

Variants

The Journal of the European Society for Textual Scholarship

15-16 | 2021 Textual Scholarship in the Twenty-First Century

Dynamic Facsimiles: Note on the Transcription of Born-Digital Works for Genetic Criticism

Dirk Van Hulle



Electronic version

URL: https://journals.openedition.org/variants/1450 DOI: 10.4000/variants.1450 ISSN: 1879-6095

Publisher

European Society for Textual Scholarship

Printed version

Date of publication: 1 July 2021 Number of pages: 231-241 ISSN: 1573-3084

Electronic reference

Dirk Van Hulle, "Dynamic Facsimiles: Note on the Transcription of Born-Digital Works for Genetic Criticism", *Variants* [Online], 15-16 | 2021, Online since 01 July 2021, connection on 05 November 2021. URL: http://journals.openedition.org/variants/1450; DOI: https://doi.org/10.4000/variants.1450

The authors

Dynamic Facsimiles: Note on the Transcription of Born-Digital Works for Genetic Criticism

Dirk Van Hulle

Abstract: Born-digital works of literature sometimes present genetic critics with so much data that it becomes possible to analyse not just the microgenesis but even the nanogenesis of small textual units such as sentence versions (textual versions of one single sentence). This 'work in progress' essay is a suggestion to visualize these nanogenetic analyses as dynamic facsimiles, a filmic replay of logged keystrokes in parallel with a transcription, by analogy with the popular and effective format of juxtaposing digital facsimiles with their transcription in digital scholarly editions of analogue works.

In 2009, Marita Mathijsen announced the end of genetic editing: "the physical circumstances in which a work comes into being nowadays have changed so much that one can speak of a new era of scholarly editing, and of a radical shift which might well herald the end of the genetic method of editing" (Mathijsen 2009, 234). But while Mathijsen predicted that genetic editing would no longer be possible in the digital age because of the lack of manuscripts, we are now in the situation that — thanks to digital forensic tools (Kirschenbaum 2016; Kirschenbaum and Reside 2013; Lebrave 2011; Ries 2018) or keystroke logging software applied to born-digital works of literature — we have so much material that the biggest challenge is not the gaps in the archival record but the abundance of data. Every single typo, every keystroke, every visit to a website, every move with the cursor, every pause is registered. This note suggests one way of visualising this material in a digital genetic edition.

Extremely Distant and Incredibly Close

In his essay "Conjectures on World Literature" (2000), reprinted in *Distant Reading* (2013), Franco Moretti introduced the notion of distant reading as a form of making use of secondary literature rather than primary literature, arguing that "literary history will become very different from what it is now: it will become 'second hand': a patchwork of other people's research, without a single direct textual reading. Still ambitious, and actually even more so than before [...] but the ambition is now directly proportional to the distance from the text: the more ambitious the project, the greater must the distance be" (Moretti 2013,

48). The ambition is expressed in terms of distance, and implicitly in terms of quantity: "the trouble with close reading (in all of its incarnations, from the new criticism to deconstruction) is that it necessarily depends on an extremely small canon" (Moretti 2000, 57). Moretti regards close reading as "a theological exercise — very solemn treatment of very few texts taken very seriously" (Moretti 2000, 57). The focus on big data, distant reading and macroanalysis in Digital Humanities seems to have the immediate effect that close reading is forced into an antonymous position and non-digital literary studies suddenly look parochial in comparison. But not all traditional forms of literary studies are microscopic or focused on close reading, and vice versa, not all digital forms of literary studies are macroscopic or panoramic. Distant reading can also be reductive in some ways, as it usually limits its "reading" to only one version of texts.

Moretti's "distant reading" is conceived as a form of "indirect reading", not unrelated to what Matthew Kirschenbaum (2007) and Kestemont and Herman 2019 refer to as "not-reading", building on a term coined by Martin Mueller, who emphasizes that this form of reading is not that new:

there are age-old techniques for doing this, some more respectable than others, and they include skimming or eyeballing the text, reading a bibliography or following what somebody else says or writes about it. Knowing how to "not-read" is just as important as knowing how to read.

(Mueller, qtd. in Kirschenbaum 2007, n.p.)

These forms of "indirect" reading indicate that the definition of "distant reading" is quite broad; it only excludes "direct reading". It would be incorrect to equate reading by means of computers with "indirect" or "distant" reading. In the study of born-digital works of literature (e.g. Bekius 2021; Kirschenbaum 2016; Ries 2018; Van Hulle 2014; Vauthier 2016), digital tools not only enable distant reading but also a form of analysis that is actually "closer" than close reading. By means of keystroke logging software as an "observational tool" it might be possible to collect what can be called nanogenetic data about literary writing processes, including *currente calamo* corrections, without interfering in the writing process itself — at least, that is how writing researchers use it (Miller and Sullivan 2006, 1). As an interdisciplinary experiment, it is interesting to apply this method from cognitive writing process research to genetic criticism, notably to the reconstruction of the writing sequence. After all, chronology is the backbone of the genetic edition.

See also Lamyk Bekius's contribution to the present issue of *Variants*.

The Writing Sequence in Genetic Editions of Analogue Works

With analogue writing processes, experiments with ways to encode the writing sequence on the level of the sentence have been only partially successful. The main obstacle is the relatively limited amount of data that can be derived from the analogue writing traces. As a consequence, the reconstruction of the writing sequence of complex writing processes involves so much interpretation that if one were to ask ten editors to make their reconstructions, they would probably all differ from each other. Editorial interpretation itself is not the problem; the problem is that, if the sequence is encoded in the markup (e.g. XML-TEI), this may easily create the impression for the reader that the reconstruction is part of what Hans Zeller called the "record" (Befund), rather than "interpretation" (Deutung) (Zeller 1955).

An example is the digital edition of the Belgian author Willem Elsschot's Achter de Schermen (Behind the Scene), a short text in which Elsschot reconstructs the genesis of one of his own texts, the introduction to his novel Tsjip (Elsschot 2007). The edition enables readers to study the development of the entire text, or sentence by sentence. On the smallest level of granularity, the numbering of the writing steps within a sentence was encoded as follows in the XML transcription: to every and <add> tag a @layer attribute was added indicating the number of the writing steps, starting from @layer="101" to the last step in the composition of a sentence. The edition offered an option to study the writing sequence step by step, visualising each step as a separate line. In the case of complex writing processes, this sometimes resulted in more than a dozen steps per sentence. In this sentence, for instance, the narrator reprimands himself concerning a vague phrasing, urging him to be more specific, to call a spade a spade, and say what his children have done when he came in after a writing session. Originally, he wrote they had pretended or acted ("gedaan") as if he had never been away. But this "gedaan" is too unspecific, too abstract. The narrator tries to express what this word actually does (or does not) accomplish, by looking for the right metaphor.



Figure 1: Elsschot's draft version of the sentence "Zeggen kerel, als je kunt" (Elsschot 2007, Letterenhuis E 285 H 5372, f. 3r).

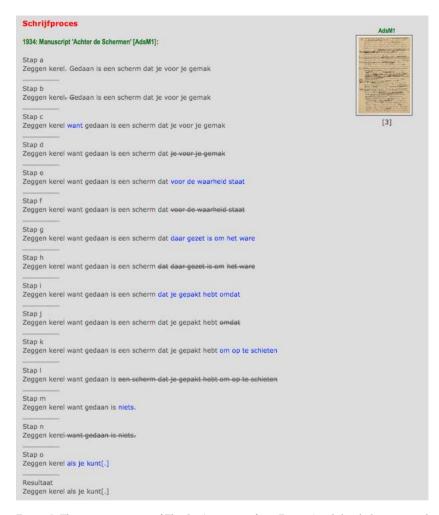


Figure 2: The writing process of Elsschot's sentence from Figure 1, subdivided into several writing steps.

In several writing steps, he initially compares it to a screen that he has used to release himself from the duty of saying what they have exactly done; a screen that he has put between himself and the truth; to make it easy for himself; to get on. Until he concludes that "gedaan" is "niets" (nothing). And even this "nothing" turns out to be too much. He crosses it out and, after several attempts, he ironically and self-deprecatingly snorts: "Say it, man, if you can." ["Zeggen kerel, als je kunt."] (see Figure 1).

We can ask ourselves whether it is worthwhile reconstructing the writing

process in such detail, but it does make one realize how much debris comes with building a story or a novel, how many actions are involved that — retrospectively — turn out to have been seemingly "unnecessary", but that were necessary nonetheless, otherwise the piece would never have taken shape. This is even more striking when the granularity of data is smaller yet, as in keystroke logging data of writing processes.

The Writing Sequence in Genetic Editions of Born-Digital Works

In and of themselves, many of the nanogenetic variants in born-digital works (often typos and cursor shifts) may seem rather meaningless, but taken together, they can help us reconstruct not just different stages in a writing process, but the actual sequence of words and the order in which they were written, letter by letter, as a process. Evidently, there is an important difference between the traces of the writing process (as in the case of Elsschot's Achter de Schermen) and the record of writing actions (by a keystroke logger). The traces of born-digital writing are to a large extent recoverable as well, but that operation requires digital forensics. As Thorsten Ries notes, "Digital forensic tools are able to recover deleted draft versions and stages of the writing process from restored files, temporary data, and system files, file structure artifacts and data fragments from archived and preserved storage media" (Ries 2018, 393). The analogue equivalent of keystroke logging software would be rather like a camera that records every pen stroke an author writes on a page, which raises the obvious question to what extent the element of "being watched" has an impact on the writing process.

A good example is Craig M. Taylor's novel Staying On (2018), which he wrote in collaboration with the British Library, documenting the entire fouryear writing process with keystroke logging software. What is often seen as an intrusion (the installation of a form of spyware on the writer's computer) was approached quite differently by Taylor. In his case, he himself was the requesting party. As he explains in a 9 November 2018 British Library blog post (Taylor 2018), he contacted the British Library before starting his book project out of two concerns: the first concern was with the perceived loss of drafts in born-digital works and the second with "the long-haul loneliness of novel writing, a process I considered in my most despairing moment as like wallpapering a dungeon" (Taylor 2018). In an unexpected way, the experiment was thus partly motivated by the sociology of writing. According to the author, "it actually did help me begin again with novel writing". He even speaks of writing in terms of collaboration: "Somehow the writing felt collaborative, not only because the software was recording me, but also because of the digital curation team who were taking the data" (Taylor 2018).

Apart from the ontological difference between traces and a recording, the types of results are also dependent on the sophistication of keystroke logging

software. Regarding the Spector Pro software, used for the project, Taylor quotes Jonathan Pledge, a curator of contemporary archives at the British Library, noting that Spector Pro was originally designed as spyware for company surveillance of employees, and as a result it is not very sophisticated as keylogging software (cf. Taylor 2018). But keystroke loggers such as Scriptlog, Inputlog or Translog do provide data that can be of interest to the study of writing processes (Bekius 2021; Leijten and Van Waes 2013; Leijten et al. 2014; Van Waes et al. 2011). And even though this type of data is based on a recording rather than on the traces themselves, it can be of help in digital genetic criticism.²

Dynamic Facsimiles

Apart from the question how to analyse the abundance of data provided by keystroke logging software, the question is also: how can we visualize it in such a way that it becomes relevant to users of critical editions? Compared to print editions, digital scholarly editions are still at an early stage of their development. But there is one feature that seems rather constant: the combination of (digital) facsimile with transcription. Combining a "document"-oriented approach with a focus on "text", this parallel presentation format appears to be an aspect of digital editions that works for most users and editors alike.

This raises the question what the "document" in a digital environment actually is. Building on Blanchette, Drucker, Kirschenbaum and others, Thorsten Ries suggests that it is "odd to still tie the term "document" to the physicality of a text carrier, although obviously the term and concept is historically derived from physical documents and graphical user interfaces are still mimicking the physical document on the screen" (Ries 2018, 397). If we want to take this to heart, I suggest we also need to look for different visualizations of the "document" in scholarly editions of born-digital material, according to the motto that the interface is an integral part of the editorial argument (Andrews and Van Zundert 2018; Bleeker and Kelly 2018; Bleier and Klug 2018; Dillen 2018; Schäuble and Gabler 2018).

Whereas most digital editions nowadays show a static digital facsimile of a scan on one side and a static transcription next to it, the equivalent of a born-digital work's genesis could present readers with a more dynamic presentation, linked to a static transcription. In this way, a scholarly editor can combine stasis with movement, a transcript of every version and a dynamic (filmic) visualization of all the keystrokes constituting a sentence.

Imagine Jane Austen writing the first sentence of Pride and Prejudice on a com-

² See for instance the research project called *Track Changes: Textual Scholarship and the Challenge of Digital Literary Writing*, a collaboration between the Huygens ING (Royal Netherlands Academy of Arts and Sciences, Amsterdam) and the University of Antwerp (Centre for Manuscript Genetics, University of Antwerp), funded by the Dutch Research Council (NWO). Project members include Karina van Dalen-Oskam, Dirk Van Hulle, Luuk Van Waes, Mariëlle Leijten, Vincent Neyt, Lamyk Bekius and Floor Buschenhenke.

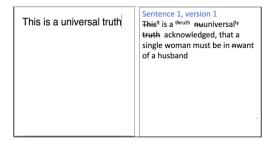


Figure 3: An animated rendition of this visualization is available in the online version of this essay. Please visit: https://journals.openedition.org/variants/1450.

puter, with keystroke logging software: a (fictitious) first draft (Figure 3) and then a revision campaign (Figure 4). The *static* visualization enables macrogenetic analysis (examining the genesis of the work in its entirety across multiple versions) and microgenetic research (the processing of a particular source text; the revision history of one specific textual instance across versions; revisions within one single version) while the *dynamic* visualization facilitates especially microgenetic and even nanogenetic analysis (relating to revisions on the level of the character, the individual keystroke or mouseclick).

The author may take their time to revise the sentence in many places, but as long as the revisions take place within the boundaries of a particular sentence, this textual unit can be regarded as one "sentence version" (Van Hulle 2019). And whoever is interested in the internal changes within this sentence version can follow this process in the dynamic, filmic visualization. Thus, it becomes possible to work with "writing footage" or a "dynamic facsimile" to study the writing process on the nanolevel. In this context, the "sentence" is broadly defined as a syntactic unit that ends with a full stop, an exclamation mark or a question mark. As soon as the author leaves the boundaries of the sentence to work on another part of the text, the sentence version is complete; as soon as the author returns to this sentence to revise it, the next sentence version begins.³

The combination of scan with transcription is referred to in French genetic criticism as a combination of "donner à voir" and "donner à lire" (Grésillon 2016, 149): the facsimile is an image rather than a text, "made for looking", whereas the transcription is "made for reading". With keystroke logging software, we can now offer editions "made for watching" — as in "watching a movie". This kind of editing comes closer to reconstructing the actual writing sequence including the pauses, and therefore also allowing for the analysis of aspects such as "fluency" or "writer's block".

In cognitive writing process studies, the two types of changes (respectively shown in examples one and two) are called "pre-contextual" revisions — "revisions made before an externalised context is completed" — and "contextual" revisions — "revisions made within a completed externalised context" (Lindgren and Sullivan 2006, 159).

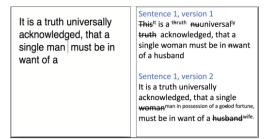


Figure 4: An animated rendition of this visualization is available in the online version of this essay. Please visit: https://journals.openedition.org/variants/1450.

By analogy with the notion of a (static) digital facsimile in a parallel presentation of scan plus transcription in scholarly editions of analogue writing processes, it seems appropriate to call the writing footage a "dynamic facsimile" because it tries to "do like" (fac simile) what happens in the digital document. And since, as discussed above, "the term and concept ['document'] is historically derived from physical documents and graphical user interfaces are still mimicking the physical document on the screen" (Ries 2018, 397), the dynamic facsimile thus mimics this mimicking. This sounds fancier than it is. The resulting interface is easy to use and to read. After all, the purpose of a genetic edition is to make the writing process accessible.

Obviously, the proposed interface is a form of modelling. It inevitably, but also purposefully, reduces the complexity of the writing process in order to try and understand it. One of the aspects it does not capture well is an author's sudden decision, for instance, to jump from the middle of one sentence back to the beginning of the story to change something in one of the first sentences. This is a challenge we are trying to find solutions for in the "Track Changes" project (see especially the PhD work-in-progress by Lamyk Bekius and Floor Buschenhenke).

If we understand the scholarly edition as an embodied argument *about* the work instead of seeing the edition only as a presentation or representation *of* the work (Eggert 2016), and if we understand the interface as an integral part of the editor's argument (Andrews and Van Zundert 2018; Dillen 2018), I suppose my proposal to work with dynamic facsimiles embodies the argument that the work of literature cannot be reduced to just a static text. What we read at any given moment is only an instantiation of a dynamic process. And therefore, it is useful to present the instantiations or versions 1) as a dynamic facsimile, 2) in combination with a system that divides this dynamic writing process in snapshots (versions of a textual unit, in this case "sentence versions") to enable users not only to zoom in on the micro- and nanolevel of the writing sequence, but also to zoom out and compare (collate) various stages of a particular textual unit on a macrolevel and study its development over time.

Very short conclusion for aspiring editors

Do not despair; the end of genetic editing is anything but near. Do not presume; there is a lot of work to be done.

Bibliography

- Andrews, Tara L. and Joris J. Van Zundert, 2018. "What Are You Trying to Say? The Interface as an Integral Element of Argument." In *Digital Scholarly Editions as Interfaces, Schriften des Instituts für Dokumentologie und Editorik*, volume 12, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, Norderstedt: Books on Demand (BoD), pages 3–34. ISBN 978-3-7481-0925-9.
- Bekius, Lamyk, 2021. "The reconstruction of the author's movement through the text, or how to encode keystroke logged writing processes in TEI-XML." *Variants* 15–16, pages 3–44.
- BLEEKER, ELLI and AODHÁN KELLY, 2018. "Interfacing Literary Genesis." In *Digital Scholarly Editions as Interfaces, Schriften des Instituts für Dokumentologie und Editorik*, volume 12, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, Norderstedt: Books on Demand (BoD), pages 193–218. ISBN 978-3-7481-0925-9.
- BLEIER, ROMAN and HELMUT W. KLUG, 2018. "Discussing Interfaces in Digital Scholarly Editing." In *Digital Scholarly Editions as Interfaces, Schriften des Instituts für Dokumentologie und Editorik*, volume 12, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, Norderstedt: Books on Demand (BoD), pages V–XV. ISBN 978-3-7481-0925-9.
- DILLEN, WOUT, 2018. "The Editor in the Interface: Guiding the User through Texts and Images." In Digital Scholarly Editions as Interfaces, Schriften des Instituts für Dokumentologie und Editorik, volume 12, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, Norderstedt: Books on Demand (BoD), pages 35–59. ISBN 978-3-7481-0925-9.
- EGGERT, PAUL, 2016. "The Reader-Oriented Scholarly Edition." *Digital Studies in the Humanities* 31(4), pages 797–810.
- Elsschot, Willem, 2007. *Achter de Schermen: Digital Edition*, edited by Peter de Bruijn, Vincent Neyt, and Dirk Van Hulle. Available from: https://www.beckettarchive.org/ads/Ads.htm.
- Grésillon, Almuth, 2016. Éléments de critique génétique: Lire les maunscrits modernes. Paris: CNRS Editions, 2nd edition.

- Kestemont, Mike and Luc Herman, 2019. "Kunnen machines (literatuur) lezen?" Tijdschrift voor Nederlandse Taal- en Letterkunde 135(4), pages 316–28.
- Kirschenbaum, Matthew G., 2007. "The Remaking of Reading: Data Mining and the Digital Humanities." In *The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*, Baltimore, page n.p. Available from: https://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf.
- ———, 2016. *Track Changes: A Literary History of Word Processing*. Cambridge, MA: Harvard University Press.
- Kirschenbaum, Matthew G. and Doug Reside, 2013. "Tracking the Changes: Textual Scholarship and the Challenge of the Born Digital." In *The Cambridge Companion to Textual Scholarship*, edited by Neil Fraistat and Julia Flanders, Cambridge: Cambridge University Press, pages 257–88.
- Lebrave, Jean-Louis, 2011. "Computer forensics: la critique génétique et l'écriture numérique." *Genesis* (33), pages 137–47.
- Leijten, Mariëlle and Luuk Van Waes, 2013. "Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes." Written Communication 30(3), pages 358–92.
- Leijten, Mariëlle, Luuk Van Waes, Karen Schriver, and John R. Hayes, 2014. "Writing in the Workplace: Constructing Documents Using Multiple Digital Sources." *Journal of Writing Research* 5(3), pages 285–337.
- LINDGREN, Eva and Kirk P.H. Sullivan, 2006. "Analysing Online Revision." In *Computer Keystroke Logging and Writing: Methods and Applications, Studies in Writing*, volume 18, edited by Eva Lindgren and Kirk P.H. Sullivan, Oxford: Elsevier, pages 157–88. ISBN 978-0-08-044934-0.
- Mathijsen, Marita, 2009. "Genetic textual editing: the end of an era." In *Was ist Textkritik? zur Geschichte und Relevanz eines Zentralbegriffs der Editionswissenschaft*, edited by Gertraud Mitterauer, Werner Maria Bauer, and Sabine Hofer, Tübingen: Niemeyer, number Bd. 28 in Beihefte zu Editio. ISBN 978-3-484-29528-5.
- MILLER, KRISTVAN S and KIRK P.H. SULLIVAN, 2006. "Keystroke Logging: An Introduction." In *Computer Keystroke Logging and Writing: Methods and Applications, Studies in Writing*, volume 18, edited by Eva Lindgren and Kirk P.H. Sullivan, Oxford: Elsevier, pages 1–9. ISBN 978-0-08-044934-0.
- MORETTI, Franco, 2000. "Conjectures on World Literature." New Left Review (1), pages 54–68.
- ——, 2013. Distant Reading. London & New York: Verso.

- RIES, THORSTEN, 2018. "The rationale of the born-digital dossier génétique: digital forensics and the writing process with examples from the Thomas Kling Archive." *Digital Scholarship in the Humanities* 33(2), pages 424–391. doi: https://doi.org/10.1093/llc/fqx049.
- Schäuble, Joshua and Hans Walter Gabler, 2018. "Encodings and Visualisations of Text Processes across Document Borders." In *Digital Scholarly Editions as Interfaces, Schriften des Instituts für Dokumentologie und Editorik*, volume 12, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, Norderstedt: Books on Demand (BoD), pages 165–91. ISBN 978-3-7481-0925-9.
- Taylor, C.M., November 2018. C M Taylor on 'keystroke logging project' with British Library. Publication Title: British Library English and Drama Blog, Available from: https://blogs.bl.uk/english-and-drama/2018/11/c-m-taylor-on-keystroke-logging-project-with-british-library.html (Accessed: 2020-07-16).
- Van Hulle, Dirk, 2014. *Modern Manuscripts: The Extended Mind and Creative Undoing from Darwin to Beckett and Beyond*. London: Bloomsbury.
- ———, 2019. "De logica van de tekstversie in digitaal geschreven literatuur." Tijdschrift voor Nederlandse taal en letterkunde/Journal of Dutch Linguistics and Literature 135(4), pages 465–77.
- Van Waes, Luuk, Mariëlle Leijten, Åsa Wengelin, and Eva Lindgren, 2011. "Logging Tools to Study Digital Writing Processes." In *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology*, edited by Virginia Wise Berninger, New York: Psychology Press, pages 507–34. ISBN 978-1-84872-963-6.
- Vauthier, Bénédicte, 2016. "Genetic Criticism Put to the Test by Digital Technology: Sounding out the (mainly) Digital Genetic File of El Dorado by Robert Juan-Cantavella." *Variants* 12-13, pages 163–86.
- Zeller, Hans, 1955. "Record and Interpretation: Analysis and Documentation as Goal and Method of Editing." In *Contemporary German Editorial Theory*, edited by Hans Walter Gabler, George Bornstein, and Gillian Borland Pierce, Ann Arbor: The University of Michigan Press, pages 17–58.