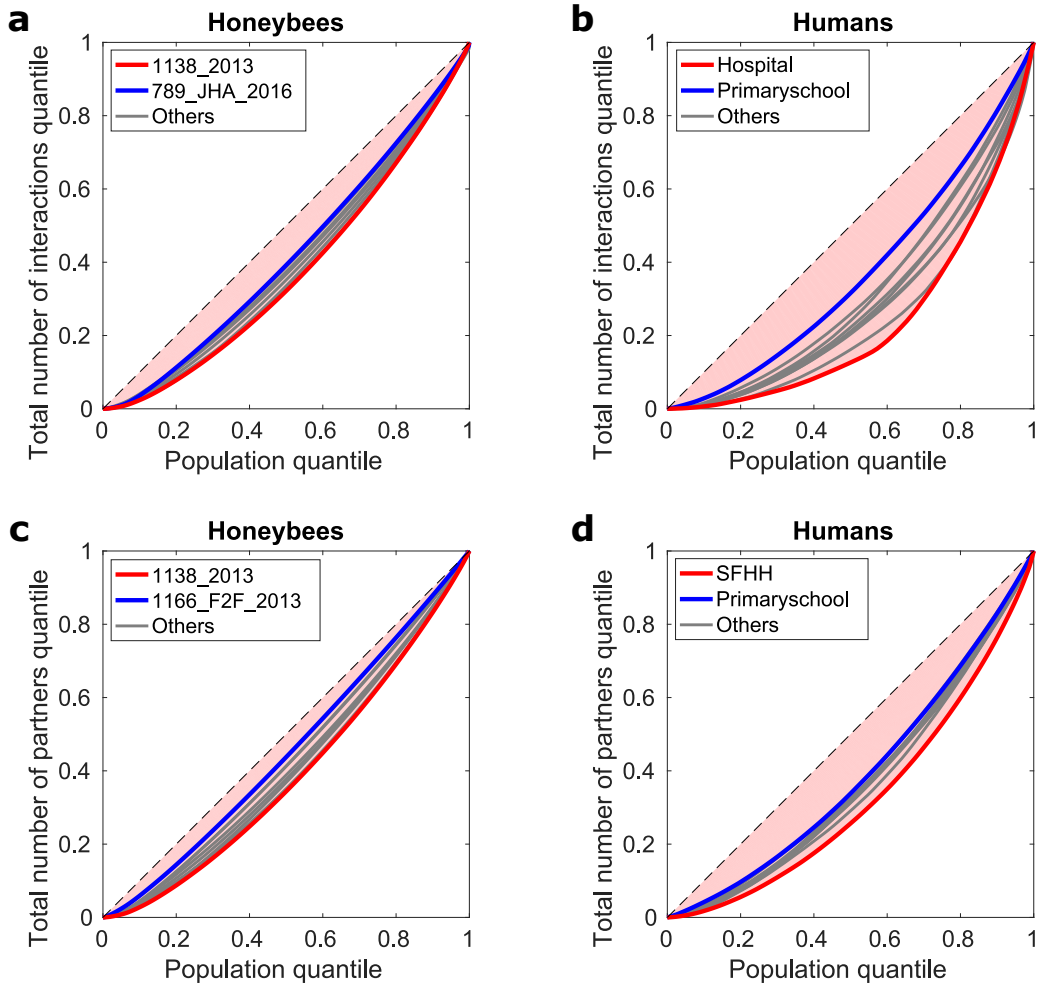

Fig. S5. Lorenz plots for total number of interactions and total number of partners for honeybees and humans. **a**, Gini coefficients for total number of interactions of honeybees are following. 1164_2013: 0.2125, 1140_2013: 0.1823, 1138_2013: 0.2526, 1174_2013: 0.2462, 1170_2013: 0.2362, 1166_F2F_2013: 0.1949, 789_JHA_2016: 0.1609, 757_JHA_2016: 0.1667. **b**, Gini coefficients for total number of interactions of humans are following. Highschool_2011: 0.3509, Highschool_2012: 0.4361, Hospital: 0.5293, Household: 0.4992, Hypertext: 0.4089, Primaryschool: 0.2595, SFHH: 0.4410, Workplace_2013: 0.4011, Workplace_2015: 0.3426. **c**, Gini coefficients for total number of partners of honeybees are following. 1164_2013: 0.1758, 1140_2013: 0.1619, 1138_2013: 0.2222, 1174_2013: 0.2081, 1170_2013: 0.1897, 1166_F2F_2016: 0.0916, 789_JHA_2016: 0.1284, 757_JHA_2016: 0.1301. **d**, Gini coefficients for total number of partners of humans are following. Highschool_2011: 0.2612, Highschool_2012: 0.2538, Hospital: 0.2828, Household: 0.2498, Hypertext: 0.2643, Primaryschool: 0.2215, SFHH: 0.3474, Workplace_2013: 0.2537, Workplace_2015: 0.2243.

pairwise heterogeneity is essential for the heavy tail to appear in the interaction time distribution, as demonstrated in this report, but whether the pairwise heterogeneity is purely a property of a pair independent of individuals forming the pair is still open for further study.

References

[S1] Gnedenko BV, Kolmogorov AN (1954) *Limit Distributions for Sums of Independent Random Variables.* (Addison-Wesley).

[S2] Bertin E, Clusel M (2006) Generalized extreme value statistics and sum of correlated variables. *J. Phys. A: Math. Gen.* 39(24):7607–7619.

[S3] Gini C (1912) Variabilità e mutabilità (variability and mutability). *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1955) ed. Bologna.*

[S4] Lorenz MO (1905) Methods of measuring the concentration of wealth. *Publications of the American statistical association* 9(70):209–219.