DISCUSSION PAPER SERIES

DP14261

COGNITION, CULTURE, AND STATE CAPACITY: AGE-HEAPING IN XIX CENTURY ITALY

Brian A'Hearn, Alexia Delfino and Alessandro Nuvolari

ECONOMIC HISTORY

ISSN 0265-8003

COGNITION, CULTURE, AND STATE CAPACITY: AGE-HEAPING IN XIX CENTURY ITALY

Brian A'Hearn, Alexia Delfino and Alessandro Nuvolari

Discussion Paper DP14261
Published 31 December 2019
Submitted 30 December 2019

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Economic History

COGNITION, CULTURE, AND STATE CAPACITY: AGE-HEAPING IN XIX CENTURY ITALY

Abstract

We re-examine the causes and interpretation of age-heaping in a case study of nineteenth century Italy. Italian census data allow us to calculate age-heaping measures by province, education, gender, and marital status. Our results validate the use of age-heaping as a proxy for human capital, but also reveal anomalies difficult to reconcile with a pure numeracy interpretation. Alongside individual cognitive ability, the census data clearly suggest a role for contextual factors in shaping age-heaping patterns. Direct evidence from Italian social and political history buttresses the case for culture and state capacity as determinants of age-heaping. Age-heaping and illiteracy are well correlated because both are reflections of an underlying process of modernisation, a process which, in nineteenth century Italy, was slow and incomplete.

JEL Classification: N33, J24

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

Brian A'Hearn - brian.ahearn@pmb.ox.ac.uk
Pembroke College, Oxford

Alexia Delfino - a.delfino2@lse.ac.uk
London School of Economics

Alessandro Nuvolari - alessandro.nuvolari@santannapisa.it
Sant'Anna School of Advanced Studies and CEPR

Cognition, Culture, and State Capacity: Age-Heaping in XIX 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

We re-examine the causes and interpretation of age-heaping in a case study of nineteenth century Italy. Italian census data allow us to calculate age-heaping measures by province, education, gender, and marital status. Our results validate the use of age-heaping as a proxy for human capital, but also reveal anomalies difficult to reconcile with a pure numeracy interpretation. Alongside individual cognitive ability, the census data clearly suggest a role for contextual factors in shaping age-heaping patterns. Direct evidence from Italian social and political history buttresses the case for culture and state capacity as determinants of age-heaping. Age-heaping and illiteracy are well correlated because both are reflections of an underlying process of modernisation, a process which, in nineteenth century Italy, was slow and incomplete.

JEL codes: N33, J24

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

1. Introduction

Historical research on human capital has had to overcome serious data constraints. Microdata are scarce, and only from the late nineteenth century is information widely available even for basic indicators such as literacy or school enrolments. For earlier periods, researchers have turned to proxy indicators that can be constructed from accessible sources. For health status, they have made use of information on heights, which reflect net nutrition in the growing years (Steckel, 1995). For education, they have calculated signature rates on marriage registers and other documents as an indicator of literacy (Cipolla 1969). For craft skills they have studied skill premia in the construction industry (van Zanden 2009). To this list economic historians have added a new proxy in recent years: age-heaping.

Irregularities in an age distribution, in particular the overrepresentation of round numbers such as multiples of ten, indicates that some individuals do not know, have reasons not to reveal, or are unable to compute precisely their true age. Instead they report a round or otherwise attractive number that is not implausible and has some salience for them. In a histogram, this results in observations being excessively "heaped" on particular ages. Such heaping was early recognised as a problem by demographers; textbooks of the early twentieth century recommended a diagnostic tool now called the Whipple index (Whipple, 1919, 196-70). Age heaping began to attract the interest of historians from the late 1970s. Studying US data, Fischer (1977, 82-86) detected a change in popular attitudes towards youth in the increasing tendency to report younger ages at critical junctures in life, e.g. 39 rather than 40. In a later study, Kaiser and Engel (1993) adopted a similar, cultural interpretation, seeing age-heaping as a measure of "time- and age-awareness" in the pre-industrial society of early eighteenth-century Russia. Thomas (1986, 128) called attention to the social meaning of age in early modern England: "… ages were reported with precision for people under twenty, because differences in age could be of considerable social and administrative importance for young persons. But when people reached adulthood their exact numerical age had much less social meaning; and it was much more vaguely reported."[1] Closer to home, from the perspective of this study, Herlihy and Klapisch-Zuber (1985; 159-82) documented a significant decrease in age-heaping over time in medieval Tuscan tax records. Rather than cultural change, they identified the growing pressure exerted by fiscal authorities collecting the data as the cause: what we would now call state capacity.

Perhaps the first *economic* historian to consider age-heaping was Mokyr (1983), who sought to assess whether Irish emigrants were positively or negatively selected. In his view, age-heaping was informative not only about social context but also about individual human capital:

[1] Thomas also commented on the popularity of ages ending in 6, which reflected "the duodecimal thinking that had been so common in Anglo-Saxon England," and saw increasingly accurate age reporting as evidence of increasing numeracy (1986, 126-8).

> "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).

Individual cognitive skills ("arithmetical abilities") and a broader cultural outlook ("respect for accuracy" and "a more serious attitude towards time") are complements in this interpretation. Subsequent work in economic history has followed this line. Comparing literacy and age-heaping in an exploration of sources and methods, A'Hearn, Baten, and Crayen concluded that "[b]oth measures offer a partial view of human capital, and both reflect not only individual but also broader social capabilities – a sort of administrative capital" (2009, 805).

In the last decade, studies of age-heaping have proliferated, covering times and places that range 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). The majority of these papers have an exclusive focus on individual human capital: their treatment of contextual factors such as the bureaucratization of everyday life (associated with the emergence and consolidation of modern states) or variation in cultural norms is limited to eliminating possible confounding influences on a valid estimate of numeracy. Thus, in regression analyses the number of censuses previously conducted might be included as a control variable for state capacity, or country fixed effects included to capture cultural differences.[2] This body of work establishes a robust correlation between age-heaping and variables such as literacy or school enrolments, justifying age-heaping as an (inverse) index of human capital (Tollnek and Baten, 2016).

In this paper we investigate age-heaping in nineteenth century Italy. The early Italian censuses reported unusually rich age data, with single-year age distributions broken down by province, gender, marital status, and literacy. In addition, the political connotations of the new country's censuses in terms of nation building generated a vigorous debate on methods, results, and interpretations, meaning census operations are well documented. Finally, Italy was (and remains) a country characterised not only by great disparities in economic performance, but also in institutions, education, and cultural attitudes.[3] This heterogeneity makes the Italian case particularly interesting for probing the nature and meaning of age-heaping. Going back to the origins of the

[2] Baten, Crayen, and Voth (2014) is an important study in which the possible effects of cultural norms and administrative procedures are explicitly considered. The authors' research design allows them to argue that age-heaping variations in early 19th century Britain reflected individual cognitive ability (which varied with early childhood nutrition).

[3] These geographical differences are such that influential interpretations depict qualitatively fundamentally different historical processes of modernisation in the North and the South (Cafagna 1988; Felice and Vasta 2015).

age-heaping literature, we consider three sets of factors influencing age-heaping: individual cognitive ability, state capacity, and culture. Our reading of the Italian evidence indicates that all three mattered in shaping age-heaping patterns.

2. Age-heaping in the Italian censuses: what the data show

Italy conducted five censuses between Unification and the First World War: in 1861, 1871, 1881, 1901, and 1911.[4] For the first two, the Ministry of Agriculture, Industry, and Commerce tabulated and published single-year age distributions (Italy 1865, 1875). The age distribution in 1871 is depicted in Figure 1. Among adults, the heaping on multiples of ten is unmistakable, as is a set of minor peaks associated with ages ending in 5. More subtle, but quite systematic, is a preference for even numbers.

Figure 1 about here

Italians' preference for terminal digits zero and five ("0 and 5 ages") means that we can characterize the extent of age-heaping using the Whipple index. This index is defined as the ratio of the actual share of 0 and 5 ages to a prediction of 20 per cent (such as would be expected in a uniform distribution of terminal digits). Multiplying Whipple index values ("*W*") by 100 for ease of presentation, we have a measure that is 100 when 0 and 5 ages just equal the expected proportion of one-fifth, and 500 when all reported ages end in 0 or 5.[5]
Five Whipple points are equivalent to 1 percentage point in the share of 0 and 5 ages. For example, if the 0 and 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 the degree of age-heaping to rise with age itself.[6] An older population will therefore display more age-heaping than a younger population, *ceteris paribus*. To facilitate comparisons of populations with different age structures, we employ a two step procedure, first calculating *W* separately for ten-year age-groups, then computing their simple average as our overall measure of heaping. Think of this two-step *W* estimate as measuring the age-heaping of the average cohort (rather than the average individual). For comparability with other studies we occasionally report the alternative *W* calculation in which terminal digit frequencies for all age groups are pooled.[7]

Table 1 about here

---

[4] The missing census of 1891 was cancelled due to a fiscal crisis (Gallo and Paluzzi, 2012, 38).

[5] Formally the Whipple index is given by $W = \frac{(n_{25}+n_{30}+n_{35}+n_{40}+\cdots n_{70})}{\frac{1}{5}\sum_{i=23}^{72} n_i} \times 100$
where n_i indicates the frequency of individuals of age *i*.

[6] See Baten and Crayen (2010) for a discussion of how age affects heaping propensity.

[7] 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 seemingly more natural 20-29.

The *W* values reported in Table 1 show clearly that 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. For comparability with other studies we restrict attention to the five central cohorts spanning ages 23-72 in calculating the average *W*'s reported in the final two rows of Table 1.[8] The overall figure of 152 places nineteenth century Italy clearly behind Northern and Central Europe in accuracy of age reporting, on a par with other nations of Europe's Southern and Western periphery, but ahead of much of the rest of the world including Eastern Europe, Latin America, and parts of Asia.[9]

A map immediately reveals the extreme geographic variation in age-heaping characteristic of the Italian case (Fig. 2). With *W*=105, Belluno province in the Northeast displays only one percentage point of excess 0 and 5 frequency, a value as low as anywhere in Europe at the time. Meanwhile in Sardinia, the province of Sassari 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.[10] We say "at least" because there is no way of knowing whether declared ages with terminal digits other than 0 and 5 are accurate; they simply are not obviously suspicious. There is an obvious North-South gradient in Figure 2, but more subtle patterns can also be discerned, both between and within the former states of pre-unification Italy. The former Kingdom of Lombardy-Venetia in the Northeast stands out as an area of uniformly low age-heaping, clearly contrasting with its neighbours. Within the former Papal State there is almost as much variation as in Italy overall: from *W*=119 in Bologna to *W*=178 in Perugia.

Figure 2 about here

After 1871, single-year age distributions were no longer published in the census volumes: the tabulations of 1881 and later censuses refer to five-year age-groups such as 30-34 and 35-39. Though such data do not allow us to compute Whipple values, we can still construct an approximate measure of heaping on multiples of ten. This alternative *T* index (*T* for "tens") is based on a ratio, like the Whipple. We compare the frequency in a five-year age-group containing a multiple of ten

[8] The slightly narrower age range 23-62 is another common choice in the literature. Ages 13-22 are typically excluded because some samples, particularly from military sources, have a great concentration of individuals at a particular age such as 18, 20, or 21. Ages 73 and above are excluded due to concern about the reliability of the data. In many studies 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 but also exaggerated. For a recent discussion see Newman (2019).

[9] Comparative international age-heaping estimates can be found in Crayen and Baten (2010), Hippe and Baten (2012), and Baten and Juif (2014). The Table 1 estimate for the 43-52 age group neatly matches Baten and Juif's figure for the 1820s birth cohort in Italy (converting their ABCC index back to the underlying *W* 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.

[10] Sassari's *W* of 230 is the ratio of 0.46 actual to 0.20 expected 0 and 5 age shares; 26 percentage points is thus the excess frequency of 0's and 5's.

with that in the successive age-group that does not, e.g. $(f_{30-34})/(f_{35-39})$, where *f* stands for frequency and the subscripts denote the age range. This ratio will exceed one if there is significant heaping on multiples of ten (relative to heaping on terminal digit 5). Of course, it will also exceed one if the population is growing and the young systematically outnumber the old, even if there is no heaping. Accordingly, we do not work with raw values of the tens ratios but express them relative to a prediction based on the overall age structure of the province.[11] When we pair *T* and *W* values for the same age-group by province, the two indices are highly correlated (ρ=+0.73, n=345).[12]

Table 2 about here

T-index values by age-group and census are presented in Table 2. Reading across the bottom row, we see that average age-heaping decreases only slightly from 1861 to 1881, before dropping quite sharply in 1901 and stabilising at that level. We can track an individual cohort by reading down and to the right along a diagonal, remembering to drop down *two* rows in 1901 because 20 years have passed since the previous column. Consider the cohort born in the 1830s and aged 20-29 in 1861. The *T* values for this group increase from 103 to 105 and then 118 as they age into their 30s and 40s, much as we might expect based on the nearly ubiquitous ageing effect mentioned earlier. Reaching their 60s in the census of 1901, we would expect a further, indeed considerably greater increase. Instead, *T* not only fails to increase, but actually *falls* by 11 points (to 107). Ten years later, in the census of 1911, *T* has resumed its pattern of increasing for a cohort with each passing decade, reaching a value of 123. We can see this pattern repeated for other cohorts: age-heaping increases as the cohort ages across the censuses of 1861, '71 and '81, drops sharply in 1901, then rises again in 1911.[13]

The odd behaviour of the *T* index turns out to be driven entirely by the South and Islands. Consider the cohorts in their 30s, 40s, and 50s in 1881. In the South, *T* falls sharply for all three cohorts: from 118 to 107 (-11), from 135 to 115 (-20), and from 143 to 121 (-22), respectively. In the North, the changes in *T* are -2,+2, and +18 for those same three cohorts, much more in line with what might be expected based on ageing.[14] Could the sharp improvement in accuracy in the South reflect a sharp improvement in numeracy there? Only in the event of a

[11] More specifically, we non-parametrically estimate a smoothed age distribution for each province, and compute from it predicted tens ratios. For more detail on construction of the index and discussion of anomalies, see A'Hearn et al. (2016), where the index is designated *R* instead of *T*.

[12] The two indices are based on different age groups, so we match those pairs with the greatest overlap: *T* for ages 20-29 with *W* for ages 23-32, *T* for ages 30-39 with *W* for ages 33-42, and so on.

[13] The lone exception is the cohort born in the 1860s between the censuses of 1901 and 1911, when it experienced a slight decrease in *T*.

[14] In the Centre the changes are +1, -6, and +9. Unlike the figures in Table 2, the sample used to calculate these *T* index values does include the provinces of Veneto and Lazio, since we observe these regions in both 1881 and 1901.

chimerical mass adult education initiative in a time and place where at most half of children aged 6-10 were enrolled in school (A'Hearn and Vecchi 2011, 200).

3. Censuses and state capacity in Italy

The funding and organisation of the Italian census left much to be desired in the early years. Financial constraints precluded sending enumerators to interview every household. Instead, a questionnaire was distributed to families by local governments, to be filled out by the head of household for later collection. In practice, as the head of household was often illiterate, it was often another household member or a neighbour who wrote out the family's responses – a potential source of inaccuracy right from the start.[15] In the event of obvious mistakes in a completed return, census workers were to call on the family, interview them, and record answers in their own hand.

Census officials were well aware of the age-heaping we have documented here – and not very happy about it. It was an unmistakeable symptom of the weakness in census procedures, and a serious problem for the development of mortality tables.[16] One response was to fine-tune census questionnaires and exhort local census workers to greater efforts. The censuses of 1861 and '71 asked for age in completed years.[17] In 1881, the questionnaire was modified to request both age and birth year, as recommended by the International Statistical Conference of 1872, and census workers were instructed to check the two for consistency (Dillon 2008, 42). In 1901, with the specific goal of eliciting more accurate responses, the question on age was dropped altogether, with only birth year – and now also birth month – being requested (Italy 1904: LVIII, 193-6). Census workers were urged to ensure that birth dates were recorded with "scrupulous precision." Only as a last resort, when birth year and month were not forthcoming, were officials to inquire about age and calculate birth year accordingly, checking this against local birth records. In 1911, finally, exact birth dates – year, month, and day – were requested.[18]

A second response was to centralise data processing. In the censuses of 1861 and 1871, it was local government (the municipality, or *comune*) that was charged

---

[15] A surviving household return from the census of 1861, posted at www.genealogiadavini.it, nicely illustrates this. All members of the Morelli household of Arezzo being illiterate, the form was completed and signed by a non-family member who failed to read the instructions carefully. In the column headed "relationship to the head of household," he filled in "Arezzo" for all family members. Ironically, he recorded ages with unnecessary precision, failing to heed the instruction to record *either* age in months (for young children) *or* age in years (for older individuals). See Appendix A

[16] Numerous efforts were made to circumvent the problem in estimating age-specific mortality by means of interpolation, e.g. Italy (1885). Useful overviews of the history of nineteenth century Italian censuses are Gallo and Palluzzi (2012), Mastroluca and Verrascina (2012) and Favero (2012). A useful summary regarding age can be found in Italy (2014) pp. 28-36.

[17] Only for the youngest children, aged under three in 1861 and under one in 1871, was age to be recorded in months. In 1861, months were actually to be measured in quarters: "*di tre in tre mesi.*"

[18] Exact birth date was requested on the summary household form. On the forms pertaining to each individual, however, age in years was reported.

with distributing questionnaires to each household, collecting completed forms, checking for errors and omissions, transcribing from the family return a set of individual records, tabulating from the individual records the distribution of ages (and other desired statistics), and forwarding the results up the administrative chain of command to the General Directorate of Statistics (*Direzione generale della statistica*) in Rome. To carry out these tasks *comuni* were to engage temporary clerks and rely on the unpaid voluntary efforts of local notables, professionals, schoolteachers, and government employees. No funding for these operations was provided by the central government.

So it was not just local government but also the cooperation of local populations, notables and common people alike, that determined the accuracy of data collection and processing in the first censuses. Writing about the upcoming census of 1871, Luigi Bodio, head of the *Direzione generale*, worried that the situation was not propitious. The clergy were alienated by the recent annexation of Rome, and

> "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…" (quoted in Favero 2001, 94; our translation).

A decade later, during preparations for the census of 1881, Bodio 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; our translation).

It was hoped that the statistics of large cities might be more trustworthy due to the greater education of urban populations. In the census of 1881, these hopes were disappointed, at least in age reporting.[19] The reasoning of census officials is of particular interest. They associated education not with cognitive skills, but with acceptance of, ideally even identification with, the nation-building purposes of the state's intrusion into private lives (Favero 2001, 98).

[19] 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 *T* index. 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.

In the census of 1901, two changes were introduced. First, *individual* (rather than family) questionnaires were filled out by household heads. Second, these forms were not processed locally, but were forwarded by the *comune* directly to Rome.[20] In this way, the operations of information processing *and their costs* were centralized in a dedicated government office for the first time (Gallo and Paluzzi 2012, 38-39).

The numeracy of Italians surely improved over the half-century from 1861 to 1911, but the increasing accuracy of age reporting in the censuses – most unequivocally the discontinuous drop in age-heaping between 1881 and 1901 – was driven by improvements in census procedures, an aspect of state capacity.[21]

The evolution of age-heaping in the census suggests that the state's ability to gather and process information, eliciting the cooperation of local institutions and individuals, varied as much geographically as over time. In the South, it proved difficult to induce *comuni* to comply with basic legal requirements such as keeping proper accounts and publishing budgets, to say nothing of effectively providing public services such as infrastructure, health, and education (Randeraad 1993, ch. 6). At the other extreme are towns in recently-annexed Veneto (in the Northeast) requesting permission to carry out *ad hoc* local censuses as a foundation for the continuously updated population registers required under Italian administrative law but not actually implemented by many towns elsewhere (Favero 2001, 57). Persistent local variation in the competence of peripheral officials, and in 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 the state among local citizens.

Table 3 about here

In Table 3 several proxy indicators of state capacity in the former states, just prior to unification, are presented. We see that the Bourbon Kingdom of the Two Sicilies was last among the major pre-unification states by statistical publications, while its land register was the only one still based exclusively on verbal declarations of landowners rather than an expert survey and valuation. A crude summary measure of administrative capacity is total tax revenue per capita. It is crude both because the structure or prosperity of some economies allows for high tax revenues even in the absence of sophisticated administrative capabilities, and because taxes are a policy choice and could be low in a well-administered state. 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.

[20] 1901 was also the first census in which mechanical data processing aids (Hollerith machines) were employed (Gallo and Paluzzi, 2012; 39 and Favero, 2012;132-133).

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

Another plausible proxy for state capacity is draft evasion. Successfully implementing mass conscription requires accurate knowledge of the identities (names, ages) and locations of all men in the relevant age range. It requires the ability to check their family status in order to process exemption requests. It requires a transparent and fair way of assigning draft numbers.[22] It requires coordination across different bureaucracies to ensure that a man on the list of the conscription-eligible has not died, already been drafted by another service (i.e. the navy), or already volunteered; does not have a brother serving at the same time (which could justify an exemption); is not simultaneously listed in more than one draft district; is truly enrolled at a seminary and studying for the priesthood (for example). And it requires some ability to identify, track down and bring to justice draft resisters. The data in Table 3, for the period immediately after unification (1863-64) when a common system prevailed throughout Italy, reveal a significant gap between the regions of the Kingdom of the Two Sicilies or the Papal States and the rest of the country. To be sure, the evidence collected in Table 3 is imperfect and incomplete, but it all points in the direction of lower administrative capacity in the parts of Italy where age-heaping was greatest.

These considerations resonate with recent contributions in political science. Lee and Zhang (2016) investigate the informational foundations of state capacity, focussing on the breadth and depth of the state's knowledge of its citizens and their activities – their "legibility". In a large sample covering 1980-2010, they find that age-heaping, both within and between countries, is well correlated with other indicators of state capacity (such as political stability or effective enforcement of the rule of law), with outcomes of government policy (e.g. tax revenues, public good provision, infant mortality, or literacy), and specific information gathering efforts like birth registration.

4. Age-heaping and illiteracy

A key piece of evidence supporting age-heaping as a proxy for individual human capital is that illiterates have a higher propensity to report approximate, rounded ages than those with some education. With separate age distributions for those who could and could not read, the census of 1871 allows us to investigate this for Italy. But can we trust the census literacy data? They were collected via the same flawed process that generated our age distributions, after all. Heads of household, or a trusted person in their stead, were to report whether each individual could read and/or write. It is easy to imagine people being embarrassed to admit they were illiterate, or embarrassed to admit that their spouse or other household members could read and write, while they could not.

[22] Conscripts were randomly assigned draft numbers; the lower-numbered individuals had to go on active service, while higher-numbered men received basic training and were then allocated to the reserves.

We can assess the accuracy of self-reported literacy by comparing census data with information from sources in which some form of test was applied. We turn first to military records. For both practical military reasons and broader social concerns, the army of Piedmont had been obliged since 1848 to run "regimental schools" for the education of serving soldiers. Upon unification, this regulation was extended to the new Italian army (Mastrangelo 2017). During conscription operations, provincial draft councils were directed to ascertain the literacy of all men who passed their physical examination and were declared fit to serve. Any conscript whose literacy was in doubt was asked to read a few words out loud and to sign his name and surname.[23] Results were published each year in the army's annual account of the year's draft, and published in retrospective summary form in the *Annuario statistico italiano* of 1905-07. Comparing tested literacy rates among conscripts aged 20 in 1872 against the self-reported literacy rates of 19-year old men in the census of 1871, we find a high correlation across provinces (ρ = 0.94).[24] Conscript literacy is roughly two points higher on average, which likely reflects a degree of positive selection. Men judged unfit for service by failing to meet the minimum height requirement did not have their literacy assessed, for example.

Figure 3 about here

A second source of evidence, one which allows us to observe both men and women, are marriage records. Newlyweds were asked to sign their marriage record, or to make a mark, if they could not write. Our data were compiled by the government and published in the *Annuario statistico* of 1905-07. They refer to brides and grooms marrying in 1872. Figure 3 plots newlywed signature rates by province against census literacy rates for men and women aged 17-32 in 1871.[25] This independent measure is highly correlated with census literacy for both men (ρ = 0.95) and women (ρ = 0.96).[26] As with conscripts, literacy is slightly higher among grooms than in the census, which is again likely due to positive selection, in this case positive selection in the marriage market. Among women, brides have a slightly *lower* literacy rate than the relevant age group in the census. This probably reflects a difference in definitions: our census literacy rates are based

[23] This description is taken from the *Annuario Statistico Italiano* of 1881, p. 250. The original text reads "*l'accertamento del grado di istruzione letteraria dei coscritti si fa invitando quelli dei quali per la loro condizione sociale si possa dubitare se sappiano leggere e scrivere, a leggere a voce alta qualche parola e a scrivere il proprio nome e cognome.*"

[24] There is a bit of variation in census literacy from one age to the next, but the correlations of conscript literacy with census literacy for any age from 15 to 25 are in the range from +0.93 to +0.95. Census literacy rates are systematically lower for men aged 20, 30, etc. (i.e. on 5/10 multiples, the heaping ages). In 1881 the correlation of conscript and census literacy rates (now referring to men aged 20-24 in the absence of single-year age data) is even higher, at +0.96.

[25] Among women, the share ever married in 1871 rises rapidly (by at least three points per year of age) from 18 to 27; the total increase is from 6 to 70%. Among men, the analogous ages of rapid increase in marriage are 22-33, the total rise from 9 to 74%. This implies that an age range of 18-33 ought to cover most newlyweds in 1872, or ages 17-32 in 1871 at the time of the census.

[26] The correlation of bride and groom signature rates with each other is +0.91. The correlation of groom signature rates with conscript literacy rates is +0.90.

on ability to read, while signature rates test ability to write.[27] The Italian primary school curriculum introduced writing only after some facility in reading had been achieved, so if boys attended school for longer than girls they would be more likely to pick up both reading and writing. Men also had more occasion to practise their signature outside of school, in business dealings for example. We might also conjecture that when a groom could not sign the marriage act, even a bride who could write might choose not to sign to avoid embarrassing her new husband.

Finally, we can explore the correlation of census literacy with earlier school enrolments. For the academic year 1862-63, school enrolments relative to the population aged 5-12, covering both public and private schools, are available for all provinces except those of Veneto and Lazio, regions not yet annexed to the Kingdom (Italy 1865b).[28] There is huge variation in enrolment rates, from just 2% of girls in the province of Siracusa (Sicily) to 104% of boys in Turin province.[29] If we pair these figures with provincial literacy rates for the same cohort, aged 15-22, in the 1871 census, the correlations are high (ρ = 0.93) for both for men and women.

The close correlation of self-reported census literacy with tested literacy from conscription and marriage register data, as well as with primary school enrolment rates, suggests that we can accept it as a reliable indicator of basic educational attainments.

It is important to bear in mind that most Italians reported their age "correctly." Even in Sassari province (Sardinia), where the Whipple index stood at 230 in 1871, 74% of individuals reported ages that were apparently accurate, that is, did not end in 0 or 5. And, as far as age-heaping allows us to infer, these 74% were just as numerate as the 94% of Torino province's population who reported a non-suspect age.[30] Age-heaping does not tell us about the numeracy skills of the typical individual. Rather, it tells us about the size of the minority who didn't know, or did not report, their age correctly.

To fix ideas clearly, imagine a world in which formal schooling is the only source of literacy and numeracy, and where schooling is binary: a child either completes school and emerges literate and numerate, or does not attend at all, remaining ignorant of both letters and numbers (so illiteracy and innumeracy are interchangeable). Suppose further that the numerate population all report their age accurately, while the innumerates all report an age ending in 0 or 5. In this

---

[27] The census age distribution column headings are "*sanno leggere*" and "*non sanno leggere.*" Regarding writing ability, among the conscripts of 1872 actually called to serve and re-examined by the army, 40% could read and write, 4% could only read, and 57% could neither read nor write (Italy 1875, 45).

[28] The source reports "pupils" rather than specifying enrolment or attendance. Enrolment is the more plausible interpretation if we compare with other sources.

[29] Rates in excess of 100% are possible when students from outside the reference age group (typically students repeating grades) or from outside the province boundaries are enrolled.

[30] For Sassari, see footnote 10. For Torino in 1871 W = 130. An analogous calculation yields 6% of individuals incorrectly reporting a 0 or 5 age.

world there is a simple linear relationship between the Whipple index and illiteracy:

$$W = 100 + 4 \cdot ill,$$

where *ill* stands for the illiteracy rate, measured in percentage points.[31]

Figure 4 confronts the model with the evidence. We treat age-groups as separate observations in order to avoid losing the considerable variation in both age-heaping and illiteracy along this dimension. The main features of the analysis are not dependent on this choice and carry through if we collapse the data into a single observation for each province and gender.

Figure 4 about here

Consider first the North-Centre (Panel A). Econometrically, the situation is straightforward here: a linear specification fits the data well; pooling restrictions for combining men and women cannot be rejected; and the OLS regression estimate has a respectable R^2 of 0.24. The point estimate for the intercept, at 89.7, is not far from the model's value of 100, but the slope coefficient is nowhere near the predicted value of 4. A slope of only 0.65 means that age-heaping rises by nothing like as much as it should if all illiterates reported 0 and 5 ages.[32] Some fraction of illiterates, call it δ, must have reported their ages accurately. Allowing for this in the model yields a modified formula for the Whipple index:

$$W = 100 + 4 \cdot (1 - \delta) \cdot ill.$$

The North-Centre slope estimate of 0.65 implies $\delta = 0.84$, meaning the vast majority of illiterates reported accurate ages, contradicting the initial assumption equating schooling with literacy and numeracy.

A quick glance at Figure 5's Panel B might suggest the situation is not too dissimilar in the South and Islands. We cannot reject the pooling restrictions for men and women; the regression intercept is not far from 100; and the estimated slope (1.07) is far below the model's prediction of 4 (implying $\delta = 0.73$). In fact, the model performs much worse in the South. The regression R^2 of just 0.08 implies very limited explanatory power.[33] And results are sensitive to the

---

[31] The share of 0 and 5 ages in the population will be $\{0.2 \cdot (1 - s) + 1 \cdot s\}$, or $\{0.2 + 0.8 \cdot s\}$, where s is the illiteracy rate measured as a fraction. To get the Whipple index, we divide by the expected share of 0 and 5 ages, i.e. 0.2, yielding $\{1 + 4 \cdot s\}$, and multiply by 100: $W = 100 + 4 \cdot 100 \cdot s$. Defining *ill* as the illiteracy rate in percentage points gives the formula in the text.

[32] Decreasing accuracy of age reporting among older individuals is not "natural" or a nuisance to be controlled for, but a substantive feature of the data. Here we want to understand how much of this effect can be attributed to greater illiteracy among older individuals. Hence, we do not include controls for age-group in the regression. Adding age-group dummies to the regression, they are highly statistically significant, and improve the R^2 to 0.35. But the amended regression conforms even less well to the predictions of the simple model: the intercept falls to 81.1, the slope to 0.56.

[33] Adding age-group dummies, which themselves are highly statistically significant, raises R^2 to 0.58, but lowers the intercept further below 100 and the slope coefficient further below 4.

treatment of Napoli province, where age-heaping is surprisingly high relative to illiteracy. Introducing a dummy for Napoli province causes us to reject the pooling restrictions. Re-estimating the regressions separately for men and women (including a Napoli dummy in both cases) does not really improve support for the model. Though slopes are considerably steeper, R^2's remain small (just 0.09 for women, largely attributable to the Napoli dummy), and the intercepts are large and negative (-193.3 for women). This is impossible to reconcile with the simple model.[34]

Figure 5 presents the data in a different way. For each province *W* is plotted separately for literates and illiterates. The simple model of school-based literacy and numeracy predicts two parallel lines of dots: *W* among literates should be approximately 100 in all provinces, while among illiterates it should be approximately 500. Only the relative shares of literates and illiterates would differ between provinces.

Figure 5 about here

As expected, age-heaping is in fact greater among illiterates in every province (except Treviso). As we have noted before, this supports a human capital interpretation. But the average gap is much smaller than predicted, because age-heaping is "too low" among illiterates. And there is surprising variation *within* the two groups. *W* ranges from 100 to 150 among literates, and from 105 to 230 among illiterates. Clearly, it is not education that explains these within-group differences. Toward the bottom of Figure 5 there are 23 mostly Southern provinces with *W* values above 125 among literates. Towards the top there are 19 mostly Northern provinces with *W below* 125 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.

5. Numeracy in everyday life

The evidence suggests that only a fraction of uneducated Italians (alongside a smaller fraction of the literate) reported heaped ages. Is it plausible to argue that this fraction were innumerate, insufficiently comfortable with numbers to understand, remember, or calculate their age? There are both *a priori* reasons and a few scraps of direct evidence to suggest we should be very cautious about this argument.

Numbers were everywhere in the worlds of both peasants and city dwellers, even in their games. In the ubiquitous game of *morra* the object is to quickly

[34] Adding age-group dummies increases the explanatory power of the regressions considerably, but does little to bring parameter estimates in line with the model. For example, the slope coefficient for women falls from 4.12 to 0.36, insignificantly different from zero.

guess a sum: the total number of fingers extended simultaneously by the players.[35] Meanwhile, the state lottery (*lotto*) ensured that from the late seventeenth century onward all Neapolitans knew their numbers, at least from 1 to 90 (Macry 1997).[36] A passion for the *lotto* extended from the Bourbon King himself to the denizens of the city's poorest slums. The poor made micro-wagers and experienced micro-gains and -losses on a regular basis. The minimum wager was only one *grano*, a fraction of the price of a loaf of bread; in practice shares as small as a twelfth of the minimum were available on a clandestine basis from the neighbourhood *lotto* shop operator, often on credit.

Printed lottery tickets specifying the numbers chosen by the bettor and the potential winnings, as well as the public display of the week's winning numbers at *lotto* shops, presumed widespread recognition of written numbers. So too did the popular books on 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 Serao's time, illegal private lotteries with minimum wagers as low as a couple of *centesimi* flourished alongside the Italian state *lotto.* Receipts, though written in pencil on dirty scraps of paper, had the canonical form of old, specifying the bettor's wager, chosen numbers, and associated payoffs (Serao 1885; 64).

---

[35] The game was once a popular motif for painters of common life scenes. Dickens vividly described a variant in his *Pictures from Italy* (1846). "The most favourite game is the national one of Mora, which they pursue with surprising ardour, and at which they will stake everything they possess. It is a destructive kind of gambling, requiring no accessories but the ten fingers, which are always—I intend no pun—at hand. Two men play together. One calls a number—say the extreme one, ten. He marks what portion of it he pleases by throwing out three, or four, or five fingers; and his adversary has, in the same instant, at hazard, and without seeing his hand, to throw out as many fingers, as will make the exact balance. Their eyes and hands become so used to this, and act with such astonishing rapidity, that an uninitiated bystander would find it very difficult, if not impossible, to follow the progress of the game. … It is never the quietest game in the world; for the numbers are always called in a loud sharp voice, and follow as close upon each other as they can be counted. On a holiday evening, standing at a window, or walking in a garden, or passing through the streets, or sauntering in any quiet place about the town, you will hear this game in progress in a score of wine-shops at once; and looking over any vineyard walk, or turning almost any corner, will come upon a knot of players in full cry. It is observable that most men have a propensity to throw out some particular number oftener than another; and the vigilance with which two sharp-eyed players will mutually endeavour to detect this weakness, and adapt their game to it, is very curious and entertaining." (Dickens 1846, pp. 45-.46). The game, which can also 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).

[36] 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 he or she had chosen would be drawn. Winnings were proportional to the amount wagered, with the ratio lowest for a just one of the bettor's numbers appearing (ca. 12:1) and highest for a triple in which all three were drawn (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 of divination). We do not challenge that interpretation, though it is worth noting that playing numbers suggested by dreams or current events was not a suboptimal strategy given the structure of *lotto* payoffs.[37] Our argument is only that the popularity of the game suggests the ability to count and to recognise and name numbers was widespread.

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 recognize 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. Federico (1986) has shown that Italian agriculture, notwithstanding its low productivity, was quite market oriented, with on-farm consumption accounting for only about 30% of total production. In this context, peasant families developed "survival tactics" based on rudimentary cost-benefit analyses, which required a modicum of numerical skill (Federico 1987). To make this essentially evolutionary argument is not to claim that the average Italian landless labourer understood compound interest, nor to deny recent 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.[38]

The tools of basic numerical competence were there to be acquired, though not necessarily in forms familiar to us today. Methods of counting with one's fingers to 100 and beyond, though no longer set out in Italian math 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.[39] In Italy as elsewhere, the monetary system had been based on 12 and 20, numbers twice-divisible by two: 12 *denari* per *soldo*, and 20 *soldi* per *lira*.[40] Twelve was also the basis of various physical measures: in southern Italy 1

---

[37] The *lotto* did not split a fixed total award among all holders of winning tickets. In that case, it would have been wise to avoid popular numbers chosen by other gamblers. Instead, the *lotto* paid a *fixed multiple* of the amount wagered to *every* winning player. The operation made big losses when popular numbers were drawn from the urns.

[38] Tollnek and Baten (2017) 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 relaxed nutritional constraints on cognitive development. For an earlier period, a student of medieval Tuscany's fiscal records is impressed by the "bottom-up" numeracy of ordinary farmers, who could estimate average yields and knew the value of their assets. At the same time, they reported their ages inaccurately. She concludes that "inaccuracies in age reporting may not necessarily reflect lack of numeracy per se; it may reflect a somewhat narrower lack of knowledge of the biological age" (Emigh, 2002; 670).

[39] The so-called "Russian peasant multiplication" method (Bowden, 1912) reduces *any* multiplication problem to a series of halving, doubling, and summation operations.

[40] In the Kingdom of Naples 12 *cavalli* per *grano* and 20 *grana* per *tarì* had played equivalent roles in the monetary system, though there were also other denominations in circulation.

palmo (.26m) was 12 *once*; 1 *botte* (524l) was 12 *barili*; and 1 *libbra* (.32kg) was 12 *once*. Attempts to introduce decimalisation beginning 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 of all disciplines mathematics was where teaching produced the most satisfactory results, perhaps because the subject suited the "speculative and calculating nature" of locals. Practical instruction in using the metric system, however, was being held back by the persistence of ancient Neapolitan measures in daily usage, he claimed (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 oral statements of age, 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 recording numbers temporarily by means of traditional systems of marks made with a piece of charcoal or a sharp object was a daily practice. So too was use of the abacus (Roggero 1994).[41] For record-keeping, tally sticks and their like were often used. 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*. The system was still common in the twentieth century in Puglia, where illiterate shepherds 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 the ears (La Sorsa, 1953). In a fascinating study Ninni (1889) reconstructed a relatively sophisticated system of recording and summing numbers employed by the mostly illiterate fishermen of Chioggia (near Venice) to keep track of the payments due them (Appendix B). A similar system of counting was also used by the peasants of the Valle della Morra in Umbria (Nicasi, 1906). These studies suggest that, even among unskilled occupations, some basic numeric knowledge was widespread.

On reflection, this degree of numerical fluency is hardly surprising. We know from modern studies that children learn to count readily, largely figuring it out for themselves before starting school. They even devise strategies for basic addition and subtraction without instruction, such as breaking 7+8 into 7+(3+5) and then rearranging this into the easier sums (7+3)+5 = 10+5 = 15 (Dehaene 1997, Ch. 5). Children of the late twentieth century whose occupations keep them out of school but involve them in monetary transactions develop their own "street math": non-standard methods for solving more complicated problems. In a classic study, unschooled Brazilian street vendors aged 10-12 performed better

[41] 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.

on several mathematical tasks than their counterparts in urban schools (Saxe 1988).[42] A few decades earlier, the Soviet developmental psychologist Luria had 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. Some of Luria's questions involved numerical 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. Because Luria wished to understand whether difficulty answering such questions was caused by a lack of basic numeracy or the inability to think abstractly, he would also pose the same question in different form, e.g. "divide thirty cookies equally among five men". Even for illiterate, never-schooled individuals, he found, "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-math, and the tally stick are clear; they are not good substitutes for proper education in basic arithmetical operations. The claim for numeracy made here is only that the typical Italian had the cognitive capacity required to comprehend, to state, to record, and to recognize the numbers relevant for human age.

6. Age-heaping and gender

A gender gap is evident in Table 1 and Figure 4. Of 345 province—age-group observations in 1871, women's age-heaping exceeds men's in 339. The mean gap is 13 Whipple points, or about 10% in proportional terms. As shown in Section 4, this is *partly* attributable to women's lower educational attainment. Focussing on women only, we find the pattern in Figure 5 repeated in 1861 and 1871: age-heaping is lower among educated women in both North and South, but illiterate northern women report their ages as accurately as educated southerners. Tracking gender differences across later censuses with the T index described earlier, we find that the gender gap dies out slowly but steadily in the North, while in the South it decreases sharply in the census of 1901 but otherwise displays no trend.[43]

The data for 1861 and 1871 allow us to break down women's age-heaping by marital status. Table 4 presents our calculations of both W levels and W gaps relative to men, i.e. $(W_{female} - W_{male})$. Two striking patterns are revealed. First, age-heaping is much greater among widows than married women. This is *not* due to widows being older on average: age-heaping is greater among widows in every age-group.[44] Similar findings regarding widows have been reported for

[42] See also Carraher, Carraher and Schlieman (1985) for another influential study of the numerical abilities of working children in Recife.

[43] Between 1881 and 1901, men's age-heaping falls from T = 147 to T = 108 (-39), while for women the decrease is from 159 to 111 (-48). The Centre is omitted from the comparison in the text. In this area the gender gap was small right from 1861.

[44] And in (almost) every province. In 334 of 345 province–age-groups, widows heap more than wives in the census of 1871. The exceptions are all in the 23-32 age bracket and all in the North-Centre (or on the border, in the case of L'Aquila province).

Britain, the US, and Ireland in this period (Földvari et al. 2012; Blum et al. 2016). Second, gender gaps are dramatically smaller in the married category than for either singles or widows and widowers. In the South, married women's ages are actually *less* heaped than married men's (making the gender gap is negative).[45]

Table 4 about here

A numeracy interpretation of these patterns would imply differential selection into different marital states. Consider marriage first. Lower age-heaping among married women would mean positive selection on human capital, operating by way of numeracy's correlation with wealth, health, and perhaps income earning capacity – characteristics valued in the marriage market. We do see this in the North, where *W* values are roughly 10 points lower among married than single women. (Since the gender gap is also roughly ten points lower among the married in the North, we can infer no analogous selection effect on men.) In the South, instead, there would be no positive selection into marriage. If anything the opposite seems to have been true, with age-heaping among married women somewhat higher than among singles. Meanwhile there is a 30 point change in the gender gap, which is roughly +20 among singles and -10 among married women. Given slightly higher levels of age-heaping among married women, this would imply strong negative selection into marriage among southern men. While not impossible, these patterns are difficult to make sense of.

Now consider selection into widowhood. The human capital interpretation would suggest powerful negative selection, in both North and South, in both censuses, in all age groups. It is not hard to imagine negative selection into *early* widowhood: women with low human capital might have married similar men, lived in poverty, and more often experienced the early death of their spouse.[46] But negative selection into early widowhood implies, *ipso facto*, *positive* selection in *later* widowhood, which we do not see in the age-heaping figures. Also, this effect should be symmetric for men, who instead show only very modest increases in age-heaping when we compare married and widowed individuals.

The status of widow depended, of course, on a woman not having remarried. Livi Bacci (1981) showed that in the 1880s many women remarried, especially those losing their husband before age 50, among whom 40% married again.[47] If remarriage positively selected women, those remaining as widows would be negatively selected. The effect could have been stronger in the North, where remarriage was less common, hence, possibly, accessible only to a more select

[45] One could worry that *W* values for some groups – young widows, say – are based on tiny samples. This is not the case. The smallest group in Table 4 is never-married women aged 63-72 in the South: 25,000 individuals, or 12 per cent of all southern women in this age bracket. Second-fewest are young widows aged 23-32 in the North: 27,000 women, or 2% of all northern women of this age.

[46] Or women with low human capital could have become second wives of older men, who would of course die earlier in the woman's life.

[47] For men remarriage rates are higher: 43 per cent (overall) and 73 per cent (under 50).

group. Livi Bacci estimated remarriage rates for widows under fifty at 31% in the North-Centre, a much higher 48% in the South and Islands.[48] He speculated that this might be due to the greater proportion of extended- and multi-family households in parts of the North-Centre, where support for widows and their children would have been available without remarriage.

But an argument for strong positive selection into remarriage faces two difficulties. The first is the contrast, in the South, between negative selection into first marriage and strong positive selection into second marriage. The second is that, as Livi Bacci's conjecture reminds us, it is not just a woman's economic desirability in the (re)marriage market that counts, but also her own desire, or need, to remarry. Legally, women who remarried could lose rights to wealth accumulated during their married life, to the family home, and even to their children. And a widow living with parents- or siblings-in-law could be pressured not to depart with her dowry by remarrying (Mazzoni et al. 2013). The net effect of these various influences is ambiguous (Breschi et al. 2009, 279).

Thus, the selection patterns required to reconcile differences in age-heaping by marital status with an individual human capital interpretation are, at best, convoluted. More plausible than strong selection on numeracy is that age *reporting* behaviour varied with marital status. The social convention in Italy was for wives to be somewhat younger than their husbands. Among marriages contracted between 1872 and 1881, the single most common marriage category in the official statistics paired a groom aged 25-29 with a bride aged 20-24 (Italy 1882, xvii).[49] To produce the observed age-heaping patterns, couples wishing to (appear to) conform to spousal age norms must have started from the husband's age, from which was subtracted a socially-conventional difference to obtain an appropriate age for the wife.[50] A socially-favoured difference of 0 or 5 years would mean the age distribution for wives inherited the degree of heaping of the husbands. An imposed difference of 1-4 years could result in married women's ages being *less* heaped than men's, since 55 for the husband could mean 53 for the wife, or 60 for the man being paired with 56 for the woman. In this way we can explain the absolute decrease in age-heaping between single and married women in the North, and the declining gender gap in age-heaping with marriage in both North and South.[51]

Upon the death of their wife, widowed men would continue to report their age, whether accurately or inaccurately, as before. Across both censuses and regions,

---

[48] These are simple averages of Livi Bacci's estimates for the Northwest, Northeast, Centre, South and Islands (1981, p. 358).

[49] Government statisticians were interested in fertility, and classified marriages in which either the groom was older than 65 or the bride older than 50 as "presumed infertile." These amounted to 1.63 per cent of the total.

[50] Referring to the charivari, Corsini (1981) reminds us that local social norms were enforced regarding age differences in remarriages.

[51] A complete account requires explaining why married men in the South reported their ages less accurately than single men. 1871 *W* values are 148 and 183, respectively, in the South, 126 and 125 in the North. Possibly, married southern men aspired to the dignity and authority associated with a paterfamilias, to which an exaggerated age might further contribute.

men's *W* values are only slightly higher, for the older age-groups, among widowers. Differently, widows would now find themselves unconstrained by their husband's age. Still mindful of the social significance of age, some women could adopt a younger age to gain an advantage in the (re)marriage market, where as we have seen, a younger woman's chances of finding a second husband were dramatically higher. She would likely settle on a not-implausible, round number. (Alternatively, a widow could choose an older age befitting a desired status as respected elder.) This explanation, which accounts for the observed pattern of widows reporting age less accurately than both single or married women, has been suggested by Földvari et al. (2012) for Britain the US and Britain on the basis of evidence from the censuses of 1851 and 1880, respectively.

In the context of Italian census reporting, widows could only exercise this agency to the full if they were heads of household. In the event that they resided as dependents with their adult children, in-laws, or other relatives, their age might be reported for them. Kertzer and Karweit (1995) analyse patterns of co-residence in Casalecchio di Reno, and find that in 1881 only 16% of widows were household heads. 50% were mothers of the head of household (almost always her son, usually married). Among widowers, by contrast, fully 79% remained heads of household.[52] Casalecchio, a rural municipality near Bologna, in a region where large, complex households were common, cannot be taken as representative of Italy; the example shows only that widows becoming dependants, in contrast with widowers, is more than just a theoretical possibility. Would not children report their mother's age accurately, though? Not necessarily. In the US and Canadian censuses of 1870/1 and 1900/1, Dillon (2008, 110) finds that parents of the head were the household members most likely to have a multiple of ten age reported, controlling for other characteristics. Certainly sons, in-laws, and other relatives did not have the same incentive as husbands to tie the age reported for a widow to their own age. Gender gaps in age-heaping deserve further study. What we can say at this stage is that in Italy, women's ages were often reported by others and were heavily influenced by social conventions.

7. Age-awareness and birthdays

There is no evidence that birthdays 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 theorizing 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 calendars of recurring religious feasts dovetailed with agricultural rhythms underlying social life. With modernisation, this equilibrium was

[52] Casalecchio's population of widows and widowers in 1881 was small: just 185. But the authors repeat the analysis and find similar patterns in 1921, with 418 widows and widowers.

disrupted and attitudes towards time changed. In the West, the rise of Protestantism weakened the cult of saints and the general orientation of life around the liturgical calendar. Meanwhile urbanisation and industrialisation cut links with natural cycles, and the rise of wage labour further focused 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 labour 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, the medieval Church had opposed their celebration as a pagan ritual. Baptism or death – either interpreted as marking the start of one's true life in God – were dates that mattered more. Calendar age was similarly unimportant in the medieval conception of a life course. The "ages of man" were developmental rather than strictly chronological, marked by visible signs of maturity and eventually senescence. Only in the modern, linear conception of time did the progressive accumulation of age, or "capitalisation of years," acquire greater meaning (Schmitt 2007). Individualism and, more speculatively, astrology may also have had roles to play in this development. Despite such changes, Schmitt finds scant evidence before the late eighteenth century of birthday celebration even among the rich and well-born, even among individuals who definitely knew their exact birth date and year.[53] 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 in printed works. Among books published circa 1850, "*Geburtstag*" and "birthday" occur more than ten times as frequently as "*compleanno*" and "*anniversario della nascita*" together in the relevant German, English and Italian collections of digitised historical texts (Table 5).[54]

Table 5 about here

[53] Carl Friedrich Gauss was motivated to devise a method for calculating the date of Easter by a desire to discover his own birthdate. His mother had not recorded the date (which lets us infer his childhood birthdays were not celebrated), but easily remembered where it had fallen in the liturgical calendar: on a Wednesday, eight days before the Feast of the Ascension. His birthdate was 30 April, 1777 (Dunnington 1955, 69). In the English speaking world, the diarist Samuel Pepys always noted his birthday, but never mentions any celebration of it.

[54] 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, with the result that a single instance of a birthday word can have a substantial impact on the relative frequency series in the early years.

The trends in birthday word frequency are rising in all three literatures, but latest and most weakly in Italian. Interestingly enough, Table 5 shows that the word "*onomastico*" (celebration of one's name day, on the feast day of the saint after whom one is named) remains more popular than *compleanno* until 1950. In southern Italy, at least in Naples, the *onomastico* was more commonly celebrated than the birthday 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" (Ferrante 2011, 221; Galasso 2012). Saint's days did not coincide with birthdays; in the case of popular given names the *onomastico* was shared with countless other individuals, and fell on a religious holiday. On a name day the individual's age was much 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 temporal, occupational, and geographic differences in age-heaping.

An interesting example of cultural differences in age awareness comes from twentieth century China. Age reporting there was "astonishingly accurate" in censuses from 1953 to 1982, even among largely 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 easily memorable, 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).[55] The 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 Uighur ethnic minority, which alongside smaller Kirgiz and Tajik groups maintained a clear cultural and linguistic separateness from the Han Chinese. Among the Uighur, "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."

8. Conclusions

The findings reported here validate age-heaping as a proxy for human capital in nineteenth century Italy. They are less supportive of its role as direct measure of

[55] The twelve-year cycle of zodiac animals ("earthly branches") interacted with a ten-year cycle of "heavenly stem" names, creating a set of 60 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).

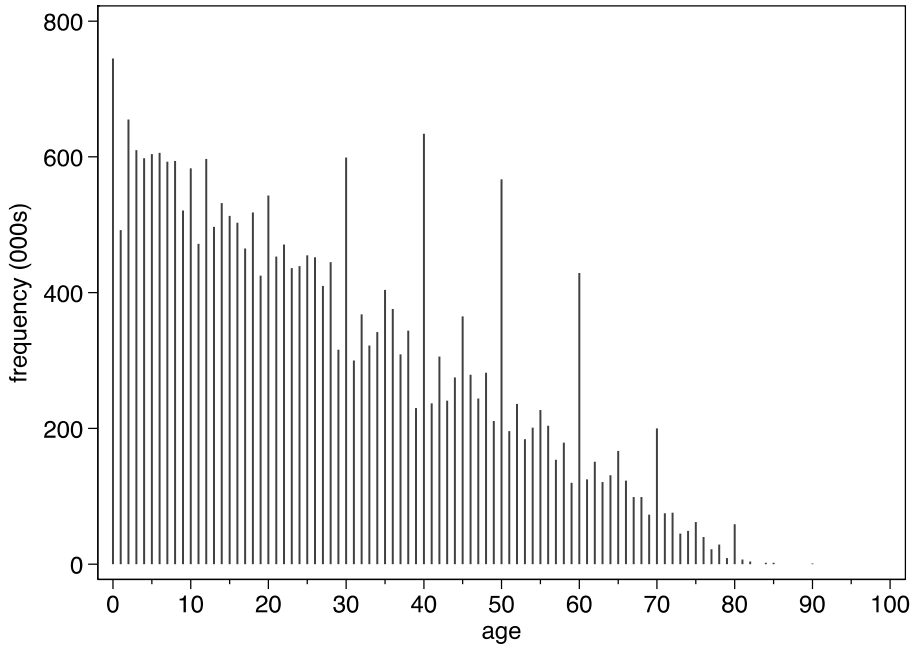
numeracy, capable of providing additional information on the distribution of cognitive abilities even where literacy data are available. A close reading of the evidence reveals too many anomalies for that: age-heaping was non-trivial, and varied widely, even among educated individuals; it varied by gender and marital status in a way hard to explain in terms of human capital; and it varied over time in a way clearly attributable to changes in the data gathering process.

Furthermore, parallels with other times and places, together with shreds of direct evidence gleaned from Italian social history, cast doubt on the assumption that a significant share of Italians, even the large pool of illiterates, lacked basic familiarity with numbers and elementary arithmetical operations. More plausible than ignorance of numbers is simple ignorance of age, or a lack of interest in reporting it accurately. There is evidence to support both hypotheses. Age-awareness was limited in much of Italy by traditional perceptions of time and attitudes regarding individualism and birthday celebration. Meanwhile, the capacity of pre-unification Italian states to gather and process information, eliciting the cooperation of their citizens, matches up well with regional variation in age-heaping. As Herlihy and Klapisch-Zuber argued more than forty years ago, culture and the state should be understood as coevolving (1985, orig. 1978, 164). From 1371 onwards "the people of Tuscany, including its humblest members, underwent an apprenticeship in measuring time… Through the Catasto of 1427, we can glimpse the means by which the commune's authority encouraged this cultural revolution, and we can discern the kinds of conscious or unconscious resistance which opposed the official demand for accuracy."

Cultural change in the direction of secularism, individualism, and linear time-perception, and institutional change in the form of administrative capacity building by the state, are both aspects of what can usefully be called modernisation.[56] The spread of education 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. It remains, on this basis, an important indicator for diagnosing and predicting economic development. But age-heaping has more to teach us, and the time is ripe to go back to the beginning and broaden the focus to encompass culture and state capacity alongside cognitive abilities, to consider age-heaping as the outcome of a historical process of social and economic development. That is the lesson of the Italian experience of the XIX century.

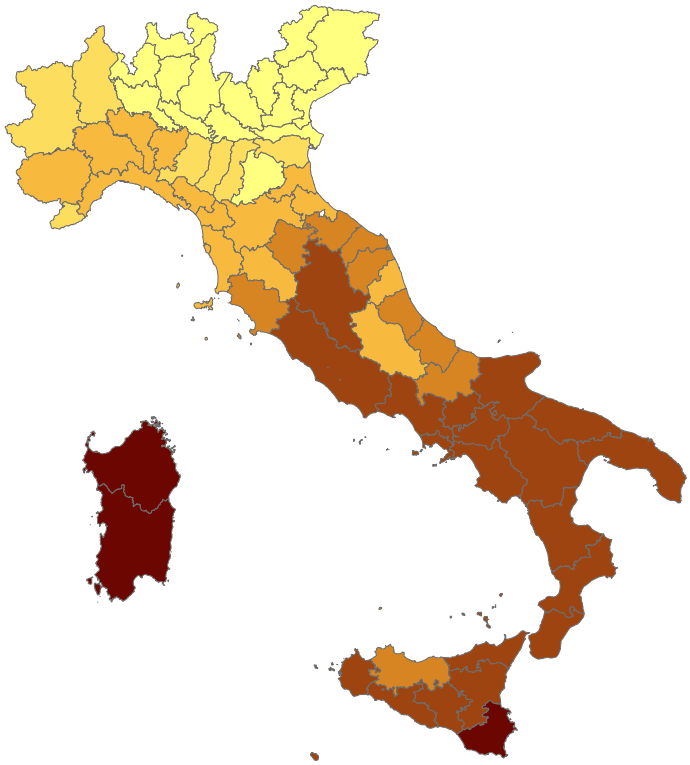
[56] We use the term "modernisation" in a descriptive way, to characterise a multidimensional, mutually reinforcing process of change, without endorsing any specific variant of sociological theory. For a useful recent take on modernisation, firmly locating its origins in economic change, see Inglehart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

Figure 1. Declared ages, census of 1871



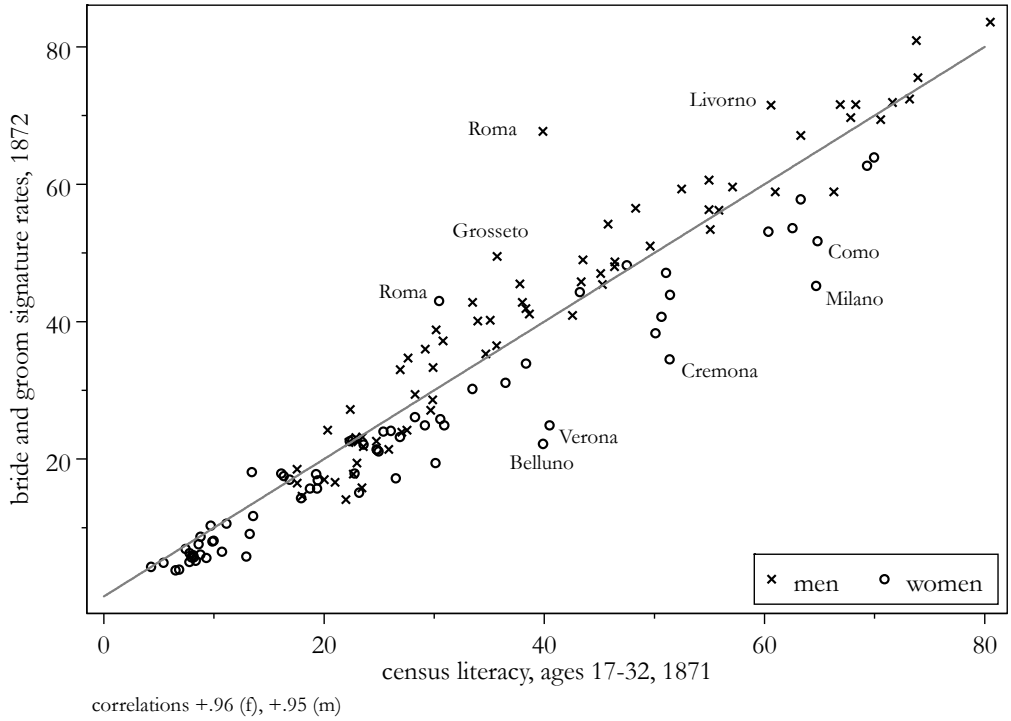
Note: Frequencies of reported ages in the Italian census of population of 31.12.1871.

Figure 2. Age-heaping by province, census of 1871



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

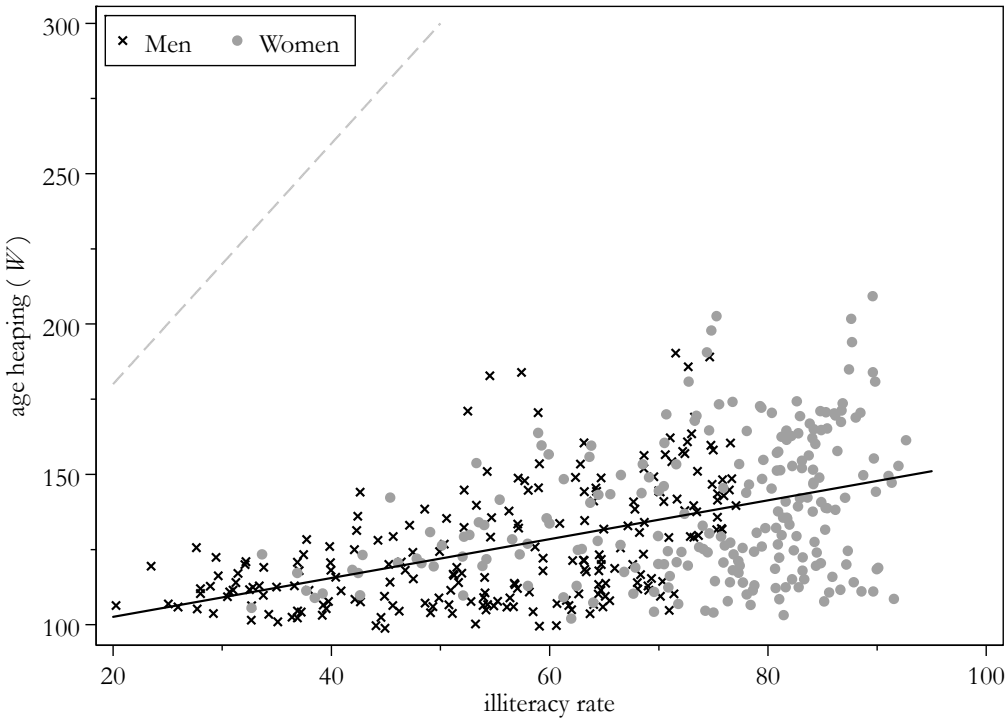
Figure 3. Literacy: census of 1871 and newlyweds of 1872



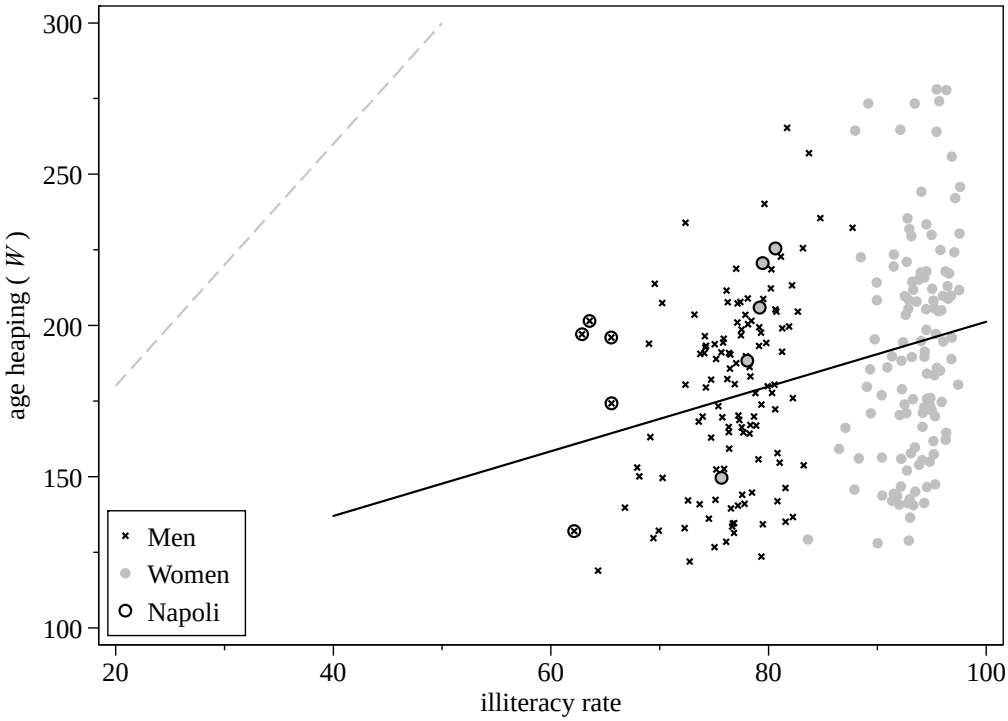
Note: signature rates for brides and grooms married in 1872 and literacy rates for men and women aged 17-32 in the census of 1871; each marker represents a province and sex; diagonal is a 45-degree line.

Figure 4. Illiteracy and age-heaping

a. North-Centre

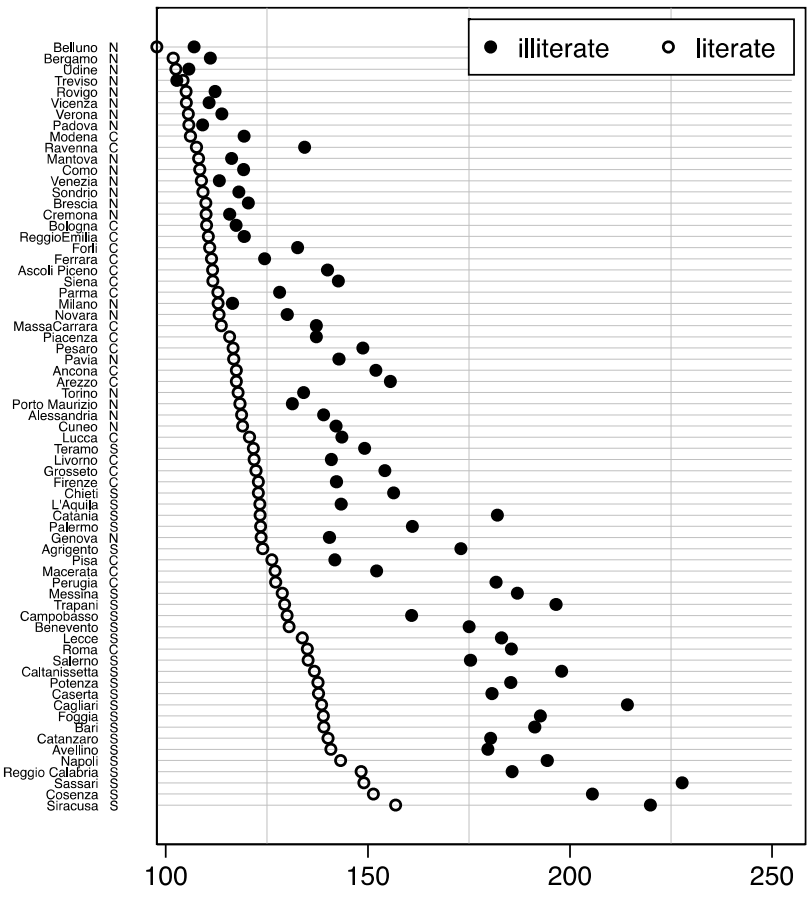


b. South and Islands



Notes: each marker represents a province-age group-gender observation; solid lines are OLS regression estimates; dashed lines illustrate the model described in the text.

Figure 5. Age-heaping among literates and illiterates



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

Table 1. Age-heaping by age-group and gender, census of 1871 (Whipple index)

	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 pooled	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 and 5 ages. *n* = number of observations, in thousands. "pooled" average pools all ages and calculates a single *W* value; "2-step" figure is a simple average of *W* values calculated separately for each age-group and reported in the table. Age groups 13-22 and 73-82 excluded from both *W* calculations and from the total *n*.

Table 2. Age-heaping across 5 censuses (*T* index)

age group	*1861*	*1871*	*1881*	*1891*	*1901*	*1911*
20-29	**103**	105	106			106
30-39	101	**105**	108		99	102
40-49	126	112	**118**		102	96
50-59	133	138	122		103	105
60-69	153	141	152		**107**	109
70-79	154	156	146		121	**123**
average*	134	130	129		106	107

Notes: *T*-index calculated at province level as described in text. Reported figures are simple averages across provinces (those of Veneto and Lazio excluded for comparability with 1861). Averages in the bottom row are simple means across age groups 30-39 to 70-79, the youngest group being excluded for comparability with 1901, when the census tabulation does not permit calculation of the ten-ratio for ages 20-29. Bold figures refer to the cohort born 1832-41.

Table 3. Indicators of administrative capacity, 1850s

	Population	Tax revenue	Draft evasion	Publishing	Cadastre	Age-heaping
Piedmont	4.6	100	3.8	100	mixed	133
Lombardy-Venetia	5.6	83	1.8	102	survey	115
Parma	0.5	59	1.8	.	survey	134
Modena	0.6	63	1.3	.	mixed	126
Tuscany	1.8	63	1.7	64	survey	146
Papal State	3.1	70	17	49	survey	153
Two Sicilies	9.3	55	7.2	41	descript.	181

Notes: Population in millions. Tax revenue: total ordinary revenue of central government p.c., relative to Piedmont's 30.46 liree; estimate for Kingdom of the Two Sicilies for continental provinces only. Draft evasion: percentage of draft dodgers (class of 1843 called to arms in 1863-1864) relative to men declared fit to serve and not granted an exemption.. Publishing: number of statistical works published 1850-59 per million population, relative to Piedmont's 8.0. Cadastre: basis of the land register as described in text. Age-heaping: Whipple index; *W* of Piedmont refers to Piedmont and Liguria only; Sardinian *W*=224.
Sources: population – Correnti and Maestri (1858, p. 381); revenue – Romani (1982, p. 404); draft evasion – Italy (1865c); statistical publication rates – Patriarca (1996, p.241); cadastre – Zamagni (1993, p. 73); age-heaping – our calculations.

Table 4. Female age-heaping (Whipple index)

	Total	*Single*	*Married*	*Widowed*	*Illiterate*	*Literate*
a. *Levels*						
North-Centre						
1861						
Average	**149**	**149**	**140**	**171**		
1871						
23-32	121	117	123	136	126	115
33-42	139	148	136	158	145	128
43-52	149	160	143	169	155	137
53-62	155	155	142	178	162	134
63-72	148	148	132	161	154	127
Average	**143**	**146**	**135**	**161**	**149**	**128**
South-Islands						
1861						
Average	**190**	**167**	**176**	**229**		
1871						
23-32	148	130	154	185	150	124
33-42	179	172	175	226	183	140
43-52	198	184	186	246	201	154
53-62	215	190	187	267	219	166
63-72	210	185	175	246	214	162
Average	**190**	**172**	**175**	**234**	**194**	**149**
b. *Gender gaps*						
North-Centre						
1861						
Average	**17**	**18**	**8**	**31**		
1871						
23-32	13	13	9	18	14	10
33-42	15	21	12	27	11	13
43-52	15	22	10	29	9	16
53-62	20	18	9	34	13	15
63-72	19	18	5	27	12	15
Average	**17**	**18**	**9**	**27**	**12**	**14**
South-Islands						
1861						
Average	**11**	**21**	***-10***	**36**		
1871						
23-32	13	14	1	39	9	9
33-42	11	26	2	55	4	7
43-52	7	24	*-9*	47	*-4*	8
53-62	16	30	*-14*	46	5	16
63-72	20	27	*-16*	40	10	13
Average	**13**	**24**	**-7**	**45**	**5**	**11**

Note: "average" values calculated as simple averages over the five age groups; gender gaps are female – male values for the same status and age; North-Centre excludes Veneto and Lazio in 1871 for consistency with 1861; age-group breakdown for 1861 omitted for clarity of presentation.

Table 5 Frequency of birthday words in published work in three languages

	1700	*1750*	*1800*	*1850*	*1900*	*1950*	*2000*
Compleanno	0.00	0.00	0.27	0.29	4.57	11.22	28.27
anniversario della nascita	0.00	0.00	0.51	1.44	2.01	2.71	3.03
Onomastico	0.00	0.00	1.87	6.47	7.10	6.86	11.15
Birthday	1.54	1.84	5.74	24.72	60.28	79.66	101.16
Geburtstag	.	7.02	13.75	20.81	42.84	120.68	212.91

Notes: relative frequency of the phrases "*compleanno,*" "*anniversario della nascita*", "*onomastico,*" "*Geburtstag*", and "birthday" in the Google Books Italian, German, and English corpora; 21-year moving averages; case-insensitive searches using Google's N-gram Viewer on 20/10/2019; raw values multiplied by 100,000 for ease of interpretation; "*Geburtstag*" value for 1700, determined by a single instance of the word in 1692, is suppressed.

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.

A'Hearn, Brian, Alexia Delfino, and Alessandro Nuvolari, "Rethinking Age Heaping. A Cautionary Tale from Nineteenth Century Italy," *University of Oxford Discussion Papers in Economic History*, no. 148 (Sept. 2016).

A'Hearn, Brian, and Giovanni Vecchi, "Istruzione," Ch. 3 in *In ricchezza e povertà. Il benessere degli italiani dall'Unità a oggi*, G. Vecchi ed. (Bologna: Il Mulino, 2011), pp. 159-206.

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 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.

Blum, Matthias, Christopher Colvin, Laura McAtackney, and Eoin McLaughlin, 'Women of uncertain age: quantifying human capital accumulation in rural Ireland in the nineteenth century,' *Economic History Review*, vol. 70 (2017), pp. 187-223.

Bowden, Joseph, "The Russian peasant method of multiplication", *The Mathematics Teacher,* vol. 5, no. 1 (1912), pp. 4-8.

Breschi, Marco, Alessio Fornasin, Matteo Manfredini, and Marianna Zacchigna, "Family Composition and Remarriage in Pre-transitional Italy: A Comparative Study," *European Journal of Population*, vol. 25 no. 3 (2009), pp. 277-96.

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

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

Correnti, Cesare and Pietro Maestri, *Annuario Statistico Italiano. Anno I 1857-1858,* Torino: Tipografia Letteraria, 1858.

Corsini, Carlo, "Why is Remarriage a Male Affair? Some Evidence from Tuscan Villages During the Eighteenth Century," ch. 26 in *Marriage and Remarriage in Populations of the Past*, J. Dupâquier, E. Hélin, P. Laslett, M. Livi Bacci, and S. Sogner, eds. (London: Academic Press, 1981), pp. 385-95.

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.

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.

Dickens, Charles. *Pictures from Italy*. London: Bradbury & Evans, 1846.

Dillon, Lisa. *On the Shady Side of Fifty: Age and Old Age in Late Victorian Canada and the United States*. Montreal: McGill – Queen's University Press, 2008.

Dunnington, G. Waldo. *Carl Friedrich Gauss: Titan of Science*. New York: Haffner Publishing, 1955.

Favero, Giovanni, "Il dibattito sull'organizzazione dei primi censimenti unitari negli atti della Giunta e del Consiglio Superiore di Statistica," *Annali di Statistica*, Series XII, Vol. 2 (2012), pp. 121-145.

Federico, Giovanni, "Mercantilizzazione e sviluppo economico italiano (1861-1940)," *Rivista di Storia Economica,* vol. 3 (1986), pp. 149-186.

Federico, Giovanni, "Contadini e mercato: tattiche di sopravvivenza," *Società e Storia,* vol. 4 (1987), pp. 877-913.

Felice, Emanuele and Michelangelo Vasta, "Passive modernisation? 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 H. *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.

Gallo, Gerardo and Evelina Palazzi, "Le trasformazioni del censimento della popolazione in Italia: i cambiamenti della macchina organizzativa censuaria negli ultimi 150 anni", *Annali di Statistica*, Series XII, Vol. 2 (2012), pp. 31-75.

Herlihy, David and Christiane Klapisch-Zuber. *Tuscans and Their Families: A Study of the Florentine Catasto of 1427*. New Haven: Yale University Press, 1985 (orig. publ. 1978).

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.

Italy. Istituto Nazionale di Statistica (Istat). *Censimenti e società. Mutamenti sociodemografici della Sicilia in 150 anni di storia*. Rome: Istat, 2014. Available at https://www.istat.it/it/files//2014/03/Censimenti-e-società-.pdf.

Italy. Ministero di agricoltura, industria, e commercio. *Popolazione*. Censimento generale 31 dicembre 1861, vol. II. Torino: Tipografia Letteraria, 1865.

Italy. *Statistica del Regno d'Italia. Istruzione pubblica e privata, anno scolastico 1862-63. Parte prima. Istruzione primaria*. Torino: Enrico Dalmazzo Tipografo, 1865 (b).

Italy. Ministero della Guerra. Direzione Generale delle Leve, Bassa-Forza e Matricola. *Della leva sui giovani nati nell'anno 1843 e delle vicende dell'esercito dal 1o ottobre 1863 al 30 settembre 1864. Relazione del Maggior Generale Federico Torre al Sig. Ministro della Guerra*. Torino: Tipografia Fodratti, 1865 (c).

Italy. Ministero di agricoltura, industria, e commercio. *Popolazione classificata per età, sesso, stato civile, ed istruzione elementare.* Censimento 31 dicembre 1871, vol. II. Roma: Tipografia Cenniniana, 1875.

Italy. Ministero di agricoltura, industria, e commercio. *Popolazione. Movimento dello stato civile . Anno XX – 1881*. Roma: Tipografia Bodoniana, 1882.

Italy. Ministero di agricoltura, industria, e commercio. Direzione generale della statistica. *Relazione sul metodo di esecuzione e risultati del censimento, raffrontati con quelli dei censimenti italiani precedenti e di censimenti esteri*, in *Censimento*

della popolazione del Regno d'Italia al 10 febbraio 1901, vol. V. (Roma: Tipografia nazionale G. Bertero e c., 1904.)

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.

Lee, Melissa, and Nan Zhang, "Legibility and the Informational Foundations of State Capacity," *Journal of Politics*, Vol. 79, no. 1 (Jan. 2016), pp. 118-32.

Livi Bacci, Massimo, "On the Frequency of Remarriage in Nineteenth Century Italy: Methods and Results," ch. 23 in Marriage and Remarriage in Populations of the Past, J. Dupâquier, E. Hélin, P. Laslett, M. Livi Bacci, and S. Sogner, eds. (London: Academic Press, 1981), pp. 347-61.

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

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

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.

Mastrangelo, Gianfranco. *L'istruzione della truppa analfabeta: le scuole di reggimento dalle Alpi al Lilibeo 1835-1913*. Rome: Nadir Media, 2017.

Mastroluca, Simona and Mariangela Verrascina, "L'evoluzione dei contenuti informativi dei censimenti della popolazione," *Annali di Statistica*, Series XII, Vol. 2 (2012), pp. 77-119.

Mazzoni, Stanislao, Marco Breschi, Massimo Esposito, and Lucia Pozzi, "Widowhood and Remarriage in Sardinia, Alghero, 1866-1925," *Population*, vol. 68 no. 2 (2013), pp. 237-64.

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

Newman, Saul, "Supercentenarians and the oldest-old are concentrated into regions with no birth certificates and short lifespans," bioRχiv *working paper*, 2019. doi: https://doi.org/10.1101/704080

Nicasi, Giuseppe, "Dei segni numerici usati attualmente dai contadini della valle di Morra nel territorio di Città di Castello," *Bollettino della Regia Deputazione di Storia Patria per l'Umbria*, vol. 12 (1906), pp. 267-276.

Ninni, A. P. *Sui segni prealfabetici usati anche ora nella numerazione scritta dai pescatori clodiensi*. Venezia: Antonelli, 1889.

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

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).

Steckel, Richard, "Stature and the standard of living", *Journal of Economic Literature*, vol. 33 (1995), pp. 1903-1940.

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

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

Tollnek, Franziska and Jörg Baten, "Farmer Families at the Heart of the 'Human Capital Revolution'? Decomposing the Numeracy Increase in Early Modern Europe," *Economic History Review,* vol. 70, no. 3 (August 2017), pp. 779-809.

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

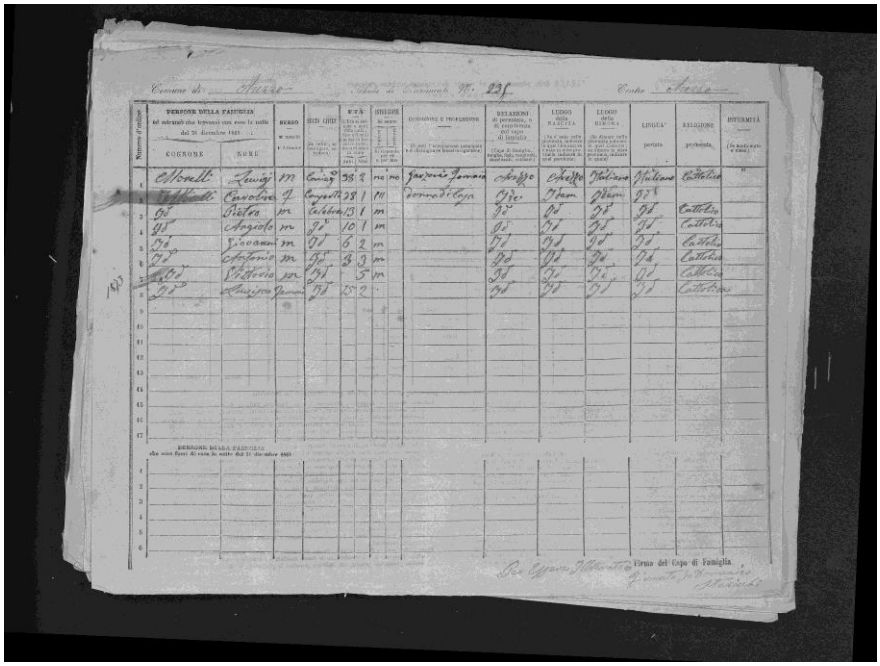
Zamagni, Vera. *Dalla periferia al centro. La seconda rinascita economica dell'Italia (1861-1990),* 2e. Bologna: Il Mulino, 1993.

Zanden, Jan Luiten van, "The skill premium and the 'Great Divergence'," *European Review of Economic History*, vol. 13 (2009), pp. 121-53

Appendix A.

The figure below reproduces a household return of the 1861 Census for the Comune of Arezzo. Source: http://www.genealogiadavini.it/genealogia/enrico/doc/arezzo/1861%20censimento%20arezzo%20s_agostino%20morelli%20luigi.jpg, accessed on 20 October, 2019.

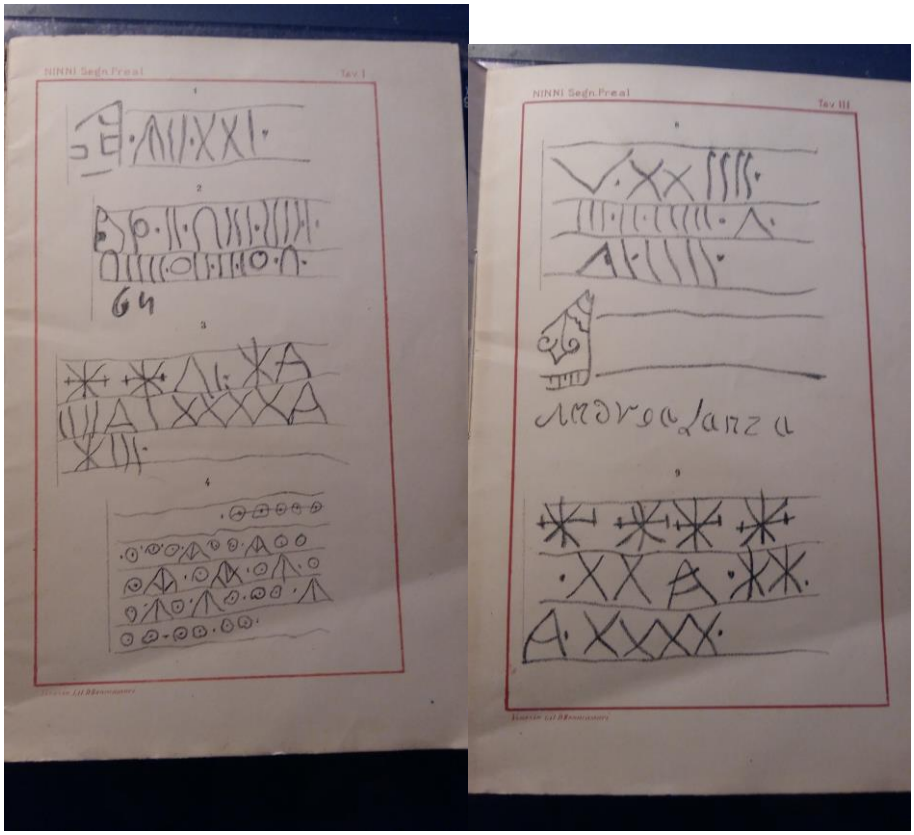
Figure A1: Household return of the 1861 Census



Appendix B

The figures below are excerpts from Ninni (1889). They were retrieved from the actual log books of Chioggia fishing boats. Each line represents the sum due an individual fisherman. The trapezoid at the beginning of each row is a sort insignia adopted by each fishermen to identify himself. The following marks are the actual numbers indicating sums due. In some cases, the owner of the boat also used Arabic numbers to indicate a total (e.g. 64 at the end of the first entry). The dots are used as a separation sign. Some symbols are identical to Roman numerals, e.g. "X" indicates the number 10. According to Ninni (1889, p. 11), the system was still very popular when he was writing because "it had the advantage of being well-known both by the literate owner of the boat and by his crew, the largest share of whom was illiterate."

Figure B1: Log books of Chioggia fishing boats



Source: Ninni (1889, tab.1 and tab.3)