The accuracy and precision of birthplace reporting in the 1851–1911 censuses: Place as a component of identity in nineteenth-century England and Wales

Joseph Day

School of Geographical Sciences, University of Bristol, Bristol, UK

Correspondence
Dr Joseph Day, School of Geographical Sciences, University of Bristol, Bristol, UK. Email: joe.day@bristol.ac.uk

Funding information
Economic and Social Research Council, Grant/Award Number: ES/R005443/2

Abstract
If migration between 1851 and 1911 is to be analysed using individual-level census data, it must be proven that individuals’ place of birth was reliably recorded. However, existing studies have primarily been concerned with establishing only the level of absolute error, rather than how such errors were introduced. Through assessing the determinants of inaccuracy in the 1851–1911 censuses, this article finds that while birthplaces were generally reported accurately, the errors which did exist were not only the fault of the individual in question, but also of the householder whose responsibility it was to fill in the census form, and of those charged with transcribing the original, handwritten schedules into a machine-readable format. By demonstrating that individuals’ ability to identify their place of birth was a function of their age and the distance they migrated from their place of birth, rather than either their intelligence or familiarity with their place of birth per se, shows that individuals’ ability to identify their birthplace accurately was a function of the relevance it held in their lives. Consequently, this article argues that individuals’ ability to identify their place of birth reflected its role as a component of their identity, becoming less relevant the further from it they became, but increasingly relevant as they aged.

KEYWORDS
birthplace, census, England, identity, nineteenth century, Wales

1 | INTRODUCTION
As a source for policymakers, demographers, sociologists, economists, geographers, and even historians of England and Wales, the census is without rival. Only global war led to its cancellation in 1941, while the 2021 census went ahead in spite of the COVID-19 pandemic. Given its pre-eminence, it is easy to forget that the census relies on millions of people making a truthful return. While it has always been an offence to knowingly provide a false return—a 44-year-old schoolteacher was fined £1 in 1841 for stating her age as 20 (Christian & Annal, 2014)—most of the information required by the census can only be verified by cross-checking it against independent sources. However, not only was this deemed unfeasible in 2021—let alone 1851—but the census would be obsolete if such data could be acquired by other means (Select Committee on Public Administration, 2014). Consequently, its reliance on the integrity of both the public and those charged with collecting the data has led many historians to view the census with suspicion, with much criticism targeted at both the reliability of female occupations (Higgs, 1987, 1995, 2007; Higgs & Wilkinson, 2016) and the accuracy of the reporting of birthplace.
of birthplaces (Hinde, 2004). While the accusation that the census did not accurately record female occupations has been comprehensively rebutted (You, 2020), the accuracy of birthplace reporting remains an open question. This article will therefore assess the determinants of error in birthplace recording and, in so doing, analyse what such errors might reveal about individuals’ birthplace as a component of their identity.

There have been five principal studies assessing the accuracy of birthplace reporting in the nineteenth-century and early twentieth-century censuses, and they have all used the same record-linked approach (Anderson, 1972; Perkyns, 1991; Tillott, 1972; Wrigley, 1975; Yasumoto, 1985). By matching individuals to either consecutive censuses or baptism registers, these studies interpreted birthplaces that had been inconsistently recorded across the sources as evidence of error. Despite some regional variations, all four studies determined that birthplaces were generally reported accurately. By Perkyns’ (1991) calculations, between just 2.74% and 5.35%—depending on how strictly a given birthplace is defined as “correct”—provided an inaccurate place of birth.

While these findings are reassuring, they are nonetheless problematic. As Armstrong (1978) observes, studies in which individuals’ birthplaces have been matched across censuses can be more truthfully described as measuring the consistency with which birthplaces were reported, rather than its intrinsic accuracy, as records are not matched to independent accounts of birthplaces such as baptism registers. However, matching individuals to baptism registers is fraught with the potential for making an incorrect match. Eriksson (2016) demonstrated that linking records from sampled data significantly increase the risk of a match being a false positive. Also, as birthplaces are important pieces of evidence in the record-linkage puzzle, if the birthplace in a baptism register does not match that in a census entry, the more conservative assumption would be that the two records refer to two different people, rather than a single individual who mis-specified their birthplace in the census (Newton & Bennett, 2020). Lastly, these studies do not assess the cause of errors.

This paper will therefore analyse the cause of errors by working backwards through the stages by which birthplace data were produced, as illustrated in Figure 1. By accounting for errors produced by those “processing” the data—from the householder to the researcher—errors produced by the individual can be more confidently inferred. Once Section 2 has outlined the data used and means by which a birthplace is defined as “incorrect,” Section 2.1 examines the errors introduced at each stage as detailed in Figure 1, starting with errors introduced by the algorithm when coding birthplaces. Next, errors introduced during transcription for the I-CeM project (Schürer & Higgs, 2020) are considered, before analysing the role of the census enumerator who copied the individual census forms into the census enumerators’ books (CEBs). Finally, the data are scrutinised for possible errors introduced by the householder, whose responsibility it was to collect accurate responses for all household members. Section 2.1 finds that while both the algorithm and the census enumerator are relatively blameless and responsible for no structural errors, there were several notable transcription errors, likely a result of similarities between county names (Herefordshire/Hertfordshire, HANTS/HUNTS, etc.). Similarly, the likely indolence of some householders caused errors by failing to enquire of household members’ county of birth directly. This leaves only the errors caused by individuals themselves to be considered in Section 2.2.

Section 2.2.1 demonstrates that while few errors can be ascribed to individuals’ level of education—as proxied by the skill implied by their occupation—both their age and the distance they migrated proved significant determinants. As individuals aged, they became increasingly likely to identify their county of birth correctly, but less able to do so the farther they had migrated from their birthplace. This observation is repeated in Section 2.2.2 which shows that individuals became less precise (London rather than Paddington for example) as well as less accurate, when describing their birthplace, the further they migrated.

Recent research has also demonstrated that individuals’ identity with, attachment to, and memory of a place, is largely determined by both their age and how long they have been resident in that place (Benson & Jackson, 2013; Degnen, 2016; Lewicka, 2008). As this suggests that individuals’ ability to identify their place of birth is a function of their place identity and place attachment, it is consequently argued here that individuals’ ability to identify their birthplace—somewhere one would expect them to be familiar with—can be taken

---

**Figure 1** Flow diagram representing the transmission of birthplace data from the 1851–1911 censuses. Source: author’s illustration
as a proxy for the extent to which individuals identify with, have an attachment to, and a memory of, that place. However, whereas such analyses use interviews and surveys to assess the importance of place to individuals (Peng et al., 2020), this article shows that place identity and attachment can be inferred simply from how a place is described, suggesting that a far wider range of evidence might be utilised to investigate the relationship between identity and place.

2 | DATA AND METHODS

In 1911, the chief clerk argued against the continued collection of birthplaces, since “a great many people did not know in which county they were born” (Committee on the Census of 1911). Figure 2 shows that the chief clerk was broadly correct in this assessment and that people were two to eight times more likely to fail to give a county, rather than parish of birth between 1851 and 1911. Consequently, where an individuals’ stated parish does not exist in their stated county of birth, it is assumed that it is the county which had been incorrectly stated.2

Figure 3 therefore quantifies the extent of such errors by showing the stacked proportions of the population that did not state either their county of birth or one that was either adjacent or non-adjacent to the county in which their stated parish of birth was located. There are a few features worth noting. First, there is a clear temporal pattern. The error rate goes up—on all measures—between 1851 and 1861, declines until 1901, and then rises slightly in 1911. This suggests that migration may have been a determinant of individuals’ ability to accurately recall their place of birth—as both the numbers migrating and distances migrated increased rapidly between 1851 and 1861. As the population became increasingly familiar with a larger geographical area, their accuracy improved, but in 1911—when enumerators no longer copied the household schedules into census enumerators’ books (CEBs), but the household schedules were instead sent directly to central offices for processing—the error rate ticked up. While the increase in the non-response rate could be attributed to transcription error—more sets of handwriting likely made the schedules more illegible than the CEBs—the increase in the numbers mistaking their county for one adjacent to it, one suggests that the errors made in 1911 would previously have been caught—and corrected—by enumerators.

The technique for matching individuals to a birthplace utilises a custom-built algorithm designed to matches birthplaces to as precise a location as possible in a geographic information system (GIS).3 As an “incorrect” county is defined as conservatively as possible, the numbers reported as having made an error is inevitably an underestimate.4

What follows are a few examples of how a birthplace may be allocated to a county other than that stated and, therefore, how a stated county is determined as being “correct” or not.

Birthplaces were transcribed—from the original manuscript returns—in three parts: [PARISH/TOWN][COUNTY][COUNTRY]. If any of this information was not provided, the relevant section was marked [BLANK]. Using an exhaustive gazetteer of placenames (with spelling variations), the algorithm sought to match the stated [PARISH/TOWN] to the gazetteer placename which had both the most matching words and characters and the fewest non-matching word

---

**Figure 2** Proportion of population that failed to provide either a county of birth or a parish of birth in their census return. England and Wales 1851–1911. Source: author’s analysis based on data from UK Data Service SN 7481 (Schüer & Higgs, 2020)
and characters (Schürer, Penkova, & Shi, 2015). In the first parse, the algorithm matches to the shortest gazetteer placename that fulfils these conditions. Consequently, S PANCRAS|LONDON| [BLANK] matches to both Pancras-Devon and St Pancras-London on seven characters, “PANCRAS,” with one redundant character, “S.” As “Pancras” is shorter than “St Pancras,” the first parse of the algorithm matches the string to Pancras-Devon. In pre-processing, all strings were matched to a standardised version of the county as stated in the birthplace string. As the “stated” county for the string S PANCRAS|LONDON| [BLANK] was London, a second parse searches for the best match closer to London than Devon, in both London and counties adjacent to it. The requirement to match to the shortest gazetteer placename is dropped for the second parse, so the match is reallocated to St Pancras-London. The county field is therefore interpreted as being “correct.” However, the birthplace string ABBOTS LANGLEY|STAFFORDSHIRE| [BLANK] is matched to Abbots Langley-Hertfordshire rather than Abbots Bromley-Staffordshire. Even though Abbots Bromley is the best match either in or adjacent to Staffordshire, it matches just six characters with seven redundant characters whereas Abbots Langley-Hertfordshire matches 13 characters with zero redundant characters. The county field is therefore interpreted as incorrect.

2.1 | Sources of inaccurate birthplace reporting: Errors in processing stages

2.1.1 | Algorithmic error

As the algorithm codes the birthplace string, any egregious errors caused by the algorithm will show up first as a birthplace string being incorrectly matched to a county other than that stated. Across all the censuses between 1851 and 1911—except for the 1871 census, of which a digitised version is not currently available—152,710,566 individual records yielded 5,294,498 unique birthplace strings. 75.2% of strings—89.7% of the population—were matched to the county as stated. Of the remaining birthplace strings, 4.2% were matched to adjacent counties—2.4% of the population—and 5.0%—0.5% of the population—were matched to non-adjacent counties; 15.6% of strings—7.3% of the population—provided no county of birth.

The average distance migrated was analysed by birthplace strings, where the county had been reassigned by the algorithm to one which was neither in nor adjacent to the county stated. The 547 birthplace strings with frequencies over 100—accounting for 13.1% of the population whose birthplaces had been reassigned by the algorithm—were checked, as implausibly high average distances would be suggestive of an incorrect match. Indeed, if the majority of those whose birthplace was matched to Stoke-on-Trent were resident in Stoke Newington, this could best be interpreted as an algorithmic error.

While the majority appear to have been correctly reassigned by the algorithm, the only egregious algorithmic error was to assign the 2,280 birthplace strings of CLIPSTONE|NORTHAMPTONSHIRE| [BLANK] and CLIPSTONE|NORTHAMPTON|ENGLAND to Clipstone-Nottinghamshire rather than Clipston-Northamptonshire. While these were corrected manually, it is illustrative of how even minor spelling errors—such as those which can occur during transcription—can radically distort meaning.

2.1.2 | Transcription error

Analysing transcription error, the census manuscripts were transcribed by a diverse group including local history volunteers, prisoners, professional transcription services in India and Sri Lanka, and followers of...
the Church of Jesus Christ of Latter-Day Saints. Whereas local historians may be more tempted to use their knowledge to make an educated guess of an otherwise illegible entry, overseers transcribing unfamiliar with the source material may be more likely to misread unfamiliar words or handwriting. Analysis of the 1881 census transcribed by the Genealogical Society of Utah (GSU) suggests, for example, that the marital status of females under 16 recorded as married was routinely changed to “unmarried” (Schürer & Woolard, 2002; Woolard, 1999). While prisoners might have deliberately transcribed “prison warder” or “gaoler” as “screw” (WalesOnline, 2003), slight mis-transcriptions—even by a single letter—for example, HANTS/HUNTS, HEREFORDSHIRE/HEREFORDSHIRE, can change a word completely (Goose, 2001).

Therefore, examining transcription error more fully, Table 1 shows the likelihood that individuals stated a different county to that in which their parish of birth was located. In Table 1, it appears that most instances of this were in those counties where their names or abbreviations were similar to others: Northamptonshire/Northumberland, Cambridgeshire/Cumberland (CMB/CUMB), Herefordshire/Hertfordshire, Hampshire/Huntingdonshire (HANTS/HUNTS). Consequently, transcription errors are concentrated in these counties. Indeed, even if one were to assume that all those errors with a likelihood index greater than 2.5 were transcription errors, this would account for just 1.5% of all errors in 1851, 1.7% in 1861, 1.2% in 1881 and 1891, 0.5% in 1901, and 0.4% in 1911. However, such errors were concentrated in those counties with similar-sounding names to others, identified in Table 1 as those with a likelihood index greater than 2.5. In these counties, transcription errors accounted for 10.9% of the total. While it might be tempting to dismiss these errors, given that they were not only detected and removed, but also relatively few, such errors were significant nationally. Instead, like evidence of other errors presented here, it is not only indicative of how errors were produced and the proportion of all errors they account for, but these types of error can then be discounted when analysing errors produced further down the data creation “chain” illustrated by Figure 1.

2.1.3 | Enumerator error

While Tillott (1972) was quick to rule out enumerators as a source of error, Perkyns (1991) noted that some parishes had twice the error rate of neighbouring one, which Higgs (2005) interpreted as evidence that some enumerators were less conscientious than others, concluding that enumerators were a major source of error. Indeed, both Anderson (1972) and Higgs (2005) point to the multiplicity of spelling variations for Cornish and Welsh places and the difficulty of spelling unfamiliar place names correctly when related on the doorstep. Therefore, as enumerators had the power to affect the accuracy of the return, it is worth investigating whether they did so.

Figure 3 has already shown that the proportion of the population of England and Wales who stated a county that was either adjacent or not adjacent to the one in which their stated parish of birth was located, increased between 1901 and 1911. The only major difference between the census in 1901 and 1911, however, was in how the data were processed. As the 1911 census was the first in which household schedules were sent directly to the Census Office for processing without requiring enumerators to first copy them into CEBs, enumerators had less opportunity to correct any errors than before (Campbell-Kelly, 1996). That the absence of enumerators in 1911 increased the error rate suggests that enumerators had a positive effect on the accuracy of the return (Tillott, 1972). To test this, it is necessary to assess the two most plausible means by which enumerators might have introduced error: unfamiliarity with the birthplace provided and careless recording.

It stands to reason that the smaller, and further away an individual’s place of birth was, the less likely it is that an enumerator would be familiar with it. Figure 4 groups the population in each census year into 10 equally-sized groups, in ascending order of the size of their self-reported settlement of birth. For example, individuals that gave their settlement of birth as “London” were in the top 10%, whereas individuals that gave their birthplace as “Camden Town”—with a population of just 12,730—were grouped between the fifth and sixth decile. However, an individual’s distance from their birthplace may have affected their own ability to recall it—as well as the enumerator’s ability to identify it—Figure 4 is standardised for both age and distance migrated. If enumerators’ unfamiliarity with an individual’s place of birth was a factor influencing the accuracy with which their county of birth was stated, one would expect the county of birth of those born in smaller settlements would be less accurately recorded than of those born in larger settlements.

Between 1851 and 1901, those born in progressively larger settlements became less likely to provide a county of birth at all. While this is perhaps understandable given that the familiarity of places such as London and Bristol would have precluded the perceived need to specify which county they were in, between 1851 and 1881, individuals also become progressively less likely to state their county of birth correctly as the stated size of their settlement of origin decreased. While this superficially suggests that enumerators—who would understandably have been less familiar with smaller places—were a source of error, one would also expect the error rate to have been negatively correlated with the size of individuals’ stated settlement of birth up until the 1901 census; the last census in which enumerators transferred—and could amend—the household schedules as they were copied into the CEBs. Instead, individuals were more likely to state their county of birth incorrectly the larger their settlement of birth was, between 1891 and 1911. That this trend continued in 1911 when enumerators no longer copied household schedules into the CEBs suggests that the renown of individuals’ birthplace was not the reason that individuals from smaller places were more likely to identify their county of birth incorrectly between 1851 and 1881. Instead, a more plausible interpretation would be that enumerators added a county of birth where they were confident of its accuracy and did not do so where they were not. Indeed, the instructions to enumerators in both 1891 and 1901 clearly directed that they “Copy from the Schedule into the other columns. All the particulars concerning the persons
### Table 1: Relative likelihood of incorrectly stating any one county of birth given the county in which stated parish of birth is observed to have been located

| Observed County of birth | Cambridgeshire | Cardiganshire | Carmarthenshire | Derbyshire | Flintshire | Hampshire | Herefordshire | Hertfordshire | Huntingdonshire |
|--------------------------|----------------|---------------|-----------------|------------|-----------|-----------|--------------|--------------|-----------------|
| Anglesey                 | 0.37           | 0.10          | 0.37            | 0.07       | 0.07      | 0.02      | 0.01         | 0.05         | 0.03            |
| Brecknockshire           | 0.28           | 0.46          | 0.21            | 0.07       | 0.18      | 0.05      | 1.77         | 1.20         | 0.18            |
| Caernarvonshire          | 0.56           | 0.77          | 1.89            | 0.21       | 0.04      | 0.03      | 0.02         | 0.06         | 0.02            |
| Cardiganshire            | 1.10           | 1.49          | 0.06            | 0.19       | 0.02      | 0.06      | 0.17         | 0.21         | 0.21            |
| Carmarthenshire          | 2.94           | 0.12          | 0.09            | 0.07       | 0.05      | 0.07      | 0.12         | 0.04         |                 |
| Cumberland               | 3.14           | 0.09          | 0.31            | 0.09       | 0.11      | 0.12      | 0.17         | 0.21         | 0.17            |
| Denbighshire             | 0.29           | 0.23          | 0.96            | 1.20       | 1.31      | 0.08      | 0.07         | 0.25         | 0.30            |
| Hampshire                | 0.16           | 0.07          | 0.11            | 0.12       | 0.29      | 0.23      | 0.48         | 1.69         |                 |
| Herefordshire            | 0.22           | 0.10          | 0.10            | 0.10       | 0.09      | 0.26      | 3.50         | 0.23         |                 |
| Hertfordshire            | 1.84           | 0.17          | 0.22            | 0.07       | 0.15      | 0.32      | 10.39        | 0.13         |                 |
| Huntingdonshire          | 0.88           | 0.18          | 0.12            | 0.04       | 0.80      | 3.85      | 0.51         | 0.22         |                 |
| Merionethshire           | 0.52           | 0.15          | 0.85            | 1.44       | 0.07      | 0.04      | 0.11         | 0.13         | 0.25            |
| Northumberland           | 0.40           | 0.15          | 0.12            | 0.22       | 1.11      | 0.15      | 0.50         | 0.40         | 0.61            |
| Northamptonshire         | 1.00           | 2.22          | 1.22            | 0.06       | 0.69      | 0.38      | 0.19         | 0.13         | 1.22            |
| Pembrokeshire            | 0.59           | 0.10          | 1.16            | 0.05       | 0.20      | 0.03      | 0.21         | 0.14         | 1.04            |
| Radnorshire              | 0.60           | 0.72          | 0.69            | 0.07       | 0.11      | 0.30      | 0.88         | 0.80         | 1.17            |
| Rutland                  | 0.05           | 0.00          | 0.13            | 0.03       | 0.38      | 0.35      | 0.34         | 0.10         | 0.04            |
| Shropshire               | 0.30           | 0.06          | 0.12            | 0.09       | 1.78      | 0.15      | 0.64         | 0.31         | 0.69            |
| Suffolk                  | 0.34           | 0.08          | 0.22            | 0.16       | 0.23      | 0.06      | 0.29         | 0.06         | 0.25            |
| Somerset                 | 0.25           | 0.10          | 0.05            | 0.12       | 0.20      | 0.05      | 0.25         | 0.14         | 1.58            |
| Wiltshire                | 0.24           | 0.04          | 0.08            | 0.08       | 0.24      | 0.20      | 0.19         | 0.17         | 0.34            |

Note: England & Wales, 1851–1911. The proportion born in the “observed” county of birth that migrated to the “stated” county of birth is taken as a proxy of the likelihood of stating it as their birth county, assuming that the likelihood which individuals had of stating any given county of birth was subject to the same distance-decay function as migration. For example, 0.06% of those born in Hertfordshire migrated to Herefordshire, so one ought to expect 0.06% of those born in Hertfordshire to give Herefordshire as their county of birth, but 0.65% did; 0.65 / 0.06 = 10.39, making individuals born in Hertfordshire 10.39 times as likely to state Herefordshire than predicted. As it is impractical to show the full 55 x 55 table, this table shows only those rows and columns where the relative likelihood of stating any one county of birth given the observed county of birth is greater than 1. The likelihoods of stating one’s actual county of birth are excluded. The likelihood of stating any given county is colour coded from red (<1 and less likely than expected) to yellow (=1 and as likely as expected) to green (>1 and more likely than expected).
TABLE 1 (Continued)

| Stated County of birth | Merionethshire | Monmouthshire | Montgomeryshire | Northamptonshire | Oxfordshire | Radnorshire | Rutland | Worcestershire |
|------------------------|----------------|---------------|-----------------|------------------|-------------|-------------|---------|----------------|
| Observed County of birth | 0.04 | 4.77 | 0.60 | 0.01 | 0.00 | 0.35 | 0.06 | 0.03 |
| 0.34 | 0.12 | 0.49 | 0.07 | 0.82 | 1.62 | 0.19 | 0.04 |
| 0.06 | 1.00 | 0.09 | 0.04 | 0.36 | 0.22 | 0.10 | 0.02 |
| 0.06 | 0.08 | 0.70 | 0.04 | 0.06 | 0.11 | 0.10 | 0.02 |
| 0.23 | 0.03 | 0.20 | 0.02 | 0.10 | 0.16 | 0.38 | 0.03 |
| 0.03 | 0.08 | 0.73 | 0.22 | 0.27 | 0.27 | 0.06 |
| 0.03 | 0.09 | 0.64 | 0.07 | 0.25 | 0.09 | 0.42 | 0.06 |
| 0.07 | 0.23 | 0.10 | 0.08 | 0.10 | 0.14 | 0.12 |
| 0.09 | 0.18 | 0.11 | 0.10 | 0.25 | 0.32 | 0.19 | 0.28 |
| 0.07 | 0.08 | 0.20 | 0.10 | 0.13 | 0.14 | 0.07 | 0.15 |
| 0.41 | 0.07 | 0.24 | 0.10 | 0.13 | 0.36 | 0.06 | 0.05 |
| 0.78 | 0.08 | 0.14 | 0.02 | 0.27 | 0.00 | 0.00 | 0.06 |
| 0.15 | 0.08 | 0.62 | 3.29 | 0.36 | 0.26 | 0.52 | 0.11 |
| 1.37 | 0.47 | 1.51 | 0.30 | 1.47 | 0.10 | 0.04 |
| 0.23 | 0.04 | 0.16 | 0.13 | 0.10 | 0.23 | 0.27 | 0.14 |
| 0.27 | 0.10 | 0.10 | 0.12 | 0.12 | 0.09 | 0.50 | 0.09 |
| 0.12 | 0.07 | 0.28 | 1.01 | 0.10 | 0.38 | 2.33 | 0.06 |
| 0.08 | 0.16 | 0.24 | 0.12 | 0.24 | 0.10 | 0.10 | 0.06 |
| 0.08 | 0.05 | 0.27 | 0.13 | 1.19 | 0.35 | 0.43 | 0.15 |
| 0.11 | 0.01 | 0.45 | 0.16 | 0.24 | 0.10 | 0.51 | 0.10 |
| 0.13 | 0.09 | 0.26 | 0.15 | 0.15 | 0.18 | 0.33 | 1.10 |

Note: England & Wales, 1851–1911. The proportion born in the “observed” county of birth that migrated to the “stated” county of birth is taken as a proxy of the likelihood of stating it as their birth county, assuming that the likelihood which individuals had of stating any given county of birth was subject to the same distance-decay function as migration. For example, 0.06% of those born in Hertfordshire migrated to Herefordshire, so one ought to expect 0.06% of those born in Hertfordshire to give Herefordshire as their county of birth, but 0.65% did; 0.65 ÷ 0.06 = 10.39, making individuals born in Hertfordshire 10.39 times as likely to state Herefordshire than predicted. As it is impractical to show the full 55 × 55 table, this table shows only those rows and columns where the relative likelihood of stating any one county of birth given the observed county of birth is greater than 1. The likelihoods of stating one’s actual county of birth are excluded. The likelihood of stating any given county is colour coded from red (<1 and less likely than expected) to yellow (=1 and as likely as expected) to green (>1 and more likely than expected).

Source: Author’s analysis based on data from UK Data Service SN 7481 (Schürer & Higgs, 2020).
returned therein, without any omission, or any abbreviation except those authorised ...” (1901 Census Returns). Although only anecdotal, the following account is at least suggestive that the average enumerator would have only added or amended information that was demonstrably incorrect, even if the response seemed implausible:

... a lady ... wrote in the column for trade or occupation “living on my own means.” The official unwisely asked for an interpretation of these words, considering that she was living in such a [poor] quarter. The lady ... informed him that she was not going to tell “whys or wherefores; that was her business ...” Now, what could the census man do with such a person; he shrugged his shoulders and went to the next house (The Derbyshire Times, 1901).

Therefore, if enumerators do not appear to have made errors accidentally, any errors they did make must have been deliberate. For example, both a conscientious enumerator—knowing the importance of the county of birth to the Census Office—who filled in blank fields with the most likely county of birth, and an inattentive enumerator—who inappropriately overused the “ditto” mark—would both be at risk of making the same error: incorrectly stating the county of residence as the county of birth. Given that most of the population remained in their county of birth, this would be both enumerators’ best guess and the most likely interpretation of an errant “ditto” mark. Also, as Higgs (2005) intimated—one would expect some enumerators to be more careless than others, and the error rate to be randomly distributed. Unfortunately however, as a GIS of EDs does not exist, it is not possible to measure this using standard measures of spatial concentration/dispersion.  

Assuming the error that enumerators were most likely to make was to erroneously state the county of enumeration as the county of birth, Figure 5 estimates the concentration of such errors in EDs by measuring the likelihood that any two individuals—whose counties of residence were incorrectly stated as their counties of birth—were located in the same ED, in what might be termed a “concentration” index. While its calculation is fully described in the footnote, a concentration index of 1 means that errors in that RSD are perfectly distributed across all EDs. Although a concentration index of around 2.5 consequently means that errors were 2.5 times as concentrated in EDs as might be supposed if the errors in an RSD were perfectly distributed between its EDs, this is approximately what might be expected if the errors were randomly—rather than perfectly—dispersed. That the concentration indices are less than 3 in much of England and Wales in both 1851 and 1861 is therefore reassuring. Enumerator error does not appear to have been a widespread cause of error. This is especially so in 1861. In 1851, however, there are several notable instances in which the concentration index was extremely high: in much of Oxfordshire, Yorkshire, and parts of Devon. 

However, the concentration index is a relative measure. It simply shows the extent to which errors in each RSD—whether there were many or few—tended to be made in the same EDs. If only one error had been made in an RSD made up of 10 EDs, the error would be
concentrated” in a single ED, producing a concentration index of 1,000. But this would obviously not be evidence that enumerators were a major source of error. Rather, only where an RSD displays both a high concentration index and a high error rate might this be taken as evidence of significant enumerator error, as the high number of errors could be attributed to a single enumerator. However, correlating the concentration index with the likelihood that where a county of birth was mis-specified, it was the county of residence that was given, not only produces an extremely low $R^2$ of 0.047 and 0.039 in 1851 and 1861, respectively, but also statistically insignificant, negative coefficients of $-0.58$ and $-0.79$. This suggests that enumerators were unlikely to have been a major source of errors, and instead, the uptick in the number of errors between the 1901 and 1911 censuses was likely because enumerators were no longer copying household schedules into the CEBs and correctly adding missing counties of birth along the way in 1911 as they had done in previous censuses. Consequently, it might be more plausible to interpret variations in the error rate between neighbouring parishes as a consequence of rate instability caused by a small sample size and random variation (Casella, 1985; Higgs, 2005; Perkins, 1991), rather than the fault of enumerators, who appear to have been largely blameless (Devine et al., 1994).

2.1.4 | Householder error

Finally, before errors which were introduced by individuals can be analysed, errors which were the fault of the householder must be accounted for. The householder was responsible for returning an accurate schedule and was indeed required to make a signed declaration to that effect. Although one would hope that this guarded against deliberate attempts to mislead, it is still possible that householders introduced errors unintentionally, by either failing to ask household members for the information directly and relying on their own knowledge or “correcting” the response of a household member. Indeed, if the householder had no adverse effect on the accuracy of others’ response, one would expect the likelihood of household members providing an incorrect county of birth to be independent of the householders’ own likelihood of doing so. However, as members of the same household are likely to have been born near one another—on a county boundary for example—they are also likely to have made the same error coincidentally.

Figure 6 therefore analyses the likelihood that household members stated their birthplace correctly where the householder had not, by the distance the household member had been born from the household head—whose county of birth was incorrectly stated—whoes county of birth was incorrectly stated—were less likely to state their own county of birth correctly than the general population, this could most charitably be interpreted as a coincidental error. If, however, household members born many miles away from a household whose stated county of birth was incorrect were still less likely to state their own county of birth correctly than the general population, this would suggest the householder was to blame.

Figure 6 does indeed show that household members’ county of birth was far less likely to be stated correctly where they were born
less than 5 km from a household head whose county of birth was also incorrectly stated. It is, however, no less noteworthy that household members born over ~30 km away from the household head were no less likely to state their place of birth correctly than anyone else. This suggests that much of the error was a function of proximity. However, it does not show whether household members born near one another made coincidental errors independently, or whether householders simply did not consult household members when filling in the return.

One way to test for this is to analyse the likelihood that individuals provided the same county of birth as the householder, regardless of the distances between their respective birthplaces. Indeed, if a householder and household member born 100 km apart provided the same county of birth, it is less plausible to interpret this as a coincidental error. Figure 7 therefore shows the proportion of household members that incorrectly stated their county of birth as being that of the household head—whose birthplace was also incorrectly stated—divided by the proportion of all household members that gave a county of birth that was the same as the household head, to produce a relative likelihood. A number greater than 1 show that household members were more likely to state the same county of birth as the householder than the population at large. If the errors made by householders and household members were solely a function of being born near one another and making the same coincidental error independently, one would expect household members to become less likely to incorrectly state the same county of birth as the household head, the farther apart the two were born. Figure 7 shows that instead, between 1851 and 1881, the farther an individual was born from the household head, the more likely they became to state the same—incorrect—county of birth as the household head. This suggests that such errors were not coincidental but, rather, householders introduced errors by erroneously stating their own county of birth as being that of other household members.

However, Figure 7 also shows that between 1891 and 1911, individuals that misstated their county of birth were not significantly more/less likely to give the householders' county of birth instead. While the precise reasons for this change are unclear, as Figure 3 showed that the proportion stating their birthplace incorrectly or not at all declined over time, it is possible that as individuals' geographical knowledge improved over time—perhaps due to a growing railway and postal network—less guesswork was needed when providing someone else's birthplace. Although Figure 7 shows that householders did indeed introduce errors—rather than the householder and a household member simply making them coincidentally—it is not clear whether this was the product of a genuine belief they had both been born in the same county, or of householders' lazily putting the first county that came to mind.

This can be tested for in Figure 8 by analysing the proportion of household members that gave a county of birth when the household head had failed to do so. If household members' county of birth was less likely to be recorded simply because the household head had not done so, this could be taken as evidence of slovenly reporting. This is divided by the total population that provided no county of birth to produce a relative likelihood. A number less than 1 show that household members were less likely to provide a county of birth than the population at large.

Two observations are worth noting. First, there was a positive gradient between the distance household members were born from householders and their likelihood of reporting a county of birth, with those born nearest the household head being the least likely to provide a county of birth. While this could be most charitably interpreted like Figure 6, as coincidental error, this seems unlikely considering the second observation. Figure 8 also shows that even when the two birthplaces were 100 km apart, household members were still ~10% less likely to report a county of birth than the population at large. This
strongly suggests that householders created errors by either guessing individuals’ details, or leaving fields blank, rather than taking the time to verify their particulars with the individual in question. Clearly, an individuals’ stated county of birth cannot be taken as a straightforward reflection of their own conceptualisation of place, as it relied so heavily on the householder both asking the individual directly, and then recording it faithfully.

The individual-level census data to which we have access today are therefore the product of many hands, and although there is scope for error to have been introduced at each of the stages outlined in Figure 1, both the coding algorithm and census enumerators appear to have been relatively blameless. Rather, the majority of errors that cannot be attributed to the individual themselves were likely the result of keystroke errors by transcribers or indolent householders. However, while the proportion of errors that can be attributed to transcribers is measurable, it is also relatively slight—approximately just 0.8% of erroneous matches were to similarly-spelt counties—whereas the errors produced by householders cannot be quite so directly inferred. Table S1 incorporates householder effects into a binary logistic regression modelling the determinants of one’s county of birth being correctly specified. This shows that there was just a 1 in 100 chance of a household member stating their county of birth
correctly where the householder had not, relative to the likelihood of doing so had the householder stated their own county of birth correctly. While these observations demonstrate that the census was a product of the process by which the data were collected and recorded, they also allow errors caused by the individual to be inferred more convincingly from the data. Such factors are explored the next section.

2.2 | Sources of inaccurate birthplace reporting

2.2.1 | Errors produced by the individual

It might be argued that individuals' ability to locate their place of birth is dependent on several factors: their geographical knowledge and capacity to place a location within a broader geographical context; the distance from their place of birth or the time since they left it and crucially; and their emotional connection to it. Indeed, it has been amply demonstrated that the more significant a place is/was to an individual's sense of identity, the more likely the memory of it is to be retained (Baddeley, 1982; Kleinsmith & Kaplan, 1963; LaBar & Phelps, 1998). As a proxy of these factors then, this section examines the effect of occupation—as a proxy of skill and the level of education and knowledge/intelligence it implies—as well as individuals' ages and the distances they migrated, as proxies for their familiarity and emotional connection to a place. While occupation is a fairly crude means of estimating individuals' knowledge and/or intelligence, it is nonetheless true that occupations which required a university-level education, or which necessitated travel—for example, train driver—might also require a better grasp of geography (Wrigley, 2010). However, occupation only appears to have influenced the type of error made, rather than likelihood of making one.

Figure 9 shows the sectoral likelihood—relative to the national occupied population—of specifying one's county of birth either incorrectly, or not at all. It is striking that primary sector workers became less likely to give an incorrect county over time, but increasingly likely to give no county at all. Tertiary sector workers on the other hand became more likely to give an incorrect county and less likely to give no county than before. As most occupations in the primary sector were unskilled, blue-collar jobs, while occupations in the tertiary sector tended to be white collar and middle-class, there may have been a social gradient to the likelihood of specifying either an incorrect county of birth, or none at all. Figure 10 therefore shows the likelihood of either specifying one's county of birth incorrectly or not at all, relative to their HiS-CAM score. HiS-CAM scores measure social interaction and as such are used to stratify occupations from the highest to the lowest social class (Lambert et al., 2013; Prandy & Bottero, 2000; Stewart et al., 1973, 1980; van Leeuwen et al., 2002). As HiS-CAM is measured on a continuous scale, it is a useful means to quickly identify the relationship between skilled/unskilled occupations and migration. The original scales were constructed from the social interactions derived from marriage registers in Belgium, Britain, Canada, France, Germany, the Netherlands, and Sweden; the “social distance” between occupations being estimated from the number of butchers that married the daughters of bakers, for example. These social distances were then transformed and

![Figure 9](image-url)
standardised onto a hierarchical scale between 0 and 100, where 50 was the mean.

Figure 10 clarifies the observation made in Figure 9. While the relatively unskilled—those with a HiS-CAM score of less than 50—were generally no more or less likely than the population at large to state their county of birth either incorrectly or not at all, in 1851, individuals were less likely to “guess” their county of birth with rising HiS-CAM scores and more likely instead to state no county of birth at all. In 1911 by contrast, the trend reverses, and individuals further up the social hierarchy became more likely to give their county of birth incorrectly, evidently preferring to “guess” their county of birth than to leave the space blank.

Given the evidence available here, however, it is not possible to say for certain why this shift might have occurred between 1851 and 1911, but it is plausible that in 1851, the more skilled individuals were, the less inclined they were to risk making an error by “guessing.” However, as the population became more migratory by 1911, and presumably more familiar with regional geographies, some may have become more inclined to state their county of birth incorrectly, evidently preferring to “guess” their county of birth than to leave the space blank.

Given the evidence available here, however, it is not possible to say for certain why this shift might have occurred between 1851 and 1911, but it is plausible that in 1851, the more skilled individuals were, the less inclined they were to risk making an error by “guessing.” However, as the population became more migratory by 1911, and presumably more familiar with regional geographies, some may have become more inclined to state their county of birth incorrectly, evidently preferring to “guess” their county of birth than to leave the space blank.

Figure 10 clarifies the observation made in Figure 9. While the relatively unskilled—those with a HiS-CAM score of less than 50—were generally no more or less likely than the population at large to state their county of birth either incorrectly or not at all, in 1851, individuals were less likely to “guess” their county of birth with rising HiS-CAM scores and more likely instead to state no county of birth at all. In 1911 by contrast, the trend reverses, and individuals further up the social hierarchy became more likely to give their county of birth incorrectly, evidently preferring to “guess” their county of birth than to leave the space blank.

Given the evidence available here, however, it is not possible to say for certain why this shift might have occurred between 1851 and 1911, but it is plausible that in 1851, the more skilled individuals were, the less inclined they were to risk making an error by “guessing.” However, as the population became more migratory by 1911, and presumably more familiar with regional geographies, some may have become more inclined to state their county of birth incorrectly, evidently preferring to “guess” their county of birth than to leave the space blank.

Therefore, while occupation and social status may have determined the type of error made, it did not have an effect on whether a mistake was made. Indeed, as one might expect individuals’ ability to recall their county of birth to have been a function of how far they migrated from it, Figure 12 shows the relative likelihood of individuals specifying their county of birth incorrectly as the distance from it increased. A score greater than 1 shows they were more likely to do so than the general population. Figure 12 shows that throughout the period, every additional kilometre migrated made it more likely that individuals would specify their county of birth incorrectly. Indeed, in 1851, individuals that had migrated 100 km were approximately 50 times more likely to state their county of birth incorrectly compared to someone that had not migrated at all. However, in 1891, migrants travelling between 10 and 20 km became more likely to specify their county of birth correctly relative to non-migrants. This trend became more pronounced in 1901 and 1911. It is likely that the growing railway network—which had grown from 14,000 km to 36,500 between 1851 and 1911—allowed short-distance migrants to regularly return home (Shaw-Taylor & You, 2018). Indeed, one would expect that as a population becomes more mobile, their geographical knowledge would also improve, if only out of necessity.

This interpretation is supported by evidence in Figure 12 that the effect of distance on the likelihood that individuals mis-specified their county of birth weakened considerably over time. Indeed, the growing postal service not only allowed migrants to stay in touch with home more easily but required they specified the destination, including the county (Daunton, 1985). By 1911, the ease of staying in touch with home—evidenced by the number of letters increasing from 15.7 letters per capita in 1851 to 75.0 by 1911—likely helped migrants recall their birthplace more accurately. Nevertheless, Figure 12 shows that while the distance effect weakened over time, it did not disappear altogether. Rather, throughout the period 1851–1911, fewer

![Figure 10](image-url)
individuals recalled their county of birth correctly the further they migrated, which probably made them less likely to maintain a connection and, therefore, the less need there was to recall their birthplace. While it is unwise to speculate as to whether individuals moved because they lacked an affinity with their birthplace, or they lacked an affinity because they had moved a long way, it seems plausible to infer that individuals’ ability to identify their place of birth was a function of their distance—and therefore attachment—to it. Individuals’ ability to identify their place of birth was therefore a function of its personal significance and either it felt a part of them, or they felt a part of it (Cohen, 1982, 1986; Snell, 2006). This interpretation is supported by evidence that both place identity—and the accuracy of birthplace reporting—improve with age (Blokland, 2001; Degnen, 2016).

Figure 13 analyses the effect that age had on individuals’ ability to identify their county of birth correctly by identifying cohorts that had likely only recently left home prior to each census, and then tracking that age group across subsequent censuses (Day, 2018, 2020). However, to remove the effect of other factors on the accuracy of birthplace reporting in each census, the proportion of
each cohort that reported their county of birth incorrectly was divided by the total proportion that did so in each census year, to produce a relative likelihood of each cohort specifying their birthplace incorrectly as they aged. Doing so shows a clear age-effect. Figure 13 shows that although earlier cohorts appear less likely to make an error than later cohorts, all cohorts became progressively less likely to state their county of birth incorrectly as they aged. Indeed, numerous recent studies indicate that individuals’ sense of identity becomes increasingly bound up with place as they age, suggesting that attachment and belonging rather than memory, education or geographical knowledge, were the key drivers determining individuals’ capacity to identify their county of birth correctly (Blokland, 2001; Degnen, 2016). It is no surprise that 70% of recreational genealogists in the USA are 50 and over and that the principal questions asked are “who am I?” and “where did I come from?” (Josiam & Frazier, 2008). Indeed, it is natural that as we age and evaluate our achievements, we wish to give our life a greater sense of meaning by contemplating our origins and hence placing ourselves into a longer-term narrative.

However, individuals’ ability to identify their county of birth should not be the sole metric by which their conceptualisation of place is judged. Rather, as the algorithm matched individuals’ birthplace to the smallest spatial unit possible given the information supplied, it is also feasible to analyse the determinants of the precision with which individuals described their place of birth. Specifically, did individuals become less precise the farther they had migrated? This will be addressed in the next sub-section.

### 2.2.2 | Precision of birthplace reporting

As stated in Section 2, the algorithm is designed to match birthplace strings to as few parishes as possible. The fewer parishes a birthplace string could be matched to, the more precisely an individuals’ birthplace is interpreted as having been stated. Figure 14 therefore shows the average number of parishes which individuals’ birthplace strings had been matched to in each year between 1851 and 1911. The upward trend is clear; individuals became less precise the farther they had migrated. Combined with evidence from Figure 12, it is evident that the act of migrating reduced individuals’ ability to identify their place of birth accurately, whether that be their ability to identify the right county or the precision with which they name their settlement of birth. This can be best observed when analysing how individuals born in London—comprised of many parishes—referred to their birthplace in 1851 and 1911, as shown in Figure 15.13

Figure 15 shows that in 1851, the average number of parishes an individual referred to as their settlement of birth increased rapidly upon migrating beyond the limits of London.14 In other words, they became more likely to simply give “London” as their birthplace. Like Figure 12, the distance effect appears to weaken over time, with London-born migrants providing far more precise birthplaces in 1911 than those moving comparable distances in 1851. Again, the expansion of the railway and postal service likely fostered a greater knowledge of place (Report of the Postmaster General on the Post Office, 1916). Yet also like Figure 12, the
distance effect had only weakened, not disappeared, by 1911. The farther individuals migrated, the less precisely their birthplace was described. As Section 2.1.4 demonstrated that enumerators can largely be absolved of blame, it seems unlikely that such imprecision was because of enumerators’ unfamiliarity with far-flung places. Rather, it is suggested here that the further individuals migrated from their birthplace, the less interaction they had with it, the less significant it became as a component of their identity, and the less they needed to recall it with either accuracy or precision.

3 | DISCUSSION

In Snell’s evaluation of the role of the parish to individuals’ construction of identity, it is tempting to think that “the parish” in question is
that of their birth. While this was indeed important for the administration of Poor Relief, it was not the only parish to which someone might have felt connected (Snell, 2006). Indeed, individuals whose graves were marked “of this parish” included those that developed a strong attachment to the parish in which they had long been resident, without having necessarily been born there (Snell, 2006). Indeed, the importance of place to identity has spawned a vast literature, comprehensively reviewed by Peng et al. (2020). Shared values and common experiences create a sense of the community to which one belongs (Strathern, 1982). Yet where individuals gained this sense of identity does not necessarily have to have been one’s place of birth. Consider the following exchange from the 1975 movie Jaws (Spielberg, 1975):

Ellen Brody: “I just want to know one thing—when do I get to become an islander?”

Mrs. Taft: “Ellen, never, never! You’re not born here, you’re not an islander, that’s it.”

Unpacking this interaction is revealing. Having committed to moving to Amity Island, Ellen is keen to join the community and be accepted as part of it. The wording of her question, “when will I become an islander,” illustrates her wish to internalise Amity as part of her identity, to be an islander, not merely an inhabitant of it. The retort from Mrs. Taft, however, signals her own conceptualisation of place and belonging. Only birth confers identity. Not only does this imply that this places her in an exclusive group but that a component of such an identity is the shared experience that comes from being an islander from birth. With such commonalities also come shared values and beliefs, which further amplifies the importance of communal belonging to a sense of individual identity. Consequently, the dialogue illustrates two important truths. That while individuals’ place of birth is important to their identity, it can also be usurped in their affection by the place in which they choose to settle long-term. Ultimately, it reveals where individuals conceive of as “home.”

This is revealed by the relationship between individuals’ age, distance migrated, and their likelihood of specifying the county of birth correctly, or their birthplace precisely. The farther individuals had migrated, the less likely they were to either identify their county of birth correctly, or describe their place of birth precisely, not because they lacked the requisite skills to do so, but because their birthplace became a less significant determinant of their identity than their place of residence. Getting one’s birthplace right was no longer important, or even necessary. Indeed, recent analyses of the interaction between place memory, place identity, and place attachment demonstrates that all three improve both with age, and the length of time they have been resident in that place (Lewicka, 2008). Clearly, individuals’ ability to identify their place of birth should not only be interpreted as a measure of the accuracy of the census as a source, but of individuals’ attachment to their birthplace.

4 | CONCLUSION

Previous investigations into the accuracy with which birthplaces in the nineteenth-century censuses were reported, sought only to establish absolute levels of accuracy, and thereby assess the reliability of the census as a source for the study of migration (Hinde, 2004; Perkyns, 1991). The present study has instead sought to determine the means which such errors were introduced. By critically assessing the journey which census data has been on to reach its present form as a digital datafile, five stages at which the exact location of individuals’ birth could be misinterpreted, were identified, and illustrated in Figure 1. By working backwards through these stages is akin to peeling back the layers of time in an archaeological excavation. The last layer produced is the first to be removed, and by removing earlier and earlier layers, the principal object of enquiry—individuals’ own conceptualisation of their birthplace—is revealed. This article has therefore identified that it was the individual and the householder who were primarily responsible for errors.

So-called “processing” errors—errors that were the result of processing the data supplied about the individual—were examined in Section 2.1. Errors created at the final processing stage—the algorithm which matched the transcribed birthplace string to a GIS—was shown not to have introduced any significant, structural errors. Occasionally, however, transcription error may have resulted in a county of birth being incorrectly reassigned by the algorithm, requiring manual correction. While these were isolated incidents, it is indicative of how even small transcription errors can distort individuals’ true birthplaces and by extension, their imputed migration path. Transcription errors were therefore investigated, and while they accounted for just 0.8% of the total, they accounted for 10.9% of the errors in those counties with similarly spelled names to others, such as Herefordshire and Hertfordshire and therefore constituted an important source of error regionally, if not nationally. Next, it was shown that errors were not a result of enumerators being unfamiliar with small, far-off places, nor were some enumerators more likely to make errors than others due to a lack of diligence. Such an analysis has demonstrated that few errors were created outside the household, which ought to reassure those critical of the census as a source for migration analyses (Hinde, 2004). Instead, the errors which remain were not created by the act of collecting and preparing the data for analysis but rather were created by those inside the household itself and as such, should not be treated simply as inconveniences, but as evidence of the role of place in nineteenth-century identity formation.

Section 2.1.4 showed that householders were responsible for introducing a significant number of errors. This is confirmed in Table S1, which shows that inattentive householder who mis-stated their own county of birth were just as slapdash with the details of others in their household as they were with their own. In other words, where householders’ own county of birth was incorrect, those of others in the household were—ceteris paribus—significantly more likely to be so also.

Having explored the extent to which errors could be attributed to data processing stages, Section 2.2.1 shows that the remaining errors could be more plausibly attributed to the individual. Having demonstrated in Figure 4 that enumerators’ possible unfamiliarity with
individuals’ birthplaces was not a source of error, that individuals’ county of birth was more likely to be incorrectly stated with each kilometre migrated from their birthplace, can be more plausibly attributed to individuals’ own propensity to recall their county of birth accurately. As the accuracy of birthplace identification exhibited no correlation with occupation—as a proxy for the level of skill and education which their occupation implied—Section 2.2 argues that both the accuracy and precision with which individuals described their birthplace was not a function of their innate ability to do so but, rather, reflected their attachment to it and its significance as a component of their identity.

This is confirmed when considering the age effect in Figure 13. This showed that despite the presumption that age fades the memory, individuals became more accurate, not less. When combined with sociological evidence that place becomes an increasingly important constituent of identity with age, the inescapable conclusion seems to be that individuals’ ability to identify their birthplace was a product of its contribution to their identity (Blokland, 2001; Cohen, 1982, 1986; Degnen, 2016; Lewicka, 2008; Snell, 2006).

Through a systematic analysis of errors in birthplace reporting, while this article has argued that the significance of individuals’ birthplace to their identity was a function of the distance migrated from it, it has—as stated at the outset—only measured individuals’ ability to name both a settlement and the county in which it was located, rather than any objective measure of their birthplace. This lacuna can therefore be remedied using record-linked data (Newton & Bennett, 2020) to refine the observations made here and analyse the extent to which individuals’ description of their birthplace changed as they migrated and aged. It was argued in the introduction that unlike the Perkyns (1991) study, birthplaces ought to form part of the evidence when conducting record-linkage, and if birthplaces in two records cannot be reconciled, they should not be linked. Happily, the Newton and Bennett (2020) dataset is indeed matched on this basis, and records are linked where the given birthplaces intersect: London and Paddington, for example, rather than London and Norwich. While such an analysis would be a measure of the consistency with which birthplaces were reported between censuses, rather than an objective measure of their accuracy—the Newton and Bennett (2020) dataset does not match records to baptism records or other third-party evidence—it will be possible to more rigorously test the conclusion drawn here: that people became more precise as they aged and less precise as they migrated.

Furthermore, this article does not show whether it was migration per se that eroded the significance of individuals’ birthplace, or whether the significance of birthplace to individuals’ identity determined migrants’ choice of destination. Given that the different dialectic regions of the UK can only be the product of relatively stable, homogeneous, populations (Martin & Moore, 2017; Trudgill, 1990, 2004), it seems most plausible that regional identities limited migration. Consequently, subsequent analyses ought to investigate the extent to which regional identities created boundaries, and whether such boundaries, in turn, limited migration. In the context of nineteenth-century England and Wales, the significance of this cannot be overstated. The urban population of England and Wales more than trebled between 1851 and 1911 while the rural population shrank by almost a million (Law, 1967). Despite such seemingly rampant urbanisation, the urban–rural wage gap persisted, suggesting that even by 1911, the appetite for urban labour had not yet been sated (Boyer & Hatton, 1997; Hunt, 1973, 1986). Testing whether boundaries created by regional cultures did indeed strangle the labour supply demanded by towns and cities would aid understanding of the complex interrelationships between migration, identity, and economic growth.

ACKNOWLEDGEMENTS

My thanks go to both my current colleagues at the University of Bristol, as well as my former colleagues at the Cambridge Group for the History of Population and Social Structure for their advice and ideas offered both informally over coffee and at more formal seminars while this paper was in progress. Special thanks go to David Manley, who not only read over a draft of this paper, but whose invaluable advice has undoubtedly made this a better paper than it would otherwise have been. Special thanks also goes to Gill Newton who first gave me the idea for this paper, as well as to Kevin Schürer with whom I worked closely while preparing the data analysed here. I also thank the two anonymous referees, whose feedback and constructive comments helped add some much-needed clarity to the structure of the paper. This research is funded by the Economic and Social Research Council (ES/R005443/2) project “Migration, urbanization and socio-economic change, England and Wales 1851–1911.” The usual disclaimers apply.

CONFLICT OF INTEREST

The author has no conflict of interest to declare.

ENDNOTES

1 The algorithm was developed by the author as part of the ESRC project: “Migration, Urbanization and Socio-Economic Change: England and Wales, 1851–1911.”

2 As those born either at sea or outside England and Wales were only statutorily required to provide a single piece of information—the county of birth or simply “born at sea,” and the technique described here requires two pieces of information—the parish and the county—to cross-validate one another, with errors being defined as those instances where the stated parish cannot be located in the stated county, those born at sea or outside England and Wales, are excluded.

3 For example, whereas someone who only gave “London” as their birthplace would be matched to all the parishes which constituted London at the time of their birth, a person that specified “Paddington” would be matched to just the two parishes which make up Paddington.

4 Given the relative imprecision of county boundaries, a liberal definition of the county was adopted. Individuals were presumed to have supplied the “correct” county if the parish they stated had ever been within the boundary of the county stated, however defined. The town of Dudley for example is an exclave of Worcestershire but is wholly within the boundaries of the county of Staffordshire. Similarly, Barnet was within the “ancient” county boundary of Hertfordshire but was returned with the “registration” county of Middlesex. Therefore, those giving their
birthplace as “Dudley, Staffordshire,” “Dudley, Worcestershire,” “Barnet, Hertfordshire,” and “Barnet, Middlesex” were all deemed to have stated their county of birth correctly. While this definition undoubtedly underestimates the true of errors, removing such false negatives is preferable to retaining false positives.

Enumerators would of course have assist householders that could not fill in their census form independently in 1911.

While enumerators might have also made mistakes as a result of mishearing a household or misreading their return, as these errors cannot be distinguished from keystroke errors occurring at the transcription stage, they are not considered here.

Where figures are specified as being age- and/or distance (migrated) standardised, the following direct-standardisation formula (Newell, 1988) was applied: \[
\frac{\sum \text{error rate in specific ED}}{\sum \text{reference population}}
\]

EDs are only available in digitised form for 1851 and 1861.

In a hypothetical RSD made up of four EDs—ED 1–4—in which 1,000 people mistook their county of residence for their county of birth; 500 were located in ED 1, 300 were located in ED 2, 150 were located in ED 3, and 50 were located in ED 4. However, the number of errors in each ED is a function of both the intrinsic error rate of each ED and its population. The effect of population can be removed—thereby isolating the effect of the error rate—by standardising the population in each ED. If the population in ED 1 was 4,000, 3,000 in ED 2, 2,000 in ED 3, and 1,000 in ED 4, they could be adjusted to the same—notional—number, say, 1,000. The population of ED 1—and the number of errors—are both divided by four to 1,000 and 125, respectively. ED 1 still has the same error rate of 12.5% but now carries equal weight as the other three EDs in the RSD. The same adjustment is made to the number of errors in ED 2 (300 ÷ 3 = 100), ED 3 (150 ÷ 2 = 75), and ED 4 (50 ÷ 1 = 50). Consequently, ED 1 now accounts for 125 of the 350 errors made in the RSD—35.7%, ED 2 accounts for 100–28.6%, ED 3 accounts for 75–21.4%, and ED 4 accounts for 50–14.3%. The probability that two such individuals would be in the same ED is given as 0.357² + 0.286² + 0.214² + 0.143² = 0.755%. However, as the likelihood that two individuals would be in the same ED goes up as the number of EDs in an RSD goes down, the probability of 0.2755 is divided by the likelihood that an individual would be in one of the four EDs were there an equal chance—1 in 4. As 0.2755 ÷ 0.25 = 0.2755 × 4, the probability that two such individuals were in the same ED is multiplied by the number of EDs, which gives a score 1.102. As a concentration index of 1 would indicate that errors were evenly distributed across the RSD, a score of 1.102 suggests errors were just 10.2% more concentrated than expected.

This is akin to the difference between statistically significant dispersion and a random distribution when testing for spatial autocorrelation using Global Moran’s I. Black squares on a checkerboard are perfectly dispersed and would produce a score of –1, whereas a random distribution would produce a higher score of 0. See Moran (1950).

Those whose stated county of birth was the same as their county of residence are excluded, as this error may be indicative of other processes.

The singulate mean age at leaving home (SMAL) is estimated for each RSD in each census year for males and females. The numbers leaving home in each census year are estimated by the proportion that were both no longer co-resident with parents and were aged within +/– two years of the regional SMAL. The SMAL is adjusted for parental mortality and the calculation is fully described in J. Day (2018). See also K. Schürer (2004).

It is for this reason that the determinants of the precision with which individuals specified their birthplace are not interrogated to the same extent as the accuracy with which they identified their county of birth. Those that referred to more than one parish as their birthplace were disproportionately urban-born—specifically London—meaning that analyses would be unrepresentative of England and Wales. The urban-born were also the least migratory. This section is therefore intended as further evidence that how individuals described their birthplace was a function of the strength of their relationship with it.

As London was expanding rapidly in this period, all those born within the 1911 boundaries of London as “London-born.”

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the author upon reasonable request.

ORCID

Joseph Day @ https://orcid.org/0000-0003-3937-6868

REFERENCES

Anderson, M. (1972). The study of family structure. In E. A. Wrigley (Ed.), Nineteenth-century society: Essays in the use of quantitative methods for the study of social data. Cambridge University Press. 10.1017/ CBO9780511896118.003

Armstrong, W. A. (1978). The census enumerators’ books: A commentary. In R. Lawton (Ed.), The Census and social structure. Routledge.

Baddeley, A. D. (1982). Implications of neuropsychological evidence for theories of normal memory. Philosophical Transactions of the Royal Society B, 298(1089), 59–72. https://doi.org/10.1098/rstb.19820072

Benson, M., & Jackson, E. (2013). Place-making and place maintenance: Performativity, place and belonging among the middle classes. Sociology, 47(4), 793–809. https://doi.org/10.1177/0038038512454350

Blockland, T. (2001). Bricks, mortar, memories: Neighbourhood and networks in collective acts of remembering. International Journal of Urban and Regional Research, 25(2), 268–283. https://doi.org/10.1111/ 1468-2427.00311

Boyer, G. R., & Hatton, T. J. (1997). Migration and labour market integration in late nineteenth-century England and Wales. The Economic History Review, 50(4), 697–734. https://doi.org/10.1111/1468-0289. 00075

Campbell-Kelly, M. (1996). Information technology and organizational change in the British Census, 1801-1911. Information Systems Research, 7(1), 22–36. https://doi.org/10.1287/isre.7.1.22

Casella, G. (1985). An introduction to empirical Bayes data analysis. The American Statistician, 39(2), 83–87. https://doi.org/10.2307/2682801

Christian, P., & Annal, D. (2014). Census: The family Historian’s guide. The National Archives.

Cohen, A. P. (Ed.) (1982). Belonging: Identity and social organisation in British rural cultures. Manchester University Press.

Cohen, A. P. (1986). Symbolising boundaries: Identity and diversity in British cultures. Manchester University Press.

Daunton, M. J. (1985). Royal Mail: The post office since 1840. The Athlone Press.

Day, J. (2018). Leaving home in 19th century England and Wales: A spatial analysis. Demographic Research, 39, 95–135. https://doi.org/10.4054/ DemRes.2018.39.4

Day, J. (2020). The process of internal migration in England and Wales, 1851–1911: Updating Ravenstein and the step-migration hypothesis. Comparative Population Studies, 44(6), 447–496. https://doi.org/10. 12765/CPOs-2020-13

Degnen, C. (2016). Socialising place attachment: Place, social memory and embodied affordances. Ageing & Society, 36(8), 1645–1667. https:// doi.org/10.1017/S0144686615000653
Yasumoto, M. (1985). How accurate is the Methley baptismal registration? Local Population Studies, 35(Autumn), 19–24.

You, X. (2020). Women’s labour force participation in nineteenth-century England and Wales: Evidence from the 1881 census enumerators’ books. The Economic History Review, 73(1), 106–133. https://doi.org/10.1111/ehr.12876

SUPPORTING INFORMATION
Additional supporting information may be found in the online version of the article at the publisher’s website.