[Supplementary material]

**Comparing ancient inequalities: the challenges of comparability, bias and precision**

Mattia Fochesato[1], Amy Bogaard[2,3,*] & Samuel Bowles[3]

[1] *Dondena Centre, Bocconi University, Via Röntgen 1, 20136 Milan, Italy*

[2] *Institute of Archaeology, University of Oxford, 36 Beaumont Street, Oxford, OX1 2PG, UK*

[3] *Behavioral Sciences, Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM, 87501 USA*

[*] *Author for correspondence (Email: amy.bogaard@arch.ox.ac.uk)*

*Received: Revised: Accepted:*

[*] *Author for correspondence (Email: amy.bogaard@arch.ox.ac.uk)*

## Measuring wealth inequality: the Gini coefficient

To represent inequality as a relationship among individuals we treat an economy as a complete undirected network, the edges of which (differences in some attribute between individuals), not the nodes (the individual attributes), are the fundamental data that motivates a standard inequality measure—the Gini coefficient (Bowles & Carlin 2018). Figure S1 shows an example, where the numbers in the circles are the wealth of the individual represented by that node, and the numbers on the arrows are the indicated pair's wealth difference.

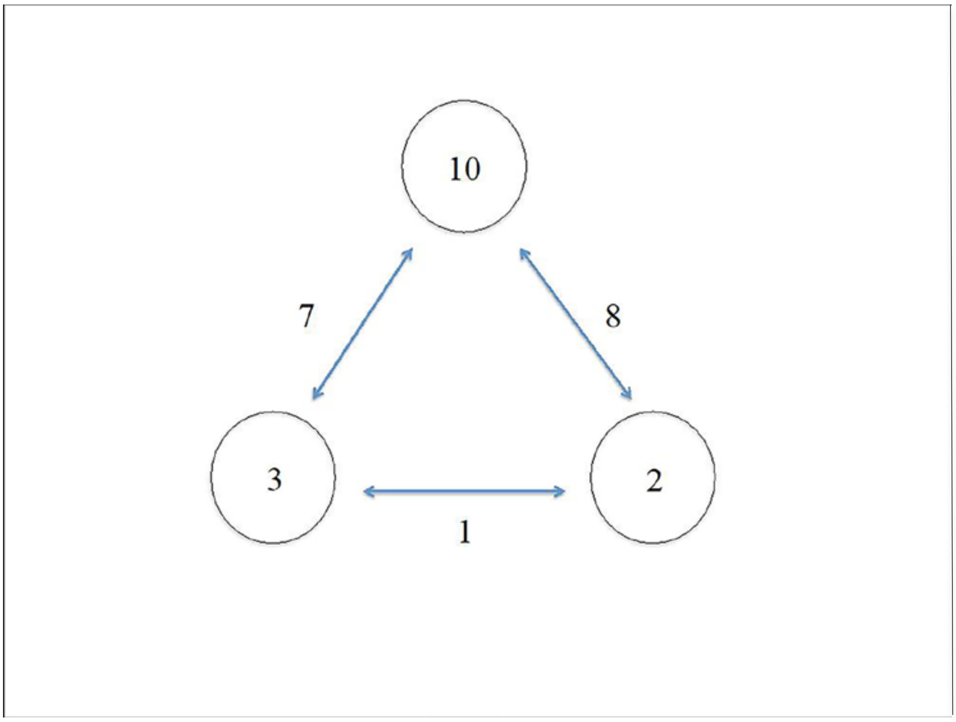

*Figure S1. Wealth differences among pairs of households. Each of the double-headed arrows indicates a unique pair, and the Gini coefficient is based on the wealth differences among all of the pairs. The Gini coefficient—one half of the average difference divided by the mean wealth—for the three-person economy shown is* $\frac{16}{3} \times \frac{1}{\frac{15}{3}} \times \frac{1}{2} = \mathbf{0.53}$

If there are $k$ members of the population then the total number of unique non-identical pairs is $(k^2 - k)/2$, which for the $k = 3$ case in the figure are shown as the three edges in the figure. Let $\Delta$ be the sum of the absolute differences among the pairs of $k$ wealth holders in a population and $\underline{y}$, the mean wealth; then we have the following expression for one-half the relative mean difference, which is the Gini coefficient:

$$G = \frac{\Delta}{(k^2 - k)/2} \frac{1}{\underline{y}} \frac{1}{2} = \frac{\Delta}{k(k-1)} \frac{1}{\underline{y}} \qquad (1)$$

The Gini coefficient is the mean difference among all pairs (the first term in the middle expression) relative to (divided by) the mean value of y (the “relative mean difference”) times one half. The algorithm conventionally used to calculate the Gini coefficient differs from equation 1 in ways that impart a downward bias that can be substantial where the number of observations is small (a common feature of feature of prehistoric datasets). We have followed Bowles and Carlin (2018) in correcting for the bias in the computational algorithm.

This representation of the Gini coefficient allows very intuitive inferences about the meaning of any particular value of this measure of inequality. Three examples illustrate this. First, from the equation above we see that the relative mean difference is simply *2G*; so, for example, a Gini coefficient of 0.35 means that the average difference between all pairs in the population is seventy percent of the mean wealth.

Second, suppose there are just two people in a population, and they are dividing a ‘pie’ representing total wealth. The portion received by the disadvantaged member of the pair (σ) is $\sigma = (1 - G)/2$ so using the same Gini coefficient as above, the smaller slice is 32.5 percent of the total, the richer of the pair receives 67.5 per cent of the ‘pie’.

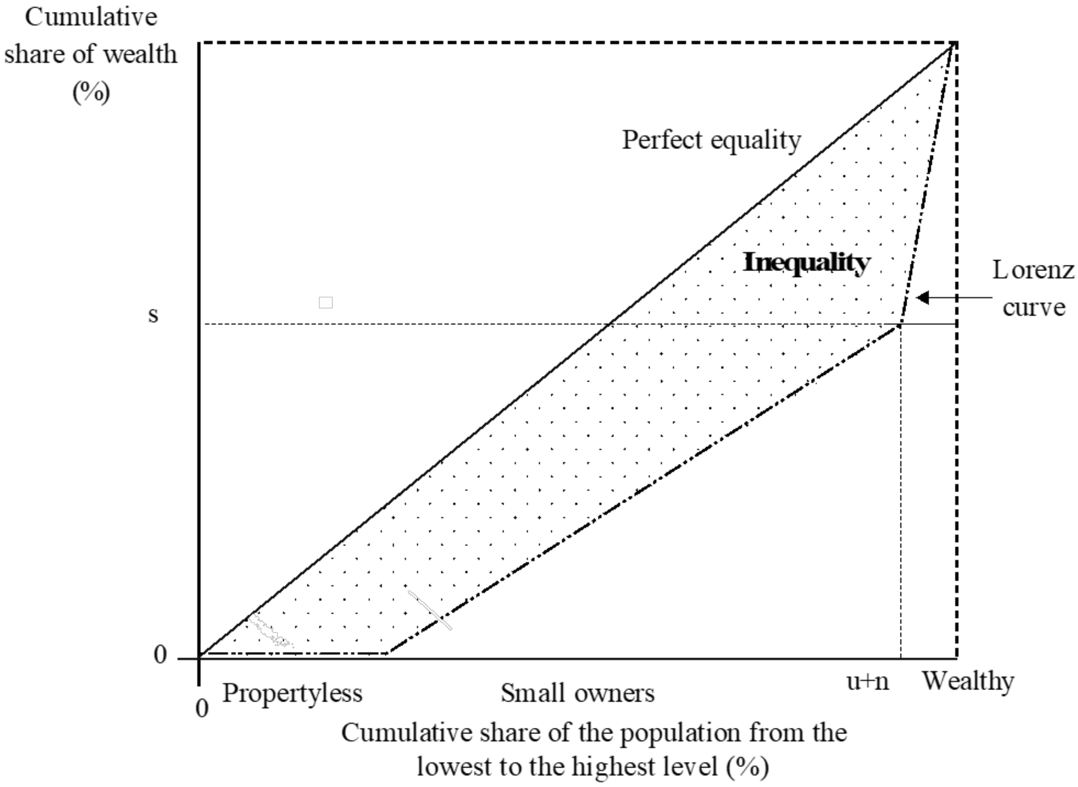


*Figure S2. A Lorenz curve for a class-divided economy. Shown in the figure are, on the x-axis, the cumulative share of population ordered from the lowest to the highest level of wealth, and on the y-axis, the cumulative share of wealth owned by the indicated population share. The 45º black solid line shows a condition of perfect equality, according to which each k-per cent cumulative share of population owns k-per cent of the cumulative share of wealth. The Lorenz curve is the black, dashed segmented line below the perfect equality line and shows the relationship between cumulative shares of wealth and population in a society made of a fraction u of propertyless and a fraction n of small wealth owners owning a fraction s of wealth. The dotted area between the condition of perfect equality and the actual distribution of wealth in the society (the Lorenz curve) is a measure of the extent of inequality. This area divided by the entire area under the perfect equality line is an approximation of the Gini coefficient appropriate for large populations.*

Finally, consider a class-divided society shown in Figure S2, in which a fraction of the population ($u$) holds no wealth at all, small wealth holders ($n$ percent of the total population) together own a fractional share, $s$, of the total wealth, and a wealthy class constituting a fraction 1-$u$-$n$ of the population own the rest of the wealth. The Gini coefficient can be expressed (Bowles and Carlin, 2018) as $G = u + n - (1-u)s$. Figure S2 shows a Lorenz curve for this population.

Using this expression for the Gini, and supposing that in the population under consideration there are none without wealth, we have $G = n - s$, and this allows us, again, to see what our average Gini of 0.35 means. If ninety-nine of a hundred wealth holders together own 0.64 of the wealth and the remaining holder owns the rest, we have G = 0.99 – 0.64. = 0.35.

We mentioned in the introduction that similar Gini coefficients can be associated with very different distributions of wealth. The Lorenz curves in Figure S3 illustrate this fact, with concentration at the top (Tell Brak) and at the bottom (Tell Sabi Abyad) of the wealth distribution accounting for similar levels of wealth inequality as measured by the Gini coefficient.

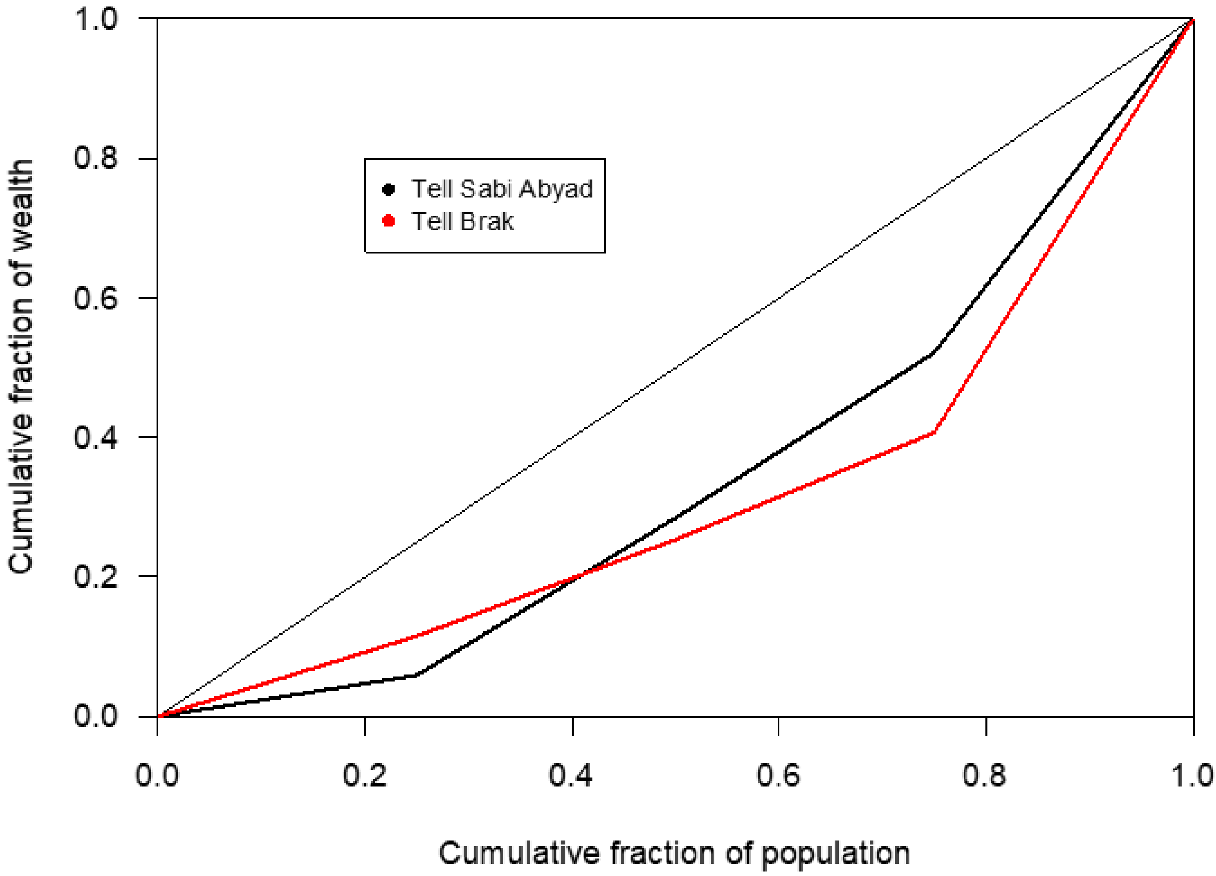


*Figure S3. Lorenz curves from two (small) populations with similar Gini coefficients. The wealth measure is house area. The Gini coefficient for Tell Sabi Abyad is 0.317 and for Tell Brak is 0.361, computed using total house area.*

**From individual to household inequality in grave wealth**

Table S1 presents the results of the method implemented to calculate between-household inequality from individual burial goods, as described in section 2 of our paper. The four sites with the greatest number of gender-identified observations are Wildcat Canyon, Berrian's Island and Sheep Creek in the Columbia Plateau dataset (Schulting 1995) and Belleview in the Hohokam dataset (McGuire 1992).

The random assortment method consists in the following steps. We first created a vector of female observations randomly ordered and then assigned each element of the vector (first to last) to an element of the males' vector ordered by increasing wealth. When females outnumbered males (or vice versa) some males (or females) were randomly drawn twice to form a couple. In the wealth assortment algorithm when females outnumber males (or vice versa), in order to not lose information, the poorest female is matched with a fictitious male with a wealth equal to the wealth of the poorest man. The last rows shows the results of the same procedure implemented among the !Kung (they provide a robustness test that we explain below.)

**Table S1. Gini coefficients of grave wealth for individual and couples. Shown in column (3) are the Gini computed only across the gender-identified individuals in each site and in column (4) those computed on couples' wealth when individuals are matched through maximal wealth assortment. Column (5) reports the average Gini coefficient across the ten rounds of random assortment. Column (6) is the mean of the Gini in columns (4) and (5). Shown in column (7) are the ratios $\mu$ in each archaeological site and in the !Kung dataset and obtained as the ratio of the average Gini of couples' wealth, column (6), and the Gini of individuals' wealth, column (3).**

| Society (1) | Site (2) | Gini on individuals (3) | Gini wealth assortment (4) | Gini random assortment (average across 10 rounds) (5) | Average Gini of couples' wealth (6) | μ (7) |
|---|---|---|---|---|---|---|
| Hohokam | Belleview | 0.698 | 0.748 | 0.533 | 0.640 | 0.92 |
| | Wildcat Canyon | 0.622 | 0.677 | 0.513 | 0.595 | 0.96 |
| Columbia Plateau | Berrian's Island | 0.522 | 0.540 | 0.372 | 0.456 | 0.87 |
| | Sheep Creek | 0.764 | 0.757 | 0.587 | 0.672 | 0.87 |
| !Kung | - | 0.219 | 0.196 | 0.148 | 0.172 | 0.78 |

We used this ratio for the individuals-to-household adjustment, treating couples' wealth as household wealth. In the southern Mesopotamia dataset (Stone 2018), for example, the Gini coefficient computed on all individual graves at Eridu during the Late Ubaid is equal to 0.445. We multiply this number by the 0.91 ratio and obtain the estimated Gini on couples' wealth equal to 0.405.

As a robustness check, we also ask, using a dataset for which (unusually) we know the actual wealth of both members of couples, if the Gini for couples obtained through our method is close to the Gini computed on the wealth of true couples. We do this using the information on individual wealth in the !Kung population from the dataset described in (Wiessner 1982) . To replicate the methods used on the Columbia Plateau and Hohokam datasets, we estimate couples' wealth as the simple sum of all the items owned by the male and the female member and we compute the Gini coefficient based on this sum for all couples, which is equal to 0.168.

We then replicate for the !Kung the couples' matching procedures used above for the archaeological datasets. In other words, we use the individual observations as if we knew nothing about who was actually paired with whom, which is the problem we confront with the burial data. We then create hypothetical couples by wealth assortment as well as through ten random assortments which, when averaged, gives a Gini coefficient equal to 0.172, which is remarkably close to the Gini computed on true couples (0.168).

**Sample size and the Gini coefficient**

Table S2 shows the skewness and bias of the three wealth distributions shown in section 3 of our paper. The degree of bias and imprecision of the estimate declines for larger sample sizes, but both are quite modest even for samples as small as 20.

**Table S2. Summary statistics of wealth distribution in three large datasets. For each of the three datasets used in the analysis, column (3) is the Gini coefficient on the total individual population, column (4) the number of observations, column (5) the third moment of the distribution (degree of right skewness) and columns (6-8) the bias, i.e. one minus the ratio of the estimated Gini to the true Gini, and the standard error (in parentheses) as a fraction of the true Gini, when the number of randomly selected individuals in the hypothetical limited dataset on which the Gini estimate is based is, respectively, 20, 50, and 150. Source: (McGuire 1992; Schulting 1995; Willführ & Störmer 2015).**

| | **Dates** | **Gini** | **N** | **Skewness** | **Bias (se) for *n* =20** | **Bias (se) for *n* =50** | **Bias (se) for *n*= 150** |
|---|---|---|---|---|---|---|---|
| **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** | **(7)** | **(8)** |
| Columbia Plateau | 2000 BC–AD 1800 | 0.623 | 498 | 2.279 | -0.003 (0.003) | -0.00004 (0.002) | -0.020 (0.001) |
| Hohokam | AD 750–1125 | 0.775 | 254 | 4.241 | -0.002 (0.003) | -0.001 (0.001) | -0.0004 (0.007) |
| Krummhörn | AD 1720–1810 | 0.803 | 3908 | 3.588 | -0.020 (0.002) | -0.008 (0.001) | -0.003 (0.001) |

**Accounting for those without wealth: using the class-based Gini coefficient to recover missing data**

If we know how many members of the population with zero wealth are missing from the data set we can make a surprisingly accurate approximation of the true Gini coefficient (that for the entire population, including those without wealth, the 'zeros').

We begin with the expression for the Gini coefficient for the 3-class society above, namely

$$G = n + u - (1-u)s$$

Here, $u$ and $n$ are the fractions of the entire population that are without wealth and 'poor', respectively.

We would like to use data on a Gini coefficient measured on the population with positive wealth (which relative to the complete population is $1-u$ in number) along with data on the fraction of the total population that are zeros to estimate the true population Gini (including the zeros). For this population there are no individuals without wealth, so we have the fraction of this population that is poor as $n/(1-u)$ and the fraction that has positive wealth as 1 (rather than $(1-u)$ in the true population). So the Gini coefficient using information from the entire population is:

$$G' = \frac{n}{1-u} - s$$

To recover the Gini for the entire population from these data, we multiply $G'$ by $(1-u)$ and add $u$. This gives us: $G = G' + u - uG'$

This expression is an approximation as it is based on a population with three homogeneous groups and, as a result, the Lorenz curve in three segments corresponding to the zeros, the poor, and the rich. We use this equation to estimate the true Gini coefficients from the Gini calculated without the zeros.

In order to check how good our approximation is, we use observed historical data in which zeros are present and perform a knock-out experiment: we eliminate the zeros, then use equation '3' above to recover an estimate of the true (zeros in) Gini coefficient, and compare this estimate with the true (zeros in) Gini. We simulate the reconstruction using 32 complete datasets (1 Florence, 4 Krummhörn, 4 Hohokam, 23 Columbia Plateau) shown in Table S3. Comparing the Gini predicted using our method and the true one, we find that the two have a correlation coefficient equal to 0.99 ($p<0.001$). The mean absolute error as a fraction of the mean of the true Gini is 0.00005.

**Accounting for those without wealth: southern Mesopotamia and Roman Italy**

We approximate the proportion of slaves in the ancient southern Mesopotamia urban population, reconstructing the number of slaves living in the city of Uruk in 3000 BC. We then apply the resulting slave ratio to the different periods (fourth to second millennia BC) covered by the inequality estimates (house space and grave goods) from southern Mesopotamia (Stone 2018).

According to Westenholz (2002), the total population of Uruk in 3000 BC numbered 40–45 000 individuals (including slaves). Taking as the total population number the midpoint 42 500 and considering that about 9000 of them were slave workers employed in the textile sector (Jacobsen 1953; McCadams 1978), the first estimate of the slave percentage of the total population is equal to 21 per cent. In addition, we add to this estimate also the slaves employed as household workers in private, public and temple households. To estimate how many of the extant 33 500 individuals (that is total population minus slaves in textiles) were household slaves, we use the estimated proportion between free individuals and household slaves provided in Diakonoff (1969). There, it is estimated that for every 100 individuals, there existed about 16 privately owned slaves (our computation from Diakonoff (1969) p. 175) employed as household workers. If we apply this ratio to the 33 500 individuals, we find that about 28 100 were free, while 5400 were slaves. Summarising, we count 14 400 slaves (textile workers and household workers), who accounted for the 34 per cent of the Uruk population. Table S4 shows how the Gini coefficients in southern Mesopotamia, fourth to second millennia BC change when the chosen fraction of zeros is 0.34.

According to the information provided in Scheidel (2011), in Roman Italy slaves represented the nine per cent of total population. This is the fraction of slaves that we use to adjust the Gini coefficients computed in the Roman Italian towns of Herculaneum and Pompeii.

**Table S3. Gini coefficients in 32 populations. The first 23 rows show the computations for the Columbia Plateau dataset, the subsequent 4 rows show the Hohokam and the remaining rows show the non-archeological datasets. Column (4) reports coefficients computed on the whole population. Column (5) shows Gini computed on the population without zeros. Note that the Gini coefficients of the whole population, column (4) of Berrian's Island, Wildcat Canyon and Sheep Creek are different than those reported for the same three sites in column (3) of Table S1. The reason is that while in Table 1, the Gini coefficients were computed only on the gender identified individuals, here the Gini are estimated on all the individuals at the site.**

| **Site** | **Sample size** | **Fraction of zeros** | **$Gini_{inc}$** | **$Gini_{exc}$** |
|---|---|---|---|---|
| **(1)** | **(2)** | **(3)** | **(4)** | **(5)** |
| 45-FE-7 | 24 | 0.583 | 0.800 | 0.489 |
| 45-ST-47 | 11 | 0.090 | 0.585 | 0.155 |
| 45-ST-8 | 15 | 0.266 | 0.478 | 0.270 |
| Congdon | 30 | 0.100 | 0.369 | 0.296 |
| Selah | 12 | 0.166 | 0.356 | 0.213 |
| Beek's pasture | 18 | 0.444 | 0.761 | 0.549 |
| Berrian's Island | 33 | 0.121 | 0.547 | 0.483 |
| Dalles Deschutes | 34 | 0.411 | 0.704 | 0.486 |
| Fish Hook Isl. II | 13 | 0.307 | 0.625 | 0.437 |
| Juniper | 22 | 0.181 | 0.572 | 0.471 |
| Keller Ferry | 12 | 0.583 | 0.743 | 0.294 |
| Koomloops | 24 | 0.041 | 0.315 | 0.284 |
| Nicoamen | 15 | 0.133 | 0.559 | 0.485 |
| Nicola Valley | 10 | 0.200 | 0.452 | 0.296 |
| Okonogan | 18 | 0.388 | 0.638 | 0.385 |
| Rabbit Island I | 11 | 0.090 | 0.390 | 0.323 |
| Rabbit Island II | 15 | 0.133 | 0.394 | 0.298 |

| **Site** | **Sample size** | **Fraction of zeros** | **Gini$_{inc}$** | **Gini$_{exc}$** |
|---|---|---|---|---|
| **(1)** | **(2)** | **(3)** | **(4)** | **(5)** |
| Sheep Creek | 38 | 0.447 | 0.750 | 0.538 |
| Sheep Island | 22 | 0.318 | 0.639 | 0.458 |
| Sundale | 19 | 0.473 | 0.774 | 0.549 |
| Whitestone | 38 | 0.342 | 0.579 | 0.352 |
| Wildcat | 32 | 0.312 | 0.667 | 0.509 |
| Yakima | 22 | 0.454 | 0.700 | 0.427 |
| 21$^{st}$ Street | 32 | 0.468 | 0.740 | 0.530 |
| 22$^{nd}$ Street | 43 | 0.348 | 0.752 | 0.608 |
| Belleview | 99 | 0.545 | 0.757 | 0.487 |
| Moreland | 69 | 0.536 | 0.812 | 0.582 |
| Krummhörn 1750 | 1066 | 0.446 | 0.766 | 0.576 |
| Krummhörn 1765 | 1553 | 0.465 | 0.775 | 0.580 |
| Krummhörn 1780 | 1984 | 0.513 | 0.797 | 0.583 |
| Krummhörn 1810 | 2354 | 0.599 | 0.829 | 0.575 |
| Florence 1427 | 9779 | 0.146 | 0.787 | 0.750 |

**Table S4. Adjusted Gini coefficients for southern Mesopotamia and Roman Italy. Shown in the table are the Gini coefficients for southern Mesopotamia and Roman Italy from Stone (2018) before the zero adjustment, and after the correction for the sample bias and couples' adjustment (only for grave goods), column (4). The fraction of the missing non-owners used in the zero adjustment is 0.34 for southern Mesopotamia and 0.09 for Roman Italy and the Gini coefficients adjusted to account for the missing zeros according to the method explained in the current section are in column (5).**

| **Phase** | **Year (midpoint)** | **Type of wealth** | **Gini adjusted by sample bias and couples excluding slaves** | **Gini adjusted by sample bias and couples including slaves** |
|---|---|---|---|---|
| **(1)** | **(2)** | **(3)** | **(4)** | **(5)** |
| Eridu Late Ubaid | 4500 BC | Grave goods | 0.405 | 0.607 |
| Kafajah Early Dynastic | 2500 BC | Grave goods | 0.409 | 0.610 |
| Kafajah Early Dynastic | 2500 BC | House size | 0.708 | 0.807 |
| Akkadian | 2250 BC | Grave goods | 0.738 | 0.822 |
| Neo Babylonian | 500 BC | Grave goods | 0.815 | 0.878 |
| Neo Babylonian | 500 BC | House size | 0.426 | 0.621 |
| Old Babylonian | 1750 BC | Grave goods | 0.819 | 0.881 |
| Old Babylonian | 1750 BC | House size | 0.494 | 0.666 |
| Italy—Herculaneum | AD 79 | House size | 0.530 | 0.572 |
| Italy—Pompeii | AD 79 | House size | 0.546 | 0.587 |

**Accounting for (non-randomly) missing wealth owners: the case of Knossos**

A problem arising with the house size dataset for Neopalatial Knossos (data from Christakis (2008) and Whitelaw (2001a, unpublished database)) is that the available data provide information on only a small fraction of the total area of the city, 13 houses in total: Palace, Little Palace/Unexplored Mansion, Acropolis House, Hogarth's House A, House of Chancel Screen, House of Frescoes, Royal Villa, SEX: North House, SEX: South House, South House, Southeast House, Southwest House, Northeast House.

Most of the 13 houses appear to have been 'elite' structures including the Palace and Little Palace/Unexplored Mansion, and likely excluding a large fraction of population (a mixture of elites/non-elites) living in the extensive urban settlement (Whitelaw 2004). For this reason, the Gini coefficient computed on the sum of living and storage space of these 13 houses is not representative of the actual distribution of house space, as also suggested by Whitelaw (2001b).

In order to have a more realistic representation of the missing population, we create a random log normal distribution of total house space for the population not excavated (in ancient societies the distribution of house space was usually strongly right skewed, making the log normal distribution appropriate.) The estimation of the Gini coefficient by this means will be sensitive to three key parameters in the random log normal distribution: the number of missing households, how unequally distributed these house sizes are and the mean size of total space they have. Following (Whitelaw 2001b, 2004, 2019), we take as the estimate of Knossos population the midpoint between 17 000 and 25 000 individuals, which translates into 6000 households (assuming that the households had a size of 3.5 adult equivalents). Assuming that such a population probably represented the middle and lower social strata, we reproduce their possible house space as a truncated random log normal distribution with mean computed from the three available observations most likely to be similar to non-elite houses (SEX and Acropolis houses). As a measure of variation, we use the average standard deviation of houses measured for two East Cretan sites with relatively complete town plans and a standardized plan of small houses, Gournia and Pseira (Whitelaw, unpublished database). We then set as minimum and maximum house, respectively, the size of the smallest house at Pseira, and the size of the Little Palace/Unexplored Mansion. Finally, we compute the Gini coefficient on the house areas of this hypothetical population.

**Comparability among different asset types**

In Bogaard *et al*. (2018) we suggested that, when living and storage spaces can be clearly identified, these two areas can be aggregated in a single measure of household wealth. Here, as explained in section 5 of our paper given the large number of archaeological cases assembled and compared, we limit our analysis to the storage space as a main measure of wealth.

We compare the wealth inequality assessed through grave goods with the one assessed through house area (living and storage space) using the archaeological cases for which the two measures are available in the same period and location. These cases are Early Dynastic Kafajah, Old Babylonian Ur and Neo Babylonian Ur in southern Mesopotamia (Stone 2018), and the late Neolithic site of Gomolava in the western Balkans (Porčić 2018).

In our comparison, we use the Gini coefficients after correction to represent couples (section 2), for sample bias (section 3), and for missing zeros (section 4). Next, we calculate for each phase/site, the ratio of the estimated inequality of household wealth (the house area including living and storage space and already corrected by sample bias, couples' and zero adjustments) to the estimated inequality of grave goods among couples. This ratio is quite similar across the four sites for which such comparison is possible: inequality in house size is approximately three quarters the level of inequality in grave goods. The implied downward adjustment in the inequality measure based on grave goods is 28 per cent with a standard error of 0.019.

The fact that grave wealth is more unequally distributed than household wealth is consistent with our signaling model of the grave wealth phenomenon, developed in section 9 of this document. In Table S5 we show the Gini coefficients from archeological data used to estimate the adjustment from grave to household wealth inequality. We use this adjustment to correct the Gini coefficients in our dataset computed on grave goods. While our estimate of the required adjustment is based on a very limited sample, we are reassured by the modest standard error, 0.019, and the fact that the two measures (Ginis based on house size and grave goods) are almost perfectly correlated ($r= 0.678$, $p < 0.001$).

**Table S5. Gini coefficients in archaeological sites and the adjustment across different asset type. Shown for each archaeological dataset for which more than one measure of inequality is available, column (1), are the Gini computed on available asset types, columns (3)–(4), and the difference between grave and house size inequality relative to inequality of grave wealth, column (5). Shown in the last row is the relative mean difference between the Gini estimated on the two asset types (standard error in parentheses.)**

| **Region—site**<br><br>**(1)** | **Period (midpoint)**<br><br>**(2)** | **Gini (house size)**<br><br>**(3)** | **Gini (grave)**<br><br>**(4)** | **Relative difference between the Gini of grave goods and the Gini of household wealth**<br>**(5)** |
|---|---|---|---|---|
| Balkans—Gomolava | 4750 BC | 0.327 | 0.498 | 0.343 |
| South Mesopotamia—Kafajah | 2500 BC | 0.610 | 0.807 | 0.244 |
| South Mesopotamia—Ur | 2250 BC | 0.666 | 0.881 | 0.244 |
| South Mesopotamia—Ur | 1750 BC | 0.621 | 0.878 | 0.291 |
| | **Average** | | | 0.281 (0.019) |

*A note on the inclusion of southern Mesopotamia data in our dataset.*

Since the southern Mesopotamian cases provided in Stone (2018) show, in some cases (Early Dynastic period, Old and Neo Babylonian), two Gini coefficients in the same period computed on two different types of assets from the same population, we add them in our dataset in the following way. We adjust both measures using the adjustments explained here and in the and we then average the two measures. We implemented this procedure also for the Balkan site of Gomolava (Porčić 2018).

**Jerf al Ahmar**

A problem arises when we aggregate living and storage spaces for Jerf al Ahmar in northern Mesopotamia (mid–late tenth millennium BC). As the only storage spaces were located in a central building (EA30), their allocation to the surrounding households of the relevant phase considered here (II/W) is uncertain (Stordeur 2015; Bogaard *et al.* 2018).

Our method to estimate a measure of aggregated wealth is similar to our algorithm for matching males and females to measure couples' grave goods inequality. It is based on two steps. We observe that the excavated area included five houses and six storage bins. We first assume that the house with the largest living area would have had the largest storage bin (in building EA30), the second largest living area the second largest storage bin and so on. Excluding from the matching procedure the smallest storage bin, we compute the Gini coefficient of the simple sum of spaces on the basis of this 'wealth matching', and find it to be equal to 0.200. We then randomly assign the five bins included to the five surrounding houses and compute the Gini coefficient on the sum of the two measures. We repeat the random assignment 10 times and get an average Gini equal to 0.187. The average Gini between the 'wealth' and random matching is 0.193, and this is the measure we use in our dataset.

We also check what the Gini coefficient would be if each household had the same proportion of storage space in Building EA30. The Gini of the sum of the two areas after equal assignment of storage is 0.187.

**The scale effect and the accuracy of the nested method**

The scale effect is estimated by comparing inequality and population at a lower-level entity (a 'village') and a higher-level entity ('district'). The scale effect for some population entity *j*, termed $\gamma_j$, is the difference between the Gini coefficient for the higher-level entity, $g_i$, and the Gini coefficient for the lower level entity, $g_j$, divided by the difference between the population of the higher-level entity, $n_i$, and the population of the lower one, $n_j$:

$$\gamma_j = \frac{g_i - g_j}{n_i - n_j} \qquad (4)$$

We assess the accuracy of the nested method to adjust for different population size using the following thought experiment. From the Columbia Plateau dataset (Schulting 1995), we select the archaeological period—the protohistoric phase—with the largest number of sites (10), which together, we assume, represent the entire set of level entities (which we will call 'villages') making up the higher level entity (the district)

Then we suppose that we have evidence on $z < 10$ of these, i.e. some but not all of the villages. How accurate a prediction of the inequalities at the district level will we produce using our estimated pure scale effect function estimated from our three datasets and shown in Figure 7 of the main text? We answer this question by selecting, for each *z* between 1 and 10, all the possible sets of *z* villages and computing the mean absolute error between the Gini predicted at the larger entity using the pure scale effect function, and the true Gini at the larger entity.

We then plot in Figure S4 the mean absolute errors computed at different number *z* of villages used for the prediction. The figure shows that the mean absolute error as fraction of the true Gini, which is already very small (0.14) when we only use one village, becomes even smaller (0.06) as the number of villages included in the prediction increases to 4. Using all 10 of the data points, the error is about 0.04.

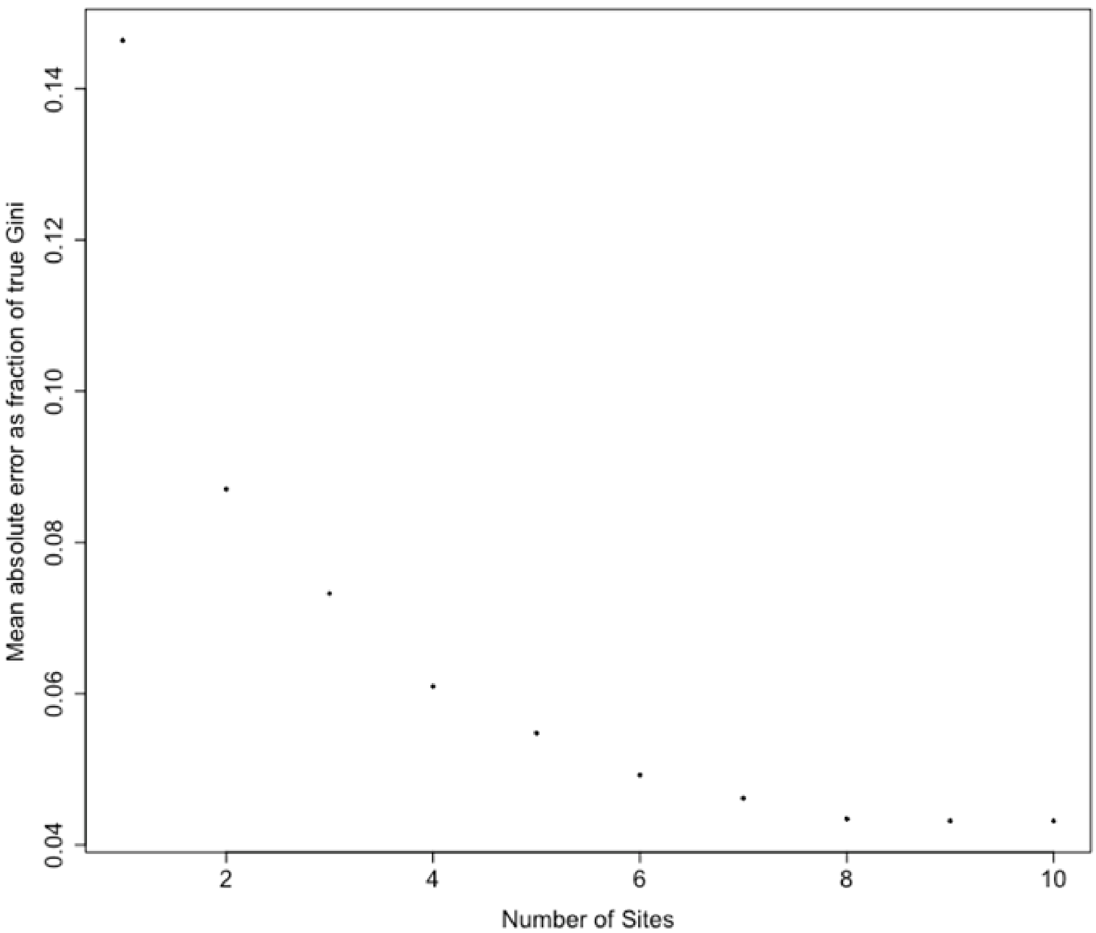


*Figure S4. The accuracy of the nested method measured using the 10 protohistoric sites in the Columbia Plateau. The vertical axis measures the mean error in predicting the Gini coefficient for the higher-level entity when using data from the number of sites shown on the horizontal axis.*

## Signaling and grave goods

For populations in which evidence of this practice is available, we treat grave goods not as a form of wealth itself but as an indicator of household wealth. We have already seen that as an empirical matter, grave goods are substantially more unequally distributed than is the floor area of dwellings, another indicator of wealth. Here we will see that if burying goods with the deceased is a form of social signaling, it is likely that inequalities in grave goods will provide an overestimate of disparities in total household wealth. Because it is the latter that is of interest (grave goods are just an indicator of total wealth) we will overestimate the degree of inequality by using measures of grave goods without an adjustment.

Here is a very simple model showing why this may occur. There is one kind of wealth and the amount held by an individual is $w$. When he dies his son (a clone of his dad, the population is asexual) inherits all of his wealth and then transforms some amount of it $g$ into grave goods, retaining the remainder $w$-$g$. If the wealth of the deceased includes stored grain, for example, some portion of that may be used to hire craftspeople to produce elaborate headdresses or ornaments.

The son gains a social esteem value $v$ for every unit of wealth that he transforms into goods buried with his late father. An individual's utility $u$ is an increasing concave function of his wealth remaining after assigning some to the production of grave goods (marginal utility is diminishing with increased remaining wealth). So, choosing ln($w$-$g$) as a suitable increasing concave function, we have for the son the following maximization problem. Choose $g$ to maximize

$$u = \ln(w - g) + vg \qquad (5)$$

Differentiating this with respect to $g$ and setting the result equal to zero to find the $g$ that maximizes the son's utility, we have the son's first order condition that determines the amount of grave wealth he should deposit as a function of the wealth he inherited from his father:

$$g = w - 1/v \text{ for } w > 1/v \text{ and } = 0 \text{ otherwise} \qquad (6)$$

In the entire wealth distribution there will be two classes of sons: those with wealth equal to or less than 1/$v$, who will deposit nothing in their father's grave, and those with greater wealth, who will deposit goods in the grave whose production required the use of an amount of wealth that increases proportionally as wealth increases. In this model those with wealth equal to or less than 1/$v$ do not signal, and those with greater wealth transform all wealth in excess of 1/$v$ into grave goods. This is consistent with the common finding that there are a significant number of individuals with no burial goods in many burial assemblages (Schulting 1995). The fraction of wealth converted to grave goods (dividing the above equation by *w*) is

$$g/w = 1-1/vw \qquad (7)$$

which increases as wealth rises. The result is that for any non-degenerate distribution (e.g. perfect equality or its opposite, one person has the entire wealth) wealth converted to grave goods (*g*) will be more unequally distributed than will be total wealth (*w*). This is because wealthy sons convert a larger fraction of their wealth to their father's grave goods than do less wealthy sons.

The reason this is the case is that marginal utility of the son's retained wealth is diminishing in its amount, while the marginal utility of the signal is a constant. If the value of the signal to the son is to advantage him in competitive status seeking it seems reasonable to assume that its marginal utility is not diminishing and could even be increasing (which were this to be the case would strengthen the results above).

**Gini coefficient for Çatalhöyük**

Our estimation of wealth inequality for prehistoric Çatalhöyük is different from the one provided in previous contributions where all observations were usually aggregated into a single population regardless of the different phases (e.g. Kohler *et al*. 2017). Here, instead, we follow the most recent phase differentiation and estimate the Gini coefficients for the two phases with more than three observations: North G and North H.

**Dataset**

The dataset assembled for the present research is part of the supplementary material (OSM_Dataset).

**References**

BOGAARD, A., A. STYRING, J. WHITLAM, M. FOCHESATO & S. BOWLES. 2018. Farming, inequality and urbanization: a comparative analysis of late prehistoric northern Mesopotamia and south-west Germany , in T.A. Kohler & M.E. Smith (ed.) *Quantifying ancient inequality: the archaeology of wealth differences*: 201–29. Tucson: University of Arizona Press.

BOWLES, S. & W. CARLIN. 2018. *Inequality as experienced difference: a reformulation of the Gini coefficient*. London: Centre for Economic Policy Research.

CHRISTAKIS, K.S. 2008. *The politics of storage: storage and sociopolitical complexity in neopalatial Crete*. Philadelphia (PA): INSTAP Academic Press.

DEMIRERGI, G.A., K. TWISS, A. BOGAARD, L. GREEN, P. RYAN & S. FARID. 2014. Of bins, basins and banquets: storing, handling and sharing food at Neolithic Çatalhöyük, in I. Hodder (ed.) *Integrating Çatalhöyük: themes from the 2000–2008 seasons*: 91–108. Los Angeles (CA): Cotsen Institute of Archaeology.

DIAKONOFF, I.M. 1969. The rise of the despotic state in ancient Mesopotamia, in I.M. Diakonoff (ed.) *Ancient Mesopotamia: socio-economic history. A collection of studies by Soviet scholars*: 173–203. Moscow: Nauka Publishing House Central Department of Oriental Literature.

HODDER, I. *et al*. 2018. Çatalhöyük Research Project. Available at: http://www.catalhoyuk.com (accessed 20 June 2019).

JACOBSEN, T. 1953. On the textile industry at Ur under Ibbī-Sîn, in F. Hvidberg (ed.) *Studia Orientalia Ioanni Pedersen*: 172–87. Munksgaard: Hauniae.

KOHLER, T.A. *et al*. 2017. Greater post-Neolithic wealth disparities in Eurasia than in North America and Mesoamerica. *Nature* 551: 619–22. https://doi.org/10.1038/nature24646

MCCADAMS, R. 1978. Strategies of maximization, stability, and resilience in Mesopotamian society, settlement and agriculture. *Proceedings of the American Philosophical Society* 122: 329–55.

MCGUIRE, R.H. 1992. *Death, society, and ideology in a Hohokam community*. Boulder (CO): Westview.

NADEL, D. 2003. The Ohalo II brush huts and the dwelling structures of the Natufian and PPNA sites in the Jordan Valley. *Archaeology, Ethnology & Anthropology of Eurasia* 1: 34–48.

PORČIĆ, M. 2012. Social complexity and inequality in the Late Neolithic of Central Balkans: reviewing the evidence. *Documenta Praehistorica* 39: 167–83.

– 2018. Evaluating social complexity and inequality in the Balkans between 6500 and 4200 BC. *Journal of Archaeological Research* 2018: 1–56. https://doi.org/10.1007/s10814-018-9126-6

SCHEIDEL, W. 2011. The Roman slave supply, in K.C. Bradley & P. Cartledge (ed.) *The Cambridge world history of slavery, volume 1: the ancient Mediterranean world*: 241–64. Cambridge: Cambridge University Press.

SCHULTING, R.J. 1995. *Mortuary variability and status differentiation on the Columbia-Fraser Plateau*. Burnaby: Archaeology Press, Simon Fraser University.

SMITH, M., T. DENNEHY, A. KAMP-WHITTAKER, E. COLON & R. HARKNESS. 2014. Quantitative measures of wealth inequality in ancient central Mexican communities. *Advances in Archaeological Practice* 2: 311–23.
STONE, E.M. 2018. The trajectory of social inequality in ancient Mesopotamia, in T.A. Kohler & M.E. Smith (ed.) *Ten thousand years of inequality: the archaeology of wealth differences.* Tucson: University of Arizona Press. https://doi.org/10.2307/j.ctt20d8801.12
STORDEUR, D. 2015. *Le village de Jerf el Ahmar (Syrie, 9500–8700 av. J.-C.): l'architecture, miroir d'une société néolithique complexe*. Paris: CNRS.
WESTENHOLZ, A. 2002. The Sumerian city-state, in M.H. Hansen (ed.) *A comparative study of six city-state cutures. An investigation conducted by the Copenhagen Polis Centre*: 23–42. Copenhagen: The Royal Danish Academy of Sciences and Letters.
WHITELAW, T. 2001a. The floor area of 207 Minoan houses, in K. Branigan (ed.) *Urbanism in the Aegean Bronze Age*: 174–79. London: Sheffield Academic.
– 2001b. From sites to communities: defining the human dimensions of Minoan urbanism, in K. Branigan (ed.) *Urbanism in the Aegian Bronze Age*: 15–37. London: Sheffield Academic.
– 2004. Estimating the population of Neopalatial Knossos. *British School at Athens Studies* 12: 147–58.
– 2019. Feeding Knossos: exploring economic and logistical implications of urbanism on Prehistoric Crete, in D. Garcia, R. Orgeolet, M. Pomadère & J. Zurbach (ed.) *Country in the city. Agricultural functions of protohistoric urban settlements (Aegean and Western Mediterranean)*. Oxford: Archaeopress.
WIESSNER, P. 1982. Leveling the hunter, in P.S. Wiessner (ed.) *Food and the status quest: an interdisciplinary perspective*: 171–91. Oxford: Berghahn.
WILLFÜHR, K.P. & C. STÖRMER. 2015. Social strata differentials in reproductive behavior among agricultural families in the Krummhörn region (East Frisia, 1720–1874). *Historical Life Course Studies* 2: 58–85.
WINDLER, A., R. THIELE & J. MÜLLER. 2013. Increasing inequality in Chalcolithic south-east Europe: the case of Durankulak. *Journal of Archaeological Science* 40: 204–10. https://doi.org/10.1016/j.jas.2012.08.017