

Images of the Child

THESIS PROSPECTUS V 2.0

CHRIS BORST

Working Document #3

Table of Contents

Context	1
Data	3
Positive Training Examples	3
Negative Training Examples	6
Test data	7
Method	8
Distinguishing texts about children from those not about children	8
Modelling each construction	9
Testing the hypotheses	10
Is there a rough chronological order observable to each construction’s flourishing?	10
What proportion of texts about children are not assigned to any constructions?	10
What proportion of texts about children are assigned to multiple constructions?	10
Do the models predict each other?	11
References	11

Context

Harry Hendrick, in his contribution (Hendrick, 1990/1997a) to the iconic and field-defining volume *Constructing and Reconstructing Childhood* (James & Prout, 1990/1997), offers an “interpretative survey” of the constructions and reconstructions of childhood in the UK, from 1800 to his present. Not surprisingly, his account has been influential. Of the chapters in *Constructing and Reconstructing Childhood*, only Prout and James’ introductory invocation of a “new paradigm” (Prout & James, 1990/1997) and Boyden’s chapter on how the Western vision of childhood has been globalized (Boyden, 1990/1997) are more cited, according to Google Scholar. Of Hendrick’s own work, much of which builds on and extends the account in this chapter (e.g., Hendrick, 1994, 1997b, 2003), no other contribution comes close to it in citations. Erica Burman’s *Deconstructing Developmental Psychology* (Burman, 2017), which appears to be the preferred contemporary text for teaching images of childhood, cites it heavily in its discussion. Indeed, while Hendrick is not actually the most cited historian of childhood, he is the sole historian included in the volume *Key Thinkers in Childhood Studies* (Smith & Greene, 2014).

He summarizes his theses thus:

In 1800 the meaning of childhood was ambiguous and not universally in demand. By 1914 the uncertainty had been virtually resolved and the identity largely determined, to the satisfaction of the middle class and the respectable working class. A recognizably ‘modern’ notion of childhood was in place: it was legally, legislatively, socially, medically, psychologically, educationally and

politically institutionalized. During these years the making of childhood into a very specific kind of age-graded condition went through several different stages or ‘constructions’ (and ‘reconstructions’). Each new construction, which was manifested in a kind of public identity, may be observed in approximate chronological order as pertaining to Rousseauian Naturalism, Romanticism, Evangelicalism, the shift from wage-earning labour to ‘childhood’, the reclamation of the juvenile delinquent, schooling, ‘Child Study’, ‘Children of the Nation’, psycho-medicine, and ‘Children of the Welfare State’. (Hendrick 1997a: 30)

His key claim is the last sentence of this paragraph, which sets out a list of 10 different constructions (to which the chapter appends an ambiguous reference to ‘the contemporary child’) which were (a) manifested in *public identities*, which may be (b) observed in *approximate chronological order*.

The first three sentences of the paragraph, however, imply difficulties for this claim. If the meaning of childhood was ‘ambiguous’ in 1800, that implies that there was more than one construction with substantive traction at that time. On Hendrick’s account at least the first four on his list were already manifest, qualifying his claim of chronological order. Similarly, if this ambiguity was ‘virtually resolved’ by 1914, such that ‘the’ (singular) identity of a “modern” notion of childhood’ was determined and institutionalized, then what does that mean for the various different constructions? On Hendrick’s account, the first eight constructions on his list preceded World War I. Should the post-1914 construction of the child be seen as a *unification* of those? A *supersession* of those? Or simply as some kind of *mix* of those earlier constructions? If they do unify, then to what extent should they be seen as *different* constructions? At the same time, if such a singular identity *is* institutionalized after 1914, then how should we construe the last two constructions in his list – and the ‘contemporary’ child? One relevant aspect of these difficulties is that the alleged constructions are clearly not all of the same kind. If Rousseauian Naturalism, Romanticism and Evangelicalism can be seen as competing worldviews, the moves against child labour and ‘delinquency’, and for compulsory schooling, are different institutional fronts, where shifting coalitions of such worldviews could be expected on both sides of each issue. Child Study and psycho-medicine name scientific programs, not clearly distinct, while ‘Children of the Nation’ and ‘of the Welfare State’ represent at best rhetorical effusions, at worst mere agglomerations. Hendrick’s account does not seem to allow for the presence of multiple, substantively conflicting constructions of childhood after 1914, perhaps even after 1880, despite itself being conceived as an intervention in that conflict.

It seems to me that there are three questions that can be productively posed with respect to Hendrick’s account:

1. Given a set of models, independently generated for each construction, and an appropriate historical test corpus, is there, in fact, a rough chronological order observable to each construction’s flourishing, as measured by the number of texts published in each year of the corpus assigned to them? (This testing Hendrick’s claim directly.)
2. Given a model distinguishing texts about children from those not about children in the test corpus, what proportion of the texts about children are assigned to at least one

identified construction and what proportion are not assigned to any? (This speaking to the issue of whether Hendrick's account misses salient constructions.)

3. What proportion of the texts about children are assigned to multiple constructions and to how many of them? Indeed, do the independently generated models of the constructions tend to predict *each other*, if tested on the other constructions' training data? (This speaking to whether Hendrick's account actually distinguishes different constructions.)

Jointly, these evaluate Hendrick's account, but also speak to the larger question of what *is* 'a construction of childhood', how can (or should) such constructions be identified and bounded?

My own hypothesis would be that Hendrick's account mostly separates stages in the development of a certain mainstream conception of childhood and does not identify any of the constructions that contested (and continue to contest) that conception. That being the case, I would anticipate that a chronological order will be observable, but that much of the debate will be missed, while most of the constructions will predict each other with considerable reliability.

Data

Positive Training Examples

As far as generating models for each construction is concerned, Hendrick helpfully offers a couple of key authors or texts for each of his alleged constructions. These can be straightforwardly used as training data. The Cambridge Neoplatonists and Locke's *Thoughts on Education* are cited as important background figures, but not as exemplars of a particular construction. So, while their texts are available, they shall be left to one side.

Rousseauian Naturalism: Rousseau's *Emile* itself, in the Foxley translation, is the obvious choice and is available from [Project Gutenberg](#). *Practical Education* by Maria Edgeworth, "a disciple of Rousseau", is also available in two volumes ([one](#), [two](#)).

Romanticism: Blake's *Songs of Innocence and Songs of Experience* is available from [Project Gutenberg](#), but Hendrick notes the distinction of Wordsworth's approach from Blake's and sees Wordsworth as the more influential. His most relevant works are all available at Project Gutenberg:

- *Lines Written above Tintern Abbey* in at least three different places ([one](#), [two](#), [three](#));
- [The Prelude](#);
- [Intimations of Immortality from recollections of Early Childhood](#); while,
- the poems grouped in an [1815 edition](#) as "Referring to the period of childhood" are split between three volumes [one](#) ('We are Seven', 'Anecdote for Fathers'), [two](#) ('My heart leaps up', 'To a Butterfly', 'Foresight', 'Lucy Gray', 'Alice Fell', 'Rural Architecture', 'The Pet Lamb', 'The Idle Shepherd Boys', 'To H.C.', 'Influence of Natural objects', 'The Blind

Highland Boy'), and [three](#) ('Characteristics of a Child', 'Address to a Child', 'The Mother's return').

Evangelicalism: John Wesley is noted as a figure in the background, whose Sermons 94, 95, and 96 of the *Sermons on Several Occasions* are available from the [Christian Classics Ethereal Library](#), amongst others. Hannah More is named as the key figure, with citations taken from her *Strictures on the modern system of female education*, available in the Eighteenth Century Collection Online via the Gale Digital Scholar Lab, again amongst others.

The Factory Child: Here Hendrick shifts from literary sources to parliamentary ones, citing the Health and Morals of Apprentices Act, 1802, and the Factory Act, 1833, as well as the 1833 Royal Commission on the Employment of Children in Factories. The debate on the Acts, in their various incarnations, is available from Hansard. The text of the Royal Commission is available in the UK Parliamentary Papers, via ProQuest, but only as a series of page images. As I understand it, ProQuest offerings are only accessible in plain text via their TDM Studio platform, the price of which the Library considers "prohibitive".

The Delinquent Child: On the legislative side, Hendrick cites the Youthful Offenders Act, 1854, as the key event for this construction, the debate for which is available from Hansard. On the literary side, he cites a text by M.D. Hill on "Practical Suggestions to the Founders of Reformatory Schools" published in an 1855 collection, a prize essay by Micaiah Hill, known for the question "Can these be *children*?", and Mary Carpenter's 1853 *Juvenile Delinquents*, all available from Gale Primary Sources and the latter two from archive.org as well.

The School Child: Hendrick's account of the school child is notable for the absence of any direct reference to a primary source, including only a handful of brief quotes of original sources taken from the secondary literature. Hendrick (1994) and (1997b), however, refer to the Elementary Education Act, 1870, and to subsequent Acts of 1876 and 1880, debates on which are, again, available from Hansard.

Child Study: Hendrick notes early articles by Darwin, Taine, and Sully, before turning to Preyer's 1888 *The Mind of the Child* and Hall's 1883 'The Contents of Children's Minds', the latter of which is usually taken as the starting point of the Child Study movement. Hall's article is accessible from the HathiTrust. Though not cited, his two-volume *Adolescence* is available from Gale, amongst others. Hendrick also cites the journal of the British Child-Study Association, *The Paidologist*, volumes 7-9 of which are available as a single file from Google Books and archive.org.

Children of the Nation: For this construction, Hendrick cites eight different pieces of legislation, on the "age of consent (1885), infant life protection (1872 and 1897), the prevention of cruelty to, and neglect of, children (1889), the school feeding of 'necessitous' children (1906), school medical inspection (and treatment) (1907), a juvenile justice system (1908), and infant welfare (1918)". Again, the debate on these is to be found in Hansard. He also makes reference to John Gorst's 1906 *Children of the Nation*, from which he names the construction, and cites two

passages by T.N. Kelynack, one from his 1910 edited collection *Childhood*, the other from *The Child Welfare Annual* (of 1916?). The first two are available from archive.org. The last is available from the HathiTrust but is deemed 'public domain' only in the U.S., so is probably easier to skip. In Hendrick (1994), to which one must refer for the full citations of these two texts, he also mentions a journal Kelynack edited, *The Child: A Monthly Journal Devoted to Child Welfare*, the first volume of which is available freely from Google Books (but not archive.org or HathiTrust).

Psycho-Medicine: For this construction, Hendrick focuses on the work of Cyril Burt, Susan Isaacs, and the Child Guidance Clinics. Burt's *The Young Delinquent* (1925), *The Subnormal Mind* (1935), and *The Backward Child* (1937) are available from archive.org, as is his 1917 work on *The Distribution and Relations of Educational Abilities* and his 1922 work on *Mental and Scholastic Tests*. The latter two include remarks on 'gifted children' that otherwise only appear in much later works (his book *The Gifted Child* actually appeared posthumously, in 1975). Similarly, Isaac's *The Nursery Years* (1929), *Intellectual Growth in Young Children* (1930), *Social Development in Young Children* (1933), and contributions to *On Bringing Up Children* (1936) are available from archive.org. With Child Guidance, matters aren't so simple. One needs to dig back into Hendrick's sources to even find specific texts cited and what one does find (e.g., Miller's 1937 edited collection *The Growing Child and Its Problems* or the Home and Schools Council's 1935 *Advances in Understanding the Child*) are not easily available – perhaps even available at all. (Note that Routledge has re-released the Miller collection in a version that is available as a VitalSource ebook, but that is not available to our Library and wouldn't be mineable in any case.) However, sources are available. Archive.org offers, e.g., a 1919 British collection entitled *An Introduction to Child Guidance*, a 1928 American collection on *The Child Guidance Clinic and the Community*, and a 1937 American manual of *Child Guidance Procedures* (note that Child Guidance began in Chicago, from which it spread to the UK, so American materials remain pertinent).

Children of the Welfare State: Hendrick returns to the legislative record for this construction, focusing on the 1946 Curtis Report and the Children Act, 1948. The Curtis Report is available from the U.K. Parliamentary Papers, via ProQuest, while the debate on the Children Act is available via Hansard.

Contemporary Children: Under this heading, Hendrick offers a quick review of a variety of developments and discussions, principally of the 1980s, but without offering any synoptic 'construction'. He refers to a much larger number of texts than for earlier periods, of which only three legislative debates (the 1985 Gillick judgement, the 1986 abolition of corporal punishment, and the 1989 Children Act) would be easily available electronically (in Hansard), for all that many can be accessed non-consumptively in the HathiTrust. As this period, roughly "the present", falls outside his list of ten constructions and reconstructions, it could plausibly be excluded. But, if we use only the terms of the parliamentary debates, this does give us at least some leverage on more recent discussions.

Negative Training Examples

To be effectively used as training data, negative examples are also needed for each of these constructions. The most typical way to accomplish this would be to take all the texts for all the constructions together in one pool, annotated for which construction they reflect. The negative examples for category *F* would then consist of the positive examples for all other categories, *G* through *P*. However, such an approach presumes that each category is actually well-differentiated from, perhaps even mutually exclusive of, all other categories. In our case, this is in doubt and is something we specifically wish to investigate. That being the case, such an approach wouldn't work. Instead, we need to model each of these constructions *independently* of one other. This poses a challenge. We need a set of cases of "not-this-construction-of-childhood" for each construction. Acquiring contemporaneous texts for each construction that do not contain constructions of childhood would be easy enough. But, obtaining contemporaneous texts containing constructions of childhood, but not the focal one, is less straightforward.

Legislative constructions: For the six constructions that Hendrick characterizes by citing pieces of legislation, we can use the arguments of those who argued *against* the legislation as our negative examples, for all that it could be the case that such arguments do not strictly oppose the construction in question, but rather give it different salience relative to other considerations. Given the frequency of 'think of the children!'-type arguments, even to challenge the salience of a construction is to problematize it.

Early literary constructions: For the three early constructions characterized by literary texts, we could contrast them – Evangelicalism seems easy to contrast with both Rousseau and with Romanticism, though it's not clear that Rousseau and Romanticism could be contrasted with each other. This choice would have the consequence that Evangelicalism would not predict Naturalism or Romanticism (and vice versa) as a matter of design. Alternatively, Locke and perhaps some of the early pediatricians ([Cadogan](#), [Buchan](#)) could conceivably be used as a contrast to all three constructions, which should ensure that each construction is being modelled independently.

Later literary constructions: Child Study and Psycho-medicine pose the greatest difficulty, most notably because they require that we go beyond Hendrick's own text. There would appear to be a narrow sense in which the two could contrast one another, the latter educational psychology being built in part on a critique of the alleged 'amateurishness' of Child Study, which actively sought the participation of parents, teachers, and other interested 'lay' people in its data collection and analysis rather than restricting such work to academics (Hendrick, 1994; Wooldridge, 1994). Yet, it is also clear that there was great continuity between the two, both in terms of personnel and in terms of the content of their work. What seems more likely to be productive is to take texts by authors who controvert important tenets of each.

Child study: Dewey's [Democracy and Education](#) opposed the thesis of 'recapitulation' that guided much of Child Study (see Cleverley & Phillips, 1986). Anti-eugenics texts like

Boas' article on '[Eugenics](#)' (at Wikisource) and Morgan's lecture '[A Critique of the Theory of Evolution](#)' opposed that common strand, while the related racist strand is perhaps best contrasted with Dubois' [The Souls of Black Folks](#) and Washington's [Character Building](#). And, texts like Goldman's 1906 '[The Child and Its Enemies](#)' and perhaps Tolstoy earlier '[Childhood](#)' could offer an anarchist response to Child Study's commitment to the 'Child of the Nation'. Lastly, it could be worthwhile to include Freud's chapter on '[Infantile Sexuality](#)' as an alternative approach to psychology. (All texts at Project Gutenberg unless otherwise noted.)

Psycho-medicine: For the latter period, two works seem like they may be most pertinent: Carl Roger's 1939 [The Clinical Treatment of the Problem Child](#) and Wilhelm Reich's 1933 [The Mass Psychology of Fascism](#) (both available from archive.org). Anderson and Mary Aldrich's 1939 *Babies are Human Beings* (published in the UK as *Understand Your Baby*) is available from HathiTrust and would also serve as a useful foil (see Hardyment, 2007).

Test data

For test data, I would propose to use the historical archive of *The Times* of London (UK), 1785-2014, consisting of some 12.4 million items, made available via the Gale Digital Scholar Lab. *The Times* is commonly referred to as Britain's 'newspaper of record' and is the traditional paper of Britain's Establishment. When it comes to assessing the construction of what might be called Britain's 'Establishment' idea (or ideas) of childhood, it seems the obvious choice. It also has the practical advantage of having published consistently since 1785, so covering the entire period of our study. None of the other major British publications can claim as long a pedigree.

That said, Gale also makes available many other British newspapers and periodicals, dating back to the 17th century, though with the conspicuous absence of the *Guardian* (available instead via ProQuest). Use of the Gale Digital Scholar Lab would permit the inclusion of all of them in this study. Bingham (2013) makes the point that the influence and centrality of *The Times* is part of its marketing, a myth of its own making, and that historical research should not take it too seriously. But, it seems to me that, for the questions to be examined in this study, there is no particular advantage to using more sources of test data. If anything, I would expect that restricting our test sample to *The Times* should be conservative and tend to offer greater support to Hendrick's Establishment-focused narrative than a more diverse sample might. If my own hypothesis is sustained in this context, it should be sustained even more strongly in a sample with a broader range of voices.

Method

Distinguishing texts about children from those not about children

This seems to me like something that could be accomplished with a simple Boolean model, selecting texts containing any of the following strings:

- a. child
- b. newborn
- c. baby OR babi
- d. infan
- e. toddler
- f. boy
- g. girl

The incidence of these in the test corpus is illustrated in Figure 1, below.

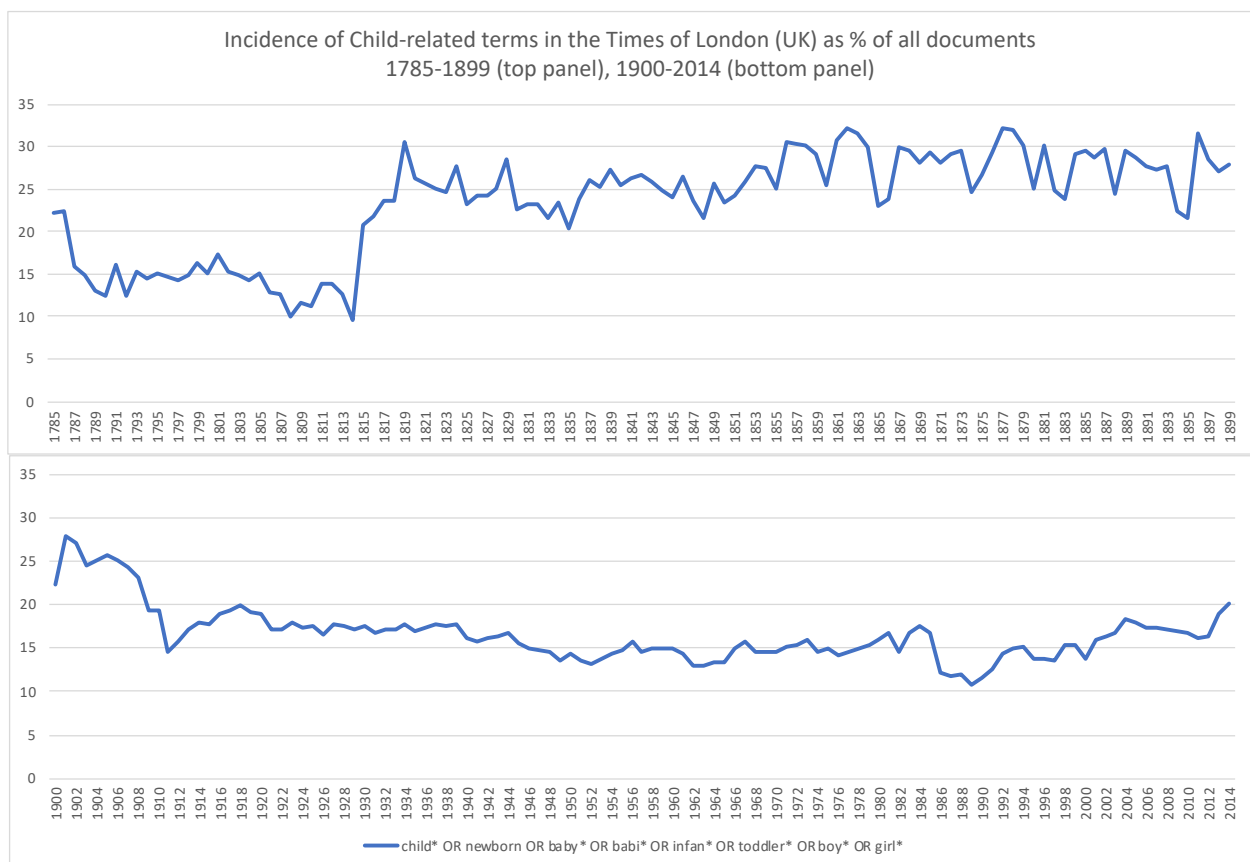


Fig. 1: Incidence of child-related terms in the Times of London (UK) 1785-2014

It will be noted that between 9.6% and 32.2% of each year of the corpus is captured by the selected strings. Moreover, between 1815 and 1908, there are no years below 20%, while

before and after that period (the focal period of transformation on Hendrick's account) there are almost no years *above* 20% - only the first two (1785-86) and, fractionally, the very last (2014).

'Son', 'daughter' and 'young' (catching itself as well as words like 'youngster' or 'youngling') are likely not specific enough, being as likely to identify adults as children. 'Pupil' could be appropriate, if not near 'eye'. 'Juvenile' or 'youth' may well tend to identify adolescents more than prepubescents, though 'juvenile' clearly is a word that was applied widely to children. 'Mother', 'father', and 'parent' could also be worth adding.

A more sensitive model could, no doubt, be created, perhaps seeded by subsamples of the texts identified and not identified by the Boolean model. But it seems unlikely that this would lead to markedly different results, given that the point is to distinguish texts that *ought* to get assigned to one or more of Hendrick's constructions of childhood from texts that presumably shouldn't.

Modelling each construction

The basic idea here is to develop logistic regression (and, for comparison, SVM) models for each of the constructions, evaluating the adequacy of the trained models on a held-out subset of the training data by the F1 measure. There are a large number of pre-processing decisions that need to be made in this context and more than one option will be tried, in search of the most predictive models. Special consideration will need to be made for dirty OCR'd text from Eighteenth Century Collection Online, U.K. Parliamentary Papers, Gale Primary Sources, archive.org, HathiTrust, and Google Books. I need to look further into best practices on that front (see, *i.a.*, Chiron et al., 2017; Hill & Hengchen, 2019; Rigaud et al., 2019; Thompson et al., 2015; van Strien et al., 2020).

I'm not sure how much detail of the intended modelling it is appropriate to include in a prospectus, but for each of the eleven constructions, the main lines of the plan would include:

- a. Break each source text into 1000-word chunks, after stripping any header and footer text from them, as well as (to the extent practicable) any paratext – book and chapter titles, running heads, page numbers, etc.
- b. Split the files 80/20 into training and development/validation sets.
- c. Filtering the samples should, I think, be a main goal, so as to minimize the size of the vocabulary being used, though allowing for (at least) bigrams and trigrams (e.g., "spare_the_rod", "4_year_old"):
 - i. Proper names and years should be removed.
 - ii. Amongst the unigrams, only nouns, verbs, adjectives and adverbs (lemmatized) should be retained.
 - iii. Average distance of terms to the child-related terms could be measured and only terms below a threshold (i.e., those closest to the focal terms) retained.
 - iv. A minimum frequency threshold could be determined.

- d. Including negation or sentiment tags may be worthwhile. I'm uncertain about this.
- e. Use of vector encoding is presumably worthwhile, though, given the observations of Underwood (2019), static embeddings like fastText or GLoVE may suffice.

Testing the hypotheses

For each text in the test corpus, determine:

1. Is it about children or not?
2. For each construction, is it predicted to be a member of that construction or not?
3. The number of constructions (0 to 11) it is predicted to be a member of.

Is there a rough chronological order observable to each construction's flourishing?

Hypothesized answer: Yes.

Test:

1. Plot the number and proportion of texts about children predicted to be a member of each construction, by year, and examine the distribution of the proportions.
2. Determine and plot mean and median year of publication of texts associated with each construction.
3. Examine order of means and medians.
4. Compare means formally by way of t -tests.

What proportion of texts about children are not assigned to any constructions?

Hypothesized answer: A large proportion. Let us say that a proportion greater than 25% is 'large'.

Test:

1. Calculate the global proportion of texts assigned to 0 constructions.
2. Also calculate the proportions for each year in the data and examine the distribution of them.
3. As a validation step, calculate the length of texts assigned to 0 constructions relative to the length of texts assigned to 1 or more constructions. These should not differ in a formal t -test of mean lengths.
4. As a validation step, calculate the global and yearly proportions of texts *not* about children assigned to 0 constructions. This should approximate 100%.

What proportion of texts about children are assigned to multiple constructions?

Hypothesized answer: If assigned to any construction, most will be assigned to more than one. Let us say that 'most' means 67% or greater.

Test:

1. Calculate the global proportion of texts assigned to 2 or more constructions.

2. Also calculate the proportions for each year in the data and examine the distribution of them.
3. As a validation step, calculate the global and yearly proportions of texts *not* about children assigned to 2 or more constructions. This should closely approach 0%.

Do the models predict each other?

Hypothesized answer: In general, yes, though Evangelicalism likely does not predict Naturalism and Romanticism, and vice versa.

Test:

1. For each construction, predict whether the training samples for that construction are assigned to each other construction.
2. Organize the proportion of samples predicted into a lower triangular matrix.
3. Let us say that one model predicts another if more than 50% of training samples of the target are predicted to be members of the predictor.

References

- Bingham, A. (2013). The Times Digital Archive, 1785–2006 (Gale Cengage). *The English Historical Review*, 128(533), 1037–1040. <https://doi.org/10.1093/ehr/cet144>
- Boyden, J. (1990/1997). Childhood and the policy makers: A comparative perspective on the globalization of childhood. In A. James & A. Prout (Eds.), *Constructing and Reconstructing Childhood*. Routledge.
- Burman, E. (2017). *Deconstructing developmental psychology* (Third edition). Routledge, Taylor & Francis Group.
- Chiron, G., Doucet, A., Coustaty, M., & Moreux, J.-P. (2017). ICDAR2017 Competition on Post-OCR Text Correction. *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 1423–1428. <https://doi.org/10.1109/ICDAR.2017.232>
- Cleverley, J. F., & Phillips, D. C. (Denis C. (1986). *Visions of childhood: Influential models from Locke to Spock* (Rev. ed). Teachers College Press.
- Hardyment, C. (2007). *Dream babies: Childcare advice from John Locke to Gina Ford*. Frances Lincoln.
- Hendrick, H. (1990/1997a). Constructions and reconstructions of British childhood: An interpretative survey, 1800 to the present. In A. James & A. Prout (Eds.), *Constructing and Reconstructing Childhood*. Routledge.
- Hendrick, H. (1994). *Child welfare: England, 1872-1989*. Routledge.
- Hendrick, H. (1997b). *Children, childhood, and English society, 1880-1990*. Cambridge University Press.
- Hendrick, H. (2003). *Child welfare: Historical dimensions, contemporary debates* ([Rev. ed.]). Policy.

- Hill, M. J., & Hengchen, S. (2019). Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study. *Digital Scholarship in the Humanities*, 34(4), 825–843. <https://doi.org/10.1093/lc/fqz024>
- James, Allison., & Prout, Alan. (1990/1997). *Constructing and reconstructing childhood: Contemporary issues in the sociological study of childhood*. Falmer Press. (2nd edition 1997, Routledge).
- Prout, A., & James, A. (1990/1997). A new paradigm for the sociology of childhood?: Provenance, promise and problems. In A. James & A. Prout (Eds.), *Constructing and Reconstructing Childhood*. Routledge.
- Rigaud, C., Doucet, A., Coustaty, M., & Moreux, J.-P. (2019). ICDAR 2019 Competition on Post-OCR Text Correction. *15th International Conference on Document Analysis and Recognition*, 1588–1593.
- Smith, C., & Greene, S. (2014). Harry Hendrick. In *Key thinkers in childhood studies*. Policy Press. <https://doi.org/10.1332/policypress/9781447308072.003.0012>
- Thompson, P., McNaught, J., & Ananiadou, S. (2015). Customised OCR correction for historical medical text. *2015 Digital Heritage*, 1, 35–42. <https://doi.org/10.1109/DigitalHeritage.2015.7413829>
- Underwood, T. (2019, July 15). Do humanists need BERT? *The Stone and the Shell*. <https://tedunderwood.com/2019/07/15/do-humanists-need-bert/>
- van Strien, D., Beelen, K., Ardanuy, M., Hosseini, K., McGillivray, B., & Colavizza, G. (2020). Assessing the Impact of OCR Quality on Downstream NLP Tasks: *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*, 484–496. <https://doi.org/10.5220/0009169004840496>
- Wooldridge, A. (1994). *Measuring the mind: Education and psychology in England, c. 1860-c. 1990*. Cambridge University Press.