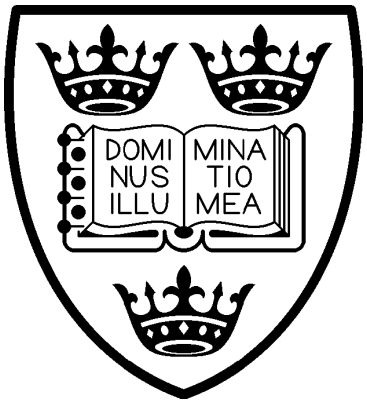


UNIVERSITY OF OXFORD

Discussion Papers in Economic and Social History

Number 148, September 2016

RETHINKING AGE-HEAPING A CAUTIONARY TALE FROM NINETEENTH CENTURY ITALY

Brian A'Hearn, Alexia Delfino, and Alessandro Nuvolari

RETHINKING AGE-HEAPING
A CAUTIONARY TALE FROM NINETEENTH CENTURY ITALY

Brian A'Hearn
(Pembroke College, University of Oxford)

Alexia Delfino
(London School of Economics)

Alessandro Nuvolari
(Sant'Anna School of Advanced Studies, Pisa)

Abstract

A swelling stream of literature employs age-heaping as an indicator of human capital, more specifically of numeracy. We re-examine this connection in light of evidence drawn from nineteenth century Italy: census data, death records, and direct, qualitative evidence on age-awareness and numeracy. Though it can stand in as an acceptable proxy for literacy, our findings suggest that age-heaping is most plausibly interpreted as a broad indicator of cultural and institutional modernisation rather than a measure of cognitive skills.

JEL codes: N33, J24

Keywords: Age-Heaping, Numeracy, Human capital, Italy

September 2016

1. Age-heaping and human capital

It is more than half a century since the concept of human capital crystallised in the work of pioneers like Mincer, Schultz, and Becker. Spreading from development to labour to economic history and more recently growth theory, it remains central to how we think about both individual welfare and aggregate economic outcomes, with no sign of interest waning. Historical research on human capital, particularly concerning skills and education, has faced serious constraints on the availability of appropriate data, however. Information on school enrolments, educational attainment, literacy rates, and (even more so) cognitive abilities is generally available only from the nineteenth century and even then typically fraught with gaps, inconsistent reporting, and other shortcomings. For earlier periods, researchers have been forced to employ considerable ingenuity in distilling suitable proxies from the existing sources.

For health status, they have made use of information on heights, which reflect net nutrition in the growing years. For education, they have turned to signature rates on marriage registers and other documents as an indicator of literacy (Cipolla 1969). And more recently a new human capital proxy has gained considerable traction among economic historians: age-heaping. A growing stream of literature has developed following A'Hearn, Baten and Crayen (2009), which exploits historical data on ages. Irregularities in an age distribution – in particular the systematic overrepresentation of, or heaping upon, popular ages – indicate that people do not know, are not able to compute exactly, or have reasons not to reveal their true age. Instead they report a round or otherwise attractive number. If this interpretation holds, age-heaping can be taken as an indicator of basic numerical skills. The literature now comprises studies covering a wide array of countries and historical periods, from Roman antiquity (Baten and Priwitzer 2015) to the preconquest Incas (Juif and Baten 2013) and beyond. Prominent contributions include Crayen and Baten (2010), de Moor and van Zanden (2010), Manzel, Baten and Stolz (2012), Hippe and Baten, (2012), and Baten, Crayen and Voth (2014).

Fischer (1977; 82-86) was perhaps the first to point to the potential of age-heaping for historical research. (It had long been known among demographers.) He examined the changing profile of age distributions in the US censuses from 1880 to 1950, and in scattered sources from the colonial period, to gauge the evolution of popular attitudes towards age. Fischer detected a growing bias towards youth (people preferring to report younger ages at critical junctures, e.g. 39 rather than 40) throughout the period he considered. In a subsequent study focused specifically on age-heaping, Kaiser and Engel (1993) interpreted it as a measure of "time- and age-awareness" in the pre-industrial society of early modern Russia. In perhaps the first use of age-heaping by an *economic* historian Mokyr (1983) suggested that it could serve as a proxy for the degree of "quantitative sophistication" of a population. In advocating this general interpretation, Mokyr remained agnostic about the exact meaning of age-heaping:

> To the extent, therefore, that the degree of age-heaping …[is]…correlated with other qualities such as arithmetical ability ("numeracy"), a respect for accuracy, or a more serious attitude toward time, age heaping measures valuable human attributes which have the potential to create important economic externalities and play a role in development (Mokyr 1983; 246).

In Mokyr's view, then, age-heaping could be both an index of the distribution of individual skills ("arithmetical abilities"), and in this way a plausible proxy for individual human capital, and a reflection of factors with a broader, contextual dimension such as "respect for accuracy" or "a more serious attitude towards time".

Subsequent studies have documented the existence of a robust correlation between age-heaping and human capital indicators such as a literacy and primary schooling enrollments, at both the individual and aggregate level (A'Hearn, Baten and Crayen, 2009; Crayen and Baten, 2010; Hippe and Baten, 2012).[1] Though a direct link between historical age-heaping and numerical problem solving ability has not been demonstrated, recent contributions increasingly elide this connection and interpret inaccurate age reporting as a straightforward and unambiguous indicator of math skills, numerical cognitive ability, or numerical education (e.g. Baten and Juif 2014; see also Tollnek and Baten 2016).[2] The success of the age-heaping approach to estimating human capital is not surprising. Data on ages are relatively abundant in many historical sources, and are relatively easy to handle. Moreover, age-heaping's potential insights into basic computational ability accord well with the emphasis in recent research on cognitive skills and abilities, rather than "inputs" such as school enrollments (Hanushek and Woessman, 2008).

In this paper, we reappraise the relationship between age-heaping and human capital using evidence from nineteenth century Italy. Nineteenth century Italy is a promising testbed for such an inquiry. The Kingdom of Italy was a newly created country, in which the national census acquired important political connotations in terms of nation building. Alongside the data of the censuses, therefore, arose an intriguing contemporary literature commenting and discussing the results and their interpretation (Patriarca 1996). Furthermore,

---

[1] A'Hearn, Crayen and Baten (2009) studied the connection between literacy and age-heaping at the individual level using data from the 1850, 1870, and 1900 US censuses. Besides a significant effect of individual literacy on the probability of reporting a popular, rounded age, their regressions also revealed an additional and significant role for the literacy rate of an individual's birth region. Hence their study provides empirical support for both a "individualistic", and also a broader, "contextualist" interpretation. The most recent contributions seems to lean more towards the former interpretation. For example, in the much-used clio-infra data bank (https://www.clio-infra.eu/), age-heaping is assuredly labelled as "numeracy" and placed alongside average years of education.

[2] A partial exception is Baten, Crayen and Voth's (2014) study of age-heaping among the English cohorts affected by high grain prices during the Napoleonic wars. The authors acknowledge the potential influence of cultural norms and administrative practices before arguing that a sharp increase such as they find for these cohorts in 1851 and '81 census data could not have been caused by such factors. They proceed to interpret age-heaping as a measure of cognitive ability.

Italy was (and remains) a country characterised not only by major geographic differences in economic performance and living standards (Vecchi 2011), but also in institutions and cultural attitudes, so much so that influential interpretations have depicted two qualitatively distinct modernization processes unfolding in the North and the South (Cafagna 1988; Felice and Vasta 2015). This heterogeneity along several dimensions makes the Italian case particularly interesting for probing the nature and meaning of age-heaping. Finally, the early censuses report unusually rich age data, with single-year age distributions by province, gender, marital status, and literacy.

The remainder of the paper is structured as follows. In the next section we examine age-heaping patterns in successive Italian censuses, and investigate their correlation with age, literacy, gender, marital status. In Section 3, using micro data from death registers, we explore similar issues at the individual level in three provinces of Southern Italy. In Section 4, we discuss our findings in the light of evidence of a more qualitative nature that sheds light on the numerical skills and age-awareness of Italians in the second half of the nineteenth century. Section 5 summarises our argument and draws conclusions for research on age-heaping.

2. Evidence from the Italian census

2.1. *Measuring age-heaping*

The census of 1871 reports single-year age distributions for all 69 Italian provinces. Figure 2.1 displays the national totals. Among adults, the heaping on multiples of ten is unmistakeable, as is a set of minor peaks associated with ages ending in 5. More subtle but quite regular is a preference for even numbers.

Fig. 2.1 about here

Italians' preference for terminal digits zero and five ("0/5 ages") means that we can characterise the extent of age-heaping using the Whipple index. This index is defined as the ratio between the actual share of 0/5 ages and the 20 per cent share we would expect from a uniform distribution across terminal digits.[3] Here we multiply Whipple index values ("*W*") by 100 for ease of presentation. *W* ranges from a theoretical minimum of 0 to a maximum of 500; with 0/5 ages just equal to the expected proportion of one-fifth, *W* equals 100. Five Whipple points are equivalent to 1 percentage point in the share of 0/5 ages. For example, if the 0/5 share increases from 20 to 21 per cent, *W* rises from 100 to 105. Complications arise due to a tendency, found in widely varying contexts, for age-heaping to rise with age itself. This is evident in Figure 2.1 if we compare the

[3] The Whipple is just one of many possible measures of the deviation from an expected smooth distribution of ages. An obvious disadvantage is the inability to capture any form of heaping that does not take the 0/5 form. On its advantages – chiefly but not only simplicity, see A'Hearn et al. (2009). Whipple himself (1919: 169-170) attributed the "Index of Concentration" to the U.S. Census Bureau.

age-groups 20-29 and 30-39.[4] In a population like that of Italy in 1871, the influence of older people is swamped by their more numerous descendants if we pool terminal digit frequencies across all ages. The resulting downward influence on *W* becomes a problem if we want to compare times and places with different population age structures . To mitigate this problem, we employ a two step procedure, first calculating *W* values separately for each ten-year age-group, then computing their simple average as our overall measure of heaping. Think of this version of *W* as measuring the age-heaping of the average cohort rather than the average individual. For comparison we occasionally report the alternative in which frequencies for all ages are pooled, which we call a "one-shot" procedure.[5]

Table 2.1 about here

The *W* values reported in Table 2.1 show clearly how age-heaping is almost absent among young people and rises with age, at least up to the 53-62 group. Also evident is a systematic gender gap indicating less accurate age reporting by women in every age group. Note that for comparability with other studies we restrict attention to the five central cohorts spanning ages 23-72 in calculating the averages reported in the final rows of Table 1.[6] The overall figure of 152 places nineteenth century Italy clearly behind Northern and Central Europe (in the sense of having less accurate age reporting), on a par with other nations of Europe's Southern and Western periphery, and ahead of much of the rest of the world including Eastern Europe, Latin America, and parts of Asia.[7]

Fig. 2.2 about here

Figure 2.2 maps the geographic variation in age-heaping, which is extreme. This is one of the interesting features of the Italian case. With *W*=105, Belluno province in the Northeast displays only one percentage point of excess 0/5 frequency, a value as low as anywhere in Europe at the time. Sassari province on the island of Sardinia instead has *W*=230, which would make it an outlier in Europe, if it were a country. In Sassari, at least 26% of individuals reported an incorrect age (230 being the ratio of 46% actual to 20% expected 0/5 ages). We say "at least" 26% reported their age inaccurately because there is no way of

---

[4] See Baten and Crayen (2010) for a discussion of age-effects on heaping propensity.

[5] On a related problem *within* ten-year age groups see A'Hearn, Baten, and Crayen (2009), who explain the choice of age intervals such as 23-32 rather than the more natural 20-29.

[6] The 13-22 group is excluded because some samples, particularly those from military sources, have a great concentration of individuals at a particular age such as 18, 20, or 21. As for the 73+ range, the concern is with the reliability of the data. In many studies, though not here, sample sizes in this age range are small. There is also evidence of a tendency for the ages of the elderly to be not merely rounded to a nearby preferred number, but also exaggerated. An effect working in the opposite direction is a possible survivorship bias favouring high human-capital individuals who report their ages accurately.

[7] Comparative international age-heaping estimates can be found in Crayen and Baten (2010), Hippe and Baten (2012), and Baten and Juif (2014). The Table 2.1 estimate for the 43-52 age group neatly matches Baten and Juif's figure for the 1820s birth cohort (when their ABCC index is converted back to its original Whipple index value). Our figures are somewhat higher than those reported in Baten and Crayen (2010); this likely results from their averaging estimates for the same cohort from later censuses.

knowing whether ages with other terminal digits are accurate; they are merely "unheaped".

There is a clear North-South gradient in Figure 2.2, but more subtle patterns can also be discerned. The former Kingdom of Lombardy-Venetia in the Northeast stands out as an area of uniformly low age-heaping, though there were significant *economic* differences between its two constituent regions. The contrast with even their immediate neighbours is clear. Meanwhile though age-heaping is clearly much more prevalent in the South (the former Kingdom of the Two Sicilies) than elsewhere, its northernmost provinces in Abruzzi-Molise would not look out of place in Tuscany. And within the former Papal State there is almost as much variation as in Italy overall: from 119 in Bologna to 178 in Perugia.

2.2. *Age-heaping and literacy*

Is age-heaping a good indicator of the cognitive abilities that we might call numeracy? The census data give us no *direct* evidence on this, but do allow us to examine age-heaping's correlation with an alternative measure of human capital, namely self-reported literacy. Table 2.2 presents Whipple index values separately for literates and illiterates. *W* values are higher among illiterates in every age-group and for both genders. The same regularity is clearly evident in the underlying provincial data: *W* is higher among illiterates in 340 of 345 age-group- and province-specific comparisons.[8] It seems clear that the tendency to report a rounded age is linked to an individual's human capital. At the same time, there are one or two surprises in Table 2.2. One is non-trivial heaping of ages among *literate* individuals, who, in addition, show the same tendency as illiterates to heap more as their age increases. And the presence of gender gaps among both literates and illiterates, in all age-groups, suggests that age-heaping is about more than just education.

Table 2.2 about here

We can learn more by disaggregating, exploiting Italy's remarkable geographic variability. Figure 2.3 plots the two indicators by age-group and province. An overall positive relationship is immediately obvious. Equally apparent is that this relationship is somewhat noisy, and highly nonlinear. This is not what we would expect based on a simple model in which education endows individuals with both literacy and a reduced probability of reporting a heaped age. There is a mechanical effect at work here, with *W* running into its practical lower bound of 100 in the North while illiteracy presses up against its ceiling of 100% in the South. Whatever the reason, this nonlinearity makes both prediction and interpretation difficult. Do the southern provinces have just a bit less human capital than the less-educated provinces of the North-Centre, as illiteracy would suggest, or is the disparity dramatic, as age-heaping would indicate? Also

[8] The number 345 results from 5 age-groups observed in each of 69 provinces. Among men considered separately, the gap is negative in just 8/345 cases, never exceeding 5 points; among women in 28/345 cases, never exceeding 9 points.

catching the eye in Figure 2.3 are two outliers; the important provinces of Rome and Naples both have *W* values at least 50 points above what would be predicted based on literacy.

Fig. 2.3 about here

Separate provincial *W* values for literates and illiterates are plotted in Figure 2.4. As reported earlier for individual age-groups, for the averages in Figure 2.4 too age-heaping is greater among illiterates in almost every province – Treviso being the only exception. What is surprising is the variation in *W* within the two groups. If a few years of school inculcated basic literacy and numeracy, we would expect similar degrees of age-heaping among literates throughout Italy. Instead, *W* ranges from 100 to 150 among literates; among illiterates the range is from 105 to 230! Clearly, it is not education that explains these within-group differences. Consider the vertical line at *W* = 125 in Figure 2.4. At the bottom there are 23 provinces (19 South, 4 Centre) with higher *W* values among literates. Towards the top there are 19 provinces (15 North, 4 Centre) with lower rates among *illiterates*. Remarkably, the sparse educational elite of the South in provinces like Napoli, Reggio Calabria, and Bari reported age less accurately than the unschooled peasants and workers of the North.

Fig. 2.4 about here

Disaggregating by gender yields further insights. First, the gender gap in age-heaping is barely correlated with the gender gap in illiteracy (Figure 2.5). Though statistically significant, the correlation coefficient is only +0.16, and the outliers are many. Second, we also find differences by marital status (Table 2.3). Married women display the same degree of age-heaping as married men, with respective *W* values of 149 and 147. But as soon as women become widows, their age reporting changes dramatically and *W* shoots up to 191. (Among men there is no similar change.) This is not due to widows on average being older than married women; we see large gaps between widowed and married women in every age group but the youngest. A plausible explanation could be that proposed by Földvári, van Leeuwen and van Leeuwen-Li (2012), namely that married women (or their husbands when reporting for them) adapted their declared age to that of their spouse. After being widowed, they would have reverted to directly reporting their own, inaccurately-known age. A problem with this explanation is that widows heaped ages much more than did never-married women. Perhaps, then, widows either understated their ages with an eye on the (second-) marriage market or exaggerated them in the hope of acquiring the dignity of advanced age, choosing in both cases round numbers. Whatever the explanation, women's age-heaping patterns clearly reflect more than just their human capital.

Table 2.3 about here

Fig. 2.5 about here

2.3. *Age-heaping across five censuses*

A further dimension along which we can explore age-heaping in the Italian censuses is time. We can track the evolution of age-heaping forward through the five censuses from 1861 to 1911.[9] Or, given that basic human capital is typically acquired early in life, we can track it backward in time by looking at cohorts of different ages in a single census, and associating their age-heaping with the decade of their birth or school age.

This second approach, recently implemented for literacy by Ciccarelli and Weisdorf (2016), is attractive as a way of gaining insights into the pre-unification period. The difficulty in doing the same for age-heaping lies in disentangling two effects that work in the same direction, both raising *W* among older individuals. The first is a possible rising trend in human capital acquisition over the decades prior to the census. The second is the aging effect discussed earlier, which implies greater heaping among older individuals even in the absence of any time trend in education. By following cohorts across censuses, watching them age, we can estimate the aging effect and implement a correction for it.

The censuses of 1861 and 1871 provide single-year age distributions for 59 provinces: all but those of the Veneto and Lazio regions. Table 2.4 reports summary statistics on the changes in *W* observed, by age-group. They do, on average, reveal the expected aging effect. For the cohort aged 23-32 in 1861, it is unmistakeable: *W* increases by 20 Whipple points on average, and the change is positive in every province. The decreasing size of the age effect at older ages, and its eventual reversal beyond age 50, though not well understood, is a common pattern documented in many sources (Crayen and Baten 2010). What is remarkable in Table 2.4 is the range of effects across provinces, which is as wide as -53 to +59 for individuals in their 60s in 1861. The average across all cohorts in the rightmost column shows that it is not just a few rogue province-age groups that generate this result. The implied cumulative effect of aging from 23-32 to 73-82 is *negative* in twenty provinces.

Table 2.4 about here

Our estimates of the aging effect are thus puzzling. In addition to the problematic cases in which the effect is negative on average, there are an equal number in which it is positive, but too large. We expect that the "between" estimate based on cohort-specific differences between 1861 and '71 will be positive, but also that it will be smaller than the biased "within" effect calculated across cohorts from the 1861 census alone. (The "within" estimate is simply the average of the differences between adjacent age-groups, e.g. $W_{43\text{-}52} - W_{33\text{-}42}$.) In twenty-one provinces, the between estimate is positive but larger than the within.[10] In only

[9] The missing census in this period is that of 1891, cancelled due to a fiscal crisis.

[10] It is possible that in these provinces the pre-unification trend in human capital accumulation was actually negative, so that the within estimate confounds two effects working in *opposite* directions. Against taking this interpretation at face value is the lack of any geographic coherence in this group of provinces. Unlike the group in which the between estimates are negative, which

18 of 59 provinces do the between estimates conform to the expectation of being positive and smaller than within.

The extreme variety in both the sign and magnitude of estimated aging effects is problematic in two ways. First, it makes it impossible to devise a correction procedure that is transparent, uniform, and reliable. We therefore give up, reluctantly, on age-corrected estimates of human capital before unification. Second, some between-census changes are not consistent with a strict human capital interpretation of age-heaping and must instead be due to changes in census enumeration procedures or families' attitudes and behaviour regarding age reporting.

We now consider tracking age-heaping forward through time, across successive censuses. A significant complication in this exercise is that the censuses of 1881 and later report adult age distributions only in terms of five-year age-groups such as 30-34 and 35-39. Fortunately, the predominance of heaping on multiples of ten means that, though we cannot compute Whipple indices, we can still construct a crude measure of heaping. This alternative index is based on the ratio of the frequency of the age-group with terminal digits 0-4 to the frequency of the successive group with terminal digits 5-9, for example $f_{30\text{-}34}/f_{35\text{-}39}$. We will use the notation R_5 for such ratios. Pooling over the entire age range from 20 to 99, the ratio of 0-4 age frequencies to 5-9 frequencies is 1.31 in 1871.

Like the Whipple, an index based on raw R_5 ratios is sensitive to the age structure of the population. If frequencies decline steadily with advancing age, R_5 will exceed one for every successive (0-4, 5-9) pair. By *how much* it exceeds 1 depends on age-specific mortality, and potentially also migration, which varied by age. To address this problem we generate predicted age-group frequencies by smoothing the distribution (using locally weighted regression), compute a set of predicted R_5 ratios from the smoothed data, and finally compute and report a final R_5 index that is the ratio of the raw value to our prediction. As with the Whipple index, we multiply by 100 for ease of presentation. An example is given in Appendix 2.

The correlation of R_5 with W across all province–age-group pairs is strong (+0.73) in the census of 1871. The relationship would be both tighter and more linear were it not for the Sardinian provinces, whose R_5 values for individuals in the 40s, 50s, and 60s are outliers – far lower than would be predicted based on W. This illustrates a weakness of R_5, which in this case fails to respond to a high degree of heaping on 0/5 ages because 0 does not dominate 5 as much as elsewhere.[11] Table 2.5 presents average R_5 values for three macro-areas across five censuses.

---

has a clear nucleus of ten provinces in the northwest, the group in which between estimates are positive but too large is highly scattered. Potenza is unique in having a negative within estimate. Because it is so tiny (-0.4), we treat it as not meaningfully different from a zero or small positive value and include it in the "positive but too large" group.

[11] Another weakness is the fact that a nontrivial number of R_5 values (53/345) are below 100. These are mostly quite close to 100 and reflect our crude index's difficulty in distinguishing the effects of heaping from the effects of mortality. There are a few anomalous, more extreme cases,

Considering first the North, we see that overall age-heaping diminishes fairly steadily in each successive census. Tracking individual cohorts through time, moving down along a diagonal in Table 2.5, that is, we see as well that the aging effect is at work in a regular way, so long as the starting *W* value is not too low (105 or more, say). The situation in the South and Islands is quite different. Overall age-heaping alternately rises and falls, rather than diminishing steadily, and there is a discontinuous drop between 1881 and 1901, without parallel elsewhere in the table. Across the first three censuses from 1861 to '81, the aging effect is strong and regular in the South, but this changes in 1901. Every cohort that we can track across the 20-year gap shows the same or *lower* age-heaping after two decades. Afterwards, normal service is resumed, with R_5 increasing for three of four cohorts between 1901 and 1911. Following the generation in their twenties in 1861, we see R_5 changes of +8 as they reach their thirties, +22 in their forties, *-20* as they age across two decades into their sixties, and +17 as they enter their seventies. 1901 stands out as a complete anomaly. The only plausible explanation is a change in census-taking procedures, or compliance with them, in the South. We return to this issue in Section 4.

Table 2.5 about here

3. Micro evidence

Census age-heaping seems to have been influenced by human capital, but also by data collection procedures, social norms, and as yet unidentified local factors. In an effort to sharpen our focus and assess the relative importance of these diverse influences, we turn now to individual-level data. This section reports the results of a exploratory study of microdata from the high age-heaping regions of the Italian South.

3.1. *Death in Naples*

In 1808 Gioacchino Murat, installed by Napoleon as King of Naples, instituted a system of vital record keeping in southern Italy. Information on births, deaths, and marriages was collected by local civil authorities, recorded in duplicate on printed forms, and retained in archives. The system was maintained by the restored Bourbons throughout the decades leading up to Italian unification. The Neapolitan death records in particular lend themselves to the sort of analysis we wish to conduct. First, they contain both direct declarations of age from the witnesses and indirect reports regarding the deceased. Second, though the dead tend to be either elderly or children (who are not included in the study), the witnesses cover quite a wide range of ages. Third, the format of the records, handwritten but on a single-page printed form, facilitates data collection. Digital images of individual records have been made available on-line via the *Antenati* ("ancestors") website maintained by the *Direzione Generale degli Archivi* of the

some of which may result from the migration of working age males distorting provincial age structures.

Ministero dei beni e delle attività culturali.[12] We study 1861, to coincide with the first Italian census.

It was a legal requirement to declare a death in order to receive permission for burial, and no fee was charged. Two adult males 21 or older were to be found among the relatives or neighbours of the deceased to declare his or her death and play the role of witness. The regulations specified that names, ages, addresses, sex, and occupation of both witnesses and the deceased be recorded, along with information on the parents of the deceased. If able, witnesses were to sign the death certificate, which provides us with a measure of literacy. The perils of using signature ability as an indicator of literacy are well known. On the one hand, a competent reader may be insecure in his penmanship; on the other, an illiterate can learn to make the pen strokes that indicate his name.

Data were drawn from three locations: metropolitan Naples, Bari province, and Reggio Calabria province. As noted in Section 2, these are among the highest age-heaping provinces in Italy, where even literate individuals reported age less accurately than the uneducated of the north. A total of 4,353 records of unique individuals aged 22 or more were obtained. They refer predominantly to witnesses rather than the deceased, not only because there were two witnesses for each death, but also because a tragically large share of the dead were infants and young children, while our study requires data on adults. Over 500 occupations were found in the death records, which have been coded into seven categories: unskilled workers (the most numerous group, including most rural occupations), enlisted soldiers, artisans, distinct dependent and autonomous service categories, "elite" occupations such as the liberal professions, and a residual group. Further details of dataset construction can be found in Appendix 3.

3.2. *Death record age-heaping*

Figure 3.1 depicts the age distribution among adults in our sample. Extreme heaping on multiples of ten is immediately obvious; they are 30% of all recorded ages. A more subtle pattern is that, unlike the census data, the second-most-favoured digit is *not* five. Six, eight, and two are all preferred to five, a pattern we discuss further in Section 4.[13]

Fig. 3.1 about here

Aggregating towns and neighbourhoods into what we can loosely call districts, we calculate Whipple index values following the procedures described in Section 2 (Table 3.1). We see the familiar pattern of age-heaping rising with age. The district average *W* values range widely, from a low of 166 in central Naples to 258 in Bari and its surroundings. They are in general higher than the provincial *W* values from the census of 1861. It is interesting to note that the periphery of

[12] www.antenati.san.beni-culturali.it

[13] Also evident is the effect of the restriction that witnesses be "at least 21 years of age", which seems to have been typically interpreted as "22 or more"

Naples looks more like distant rural provinces than the nearby city. The *pattern* of heaping also varies. The popularity of 6 evident in Figure 3.1 derives from Bari and Reggio provinces, not Naples, for example.[14] Because the Whipple index is insensitive to the preference for sixes, eights, and twos, we present an alternative age-heaping index in Table 3.1: the sum of the absolute deviations from ten per cent, of each terminal digit's share. This index can vary between 0 and 180, and is calculated using a one-shot procedure due to small sample sizes. The proportional variation in the alternative index is greater, but the rank order of districts by age-heaping is almost identical.

Table 3.1 about here

The final row of Table 3.1 gives the signature rates for our adult male witnesses. They range from an extraordinary 83% in Naples city to an equally extraordinary 3% in the Calabrian districts. There is a general alignment of low age heaping with high signature rates, but the relationship is certainly not monotonic. Naples' immediate hinterland stands out with high signature rates and high Whipple index values at the same time. Altamura has easily the greatest age-heaping (by either index), but a signature rate that is the median.

3.3. *Individual age reporting*

Heaping on multiples of 10 in our sample is extreme. Since the majority of these observations must be incorrect, predicting which individuals report "heaped" ages is tantamount to identifying who did not report their age accurately. We do this by means of logit regressions in which the dependent variable is a dummy for ages ending in 0 or 5 (where 5 is included for comparability with other studies although it is not popular in our sample) and the explanatory variables are age-group, sex, occupation, geographic region, and signature ability. Table 3.4 presents our estimates of the marginal effects of these regressors in several specifications. The reference group is an illiterate, unskilled, male witness aged 22-31, from Bari or Barletta district.

Table 3.2 about here

Columns 1 and 2 report estimates of a baseline model without signature ability, referring to the full sample of 4,353 and the narrower subset of witnesses, respectively. In both cases we observe the same pattern of age, geographic, and occupational effects. Age-effects follow a hump-shaped pattern with the highest probability of reporting a "heaped" age in the 52-61 age range. Potentially of interest in its own right, age-group is included here primarily as a control. Geographic effects are also strong and significant. Relative to the reference district of Bari-Barletta, the inhabitants of other agricultural districts, including peripheral Naples, are 7 percentage points more likely to report a 0/5 age.[15]

[14] At the level of individual towns, we find sixes to be as high as 30% or more of all ages in five cases, as low as 0% in three others.

[15] In a regression with six district dummy variables, all are associated with a statistically significantly higher probability of reporting a 0/5 age compared with Naples city. One cannot

Meanwhile, residents of Naples city are four and half points *less* likely to report such ages. These effects are meaningfully large compared to the variation across districts in raw probabilities of reporting a heaped age: 31% in Naples, 40% in Bari-Barletta, and 47% in the other rural districts. Turning to occupational effects, all groups have markedly lower probabilities of reporting a heaped age than unskilled labourers. The effects are large and mostly statistically significant: 5-6 percentage points for service workers, 7 points for artisans, 11 points for the socio-economic elite. Gender differences can be read from the female occupational effects, which indicate that, outside the social elite, women were more likely to have 0/5 ages reported than the lowest-skilled men.

The models in Columns 3 and 4 add signature ability to the regressors. Signature ability itself has a powerful and statistically significant effect, lowering the probability of a heaped age by 12 percentage points in Column 3. It also lowers the size of several other effects. All occupational effect estimates are diminished, by about half, and the Naples effect reverses sign. On the other hand, the regional effect of agricultural districts other than Bari-Barletta strengthens to more than 9 percentage points. Whatever is going on in these areas, it is not explained (or not *fully* explained) by either occupational structure or literacy. And artisans remain a statistically significant 4 points less likely to report a heaped age than unskilled workers, even after controlling for signature ability. For the socio-economic elite, the effect is 5 points, with an associated p-value of 0.104. Column 4 adds town fixed effects to the model, treating the neighbourhoods of greater Naples as separate towns. The town effects are themselves large and highly significant. Their inclusion reduces the predictive power of signature ability by half in Column 4, and causes most occupational effects to strengthen. Those for artisans and the elite are statistically significant at 5%.

Table 3.3 presents estimates of the "signature" model of Table 3.2 for distinct sub-samples.[16] In the agricultural regions, signature ability has a powerful and statistically significant effect, reducing the probability of reporting a 0/5 age by roughly 16 percentage points. Occupational effects are somewhat inconsistent and never statistically significant. Naples city looks completely different. Here signature ability has much less predictive power and is not statistically significant. It is instead occupational effects that are strong in Naples. For craftsmen and the elite, the effect is statistically significant and, at negative 15-16 points, as powerful as signature ability is in agricultural districts. The same patterns remain when town/neighbourhood effects are added to the model. (The local fixed effects themselves are jointly highly significant except within Naples city.) Where signature ability was super-abundant, in central Naples, it has little predictive power; occupation and neighbourhood matter more. Where it was scarce, in the provinces, signature ability is the most reliable predictor.

Table 3.3 about here

---

reject the null that the Bari and Barletta effects are identical; the same is true of the remaining four rural districts Altamura, Palmi, Reggio, and the Neapolitan periphery.

[16] A test of pooling restrictions fails to reject the null that the slope coefficients in the regressions are the same across the three sub-samples, but in a model in which all coefficients are allowed to differ several interaction effects are statistically significant.

4. Of numbers and numeracy, birthdays and bureaucracy

Several of our findings sit uncomfortably with the idea of age-heaping as a measure of numeracy. In this section we present further evidence against this interpretation and develop the alternative argument that the correlation of age-heaping and illiteracy in Italy arose from common roots in a process of administrative and cultural modernisation.

4.1. *The State*

The sharp drop in age-heaping in southern provinces between 1881 and 1901 documented in Table 2.5 must reflect a change in census operations. It is not plausible to argue that the numeracy of southerners of every age improved in this, and only this, period. Census officials were well aware of age-heaping – and not very happy about it. It was an instantly recognisable symptom of a general weakness in census procedures, and a specific problem for the development of mortality tables.[17]

One response was to fine-tune census questions and exhort local census workers to greater efforts. The mechanics of the census involved distribution of forms to each family, to be filled out by the head of household for later collection. In practice, as the head of household was often illiterate, it was frequently another household member or neighbour who wrote out the family's responses, a potential source of inaccuracy right from the start. In the event that a family failed to complete the form properly, census workers were to call back , interview the family, and record answers in their own hand. The instructions of 1881 specifically directed local enumerators to check reported birth year and age for consistency. In 1901 instructions to ensure age data were recorded with "scrupulous precision" were reiterated. Moreover, the census question was changed with the specific goal of eliciting more accurate responses; age was no longer asked about at all, but only the month and year of birth.[18] Only as a last resort should age in years be recorded, from which enumerators were to calculate birth year, which they were further instructed to check against local birth records.

A second response was to centralise processing of the individual family census forms. Though overseen by the *Direzione generale della statistica* in Rome, it was local government (the *comune*) that was responsible for distributing one form to each household, collecting completed forms, checking them for errors and omissions, remedying these, transcribing from the family form a set of individual records, tabulating information such as the distribution of ages from the individual records, and forwarding the results up the administrative chain of command. To carry out these tasks *comuni* were to engage temporary clerks and

[17] Numerous efforts were made to circumvent the problem by means of interpolation, e.g. Italy (1885).

[18] Census of 1901, Volume 5, *Relazione sul metodo,* pp. LVIII, 193/4, 196.

rely on the unremunerated voluntary efforts of local notables, professionals, schoolteachers, and government employees. Importantly, no funding was provided by the central government to finance these operations.

The results were highly unsatisfactory from the point of view of officials in Rome. In the preparations for the census of 1881, Luigi Bodio, head of the *Direzione generale di statistica*, complained about age distributions in earlier censuses specifically.

> "I could cite the names of numerous *comuni* whose tabulations of the population by age were written from the imagination; but they were invented poorly, because in the gradation by age the number of elderly appeared greater than that of the middle-aged. If the municipal employees charged with carrying out that enumeration had invented with artfulness, perhaps none of us could have detected the falsification." (Favero 2001; 138).

The census of 1901 was the first in which individual records were filled out by household heads and transmitted directly to Rome by the *comune*.[19]

It was not just local government but also the cooperation of local populations, notables and common people alike, that determined the success of data collection. Writing about the upcoming census of 1871, Bodio wrote that the situation was not propitious with the clergy alienated by the recent annexation of Rome and because of

> "the suspicion that has grown among the people that the census must serve fiscal purposes. And they are not wrong. … The criteria for distributing taxes, conscription, etc. are based on the number of inhabitants. And the poor multitudes feel the burdens, more than the honours, of being citizens of the state. (…) With the census, it is a question of fighting a great battle: a battle against ignorance, against the devices and mechanisms of concealment of taxpayers…" (Favero 2001; 94).

It was hoped that the statistics of large cities might be more trustworthy due to the greater average educational levels of urban populations. In the census of 1881, these hopes were disappointed, at least in age reporting.[20] What is interesting is the reasoning about education by census officials. It was believed to increase people's acceptance of, even identification with, the nation-building purposes of the state's intrusion into private lives (Favero 2001; 98).

---

[19] 1901 was also the first census in which mechanical data processing and calculation aids were employed (to a very limited extent).

[20] In the census of 1881, separate age-distributions were published for the 69 provincial capitals. These had the single-year format of earlier censuses, which was otherwise discontinued and would disappear altogether in 1901. A comparison of the capital cities with the relevant provincial totals is possible on the basis of the R_5 indicator of Section 2. This shows that the expected urban effect on age-heaping was significant in some provinces, particularly in Sicily and Sardinia, but on average small, and not infrequently of the "wrong" sign.

We call the state's ability to monitor activity, gather and process accurate information, implement policy, and elicit the cooperation of local institutions and individuals in this process "administrative capacity". The best explanation for the discontinuous improvement in census age-heaping in 1901 is an improvement in the Italian state's administrative capacity.[21] And this varied as much geographically as temporally. The prefects, the central administration's representatives in the provinces tasked with monitoring and controlling the actions of local government, encountered difficulties everywhere, but especially in the South. Here it proved difficult to induce many *comuni* to comply with basic legal requirements such as keeping proper accounts and publishing budgets, to say nothing of effectively providing mandated public services in areas such as infrastructure, health, and education (Randeraad 1993; ch 6). At the other extreme are cases from the north in the late 1860s, in which towns in recently-annexed Veneto requested permission to carry out *ad hoc* local censuses as a foundation for the sort of continually updated population registers required under Italian administrative law but not actually implemented by many towns in other parts of Italy (Favero 2001; 57).

Persistent local variation in the competence of peripheral officials and their commitment to the priorities of the central administration derived from historical traditions in the pre-unification states; the same was true of the attitudes toward state information gathering among local citizens. Tracing the history of public statistical inquiries, Patriarca (1996) finds some common features in the efforts of statesmen and scholars throughout pre-unification Italy.[22] Enlightenment-era traditions of political arithmetic and political economy were well represented in the eighteenth century, and not just in the North. The entire peninsula then shared the experience of Napoleonic administrative bureaucracy in the period of French domination, and much of this apparatus was retained by restoration governments. "(E)very state … had kept statistics and had its own heroes" (Randeraad 2010; 129). Yet there were also real differences.

The primacy of Habsburg Lombardy-Venetia was generally acknowledged. The most prominent early nineteenth century statisticians, Melchiorre Gioia and Gian Domenico Romagnosi, were both active in Milan, as was Pietro Maestri, author of the first *Annuario statistico italiano* and later the founding head of Italy's *Direzione generale di statistica*. And it was here that the greatest number of statistical works were published per capita (Table 4.1). In terms of practical results, the land register (*catasto*) of Lombardy was considered a triumph at the time and in retrospect. Completed in 1760, the *catasto teresiano* (after Empress Maria-Teresa) was based on a comprehensive and detailed mapping of landholdings conducted by Dutch surveyors under the supervision of imperial officials from other provinces. This was, by a margin of decades, an absolute first

[21] Crayen and Baten (2010; 90) find that variables likely to correlate with administrative capacity (state antiquity, the number of censuses previously conducted, andschool enrolments) lower age-heaping in a large, international sample. The strongest effect is that of school enrolments.

[22] These two paragraphs are largely based on Patriarca's *Numbers and Nationhood*, particularly Chapter 4, "Official Numbers".

in Italy. Though for a long time the data on population, crime, and the economy collected by the Austrian administration were mostly reserved for internal use, by 1846 the regular publication of "monumental compilations in strictly tabular form" had commenced (Patriarca 1996, p. 119).

Table 4.1 about here

In a series of regional case studies Patriarca (1996) recounts the difficulties experienced by statistical offices and commissions. Uncooperative local authorities were reluctant to provide information, while higher authorities were at times reluctant to divulge it, or to tolerate the liberal or patriotic tendencies of their statisticians. Nonetheless important studies were undertaken and at least some statistics regularly published in Lombardy-Venetia, Piedmont, and Tuscany. Things were different in the South. In Sicily the formal institutional set-up appeared ideal. The restored Bourbons assigned their intendants (prefects) in separatist-minded Sicily assistants for the gathering of statistics, and in 1832 created a *Direzione generale di statistica* in Palermo, staffed by an extraordinarily ambitious and active group of individuals including Francesco Ferrara, later one of Italy's most famous economists. But the actual results were meagre indeed. The *Direzione* statisticians encountered lack of cooperation or outright resistance at every level, from the intendants, from local authorities, from the social elite, and from the people. Without authority over intendants and mayors, their efforts were doomed to fail (Patriarca 1996; 88-95).

There is evidence of something similar in the death records analysed in Section 3, where the extent and pattern of heaping varied from one town to the next. Ferdinando Maria Jusseppi was mayor of three small towns in Reggio Calabria province (district of Palmi) in 1861.[23] Fully 63 per cent of recorded ages in "his" towns were 0/5 ages. In addition every one of 13 deceased women had the occupation *filatrice* (spinster), regardless of age, marital status, or other circumstance. Similarly, of 51 males, both witnesses and deceased, 48 had the occupation *bracciale* (labourer). Spinster and labourer seem to have been automatic defaults. Meanwhile, the signature rate among 47 witnesses was identically zero – even for the lone individual identified as a property owner. Minimal information was recorded about the deceased and their relatives. One has the impression that in many cases ordinary people in Jusseppi's towns may not even have been asked their age, occupation, and whether they could sign.

Turning to the indicators of administrative capacity in Table 4.1, we see that the Bourbon kingdom was last among the major pre-unification states by statistical publications, while its *catasto* was the only one still based exclusively on the verbal declarations of landowners rather than a survey. Lacking a mechanism for valuing properties themselves, southern fiscal authorities were also forced to rely on landowners' declarations of rental rates to establish their tax assessments. A crude summary measure of administrative capacity is total tax revenue per capita. It is crude both because the structure or prosperity of some

[23] Also appearing in the death records are a deputy mayor (*supplente delegato*) and clerk (*cancelliere*), both of whom shared the Jusseppi surname with the mayor.

economies allows for high tax revenues even in the absence of sophisticated information-gathering capabilities, and because taxes are a policy choice. Still, the existence of a North-South gradient, with revenues per capita distinctly lowest in the Kingdom of the Two Sicilies, is hard to overlook. To be sure, our evidence is flawed and incomplete, but it all points in the direction of lower administrative capacity in the parts of Italy where age-heaping was greatest.

4.2. *The People*

There are both a priori reasons and a few scraps of direct evidence to suggest that most nineteenth century Italians, north and south, possessed at least rudimentary mathematical knowledge. Numbers were everywhere in the world of both peasants and city dwellers, even in their games. In the ubiquitous game of *morra*, once a popular motif for painters of common life in Italy, the object is to quickly guess a sum: the total number of fingers extended simultaneously by the players.[24] Meanwhile, the state *lotto*, or lottery, ensured that from the late seventeenth century all Neapolitans knew their numbers from 1 to 90 (Macry 1997).[25] A passion for the *lotto* extended from the Bourbon King himself (payment of whose occasional winnings is noted in the *lotto* administration's records) to the denizens of the city's slums. The poor made micro-wagers and suffered (enjoyed) associated micro-losses (gains) on a regular basis. The minimum wager was only one *grano*, a fraction of the price of a loaf of bread, and in practice it was possible to buy shares as small as a twelfth of the minimum, often on credit extended by the neighbourhood *lotto* shop operator (despite an official prohibition). Printed lottery tickets specifying the numbers selected and the potential winnings for each wager, as well as the public display of the week's winning numbers at *lotto* shops, presumed widespread understanding of written numbers. So too did the popular books illustrating the numerological interpretation of dreams ("*smorfia*"). Some of these *smorfia* manuals were intended for illiterates and matched illustrations of dreams with the associated lucky numbers written in Arabic numerals. Writing in the 1880s, the celebrated journalist Matilde Serao noted that the *smorfia* was so ingrained in Neapolitan habits of thought that a colloquial expression for crazy was to call someone "*nu vintiroie*" – a twenty-two; to call a woman a seventy-eight was something worse (Serao 1885; 57). In this period, illegal private lotteries with minimum wagers as low as a couple of pence flourished alongside the state *lotto.* Receipts, though written in pencil on dirty scraps of paper, had the canonical form, specifying the bettor's wager, chosen numbers, and associated payoffs (Serao 1885; 64).

---

[24] Similar to the game *rock, paper, scissors* widely played in the English-speaking world, players with fists extended before them simultaneously extend from 1 to 5 fingers. The game, which can involve more than two players, was very popular in Naples in the first half of the nineteenth century according to Rocco (1857). De Sivo (1865) goes so far as to claim *morra* as the root of word *camorra*, i.e. the Neapolitan mafia. On the game's popularity in the southern region of Abruzzo today, see Spitilli (2012).

[25] In the *lotto*, a random number between 1 and 90 was drawn from each of five urns. The gambler's wager was that one or more of three numbers chosen by him or her would be drawn. Winnings were proportional to the amount wagered, with the ratio lowest for a single number appearing (ca. 12:1) and highest for a triple (as high as 5000:1, depending on the period). In the years leading up to 1861, *lotto* drawings took place weekly.

Of course, superstitions like the *smorfia* are often cited as evidence of a pre-modern, non-rational attitude, just as addiction to the *lotto* is supposed to be evidence of a passive, fatalistic belief that one's future is a matter of chance, impervious to control or plan (though possibly susceptible to divination). We do not challenge that interpretation here, though it is worth noting that playing numbers suggested by dreams or recent events was *not* irrational given the structure of *lotto* payoffs.[26] Our argument is only that the popularity of the game suggests the ability to count and to recognise and name numbers was widespread.

And of course there were matters more serious than gambling that required basic numeracy. Near the subsistence minimum, there is no room for errors such as a sharecropper failing to recognise that his share of 86 sacks of grain is 43, a shepherd leaving behind uncounted sheep in a mountain pasture, or an urban household being cheated by the baker or grocer. To make this essentially evolutionary argument is not to claim that the average Italian landless labourer understood compound interest, nor to deny modern findings that poverty-induced stress impairs cognition (Mani et al. 2013). It is merely to assert that there were strong incentives for a family to acquire a minimal competence in numerical reasoning.[27]

The tools required for such competence were there to be acquired, though not necessarily in forms familiar to us today. Methods of counting to 100 and beyond on one's fingers, though no longer set out in Italian maths textbooks in the nineteenth century as they had been earlier, remained common in practice (Roggero 1994). And traditional currency and measurement systems were designed to facilitate quick mental calculations based on halving and doubling.[28] In Italy as elsewhere, the monetary system was based on 12 and 20, numbers twice-divisible by two: 12 *denari* per *soldo*, and 20 *soldi* per *lira*.[29] Twelve was also the basis of various physical measures: in southern Italy 1 *palmo* (.26m) was 12 *once*; 1 *botte* (524l) was 12 *barili*; and 1 *libbra* (.32kg) was 12 *once*. Attempts to introduce decimalisation beginning under French domination in the Napoleonic period were resisted. As late as 1888 the school inspector for the *circondario* of Lagonegro (south of Naples, in Basilicata), wrote that teaching in mathematics was producing the most satisfactory results of all disciplines, perhaps because the subject suited the "speculative and calculating nature" of locals, but that practical instruction in using the metric system was held back by

[26] The *lotto* did not pay out a fixed total divided among all holders of winning tickets. (In that case, one would not want to play popular numbers chosen by other gamblers, if one correctly understood that the probabilities of all numbers being extracted were equal.) Instead, it paid a fixed multiple of the amount wagered to every winning player, the total payout thus being variable.

[27] Tollnek and Baten (2012) make a similar argument for farmers, who not only had the incentive to develop numeracy skills, but were able to pass these on to their children, who benefited from an abundant diet that removed nutritional constraints on cognitive development.

[28] The so-called "Russian peasant multiplication" method reduces *any* multiplication problem to a series of halving, doubling, and summation operations.

[29] In the Kingdom of Naples 12 *cavalli* per *grano* and 20 *grana* per *tarì* played equivalent roles in the monetary system, but there were other denominations in circulation such as the *carlino* of 10 grana and the *piastra* of 12 *carlini*.

the persistence of ancient Neapolitan measures in daily usage (Italy 1888). Part of the explanation for this persistence is likely to be reliance on established routines for mental calculations. Traditional measurement systems may also explain some details of the heaping patterns in reported ages, such as the preference for sixes, i.e. half-twelves (Thomas 1987). A similar holdover was the use of twenties in reporting ages, e.g. "three-twenties" for sixty, especially in the south (Menninger 1977, 49; Price 1991, 465).

Pen and paper were not common in nineteenth century Italy, nor the years of practice required to achieve proficiency with quill and ink, but it was common to record numbers temporarily according to traditional systems of marks made with a piece of charcoal or a sharp object, and use of the abacus remained common into the nineteenth century (Roggero 1994).[30] For record-keeping, recourse was had to tally sticks and their like. Shepherds had employed these for centuries throughout Europe, and they continued in use into the nineteenth century. Tally sticks recording a transaction and split down the middle into a matched pair, one half for each party, were recognised as legal documents in Napoleon's *code civil*, and the system was still common in Puglia in the twentieth century. Illiterate shepherds there not only recorded commercial transactions in this way, but numbered their sheep analogously, following rules for recording units, tens, hundreds and thousands by making incisions into the upper and lower lobes of both ears (La Sorsa, 1953).

We know from modern studies that children learn to count readily, largely figuring it out for themselves before school, and even devise strategies for basic addition and subtraction without any instruction, such as breaking 7+8 into 7+(3+5) and rearranging this into the two easier sums 7+3=10 and 10+5=15 (Dehaene 1997, Ch. 5). Children whose occupations keep them out of school but involve them in monetary transactions develop their own "street maths": non-standard methods for solving more complicated problems. In a classic study, unschooled Brazilian street vendors aged 10-12 performed better on several mathematical tasks than their counterparts in urban schools (Saxe 1988). The Soviet developmental psychologist Luria studied cognition among rural Uzbeks in 1931-32. His interviews vividly revealed an inability or unwillingness to think abstractly and outside the range of personal experience among the uneducated members of this population. Some questions involved problem-solving, for example computing the time to reach a destination by bicycle, given that it took 30 minutes by foot and the bicycle was five times as fast. In such cases Luria checked the subject's numeracy by posing the same question in different form, e.g. "divide thirty cookies equally among five men". Even for illiterate, never-schooled individuals, "simple computations used in everyday practical affairs presented no special difficulties" (Luria 1976; 126). It would be astonishing if ordinary Italians of the nineteenth century failed to acquire this minimal sort of numerical competence. The limits of finger-counting, street-maths, and the tally stick are clear, and the claim for numeracy made here is only that the typical

[30] Here "abacus" refers to placing tokens on lines drawn on paper or in the sand, or incised in a board, rather than moving beads along wires in a frame. Roggero (1994) notes that by 1800 abacus methods had mostly disappeared from textbooks and schools, but must have remained in common use.

Italian had the cognitive capacity required to comprehend, to state, to record, and to recognise the numbers relevant for human age.

Contemporaries took Italy's high rates of illiteracy very seriously. It was a source of national embarrassment. Though well-aware of age-heaping, as we have seen, they did not deplore it as further evidence of the population's tragically underdeveloped cognitive abilities. They reported instead, in a matter-of-fact way, that it resulted from ignorance of precise age. "The relative maxima … at the ages indicated by numbers which end with 0 or 5, are due to the ignorance or the negligence of the individuals being surveyed, who in large part do not remember their age exactly, but when asked about it round the numbers," we read in one report (Italy 1885b; 87). Inaccuracies, according to the general report on the census of 1881, derived from rounding, deliberate understatement (among women), exaggeration (among the elderly), and inconsistent understandings of whether current age meant years completed or the year currently underway. In general, people simply did not think of age in terms of single years, "relatively small units of measure not used in practice by many of those interviewed" (Italy 1885a; xxxix).

Today, at least in rich countries, birthdays and bureaucracy offer constant reminders of our age and birth year. In nineteenth century Italy, especially in the South where age-heaping was so prevalent, neither played an important role. The absence of the state from much of daily life – no compulsory schooling, no voting rights, no regular policing, no old age pensions, no taxation of income – limited the average person's interactions with the administrative apparatus, such as it was, and meant that proof of identity was never required, files of individual data never compiled. When interaction with government officials was required, as in the case of vital records declarations, approximation was undetectable (Section 3). The state generated no incentive to know one's age accurately.

What of birthdays? There is no evidence that they were celebrated in ordinary Italian families in this period. A modern invention, or rediscovery, the celebration of birthdays has been linked with changing perceptions of time by Jean-Claude Schmitt (2007). A staple of sociological theorising since late nineteenth century pioneers such as Durkheim and Simmel, the argument is that the marking of time is a socially embedded process that serves the function of coordinating group activities (Sorokin and Merton 1937). In medieval Europe, the Catholic and Orthodox churches' cyclical calendars of recurrent religious feasts dovetailed with agricultural rhythms underlying social life. In the West, the rise of Protestantism weakened the cult of saints and the general orientation of life around the liturgical calendar, while urbanisation and industrialisation cut links with natural cycles, created coordination problems, and focussed attention on the value of time (Schmitt 2007, Thompson 1967). As Kaiser and Engel (1993; 836) note, "… (I)n urban industrial society 'time rules life' … (a)nd age-awareness, together with numerous other time calculations built into labor and recreation, is very high". They interpret age-heaping in early modern Russia as an indicator of the slow diffusion of the modern calendar, the clock, and "merchant's time" there.

With regard to birthdays specifically, the medieval Church had opposed their celebration as a pagan ritual. If a particular day mattered, it was the day of either baptism or death – either interpretable as marking the start of one's true life in God. Calendar age was similarly unimportant in the medieval conception of a life course. The "ages of man" were developmental rather than strictly chronological, and they brought one back eventually to where one started. In the modern conception of time, linear and progressive, the progressive accumulation of age, or "capitalisation of years", acquired greater meaning too (Schmitt 2007). Individualism and, more speculatively, astrology may also have had roles to play. Despite such changes, Schmitt finds scant evidence of the celebration of birthdays even among the rich and well-born, even among individuals who definitely knew their exact birth date and year, before the late eighteenth century.[31] Birthday celebrations spread among the aristocracy and bourgeoisie only in the nineteenth century, and did not reach the working classes before the twentieth (Schmitt 2012, 73-92).

Italy seems to have lagged behind northern Europe in adopting the birthday. Impressionistic evidence is provided by the frequency of birthday words appearing in printed works. Among books published circa 1850, "*Geburtstag*" occurs more than ten times as frequently as "*compleanno*" and "*anniversario della nascita*" together in the relevant German and Italian collections of digitised historical texts. "Birthday" occurs more than twenty times as often in the British English collection.[32] The trends are rising in all three languages, but latest and most weakly in Italian. And in southern Italy, at least in Naples, it was celebration of one's saint's day (*onomastico*) that remained the norm as late as 1960, as we are reminded in a recent novel: "Although custom had it that it was the *onomastico* that should be celebrated – birthdays were then considered irrelevant – the Sarratores and Nella insisted on arranging a little party in the evening" (Ferrante 2011, 221; Galasso 2012). Saint's days did not coincide with birthdays ordinarily; in the case of popular given names they were shared with countless others and might fall on a religious holiday. The life of the individual and his or her age is less in focus than in a birthday celebration. To the extent that celebration of individual age achieved on birthdays was a modern cultural practice diffusing from north to south, from rich to poor, and from city to country in the nineteenth century, we have a natural explanation for occupational and geographic effects on age-heaping.

An interesting parallel comes from twentieth century China. Age reporting there was "astonishingly accurate" in censuses from 1953 to 1982, even among largely

[31] Goethe is a well-known exception, whose birthdays became something of a public event as his celebrity grew. A well-known case which is more typical is Pepys, who always noted his birthday in his diary but never mentions any celebration of it.

[32] The evidence is drawn from Google's "Ngram Viewer". Two sources of uncertainty or slippage mean this evidence must be taken with a grain of salt. The first is that common cultural practices need not be reflected in published books. It is easy to imagine publishing being dominated by novels in one country and by religious books in another. The second is uncertainty about what works are included in Google's language-specific collections of digitized books. The Italian and German collections are much smaller than the English.

illiterate, rural groups. Jowett and Li's (1992) explanation is based on the practice of *naming* years after the familiar animals of the Chinese zodiac, making birth year an indelible part of individual identity. In general this information alone was sufficient for a census enumerator to infer an individual's age, but the calendrical system based on names actually included further elements allowing for a more precise translation into numerical ages and dates (for which purpose officials carried a ready conversion table).[33] An exception proving the rule was the relatively literate, western province of Xianjiang, where age-heaping was quite pronounced. The population here was dominated by the Uygur ethnic minority, which alongside smaller Kirgiz and Tajik groups maintained a clear cultural and linguistic separateness from the Han Chinese. Among the Uygur, "birthdays are not particularly important days and they are not accorded celebrations and special observance" (Jowett and Li, 1992; 440). Jowett and Li (1992; 427) do not infer that these groups are innumerate, but rather that they "neither know, nor care, how old they are."

5. Conclusions

The promise of insights into the distribution of human capital in the past has prompted a profusion of age-heaping studies.[34] The evidence of nineteenth century Italy suggests it may be misleading to interpret age-heaping as a direct measure of numerical cognitive skills, however.

Italian census data exhibit the same broad correlation of illiteracy and age-heaping that has been found in many other studies. But they also throw up a number of anomalies that are not so easily reconciled with a cognitive ability interpretation: i) age-heaping varies quite widely even among *literates*, as well as among illiterates; ii) age-heaping varies by gender in a way that is poorly correlated with literacy and clearly tied up with marital status; iii) age-heaping varies with age in a way inconsistent with a straightforward human capital interpretation (a common finding treated as a mere nuisance in the literature); and iv) age-heaping varies across censuses and regions in a way that can only be explained in terms of the efficacy of data gathering procedures. An exploratory study of micro-data from southern Italy shows much the same: witnesses who did not sign death records were more likely to report 0/5 ages, but even controlling for this measure of literacy other factors such as occupation and place had a significant role – in some cases more important role – in predicting inaccurate age reports.

Parallels in analogous contexts, together with scraps of direct evidence gleaned from Italian social history, cast doubt on the assumption that most Italians, even

[33] The twelve-year cycle of zodiac animals ("earthly branches") interacted with a ten year-cycle of "heavenly stem" names, creating a 60-year set of unique birth-year identifiers. Within years, 24 "solar seasons" named after prevailing weather conditions were distinguished. "Thus, someone born at the time of the white dew in a geng shen [monkey] year, has a date of birth in early September 1920 or 1980" (Jowett and Li, 1982; 429, "[monkey]" added).

[34] By our count there have been at least fifteen publications on age-heaping since 2009, in addition to a number of working papers.

the large pool of illiterates, lacked basic familiarity with numbers and the most elementary arithmetical operations. More plausible than ignorance of numbers is simple ignorance of age among Italians, or a reluctance to make the effort to report it accurately. And there is evidence to support both hypotheses. Age-awareness was limited in much of Italy by pre-modern perceptions of time and attitudes regarding the celebration of birthdays. And evidence on the capacity of pre-unification Italian states to gather and process information, eliciting the cooperation of their citizens, corresponds well to regional variation in age-heaping.

Cultural change in the direction of secularism, individualism, and linear time-perception, and institutional change in the direction of administrative capacity building by the state are both aspects of what used to be called modernisation.[35] The spread of literacy too forms part of this bloc of interrelated social, economic, cultural and political changes. Age-heaping, on this Italian-inspired reading, correlates with illiteracy because both are reflections of slow and incomplete modernisation. Just as its advocates have contended, age-heaping is therefore a reasonable *proxy* for education, literacy, and perhaps numeracy. What it *directly* measures, we argue, is not numeracy but a broader mix of contextual factors that are just as interesting for diagnosing and predicting economic development.

[35] We use the term "modernisation" in a descriptive way, to characterise a multidimensional, mutually reinforcing process of change, without endorsing any specific variant of sociological modernisation theory. For a useful recent take on modernisation, firmly locating its origins in economic change, see Inglehart and Welzel (2005).

References

A'Hearn, Brian, Jörg Baten, and Dorothee Crayen, "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital," *Journal of Economic History*, vol. 69 no. 3 (2009), pp. 783-808.

Baten, Jörg, Dorothee Crayen, and Hans-Joachim Voth, "Numeracy and the Impact of High Food Prices in Industrializing Britain, 1780-1850" *Review of Economics and Statistics*, vol. 96 no. 3 (2014), pp. 418-30.

Baten, Jörg and Dácil Juif, "A story of large landowners and math skills: Inequality and human capital formation in long-run development, 1820–2000", *Journal of Comparative Economics*, vol. 42 (2014), pp. 375-401.

Baten, Jörg and Stefan Priwitzer, "Social and intertemporal differences in basic numeracy in Pannonia (first century BCE to third century CE)", *Scandinavian Economic History Review*, vol. 63 no. 2 (2015), pp. 110-34.

Cafagna, Luciano, "Modernizzazione attiva e modernizzazione passiva", *Meridiana*, vol. 2 (1988), pp. 229-40.

Ciccarelli, Carlo and Jacob Weisdorf, "The Effect of the Italian Unification on the Comparative Regional Development in Literacy, 1821-1911," unpublished manuscript, 14.07.2016.

Cipolla, Carlo Maria. *Literacy and Development in the West*. Harmondsworth: Penguin, 1969.

Crayen, Dorothee and Jörg Baten, "Global trends in numeracy 1820–1949 and its implications for long-term growth," *Explorations in Economic History*, vol. 47 (2010), pp. 82-99.

Dehaene, Stanislas. *The Number Sense. How the Mind Creates Mathematics*. Oxford: OUP, 1997.

de Moor, Tine and Jan Luiten van Zanden, "'Every woman counts': a gender-analysis of numeracy in the Low Countries in the early modern period", *Journal of Interdisciplinary History*, vol. 61 (2010), pp. 179-208.

De Sivo, Giacinto, *Storia delle Due Sicilie dal 1847 al 1861.* Verona, 1865.

Favero, Giovanni. *Le misure del regno: Direzione di statistica e municipi nell'Italia liberale*. Padova: Il Poligrafo, 2001.

Felice, Emanuele and Michelangelo Vasta, "Passive modernization? The new Human Development Index and its components in Italy's regions (1871-2007)", *European Review of Economic History,* Vol. 19 (2015), pp. 44-66.

Ferrante, Elena. *L'amica geniale*. Rome: Edizioni e/o, 2011.

Fischer, David Hackett. *Growing Old in America.* Oxford: Oxford University Press, 1977.

Földvári, Peter, Bas van Leeuwen, and Jieli van Leeuwen-Li, "How did women count? A note on gender-specific age heaping differences in the sixteenth to nineteenth centuries," *Economic History Review*, vol. 65 no. 1 (2012), pp. 304-13.

Galasso, Giuseppe, "E il compleanno uccise l'onomastico," *Corriere della Sera*, 19 March, 2012.

Hanushek, Eric and Ludger Woessman, "The Role of Cognitive Skills in Development", *Journal of Economic Literature,* vol. 46 (2008), pp. 607-68.

Hippe, Ralph and Jörg Baten, "Regional inequality in human capital formation in Europe, 1790–1880" *Scandinavian Economic History Review*, vol. 60 no. 3 (Nov. 2012), pp. 254-89.

Inglehart, Ronald and Christian Welzel. *Modernization, Cultural Change and Democracy. The Human Development Sequence.* Cambridge: Cambridge University Press, 2005.

Italy. Ministero della Pubblica Istruzione. *Relazione annuale del regio ispettore scolastico, circondario di Lagonegro*. Archivio centrale dello Stato, Roma. MPI, Direzione generale dell'istruzione primaria e popolare 1860-96, busta 299.

… . Ministero dell'Agricoltura e dell'Industria. Direzione generale della Statistica. *Censimento della popolazione del Regno d'Italia al 31 dicembre 1881. Relazione generale e confronti internazionali*. Roma: Tipografia Eredi Botta, 1885a.

… . Ministero dell'Agricoltura e dell'Industria. Direzione generale della Statistica. *Sulla classificazione per età della popolazione del regno secondo il censimento 31 dicembre 1881*. (*Estratto dagli Annali di statistica, serie 3a, vol. 14o*). Rome: Tipografia dei Fratelli Bencini, 1885b.

Jowett, John and Yuan-Qing Li, "Age-Heaping: Contrasting Patterns from China," *GeoJournal*, vol. 28 no. 4 (1992), pp. 427-42.

Juif, Dácil and Jörg Baten, "On the Human Capital of the "Inca" Indios Before and After the Spanish Conquest. Was there a "Precolonial Legacy"?', *Explorations in Economic History,* vol. 50 (2013), pp. 227-41.

Kaiser, David and Peyton Engel, "Time- and Age-Awareness in Early Modern Russia," *Comparative Studies in Society and History*, vol. 35 no. 4 (1993), pp. 824-39.

La Sorsa, Saverio, "Storia e folclore della mena delle pecore in Puglia," *Archivio Storico Pugliese*, vol. 6 (1953), pp. 487-99.

Luria, Alexander. *Cognitive Development. Its Cultural and Social Foundations*. Cambridge, Mass: Harvard University Press, 1976.

Macry, Paolo. *Giocare la vita. Storia del lotto a Napoli tra Sette e Ottocento.* Roma: Donzelli, 1997.

Manzel, Kerstin, Jörg Baten, and Yvonne Stolz, "Convergence and divergence in numeracy: the development of age-heaping in Latin America from the seventeenth to the twentieth century", *Economic History Review,* vol. 65 no. 3 (2012), pp. 932-60.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, "Poverty Impedes Cognitive Function," *Science*, vol. 341 (2013), pp. 976-80.

Menninger, Karl. *Number Words and Number Symbols: A Cultural History of Numbers*. Cambridge, Mass.: MIT Press, 1977 (first German edition 1934).

Mokyr, Joel. *Why Ireland Starved: A Quantitative and Analytical History of the Irish Economy, 1800-1850,* London: Allen & Unwin, 1983.

Patriarca, Silvana. *Numbers and Nationhood. Writing Statistics in Nineteenth Century Italy.* Cambridge: Cambridge University Press, 1996.

Sorokin, Pitirim and Robert Merton, "Social Time: A Methodological and Functional Analysis," *American Journal of Sociology*, vol. 42 no. 5 (1937), pp. 615-29.

Patriarca, Silvana. *Numbers and Nationhood. Writing Statistics in Nineteenth-Century Italy*. Cambridge: Cambridge University Press, 1996.

Price, Glanville, "Romance," Ch. 13 in *Indo-European Numerals*, Jadranka Gvozdanovic, ed. (Berlin: Mouton De Gruyter, 1991), pp. 447-96.

Randeraad, Nico. *Authority in Search of Liberty. The Prefects in Liberal Italy*. Amsterdam: Thesis Publishers, 1993.

Randeraad, Nico. *States and Statistics in the Nineteenth Century. Europe by Numbers*. Manchester: Manchester University Press, 2010.

Rocco, Emanuele. "Il gioco della mora" in Francesco de Bourcard (ed.), *Usi e Costumi di Napoli e contorni,* vol 1. Napoli, 1857.

Roggero, Marina, "Conti sulle dita, calcoli a penna. L'aritmetica elementare a fine Settecento," *Studi Storici*, vol. 35 no. 4 (Oct.-Dec. 1994), pp. 1039-60.

Romani, Mario. *Storia economica dell'Italia nel secolo XIX, 1815-1882*, Sergio Zaninelli, ed. Bologna: Il Mulino, 1982.

Saxe, Geoffrey, "The Mathematics of Child Street Vendors," *Child Development*, vol. 59 (1988), pp. 1415-25.

Schmitt, Jean-Claude. *L'invenzione del compleanno*. Bari: Laterza, 2012.

… , "L'invention de l'anniversaire," *Annales. Histoire, Science Sociales*, vol. 62 no. 4 (2007), pp. 793-835.

Serao, Matilde. *Il ventre di Napoli*. Milano: Fratelli Treves, 1884.

Spitilli, Gianfranco. "Di passata in passata. Il gioco della morra nell'Abruzzo settentrionale," in *Giochi, sport tradizionali e società. Viaggio tra la Valle d'Aosta, l'Italia e l'Unione Europea*, Anna Maria Pioletti, ed. Aosta: Musumeci Editori, 2012.

Thomas, Keith, "Numeracy in Early Modern England: The Prothero Lecture," *Transactions of the Royal Historical Society*, series 5, vol. 37 (1987), pp. 103-32.

Thompson, Edward P., "Time, Work-Discipline, and Industrial Capitalism," *Past and Present*, vol. 38 no. 1 (1967), pp. 56-97.

Tollnek, Franziska and Jörg Baten, "Farmer Families at the Heart of the Educational Revolution: Which Occupational Group Inherited Human Capital in the Early Modern Era?" *EHES Working Paper* no. 33, 2012.

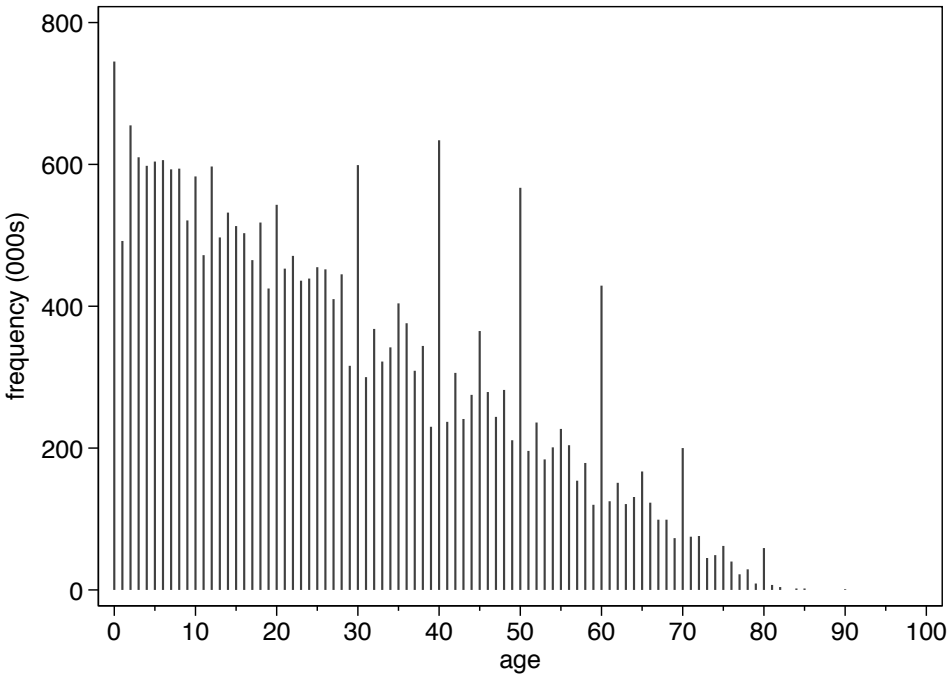
… , "Age-Heaping based human capital estimates," in Diebolt, Claude and Michael Haupert (eds.), *Handbook of Cliometrics* (Heidelberg: Springer-Verlag, 2016).

Vecchi, Giovanni (ed.). *In Ricchezza e in Povertà. Il benessere degli Italiani dall'Unità ad oggi.* Bologna: Il Mulino, 2011.

Whipple, George Chandler. *Vital Statistics. An Introduction to the Science of Demography.* London: John Wiley, 1919.

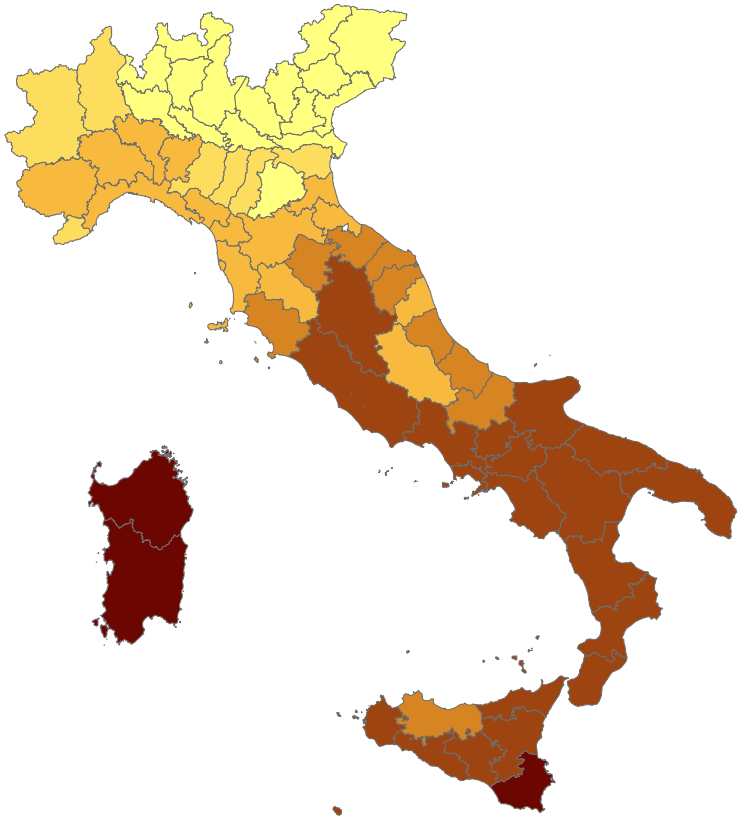
Zamagni, Vera. *Dalla periferia al centro. La seconda rinascita economica dell'Italia (1861-1990)*, 2e. Bologna: Il Mulino, 1993.

Fig. 2.1. Declared ages, census of 1871



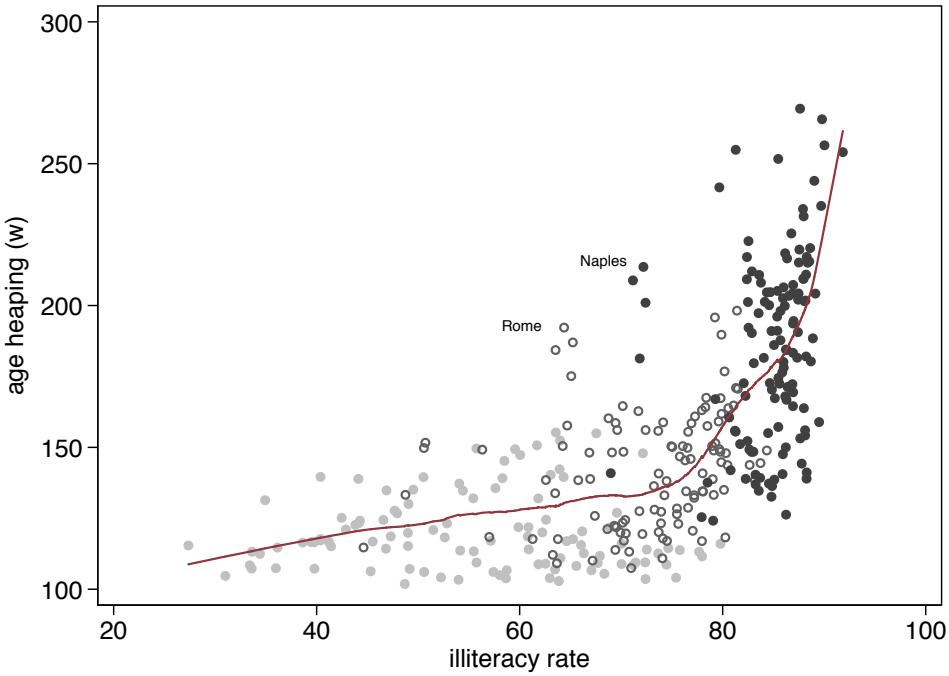
Notes: frequencies of reported ages in the Italian census of population of 31.12.1871.

Fig. 2.2. Age-heaping by province, census of 1871



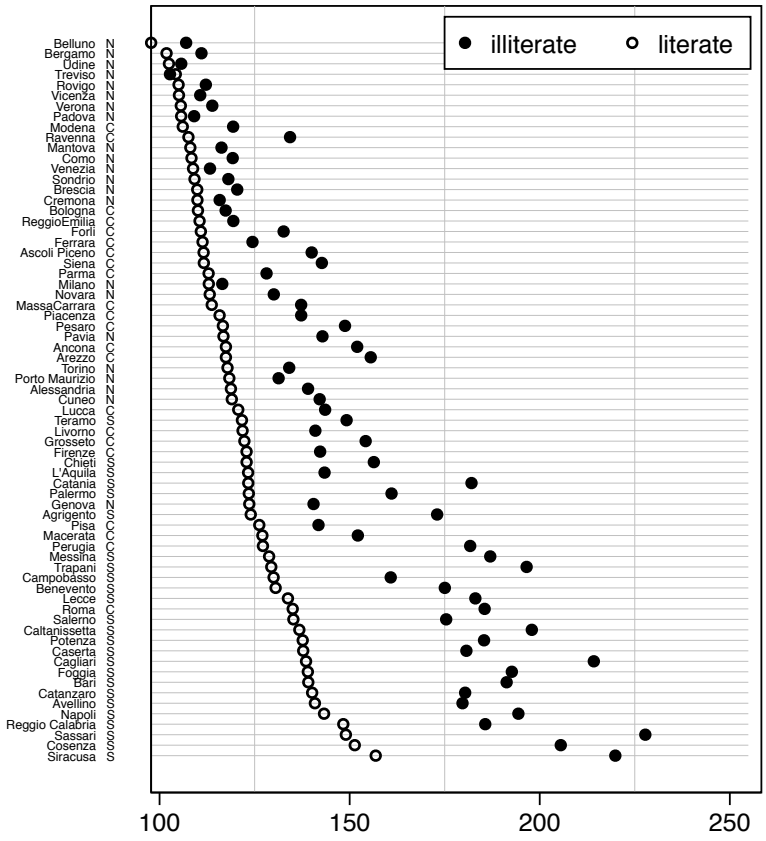
Note: Whipple index class by province, darker colours indicating higher *W* values. Classes are 105-19, 120-31, 132-45, 146-65, 166-202, and 203-30.

Fig. 2.3. Age-heaping and illiteracy, by province and age-group



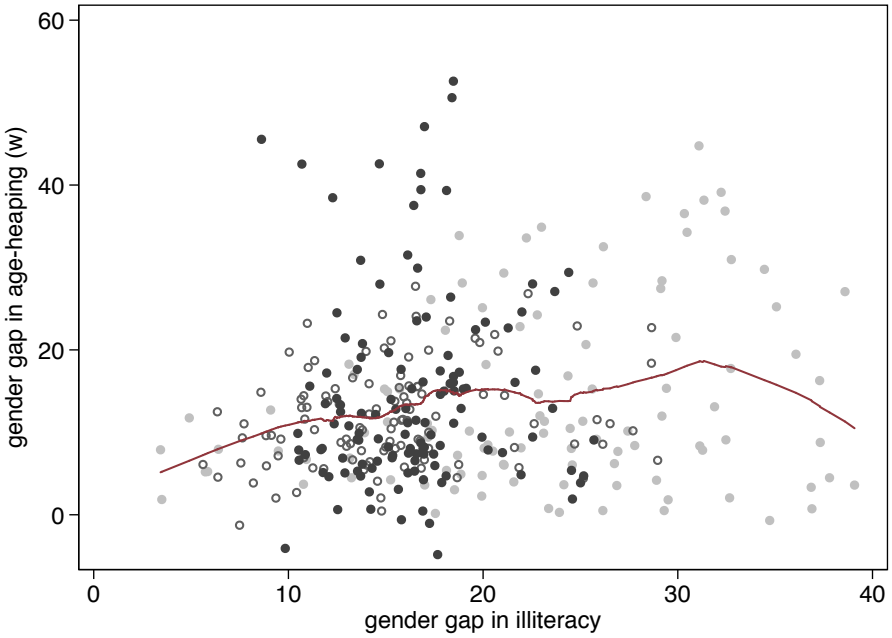
Notes: Points represent Whipple index and illiteracy pairs for a specific province and age-group. Light grey markers indicate North provinces, hollow markers Centre, dark markers South. Ages 23-72. Line based on locally weighted regression with a bandwidth of 30% of sample.

Fig. 2.4. Age-heaping among literates and illiterates



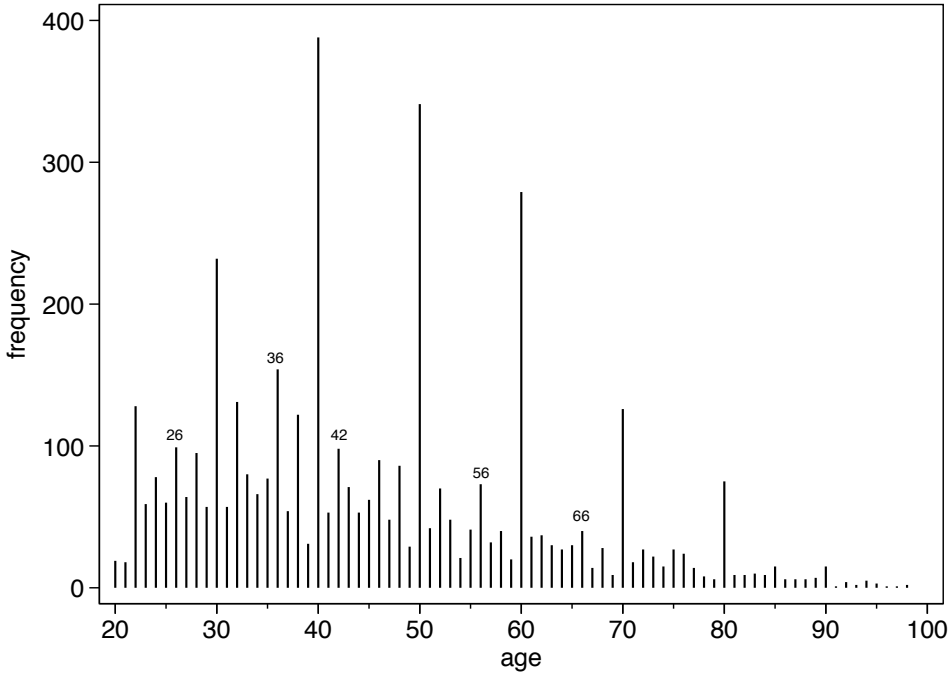
Note: Whipple index values, ages 23-72, by self-reported literacy, census of 1871.

Fig. 2.5. Gender gaps in illiteracy and age-heaping



Notes: Gender gaps defined as female minus male values of *W* and illiteracy. Points represent a specific province and age-group. Light grey markers indicate Northern provinces, hollow markers Centre, and dark markers South. Ages 23-72. Line based on locally weighted regression with a bandwidth of 30% of sample.

Fig. 3.1. Age frequencies in Southern death records



Note: age frequencies of witnesses and deceased aged 20 or more; sample of death records in three Southern provinces.

Table 2.1. Age-heaping by age group and gender, census of 1871

	women		*men*		*total*	
age group	*W*	*n*	*W*	*n*	*W*	*n*
13-22	110	2480	105	2442	107	4922
23-32	131	2130	118	2076	125	4206
33-42	154	1747	142	1758	148	3505
43-52	167	1436	155	1467	161	2903
53-62	177	973	157	995	167	1967
63-72	169	565	148	596	158	1161
73-82	178	189	154	204	166	393
average 2-step	159	6850	144	6892	152	13742
average 1-shot	154		140		147	

Notes: *W* is the age-group and gender specific value of the Whipple index, calculated using national population totals; 100 = no heaping on 0/5 ages. *n* = number of observations, in thousands. "1-shot" average pools all ages and calculates one *W* value; "2-step" figure calculates *W* separately by age-group, then takes the simple average of these figures. Age groups 13-22 and 73-82 excluded from both calculations and from the total *n*.

Table 2.2. Age-heaping by age group and literacy, census of 1871

	women		*men*		*total*	
age group	*literate*	*illiterate*	*literate*	*illiterate*	*literate*	*illiterate*
13-22	105	113	101	107	103	110
23-32	116	137	107	127	110	132
33-42	129	162	119	156	122	159
43-52	138	174	126	173	130	173
53-62	138	185	125	177	129	182
63-72	131	176	120	165	123	171
73-82	133	185	125	169	127	178
average	130	167	119	160	123	164

Note: *A*ge-group and literacy specific Whipple index values calculated using national population totals. The overall *W* reported in the final row is computed with the two-step procedure, excluding age groups 13-22 and 73-82.

Table 2.3. Age-heaping by age group, marital status, and gender

	women			*men*		
age group	*never-married*	*married*	*widowed*	*never-married*	*married*	*widowed*
13-22	110	116	116	105	77	96
23-32	121	135	161	108	131	129
33-42	157	150	190	133	143	145
43-52	168	157	202	146	156	158
53-62	168	157	213	144	158	166
63-72	161	146	189	139	148	152
73-82	160	150	191	137	158	154
average	155	149	191	134	147	150

Note: Marital status, age-group and gender specific Whipple index values calculated using national totals.

Table 2.4. Whipple index changes by cohort, 1861-71

	20s - 30s	*30s - 40s*	*40s - 50s*	*50s - 60s*	*60s - 70s*	*average*
min	4.3	-7.3	-35.1	-76.6	-52.7	-23.8
mean	20.1	12.6	6.6	-9.9	-3.1	5.3
max	54.3	49.6	45.6	21.9	58.7	40.1
positive	59/59	53/59	39/59	11/59	22/59	39/59

Note: Summary statistics on the difference between the Whipple index in the census of 1871 and 1861 for the same cohort in a given province. "20s-30s" column is the cohort aged 23-32 in 1861 and 33-42 in 1871; similarly for other columns. "Positive" = share of provinces in which observed inter-census change in *W* is greater than zero. "Average" column gives the simple average (across provinces) of the averages (across age groups).

Table 2.5. R_5 age-heaping values by census and cohort

	1861	*1871*	*1881*	*1891*	*1901*	*1911*
North						
20s	109	100	104			107
30s	100	104	98		101	102
40s	112	103	107		98	98
50s	113	115	104		96	97
60s	134	118	119		109	99
70s	136	145	119		123	125
average*	*119*	*117*	*109*		*105*	*105*
decennial chg.		*-2*	*-8*		*-2*	*-1*
Centre						
20s	95	106	110			103
30s	97	96	103		97	101
40s	123	106	104		104	93
50s	119	132	109		103	105
60s	136	125	135		97	104
70s	146	142	122		118	111
average*	*124*	*120*	*115*		*104*	*103*
decennial chg.		*-4*	*-5*		*-5.5*	*-1*
South-Islands						
20s	105	106	103			107
30s	106	113	118		100	102
40s	135	121	135		103	98
50s	156	155	143		107	108
60s	178	167	185		115	119
70s	170	175	182		121	132
average*	*149*	*146*	*153*		*109*	*112*
decennial chg.		*-3*	*+7*		*-22*	*+3*

Note: R_5 index calculated as described in text. Reported figures are simple averages of provincial values. Averages in italicised rows exclude 20-29 age-group for comparability with the census of 1901. Veneto and Lazio regions excluded in all years for comparability with 1861.

Tab. 3.1. Age heaping and signature rates, ages 23-72, by district

	Napoli			Bari			Reggio	All
	centre	periphery	Altamura	Bari	Barletta	Palmi	Reggio	
23-32	119	159	213	148	214	183	197	157
33-42	178	230	225	196	170	241	250	207
43-52	167	214	271	231	208	272	271	226
53-62	199	322	352	220	310	271	272	255
63-72	165	277	229	240	214	235	257	223
73-82	181	346	71	245	250	182	250	244
average	166	240	258	207	223	240	249	214
census	192			192		202		
altern. index	35	51	86	50	63	73	71	52
n	*940*	*622*	*310*	*1,012*	*301*	*397*	*341*	*3,923*
signature	*83*	*61*	*14*	*15*	*10*	*3*	*3*	*34*

Notes: Whipple index values calculated as described in the text; witnesses and deceased combined. "Census" is the *W* value from the 1861 census for the relevant province. "Alternative index" is the sum of the absolute deviations from 10% for each terminal digit's share of all ages. "*n*" refers to ages 23-72. "Signature" is the share of witnesses who signed the death record.

Tab. 3.2. Marginal effects on the probability of reporting a 0/5 age

	all		witnesses		signature		town fe	
artisan	-0.070	***	-0.076	***	-0.042	*	-0.053	**
enlisted	-0.077		-0.095		-0.051		-0.044	
other/miss.	-0.087	*	-0.055		-0.021		-0.035	
elite	-0.113	***	-0.121	***	-0.053		-0.061	**
service dep.	-0.049		-0.042		-0.004		-0.014	
service indep.	-0.065	***	-0.061	**	-0.021		-0.030	
fem. craft	0.074							
fem. elite	-0.176	***						
fem. missing	0.015							
fem. serv.	0.277							
fem. unsk.	0.021							
age 32-41	0.076	***	0.049	**	0.044	*	0.031	
age 42-51	0.111	***	0.090	***	0.089	***	0.082	***
age 52-61	0.156	***	0.115	***	0.109	***	0.118	***
age 62-71	0.099	***	0.048		0.046		0.077	*
age 72+	0.066	*	-0.193	***	-0.207	***	-0.165	**
Naples city	-0.047	**	-0.037		0.020			
Other rural	0.073	***	0.082	***	0.094	***		
deceased	0.010							
signature					-0.117	***	-0.062	**
n	4,353		3,140		3,140		3,140	

Note: estimated marginal effects, evaluated at zero for all variables, from logit regressions predicting an age ending in 0 or 5. Stars indicate statistical significance at 1, 5, and 10%. The χ^2_{88} test statistic for the null that the town fixed effects in Col. 4 are jointly zero is 225.6, p-value 0.0000.

Tab. 3.3. Marginal effects on probability of reporting a 0/5 age

	Naples	Bari-Barletta	Other agricult.
artisan	-0.146 *	-0.043	-0.009
enlisted	-0.149		0.079
other/miss.	-0.090	-0.044	-0.001
elite	-0.161 *	-0.014	-0.034
service dep.	-0.084	0.136	-0.100
service indep.	-0.078	0.043	-0.057
fem. craft			
fem. elite			
fem. missing			
fem. serv.			
fem. unsk.			
age 32-41	0.167 ***	-0.025	0.028
age 42-51	0.143 **	0.034	0.095 **
age 52-61	0.172 **	0.016	0.150 ***
age 62-71	0.032	-0.058	0.154 **
age 72+	0.126		-0.233 **
signature	-0.045	-0.167 ***	-0.158 ***
n	722	1,105	1,303

Note: the table presents estimated marginal effects, evaluated at zero for all variables, from logit regressions predicting an age ending in 0 or 5. Stars indicate statistical significance at 1, 5, and 10%.

Table 4.1. Indicators of administrative capacity, 1850s

	population	*revenue*	*publish-ing*	*cadastre*	*age-heaping*
Piedmont-Sardinia	4,600	100	100	mixed	133
Lombardy-Venetia	5,600	83	102	survey	115
Parma	500	59		survey	134
Modena	600	63		mixed	126
Tuscany	1,800	63	64	survey	146
Papal State	3,100	70	49	survey	153
Two Sicilies	9,300	55	41	descriptive	181

Notes: Publishing – number of statistical works published 1850-59 per million population, relative to Piedmont's 8.0. Revenue – total ordinary revenue of central government p.c., relative to Piedmont's 30.46 lire; estimate for Kingdom of the Two Sicilies for continental provinces only. Cadastre – basis of the land register as described in text. Age-heaping – Whipple index; *W* of Piedmont-Sardinia for Piedmont and Liguria provinces only; Sardinian *W*=224. *Sources*: population – Correnti and Maestri (1858, p. 381); revenue – Romani (1982, p. 404); publishing – Patriarca (1996, p.241); cadastre – Zamagni (1993, p. 73); age-heaping – our calculations.

Appendices

1. *How age structure can affect the Whipple index*

Consider an age distribution in which frequency strictly declines with age, there is no heaping, and age-groups are defined on intervals such 30-39.

If f_i denotes the frequency of the i^{th} terminal digit, the Whipple index for a ten-year age-group is defined as:

$$W = \frac{f_0 + f_5}{\sum_{i=0}^{9} f_i} \Big/ \left(\frac{1}{5}\right) ,$$

omitting the multiplication by 100 used in the text. It is useful to break down the sum $\sum_{i=0}^{9} f_i$ into two components: $\Sigma_1 = \sum_{i=0}^{4} f_i$, and $\Sigma_2 = \sum_{i=5}^{9} f_i$. We can now re-write the equation for W as follows:

$$W = \frac{5 \cdot f_0 + 5 \cdot f_5}{\Sigma_1 + \Sigma_2}.$$

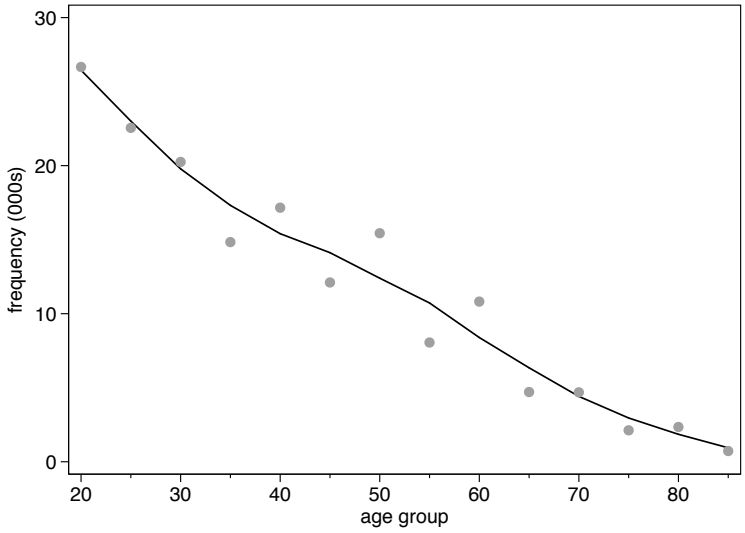
By assumption frequency declines steadily with age, so $f_0 > f_1 > f_2 \ldots > f_9$ in our ten-year age-group. This implies $5 \cdot f_0 > \Sigma_1$ and $5 \cdot f_5 > \Sigma_2$, so that $5 \cdot f_0 + 5 \cdot f_5 > \Sigma_1 + \Sigma_2$, which in turn means $W > 1$, despite there being no heaping by assumption.

For an age-group defined on an interval such as 23-32, instead, $f_3 > f_4 \ldots > f_9 > f_0 > f_1 > f_2$ and the above inequalities are no longer satisfied (at least not necessarily).

2. *Computing the R_5 index*.

As described in the text, the R_5 index is based on the ratio of the frequency of a "0-4" age group to that of the successive "5-9" group: $f_{0\text{-}4}/f_{5\text{-}9}$. Because this ratio will always be greater than one if frequencies decline with age, we express the R_5 index not in raw form but as a ratio of ratios: the actual $f_{0\text{-}4}/f_{5\text{-}9}$ ratio relative to what is predicted from a locally weighted regression of frequencies on age. The figure and table below illustrate the procedure for the province of Agrigento in 1871.

Actual and smoothed age-group frequencies, Agrigento



Note: grey markers indicate observed frequencies for each five-year age-group. The line is the smoothed distribution resulting from locally weighted regression, with bandwidth at 50% of sample

R_5 ratios for Agrigento, 1871

		R5 ratios	
age group	*observed*	*predicted*	*ratio*
20-29	1.18	1.15	103
30-39	1.36	1.14	119
40-49	1.42	1.09	130
50-59	1.92	1.16	166
60-69	2.30	1.32	174
70-79	2.21	1.49	148
average	1.73	1.23	140

Note: R_5 values calculated as described in the text. The average reported in the final row is the simple mean across all age groups.

The interpretation of a value like 140, Agrigento's overall average in the table is that 0-4 ages predominate over 5-9 ages by 40% more than we would expect, given the age structure of the province's population. Note how R_5 heaping rises with age, up to a point, just as Whipple index heaping does.

3. Construction of the death record dataset

Data were drawn from twelve neighbourhoods of greater Naples (Avvocata, Chiaia, Fuorigrotta, Miano e Marianella, Montecalvario, Posillipo, S. Carlo all'Arena, S. Ferdinando, S. Giuseppe, S. Lorenzo, Stella, and Vomero) and 8l towns ("*comuni*") in the rural provinces of Bari and Reggio. In the provinces, an informal sampling strategy was adopted based on the alphabetical ordering of towns and aimed at broad coverage. Towns with poor quality digital images were not sampled, but those with poor handwriting were included (at

considerable effort!). The most difficult handwriting was typically not sloppy, but rather over-stylised in the service of elegance.

In each town or neighbourhood, data collection proceeded chronologically starting with 1 January 1861 and continued until i) a reasonable number (a few tens) of useful observations was obtained, ii) the end of calendar year 1861 was reached, or iii) strongly diminishing returns set in due to repetition of the same witnesses. Some towns were "oversampled" when they yielded a wide range of occupations, suggesting a degree of socioeconomic differentiation. Our sampling procedure is neither truly random nor stratified. Even if it were, we would face the difficulty that witnesses are not a random selection of adult males. For this exploratory study, the aim was to cover, at low cost, the widest possible range of circumstances and individuals, rather than assemble a representative sample.

Among witnesses, it was common for the same individuals to come forward (or be chosen) repeatedly to declare deaths. Identifying perfect matches on surname, first name or initial, age, occupation, and locality is straightforward, but throws up some false matches because the range of occupations, names, and surnames was often very narrow in small towns. (For Reggio Calabria province, complete first names were entered in the database. After deleting duplicates, the top four given names – Giuseppe, Domenico, Francesco, and Antonino, in that order – accounted for 40% of the total for males of all ages including infants. Surnames are more varied overall, but not necessarily within individual towns.) We assumed all matches were valid and retained only one of the duplicate records for analysis. The same rule was applied to partial matches in which all variables except age differed. It is likely that many of these cases represent the same individual reporting different ages on different occasions (or having different ages invented for him by the registrar). This would make these cases particularly interesting to study, if we could be sure of their inaccuracy. As we cannot, we retain only the first instance of such partially-matched observations for further study. The table below gives the number of observations available.

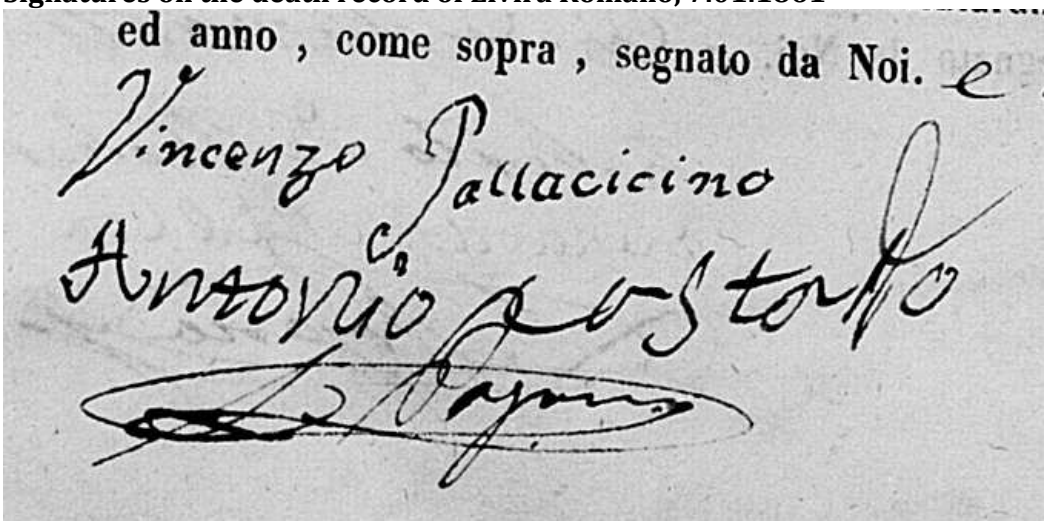
Observations in the micro database

province	records	persons	unique	age 22+
Napoli	1,203	3,609	2,392	1,806
Bari	857	2,571	2,236	1,754
Reggio	336	1,008	919	793
Total	2,396	7,188	5,547	4,353

We rely on witness signatures as a measure of literacy. The figure below illustrates the signatures on the death record of six-year old Elvira Romano, who died in the Avvocata neighbourhood of Naples on 7 January, 1861. At the bottom is the stylised, elegant signature of Luigi Pagano, the local councillor in charge of vital records. At the top is the name, neatly printed, of Vincenzo Pallacicino, a 41 year old white collar municipal employee. Interestingly, this unusual surname is spelled differently by Pagano (or his clerk) elsewhere in the document, where it is rendered as Pallavicino, a more common family name; one wonders which is correct. Finally, in the middle is the childlike, only partly legible effort of the 60 year old Antonio Castaldo, whose occupation, incredibly, is "*scrivente*", or scribe.

One is reminded of Pirandello's story *L'altro figlio*, in which an illiterate Sicilian peasant woman entrusts a neighbour to write letters to her emigrant sons. The letters go unanswered for years. The illiterate neighbour, it turns out, was merely miming the action of writing, making a few illegible strokes of the pen on the page.

Signatures on the death record of Elvira Romano, 7.01.1861



As noted in the text, some 500 occupations were observed. The most common were various types of unskilled labourers. Next most common were craftsmen or artisans, a designation interpreted broadly to include building workers, millers, and butchers. The range of this group is wide, encompassing compositors, jewellers, and watchmakers at one end and house painters at the other. Two categories of service workers were distinguished. The first includes jobs with some scope for self-direction, decision making, and negotiation. Included were a wide range of occupations including innkeepers, coffee sellers, barbers, musicians, carters, and even bootblacks. The second service group comprises dependent positions with little autonomy – even if skill or status was involved – such as coachman, cook, messenger, waiter, or domestic servants. An "elite" group included property owners, merchants, clerics, military officers and professionals. Small residual categories for enlisted soldiers and other or missing occupations round out the categories. Separate female occupational categories were also defined, but for the most part they have little real informational content. Among adults 22 and older, most women (53%, and a large majority among the married) have no occupation at all. 27% are in the craft category, but these are almost all *filatrici*, which as discussed in the text seems to have been a default occupation for any unmarried woman. 13% of women have unskilled occupations, but again these are mostly the generic *contadina* or peasant.

UNIVERSITY OF OXFORD DISCUSSION PAPERS IN ECONOMIC AND SOCIAL HISTORY

are edited by

Rui Esteves (Brasenose College, Oxford, OX1 4AJ)
Gabriel Geisler Mesevage (Brasenose College, Oxford, OX1 4AJ)