DHQ: Digital Humanities Quarterly
2019
Volume 13 Number 4

Finding and Interpreting Arguments: An Important Challenge for Humanities Computing and Scholarly Practice

Andrew Ravenscroft <a_dot_ravenscroft_at_uel_dot_ac_dot_uk>, Cass School of Education and Communities, University of East London, UK
Colin Allen <colin_dot_allen_at_pitt_dot_edu>, History and Philosophy of Science, University of Pittsburgh, USA; Philosophy, Xi'an Jiaotong University, Xi'an, China

Abstract

Skillful identification and interpretation of arguments is a cornerstone of learning, scholarly activity and thoughtful civic engagement. These are difficult skills for people to learn, and they are beyond the reach of current computational methods from artificial intelligence and machine learning, despite hype suggesting the contrary. In previous work, we have attempted to build systems that scaffold these skills in people. In this paper we reflect on the difficulties posed by this work, and we argue that it is a serious challenge which ought to be taken up within the digital humanities and related efforts to computationally support scholarly practice. Network analysis, bibliometrics, and stylometrics, essentially leave out the fundamental humanistic skill of charitable argument interpretation because they touch very little on the *meanings* embedded in texts. We present a problematisation of the design space for potential tool development, as a result of insights about the nature and form of arguments in historical texts gained from our attempt to locate and map the arguments in one corner of the Hathi Trust digital library.

Prologue: From Distant Reading to Close Reading

"We speak, for example, of an 'angry' wasp." This sentence appears in the first edition of Margaret Floy Washburn's textbook *The Animal Mind: An Introduction to Comparative Psychology*, published in 1908. It occurs as part of an argument she presents against the anthropomorphic idea that we humans can use our introspection of anger to understand the emotions of organisms so physiologically and anatomically different from us. One suspects that Washburn, whose story deserves more space than we can give it here, was intimately familiar with anger. She was the first woman to earn a PhD in psychology in the United States — albeit not from Columbia University, where she wanted to study. Columbia were unwilling to set the precedent of admitting a woman for doctoral studies. Instead she received her degree from Cornell University, where she was accepted to the Sage School of Philosophy under the mentorship of Edward B. Titchener, the pioneering psychologist who pursued a combined introspective and experimental approach to the human mind. Washburn's textbook would go through four editions, spaced roughly a decade apart, spanning one of the most consequential periods for psychology in its protracted separation from philosophy as a new experimental discipline. After World War II, Washburn's book faded from view. We discovered it in the digital haystack of the Hathi Trust with the assistance of computational methods we deployed to help us locate argumentative needles such as the sentence leading this paragraph, the kind of process one of the present authors describes elsewhere as "guided serendipity" [Allen et al. 2017]. 1

Our goal in this essay is to urge more attention in the digital and computational humanities to the important scholarly practice of interpreting arguments. We describe what we learned from our attempt to take an argument–centered approach to humanistic enquiry in a big digital repository. We acknowledge that the methods and approach we adopted represents an initial attempt to explore a complex digital humanities problem, and can be improved upon, as one of our main aims is to draw attention to this problem and spur further work in this area. We believe we have provided a road map to guide future work — or, at least, an analogue to one of those early maps of the world drawn by explorers, no doubt distorting the major land masses, but better than nothing. If not dragons, wasps lie here, and although much of the work described here involved good old-fashioned human interpretation, our discovery of Washburn's textbook and the "angry" wasps therein can be credited to the power of the computational methods we used to locate arguments about the anthropomorphic attribution of mental qualities such as anger to nonhuman animals. 2

Some of our work has been previously outlined in other publications that focused on our multi-level computational approach [Murdock et al. 2017] and a technical investigation of the challenges of automated argument extraction [Lawrence et al. 2014]. Here, for the first time, we provide more detail about the human component of argument identification, extraction and representation scaffolded by the use of topic models to find relevant content. Through a two-stage, topic-modeling process, we drilled down from a book-level model of a large corpus (too large to read in a decade) a page-level model of a smaller subcorpus (still representing at least a year's reading). This allowed us to select a few dozen pages from six books containing arguments that were mapped in detail within a few weeks by a team member with no prior expertise in psychology or the history & philosophy of science. The argument maps produced by this step of human interpretation allowed us to identify statements that could be fed back into a third level of topic modeling, drilling down to the level of sentences in a single book. In this way we were able to discover other relevant arguments within the same text, including the one about "angry" wasps and another about the cognitive powers of spiders. 3

Automated argument extraction, also known as argument mining, has significant challenges and remains a holy grail of artificial intelligence research (e.g., see Mochales and Moens, 2011; ACL, 2018). Our approach contributes only minimally to solving that problem [Lawrence et al. 2014] and, in fact, we doubt it is truly solvable with existing methods. Nevertheless, we propose that the digital humanities should invest more effort in developing argument-centered approaches to computational text analysis. We could be provocative and say that stylometrics and bibliometrics are the low-hanging fruit of digital humanities, and it is time for the digital humanities to take up challenges that may be harder, but which have more real-world impact. The skill of interpreting arguments is a cornerstone of education, scholarship, and civic life. Arguments are fundamental to human meaning making and to the maintenance, and reform of social norms. Even if the field of artificial intelligence is a long way from being able to properly interpret arguments in context, humanities scholars can use tools that are not so far out of reach to assist in their analysis and interpretation of the arguments that structure discourse in both academic and public domains. Interpreting arguments as they appear in historical documents brings them alive, allowing scholars, students, and citizens to understand their relevance for current issues. But before the arguments can be interpreted, they must first be found. As we demonstrate in this paper, available computational methods can strongly assist with that. 4

Exploring Arguments in the Digital Sphere: Animal Minds as a Proxy Domain

We focused on the early 20th Century debate about animal minds because, in the aftermath of Darwin's revolutionary effect on biology, it was a particularly fertile arena for historically important arguments that were still poised between scientific and literary styles of writing, and also for the pragmatic reason that it fitted our prior expertise in psychology, ethology, and philosophy of cognitive science. The debate remains lively in academic circles more than a century since Washburn published her book, and it is, of course, important to the ongoing public debates about animal welfare and animal rights. A close reading of Washburn's text reveals to a modern reader a mixture of familiar and unfamiliar arguments, many of which deserve revisiting today. Our work also led us to five other texts (described below), which present a similar mixture of the familiar and the unfamiliar. Anyone who engages closely with the arguments in these books learns much about the trajectory that psychology in the English-speaking world was on, and also comes to understand how current debates about animal minds are dependent on the paths laid down these earlier authors. 5

The late 19th century and early 20th century was a period of significant development for psychology that was characterised by important and competing arguments. Experimental methods were on the rise, and psychologists, who had often been housed in the same university department as the philosophers, were professionalising, forming their own associations and journals, and their own departments. Philosophy could be seen as retreating from the arguments based on experimental evidence increasingly favored by psychologists, while psychologists were wondering which of their received concepts and theories should be jettisoned, and which could form the basis of further empirical investigation. Such questions were particularly acute in animal comparative psychology. On the one hand, Darwin's theory of evolution exerted a strong pull towards the idea of mental continuity between humans and animals. On the other hand, many Darwinians were seemingly content with anecdotal evidence of animal intelligence to make their case on analogical grounds to human behaviour, leading experimentally inclined psychologists to reject such anecdotes and analogies as "anthropomorphic". Even as the disciplines of psychology and philosophy were formally disassociating themselves, philosophical arguments about the "proper" way to study animal psychology were becoming even more prominent among the psychologists themselves. 6

While comparative psychology in the immediate post-Darwin era was a particularly fertile era for the interplay between philosophy and science, the domain we selected is not special. It serves as a proxy for any domain where interpretation remains open and debate inevitably ensues. The lessons learned from our attempt to find and interpret text about anthropomorphism in comparative psychology generalise to other domains. There is no substitute for reading the relevant texts closely, but there is similarly no substitute for computational *distant reading* of such a massive repository as the Hathi Trust in order to select which texts are the best candidates for close reading and extraction of their arguments. 7

The skills involved in interpreting arguments are essential in supporting and developing critical thinking and writing skills – even, and especially, where digital media predominate (e.g., Wegerif, 2007; Ravenscroft and McAlister, 2008; Ravenscroft 2010; Pilkington 2016). The volume and variety of this digital sphere provides opportunities for thinking, learning and writing within and across educational, professional and civic contexts. Across these contexts the need to identify, understand, and critically compare arguments is particularly important today to counteract a discourse in which accusations of 'fake news' and appeals to emotion are used to promote simplistic, insufficiently contextualised arguments and propositions, often overriding well evidenced and supported positions on a subject. There is a pressing need to support and promote scholarly practices focused on identifying, understanding and comparing written arguments that can occur within texts in massive data or document repositories. 8

The availability of massive document collections transforms the scale and complexity of the tasks of searching for and interpreting arguments, but these collections hold out great potential for understanding the academic and broader cultural contexts in which these arguments were historically and are presently situated. A key inspiration for our approach was to help inexperienced scholars simulate the way an experienced or expert scholar moves from macro-level views of document collections to micro-level close reading and interpretation of the key arguments in particular texts. 9

Of course, there will always be ethical issues, linked to any sociological and political framing around decisions about which digital collections to focus on. For example, the extent to which these may or may not be not-for-profit and available to the public. In our case, we worked with the HathiTrust collection, because it is a consortium of mostly public state universities – spearheaded by Michigan, Illinois, and Indiana – who retain ownership of the scanned content, up to the limits of the applicable copyright laws, although Google supported work to accelerate the scanning of these materials. The original proof-of-concept tool-set that we are proposing and discussing in this article is aimed at gaining insights, both conceptual and technological, about finding and interpreting arguments in digital repositories of any kind in principle. Therefore this work is aiming to be relatively generic in its positioning around what repository to focus on, although for pragmatic reasons also, the HathiTrust was particularly suitable because project members, and one co-author (Allen), were working at Indiana University at the time of this project, which facilitated the cooperation with the HathiTrust Research Center. 10

Investigation by Design

Our approach was also inspired by prior work on the methodology of "Investigation by Design" by one of the present authors [Ravenscroft and Pilkington 2000]. This work was originally developed to model and simulate collaborative argumentation practices [McAlister et al. 2004] [Ravenscroft 2007] leading to learning and conceptual development [Ravenscroft and Hartley 1999] [Ravenscroft 2000]. A key idea behind this approach is that technology which effectively enhances scholarship and learning practices should balance existing practices with the technological possibilities for enhancing that practice. In other words, we should not try to fundamentally disrupt the way that people approach texts, but seek to amplify and enhance their processes and practices so as to support more powerful learning and scholarly interpretation across a wider variety of contexts. In our application the existing practice consists of skimming texts for arguments followed by close critical reading of them, and the technological enhancements are (1) topic modelling to improve the searching and (2) argument mapping to improve the identification, analysis and interpretation of the arguments. The semi-formal nature of the mapping tool used in the second component forced us to reflect on what is required of close critical reading during the analysis, construction and representation processes. Furthermore, we believe the level at which we have designed our approach satisfies what Edwards et al. (2007) refer to as "below the level of the work", i.e., a level where "Neither the exact implementation of standards, nor their integration into local communities of practice, can ever be wholly anticipated" [Edwards et al. 2007, 16] (see also Edmond 2018). 11

Consider the challenge facing learners and researchers confronted with massive, digitised document collections that are not readily browsable in the way that shelves of library books once were. For one thing, many of the books 12

of interest have been physically shifted to deep storage facilities and must be called up one-by-one rather than whole shelves at a time. (In a recent article, Jennifer Edmond (2018) laments the loss of serendipity this entails.) For another, the digitised collection represented by the HathiTrust Digital Library is an order of magnitude larger than any single library collection, so what was one shelf may have become the digital equivalent of ten. When browsing shelves of physical books, readers might pull a book off the shelf, sample a few pages from the book, and decide whether to put it back or to check it out of the library for closer reading. In the digital library, that decision takes on a different character: on the one hand there is a sense in which we don't have to put anything back as we can carry out macroscopic analyses of very large numbers of texts; on the other hand we must still make selections for the closer readings that provide valuable insights that are currently beyond the reach of algorithms.

It is our view that a tool that links searching of massive document collections to close critical reading of key arguments therein would have significant value across educational contexts. It could make the practices of experienced scholars more systematic, efficient and powerful. Perhaps more importantly, it could empower and support less experienced learners to engage in systematic critical thinking and reasoning linked to identifying and understanding arguments, which is a well-attested challenge throughout education (e.g., see Ravenscroft et al., 2007; Andrews, 2009; Ravenscroft 2010). Although previous research has shown the value of argument mapping to support greater "sense making" and learning in general, this work has involved "standalone" mapping tools [Kirschner et al. 2012] that do not link the maps to the larger textual and intellectual context in which they arise. 13

At the time we conducted the work upon which we base our discussion here, public access to the HathiTrust Digital Library was restricted to the approximately 300,000 volumes outside copyright and in the public domain in the United States. The HathiTrust now provides non-consumptive access to over 17 million volumes (as of November 2019), increasing the challenge of identifying key texts from unreadable quantities of text for the purpose of close reading and argument extraction, making it even more important to develop techniques and tools such as those we discuss here. A primary challenge at this scale concerns how to identify and compare argumentation and arguments within and across texts, in a way that is analogous to the way a scholar works, moving from a macro-level view of texts to the close critical reading of particular arguments within and across texts. This work (whose technical details are reported by McAlister et al. 2014 and Murdock et al. 2017) represented the first time that topic modeling and argument mapping had been combined in a process that allowed a scholar to identify pages within texts that should be fed into the argument mapping task, both necessitating and supporting a close critical reading of those texts by the individual engaged in the process. This work, through ostensibly technical research combining Big Data searching and AI techniques, included a broader exploration of the possibilities for integrating science mapping and visualization, along with an initial attempt at argument extraction [McAlister et al. 2014]. In this paper we provide a detailed critical examination of the nature and form of arguments that were identified in the texts, and we consider the centrality of the interpreter and the interpretative processes in extracting these arguments given their historical and cultural contexts. This critical examination supports our wider reflections on the role of such technical methods in supporting the identification, interpretation and comparison of important historical arguments. These reflections provide the basis for our 'bigger vision' concerning the important challenge of understanding arguments via the digital humanities, and the broader implications for any field where identifying and interpreting digital arguments is important, or vital. 14

Searching and interpreting as a pedagogical practice: The challenge of identifying, analysing and understanding arguments in texts

Texts do not give up their meanings easily, and different branches of the humanities bring different interpretative strategies to bear on the very same texts. For instance, philosophy students and scholars seek to understand conceptual frameworks and arguments that are typically not fully explicit in the texts they study. History students and scholars studying the very same texts may seek different kinds of clues to assist in their interpretations, such as facts about the social and cultural milieu in which they were written, or the specific contacts and experiences that led to particular acts of authorship. Literature students and scholars may focus on narrative structure in those texts, and the extent to which a given piece of work follows or flaunts literary conventions. 15

When the goal is also to exploit large datasets in support of traditional humanities research and learning, it is necessary to answer the question of how computational methods might help these kinds of students and scholars alike. For instance, consider the history scholar or student who already knows the biographical details of a 19th Century author, but wants to understand the narrative or argumentative structure of specific passages in that author's work. Scientometric methods such as the analysis of co-author and citation networks [Shiffrin and Börner 2004], and text mining methods such as named entity recognition [Nadeau and Sekine 2007] may provide hints 16

about influences on a given author, but unless these are linked to more powerful tools for textual analysis and critical work, the role of these methods is limited to very early stages of investigation for scholars pursuing disciplinary research within the humanities. Likewise, while search engines may be useful for discovering and retrieving individual documents and even key passages, they do not help with the interpretative task of distinguishing between passages where an author is accepting a particular concept, making a particular argument, or following a particular convention, and passages where those concepts, arguments, and conventions are being attacked or rejected.

17 To serve scholars and their students well, it is necessary to develop techniques for deeper analysis of the texts they care about. Sophisticated quantitative analysis of the full contents of texts will be needed. But computational methods alone will not suffice. Progress towards more effective use of massive text repositories will require a combination of computational techniques, digital curation by experts, and a better understanding of the way texts are critically understood and used in scholarly practices. No single method alone holds the key. Researchers and students need to be able to engage with the texts and discuss them with peers. Students and interested amateurs can in turn benefit from the discussions among experts if those can be adequately summarised and represented. People participating in debates may benefit from being able quickly to locate sources, both ancient and modern, that support or controvert their positions. There are many open research questions here about the design of effective systems that can serve scholars, and facilitate the representation of their knowledge in ways that others, experts and non-experts alike, can make use of in their critical engagement with the texts.

From Massive Document Repositories to Argument Identification

18 It is somewhat self-evident that massive document repositories offer access to an unparalleled number of texts across historical and disciplinary dimensions, opening up new possibilities for learning and scholarly activity. But, in practice, with so much choice about what to read, how do we decide which texts and parts of texts to focus on? And similarly, how can we focus on the key arguments within these texts to support the close reading and understanding of them? This is not just valuable in itself, it also counters the practice of reading texts in a fast, superficial and uncritical way, which is the temptation when we have access to such a massive quantity of text and information.

Why topic modelling to locate arguments?

19 Previous attempts at automated argument identification (e.g., Moens et al. 2007) have focused on key words and phrases which may indicate the introduction of premises (“for this reason”, “in virtue of”, etc.) or conclusions (“hence”, “therefore”, etc.). However, given a) the enormous variety of such markers, b) the historically shifting patterns of usage, and c) how many arguments are presented without such markers, such approaches can have significant limitations. Even when enhanced to use grammatical structure [Palau and Moens 2009] they face the additional weakness that that they do not capture the semantic content of arguments.

20 The set of documents accessible via the HathiTrust provide a robust test of our approach, as particular difficulties of understanding arguments from this historical era are: a) not all the content is congruent with the style of scientific thought and writing that we have come to expect in the modern era (e.g., the heavier reliance on anecdotal evidence in earlier times); b) the language used even in scientific publications is indirect, and verbose compared with its modern-day equivalent (e.g., there may be long digressions), and c) what passes for acceptable argument may well have been different in that era (e.g., the variety of rhetorical strategies). This problematisation contrasts significantly with other formal approaches to argument modeling, that have focused on articles with a modern, formulaic structure, e.g., in legal contexts [Moens et al. 2007] or in the context of “modern” scientific articles [Teufel and Kan 2009] [Merity et al. 2009] where “Introduction”, “Results”, “Conclusions” etc., are explicitly identified. The type of texts we were interested in were historically and scientifically important, but written in a common and more natural style, so we were deliberately giving ourselves a hard problem, but one with high authenticity and relevance. The task of understanding, identifying and mapping arguments in these more “free running” social science or philosophical (and historical) texts could be considered an “order of magnitude” more challenging than previous work into argument mapping (e.g., Lawrence et al., 2012; Kirschner et al., 2012).

21 Most scholars are interested in arguments not simply for arguments’ sake, but because of the underlying topics and issues that are addressed in those arguments. Computational methods offer a variety of ways for capturing semantic relations in text. Some, such as Latent Semantic Analysis (LSA) [Landauer and Dumais 1997] are good at capturing word-document relations, others are good at capturing word-word relationships (e.g., Word2Vec). For argument analysis, however, the right “chunks” for analysis are somewhere between words and whole documents.

We chose to explore LDA (Latent Dirichlet Allocation) topic modelling [Blei et al. 2003] as a means to find appropriately-sized, content-rich sections of text within books, which could then be subjected to further scrutiny for argument analysis and mapping. *So, our assumption was that the parts of texts that were rich in a particular topic would also be rich in the arguments that included that topic, and that assumption would be tested through our design and its application in the target domain.*

22 LDA topic modelling (LDA-TM) is by now a familiar technique in the digital humanities. It uses machine learning to represent documents as mixtures of "topics" and these are represented as probability distributions of the words in the corpus on which the model is trained. The training process automatically assigns probabilities to the topic-document and word-topic distributions in such a way that a relatively small set of topics (set by the modeler via a hyperparameter *K*) can account for the word distributions found in a relatively much larger set of documents comprising the corpus. As such, then, topic models accomplish a form of data compression, enabling common themes to be identified within a large corpus. Appropriate selection of the hyperparameter *K* for the number of topics depends on various factors including the size of the corpus and the pragmatic goals of the scholars using the model. As described in more detail below, we explored several different values of *K*, and settled on a number of topics that served our goal of identifying passages of interest for our argumentative analysis and interpretation. Also described in more detail below is the process we followed to select among and within the books. We made a number of design choices which reflected our pragmatic aim of designing a prototype toolkit that could demonstrate proof of concept, rather than pursuing a systematic investigation of the space of all possible measures and methods.

23 Going beyond the previous overview of our work by Murdock et al. (2017), here we focus in more detail on the pedagogical practice, through the link between the original *drill-down* topic modelling work and the nature, form and structure of the many arguments contained in these texts from the digital library. The detailed interpretation of the texts leading to semi-formal representation of the found arguments allow us, in this paper, to assess the importance and relevance of the "discovered arguments", and to problematise the design space.

Topic Modeling and Selection of Texts

24 Automated selection from large volume sets is necessary because one cannot hope to inspect by eye the whole collection. For example, although a standard keyword search in the HathiTrust collection, using "Darwin", "comparative psychology", "anthropomorphism", and "parsimony", reduced over 300,000 public domain works to a list of 1,315 volumes, this many books is on the order of Charles Darwin's entire personal library, accumulated and read over several decades. To help us to decide "what to read?" we chose to adapt topic modeling to our purposes. This technique is useful for information retrieval because it allows a higher level of semantic abstraction than keyword searching.

25 LDA topic modelling (LDA-TM) was first introduced by Blei et al., (2003), and it has been subsequently deployed in a variety of applications [Wei and Croft 2006] [Heinrich 2009] [Medlar and Glowacka 2017], including applications in the humanities [Tangherlini and Leonard 2013]. A key innovation of our approach is that we adopted a multilevel approach to a scholarly workflow [Murdock et al. 2017]. We first applied LDA-TM to these 1,315 volumes treating each book as a document. [1] The resulting topic model was scanned by a person who selected thresholds on the topics[2] to extract 86 volumes from the original 1,315, as those most closely related to our focus on anthropomorphism and animal minds. Amongst other advantages, the topic models allowed us to disambiguate discussions of anthropomorphism in the animal context from uses of the term in the context of comparative religion, allowing us to drill down efficiently to the most relevant materials. We then re-applied LDA-TM to these 86 volumes treating every page as a document. A further step of topic-model assisted selection rated books according to the number of pages containing a high density[3] of the topics we were interested in. This yielded six books of central interest for our argument analysis. It was notable that none of these texts appeared in the first ten results of libraries standard keyword searching.

26 The six volumes selected by the methods described above each discuss our chosen topic of Animal Psychology:

1. *The Animal Mind: A Text-Book of Comparative Psychology* by Margaret Floy Washburn, 1908
2. *Comparative Studies: Psychology of Ants and of Higher Animals* by Eric Wasmann, 1905
3. *The Principles of Heredity* by G. Archdall Reid, 1906
4. *General Biology* by James G. Needham, 1910
5. *The Nature & Development of Animal Intelligence* by Wesley Mills, 1898

6. *Progress of Science in the Century* by J. Arthur Thomson, 1908

Selection of rated pages and argument maps

We decided to adopt the visual argument mapping approach for a number of related reasons. Previous research has strongly supported the value of argument mapping for: greater "sense-making" of argumentative texts [Kirschner et al. 2012]; providing standardized and comparable semi-formal and visual representations to support the investigation and analysis of arguments generally [Reed at al. 2007]; and, providing visual representations that could be rendered into a generic computational format, the Argument Interchange Format (AIF), see Chesnevar et al. (2006), that can be re-used and shared between applications. In our case, this meant that the argument mapping approach that we adopted (see below) supported the close critical reading of the text selections through an argument "lens" and provided a standard representational scheme that could be applied across the different texts, showing the "found" arguments in each. Once mapped, these representations can be potentially re-used and shared in further argument inquiry or tool development. Further details of the mapping tool and process, and how it was used to interpret the texts and arguments that are specific to our study are provided below. 27

The rating of pages according to their loading on topics of interest was taken as an indicator of material worthy or argument analysis and mapping, but these were not used to limit arguments that started before or ended after the rated pages. Thus, each argument selected by the person doing the mapping spanned rated pages, but may also have spanned unrated bordering pages occasionally. Also, not all rated pages that dealt with the chosen topic contained argument. Table 1 (below) shows the *Pages* that were selected from each *Volume*, following our topic modelling approach, and also the number of *Maps* for each *Volume*. This shows that the first three of the listed volumes, according to our topic modelling returns were potentially "argument rich", with their arguments therein creating 15, 10 and 8 maps respectively. For *The Animal Mind,* which contained many more rated pages than listed in the table, we chose to limit our analysis to 40 pages constituting the largest blocks of contiguous pages containing pages with greater than 90% loading on the topics of interest. 28

The latter three in the list were potentially less rich in argument, creating 2, 5 and 3 maps respectively. This difference indicates the variability in writing style during this historical period, with some texts showing clearer lines of argument than others. *General Biology* is a textbook that follows a more didactic, less argumentative style, and differs from Washburn's psychology textbook, in that the there is a less controversial set of accepted "facts" to present. The fifth text is based on predominantly personal observation, so, it is a piece of anecdotal comparative psychology, and not concerned with the methodological questions that lead to the argumentative structure of Washburn's book. The final text has fewer arguments because it is a "pop-science" book and is more engaged in telling a triumphal narrative of scientific progress, rather than dealing with controversies in the field. It does have a section on animals that emphasises the discoveries that seem to show how intelligent they are, so it does not aim for the sort of complex analysis that is provided by Washburn. So, considering these findings lends support to our assumption that the "topic rich" texts according to our topic modeling method also approximate the degree to which the content is "argument rich". 29

Volume	Maps	Pages
The Animal Mind	15	13-16, 16-21, 24-27, 28-31, 31-34, 58-64, 204-207, 288-294, total = 40 pages (original page numbering)
The Psychology of Ants	10	Preface, 15-19, 31-34, 48-53, 99-103, 108-112, 206-209, 209-214, total = 37 pages (renumbered; Original page numbering masked by a bug.)
The Principles of Heredity	8	374, 381, 382, 385, 386, 390, 394, 395, total 10 pages (renumbered)
General Biology	2	434-435, 436 total = 3 pages (original page numbering)
The Nature & Development of Animal Intelligence	5	16-18, 21-26, 30-32 total = 12 pages (renumbered)
Progress of Science	3	479-484, total = 6 pages (renumbered)

Table 1. Page lists of analysed pages from selected volumes

The argument content was mapped using OVA+ [Janier et al. 2014] an application which links blocks of text using argument nodes. OVA+ (ova.arg-tech.org) provides a drag-and-drop interface for analysing textual arguments that it is designed to work in an online environment, running as a HTML5 canvas application in a browser. This particular 30

tool was chosen because it builds on the established work in argument diagraming and mapping referred to above [Reed at al. 2007] and because it is also a widely used argument mapping tool that also incorporates and generates the standardised Argument Interchange Format (AIF) that has been used by many other projects in computational argumentation [Lawrence et al. 2012]. Using this tool each argument is divided into propositions and marked up as a set of text blocks. These text blocks containing propositions were linked to propositions that they support, or undercut, to create an argument map, such as the one below (e.g., Figure 1).

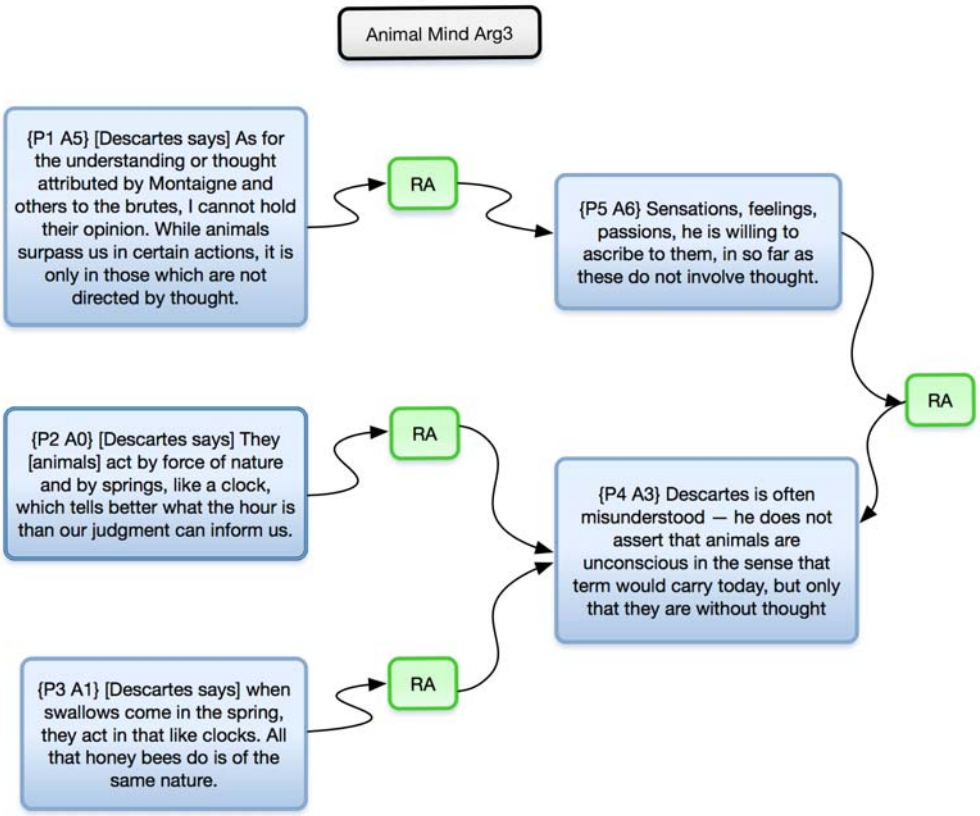

Figure 1. Argument Map of Argument 3 (Arg3) of *The Animal Mind* by Washburn (1908)

Argument Interpretation: Identification, Analysis, and Mapping

To identify the form and structure of the arguments contained in the selected texts we adapted a generic approach for manual argument analysis described by Lawrence et al. (2014). Through considering this work we developed a bespoke rubric that standardised and described the interpretative process that linked the analysis of our "historical" texts to the argument format of the mapping tool. This was informed by the members of the team with expertise in the humanities, who were familiar with the styles of writing about this topic for this historical period, and the researcher who was performing the mapping process. This was important in our case because, as mentioned earlier, the "natural" arguments contained in these texts, demanded more sophisticated interpretation compared with other applications where the arguments were more clearly defined. The full detail of this interpretative rubric can be accessed online https://bit.ly/35CshTD. To summarise it for the purposes of this paper:

1. **Initial Reading:** Read through the selected text to get a broad-brush overview of the nature and meaning of the arguments in play
2. **Argument Identification:** Mark beginnings and ends of major argumentative chunks (could span multiple pages) from where topic/conclusion is introduced to where it is concluded. This may be informed by linguistic identifiers (e.g., "because", "therefore", "suggesting that" etc.) where these are present
3. **Argument Segmenting:** For each paragraph, select zero or more sentences or whole-clauses that best summarise the arguments in this text. Unless they also contribute to arguments made by the author of the text, do not select sentences or clauses from reports of arguments or other non-argumentative materials, e.g., background information. (Mark zero if the paragraph is entirely non-argumentative, e.g., descriptive or providing background context.)

4. **Structuring:** Link the elements together with *relations* that show the direction of reasoning, from premise to conclusion, and whether premises are supporting or counter-argument (*attack*) relations
5. **Mapping:** Transfer the highlighted sentences, text and their relations to OVA+ and review and interpret for accuracy and representativeness

Through interpreting and mapping the identified arguments in these texts the researcher produced the 47 OVA+ maps covering the selections from the six volumes, which can be viewed online[4], with the maps sequentially numbered and linked to each volume. An example is included in Figure 1. It shows an argument from the first text, Argument 3 (Arg 3) from *The Animal Mind*. The links drawn on the maps between propositions are of two types – supporting and counter-supporting (links labeled RA and CA respectively). Although OVA+ supports more link-types these were not used in this study. Instead we paid particular attention to interpreting the meaning and representing the sub-components of the argumentative text. Conclusions must be supported by at least one premise. Often the maps have sub-conclusions leading to main conclusions. Propositions that expand or explain other propositions are seen as lending support to them. A link connecting two propositions always links from one to another, with an arrow showing direction, where a supporting premise links to (points to) a conclusion or sub-conclusion. 32

The argument map (Arg 3) above contains text taken from *The Animal Mind* by Washburn (1908). The argument consists of 3 propositions (in the large boxes on the left) that support two related conclusions (in the large boxes on the right). The “RA” boxes contained in the directional arrows demonstrates that the propositions on the left (P1, P2, P3) support the conclusions on the right (P5, P6), where the latter are also interconnected, as indicated through pointing to a shared relation (an *RA*). In this example P2 and P3 combine to support the conclusion P4. The close reading of the content of this argument would emphasise Washburn’s sensitivity to the contrast between Descartes’ view and that of his predecessor Montaigne, and her emphasis on his use of the exquisite functioning of the behaviour of diverse species of animal as evidence for a sophisticated view of the relationship between consciousness and thought (one that is often obscured in current presentations of Descartes’ views on animal minds) along with her sensitivity to the shifting meanings of these terms over the centuries. How this particular argument fits into the more extensive close reading of the arguments is covered in the next section. 33

This approach was particularly appropriate for the volumes that we analysed, where, in some cases, the same topic is pursued for a complete chapter and so there are opportunities to map the extended argument. Given the way the arguments were differentially expressed, with some text being more easily mapped compared to others, the mapping process was quite sophisticated, yet followed the standardised rubric to maintain consistency of interpretation. 34

Interpreting identified arguments to support better understanding and learning

This “deep” identification, representation and interpretation process linked to the subsequent argument maps, including careful reading of the identified texts provided a “double lens” onto the arguments that provided a stronger interpretative platform than if these methods had not been applied. The identification, representation and mapping process was performed by a researcher who was familiar with the basics of argument mapping, who was neither a domain expert in comparative psychology nor experienced with extracting arguments from this kind of textual material.[5] Below we describe his interpretations. In the descriptions below, for accuracy and evidence, we refer to the argument maps that the descriptions refer to that are accessible online (see footnote 4), as there isn’t the space to display them in this article. The importance and level of scholarly merit and detail of these argument interpretations is the test of our approach. In particular, we were interested in whether a researcher who knew nothing about the domain could be supported through sophisticated and deep reading of the arguments when guided by the topic models and the argument mapping process. A sample and summary of the subsequent close readings and argument descriptions of the first two volumes, which generated the most maps (15 and 10 respectively), are given below for the purposes of this paper. And these descriptions are then followed by a summary of the interpretations across the texts to demonstrate how the arguments in the individual texts could be considered collectively to improve the understanding of the topic (of Animal Psychology) in general. The full close readings of the 47 Maps linked to the six volumes is given in McAlister et al. (2014).The descriptions below have been paraphrased and condensed from the original, with material enclosed by square brackets representing additional qualifying comments introduced by the present authors. 35

Volume 1 analysis – *The Animal Mind*, 1908

[In this first edition of her textbook, destined for four editions] Washburn sets the context for the debate on animal 36

consciousness. She meets the charge that animal psychology is necessarily anthropomorphic straight away, and admits there is a problem (Arg1). She introduces Montaigne's arguments for animal intelligence based upon the similarity of human and animal behaviours (Arg2) and follows with Descartes's opposing argument, that animals are clock-like machines, with no capacity for thought (Arg3). Washburn next presents Darwin as arguing on the basis of analogical claims, such as that animals reason because they are "seen to pause, deliberate and resolve". She asserts that Darwin's aim of defending his theory of evolution in face of ongoing controversy about the mental and moral gulf between man and animals, means that his claims cannot be taken at face value (Arg4). In contrast many physiologists argue that psychic interpretations are less preferable than biological explanations of animal behaviour in terms such as tropism [unconscious reaction to stimulation] (Arg5).

37 Washburn next summarises three main anti-mentalist camps or positions in the field (Arg6). She criticises the physiologists, the first camp, for ignoring or simplifying phenomena to fit a predetermined theory, and she argues that their approach yields a *reductio ad absurdum* when applied to human behaviour (Arg7). Washburn outlines the arguments of ant expert Erich Wasmann [see next section], representing the second camp. Wasmann's definition of intelligence explicitly excludes animals on the grounds that they act only on instinct. He readily generalises from ants to all animals, stating that ants are superior to other animals (Arg8). The third camp is represented by Bethe [who belongs to an ultra-Cartesian group], holding that animals lack even sensation. Washburn identifies an inconsistency between his acknowledging that modifiability of behaviour is an indicator of consciousness, while considering this improper if applied to animals. He condemns all psychology as subjective and unknowable, and asserts that only chemical and physiological processes should be the object of scientific investigation (Arg9 and Arg10).

38 Washburn argues for a cautious approach to animal psychology, acknowledging pitfalls and problems but seeking scientific methods to overcome them (Arg11). She introduces Lloyd Morgan's [famous] Canon whereby the simplest level of psychic faculty for an animal should be assumed that can fully explain the facts of a case. She argues that the choice may not always be the right one, but at least it reduces anthropomorphism by compensating for a known bias (Arg12). Washburn next argues against Loeb's suggestion that "learning by experience" is a conclusive criterion for mind, but cautions that absence of proof does not amount to disproof. She maintains that rapid learning practically assures mind, but holds that great uncertainty remains about consciousness in lower animals (Arg13 and Arg14). Morphology and similarity of animals' physiology to humans' must be taken into account in deciding if an animal is conscious or not, and degrees of similarity indicate a gradation of consciousness, from lower to higher animals, with no possibility of drawing a sharp line between animals with and without consciousness (Arg15) [McAlister et al. 2014, 24–5].

Volume 2 analysis - Psychology of Ants, 1905

39 [Eric Wasmann was a Jesuit priest and naturalist, publicly renowned for his books about the variety of amazing ant behaviours.] Wasmann's concept that "intelligence is a spiritual power" leads him to the claim that if animals had this spiritual power "they would necessarily be capable of language". Animals don't speak, so animals don't have intelligence (Arg1). He supports his views of ants by reference to observations made by Aristotle, Stagirite, St Augustine, [and Wasmann's contemporary naturalist] Dubois-Reymond (Arg3). Wasmann denigrates suggestions by 'modern sociologists' that ant "states" and human republics can be equated, explaining that class differences arise from 'conditions of life' or 'intelligent' free choice in Man, but ant castes arise from organic laws of polymorphism [multiple body forms] (Arg4). Wasmann asserts animal intelligence is really sensile cognition and sensuous experience, but if higher animals are credited with intelligence, it would be inconsistent to deny ants the same (Arg5). He argues that ants achieve a more perfect level of social cooperation than even the higher vertebrates, such as apes (Arg7).

40 Wasmann criticises Darwin for his anthropomorphic stance towards the 'silence and obedience' of a group of baboons, which Wasmann reinterprets as 'fidelity and obedience', and takes to imply 'reasonable, voluntary subjection to the demands of duty and authority'. He argues that the more likely explanation is "the instinctive association of certain sensile perceptions with certain sensile impulses" (Arg6). This association removes the need to allow animals thought; instead, instinct is a sufficient explanation (Arg10). The author explains that instinct has two elements, 'automatism' of behaviour (generally found in lower orders of animals) and 'plasticity' of behaviour (generally found in higher orders). Because the architecture of ants' nests varies from species to species even when the physical attributes of the ants are highly similar, he argues that a simple explanation of the variety of architecture linked to physical attributes will not do; rather the decisive factor is the psychic disposition of the ant species (Arg8). Wasmann maintains that while ants 'verge on heroic unselfishness' towards their young, only 'Man' is conscious of

duty and the morals of parental love. Although he admits that some aspects of motherly love in humans are instinctual, motherly love cannot be attributed to animals because it is 'spiritual', based on awareness of duty that is unique to humans (Arg9). [McAlister et al. 2014, 25–7]

Summary of interpretation of arguments across six volumes

The section above demonstrates a sophisticated close reading of a sample of the arguments in the two selected texts, through incorporating the mapping approach into the interpretation process. For example, the comparison and contrast afforded by Washburn's survey of the arguments in the literature and her attempt to articulate a good scientific methodology for comparative psychology. This contrasts with Wasmann's more polemical and theological approach to the perfection of behaviour through instinct, which reveals that despite Darwin's work, published nearly 50 years earlier, much of the controversy revolves around whether humans have a special, perhaps God-given position, separate from the animal world. 41

A number of historically important themes emerged from the interpretation of the arguments in the six volumes that are given in full in McAlister et al., (2014). These demonstrated the ability of our selection and argument mapping methods to allow a reader, who was previously unfamiliar with the scholarship in this area, to zero in on the relevant passages and then acquire an understanding of the key themes, which is a measure of the success of those methods. Although it was not a primary goal of our project to produce new insights into the domain-specific content, these would somewhat hopefully and inevitably emerge from the close critical reading of the key arguments. So it is worth making some concise, content-specific remarks here about two of the themes that emerged from the six volumes, to demonstrate the potential value of the proposed approach. 42

(i) *Animal Flexibility*. All the authors, evolutionists and non-evolutionists alike, were willing to recognise hitherto unacknowledged flexibility and variability in behaviour of individual animals. They all identify the same extremes – excessive anthropomorphism on the one hand, and the conception of animals as automatic reflex machines on the other – but each claims the middle ground for their quite different positions! Even Wasmann, the lone anti-evolutionist in our sample, denies that individual ants are reflex machines, claiming that the flexibility of individual ants is of a "psychic variety" not "mechanical automatism", although he attributes this flexibility to "instinct" not reason. 43

(ii) *Developmental Approaches*. Three of the authors, Mills (1888), Reid (1906), and Needham (1910), explicitly advocate a developmental approach to the study of animal mind, operating within the framework of a strong nature-environment distinction (corresponding to today's "nature-nurture" distinction). They make the case for comparative developmental studies, particularly experimentally rearing animals in isolation. 44

Although the accounts (above) of the interpretation of the arguments are relatively concise, they demonstrate a successful *close reading* of the arguments located within the selected texts. And while the themes discovered should be compared with scholarly treatments of the same (e.g., Richards, 1987), nevertheless we believe that despite the variations in language (vocabulary and style), the crisscrossing overlap among the arguments discovered in these books indicates that our methods identified pages that were thematically relevant to tracking the scientific and philosophical debates about anthropomorphic attributions to animals in the late 19th and early 20th centuries. This provides confidence in our claim that the big-data analytic technique of topic modeling, linked to argument mapping, can support close reading of texts in a content-relevant, argument-guided way. 45

Discussion

The approach described in this article offers an initial prototype of a design for scholarly interaction with technology that begins with topic model-assisted search of massive document repositories and leads to close critical reading of the arguments in the texts therein. It has also produced important insights about the way these arguments are "rendered" and interpreted by a person new to such historical texts and work in the humanities. The automated content selection and categorization work described in this article demonstrated the feasibility and reliability of large-scale, fined-grained topic-based categorization across a range of topics in science and philosophy using documents defined at a variety of scales (whole books, book pages, and individual sentences in books). Categorization and selection are essential first-steps in the scholarly process of identifying further structures, such as arguments, in large data sets. Although it might have been possible to construct more sophisticated keyword searches using Boolean operators to identify the same pages of interest for our analysis, this would have required painstaking trial and error, whereas the topic modeling provided a relatively straightforward semi-automatic approach to narrowing down. A number of insights emerged from performing the human interpretation of texts that were delivered by our 46

topic modeling techniques and then mapped in argumentative terms through the argument-mapping tool OVA+.

47 Topic modeling was clearly successful in identifying the texts (chapters and pages) that contained the 'stuff' of arguments linked to the keywords and topics that were searched for, strongly supporting our assumption that we could approximate *topic rich* texts as also being *argument rich*. These could be sorted through rankings that allowed just the *topic rich* texts to be the focus of further analysis. This is very valuable in itself, as it allowed us to identify and extract 6 argument rich texts from a big data text repository (*HathiTrust*). Secondly the (human) argument identification and analysis produced 47 argument maps (in OVA+), that provided interpretations from six volumes, that also showed how the type and degree of argument in historical texts can be quite different, with the different texts producing different amounts of argument maps (ranging from fifteen to two). So, the quantitative and qualitative methodologies that we developed also enabled us to represent and distinguish different levels of argument within texts in a broad-brush way. The outcome is a set of powerful descriptive and comparative interpretations of arguments within and across texts, and linked to particular authors (see McAlister et al., 2014 for a full account).

48 Furthermore, we were able to leverage the human-constructed argument maps against a micro-level topic model trained on a single book with each sentence treated as a "document". Such an approach to Washburn's *The Animal Mind* led us from sentences represented in the maps to sentences in other parts of the book that were judged similar within the model and despite being wholly disjoint in vocabulary, including the "angry" wasp. Close reading was essential to determine why certain sentences were selected by this method. For example, the relevance to anthropomorphism of the sentence, "This, of course, does not refer to the power to judge distance," was not immediately evident. The context of this sentence in Washburn's footnote on p.238 is as follows:

> Porter observed that the distance at which spiders of the genera *Argiope* and *Epeira* could apparently see objects was increased six or eight times if the spider was previously disturbed by shaking her web. This, of course, does not refer to the power to *judge* distance. [Washburn 1908, 238] [Italics in original.]

Here, then, we see Washburn cautioning the reader not to jump to a high-level interpretation of the spider behaviour. The spiders may perceive objects at various distances but they don't judge it, where judgement is understood to be a high-level cognitive capacity. This belongs to a more elaborate argument against anthropomorphically over-interpreting the behavior of species remote from humans.

49 To summarise, here are five key points from this study:

1. We have demonstrated that topic modelling finds topic-rich text that is also potentially argument rich and worthy of careful argumentative analysis.
2. Mapping these topic-rich regions of historical texts using a computerised mapping tool (OVA+) and a suitable rubric supports, and necessitates, close critical reading of the arguments and the texts.
3. The argument mapping was often a complex process, needing interpretation and sometimes "gap filling" by the mapper, but this was cognitively valuable in supporting argument identification, representation and understanding linked to close critical reading. Some types of argument, e.g., historical arguments, are not simply latent and waiting for identification and representation. Rather, the arguments "come alive" through interpretation and the processes of mapping and then writing about them.
4. The exercise of mapping the arguments required critical reading by the non-expert. It manifestly contributed to his deeper understanding of the arguments and their scientific and philosophical contexts than simply reading the books alone without the scaffolding we provided. This is evidenced by his accounts covering all the found arguments and the summary and comparison of all of these (see also McAlister et al., 2014).
5. Further development of this approach should accept points 1-4 above, and emphasise support for the process of understanding, representing and refining argument representations and related conceptualizations. This means those who design such tools should focus more on the cognitive processes of actively reconstructing arguments from complex texts, rather than assuming that arguments might simply be identified and *extracted* from a frame provided by grammatical and terminological markers of arguments.

Critique and Further Work

Our emphasis on investigating and testing the feasibility of our computational tools to support existing scholarly practices of identifying and understanding arguments in digitised texts has meant that thus far we have deliberately prioritised validating technical possibilities over systematic empirical testing with different texts and/or different scholars. This suggests the need for further research that would incorporate technical and empirical strands into the development of the human-computer interaction. 50

The technical implications are that the next tool-set, should more closely connect the topic modelling to the argument mapping. Robust tools for topic modeling already exist in the form of *MALLET* [McCallum 2002] and the *InPhO Topic Explorer* [Murdock and Allen 2015]. The latter is also well integrated with the *HathiTrust Digital Library* so that now even copyrighted materials may be modeled (http://inpho.github.io/topic-explorer/htrc.html). However, these tools need to be better integrated with tools for visually structuring argument maps such as OVA+ so that the scholarly work potentially enhanced by these tools becomes more seamless. The system should scaffold the interpretation process from identified texts to argument mapping, as this reasoning and re-representation process is cognitively valuable in achieving better understanding of arguments. Similarly, once the text is identified and the related maps are produced, other scaffolding or visualization techniques could assist coordinating between these two related representations of argument, and among the different representations produced by learners having diverse interests and goals. In this respect, further work could draw upon the large body of work into the use of external representations for learning [Ainsworth 2006]. 51

Once a more integrated and user-friendly version of the toolkit is developed, it would support more systematic empirical investigation of the interaction between user and machine. Our hypotheses are that compared to unassisted argument identification and understanding, this approach would: find the argumentative parts of relevant texts much faster and with greater accuracy; scaffold deeper understanding; and, provide flexible and permanent representations that could be reflected upon, extended and re-used. Further and more generally, future work will accept the need to move towards an environment for constructing and developing representations of argument rather than simply mapping them. 52

The above appears a sensible conceptualization for future work, because through implementing our methods it became apparent that arguments were rarely neatly and clearly structured and defined explicitly in the texts. The historical distance to these texts, and the shift in academic writing styles over the past century served to make the task of *extracting* the arguments even more challenging. Indeed, rather than being set structures transmitted through the texts, instead these arguments came alive through the practice of interpreting, understanding and (re)constructing them. This raises the questions, “Do arguments actually exist in clearly defined forms within (certain) texts? Or, do arguments only take form when readers focus on understanding them?” When today’s reader encounters the seemingly verbose yet strangely enthymematic nature of yesterday’s arguments, what can we learn about the interaction between readers and texts, and about the minds of the authors and their original readers? 53

While these questions are too big to be answered by our original study, their potential validity as important questions are, we argue, supported. The notion that textual arguments are constructed through human interpretation is also supported by the observation that argument structure is notoriously difficult for people, even after training, to determine (see Oppenheimer & Zalta 2011, 2017 and Garbacz 2012 for an interesting example of disagreement among experts about how to formalise Anselm’s famous *ontological argument* in way that is adequate for computational validation). Of course, this should come as no surprise when even textbooks of argument analysis disagree with one another on the simplest of real-world examples. Yet the goal of using texts to construct arguments that satisfy disciplinary canons of interpretation of those texts defines an important scholarly activity. The abstraction provided by such efforts provides a regulative ideal that aids comprehension of difficult texts, and the representation of these abstractions in artifacts such as argument maps provides concrete targets for collaborative meaning making and deeper discussions about alternative interpretations of complex texts. The skill of generating such maps and interrogating their meanings is a legitimate aspect of mental agility and perspective taking, supporting a more sophisticated view of knowledge. The development of these skills, and the tools that support them, is essential for informed citizenship, particularly in our contemporary social media milieu. 54

Design investigations such as the one we have described here must remain mindful of the reconstructive nature of argument extraction.[6] Despite the claims of some A.I. proponents, computer scientists seem a long way from being able to design algorithms that match the interpretive skills and subtlety of human readers. Nevertheless, we believe we have supplied one proof of the concept that machine learning applied to big data sets can support this essential aspect of human scholarship by supplying tools for both discovery and representation of specific arguments in a specific content domain. And if we now return to the broad context of critical reading and writing in which our 55

research is placed, we argue that we have made significant technical and conceptual steps in moving towards tools that could enhance and empower this process for learners and scholars alike. This is particularly important in our contemporary digital landscape, where there is arguably an increasing need within the academy and without, to identify and understand reasoned and evidenced argument, to combat, for example, just simply “agreeing” or “disagreeing”, or “liking”, or not, simple emotive propositions and arguments.

In the application of digital tools to the humanities, we must also be mindful that high-sounding rhetoric about civic engagement, the democratization of scholarship, etc., can be undermined by the facts surrounding the choice of sources and limitations of access to the materials analysed. In our case, because of the association between HathiTrust and Google Books, some may worry (incorrectly in our estimation) that, despite its origins and continuation in publicly-funded universities, the HathiTrust nevertheless represents the sort of corporatisation of higher education that some find undesirable. Whereas we accept that there will always be challenging issues concerning which repositories to focus on, from a scholarly practice perspective our position is clear. We want to improve and democratize the scholarly practice of finding and interpreting arguments, so that argumentative and critical meaning making is potentially more inclusive, in addition to supporting deeper inquiry for those who are already engaging in such practice. 56

Conclusions

The research described in this article tackled a complex problem of how to investigate and design a technological platform that empowers and supports, or *scaffolds*, humanistic practices guiding a non-expert to perform the kind of search, argument identification, and interpretation of an experienced or expert scholar. We investigated this within our approach through using ‘drill-down’ topic modelling to move from macro-level views of a big data document repository, through identifying the main areas of interest in specific texts, then subjecting these areas to close critical reading through semi-formal argument identification, analysis and interpretation. We were also able to show how, with the argument analyses in hand, a further drill down to topic models at the sentence level of individual books could help identify content that had not been originally selected. This investigation has also provided insights into the nature, form and structure of arguments in historical texts, and how these features can be difficult to neatly isolate and also be variable, and require the human to “fit the pieces together”. This work provides an important problematisation of the design space for future tool development that should arguably focus, not on automatically extracting arguments, but instead focus on how to better interrogate, manipulate and understand them: a practice that has increasing importance and relevance within and without the academy. 57

Edmond notes that the digital tools currently available to humanists, focused as they are on text, do not fully reflect the much broader information gathering practices of humanists, which, in her phrase, remain “stubbornly multimodal” [Edmond 2018]. She argues that a certain kind of productive distraction, following leads where they may, is essential to scholarly creativity in the humanities. With respect to staying ensconced in the world of (digitised) text we are guilty as charged, unfortunately unimodal. The digital library is our easily-accessed tree, even if we would push digital humanist towards higher-hanging fruit. But we would argue that the approach we have outlined addresses some of the problems she outlines that arise from changes in the way libraries are organised in this era of digitised texts and catalogues. While we agree that “remote storage and electronic catalogues diminish the likelihood for serendipity” for reasons we already mention, we believe we have outlined a digital research environment for argument-based analysis in which serendipity arises. Following the traces provided by topic models led to sampling a few books in more detail, and then to the wasps, spiders, and amoebae that occupied the thoughts of comparative psychologists a century ago: creatures that have all re-emerged in the 21st century in discussions of non-human forms of cognition. The selections were assisted but not forced, allowing the individual scholar to follow whatever leads looked promising in light of whatever background information the scholar has gleaned from other sources. Guided serendipity resulted, and thus the “angry” wasp was found. 58

Acknowledgements

The research reported in this article derives from a project that was funded by the 2011 International Digging into Data Challenge. The project, entitled “Digging by Debating: Linking Massive Data Sets to Specific Arguments”, was co-funded in the UK by Jisc, the Economic and Social Research Council (ESRC), and the Arts and Humanities Research Council (AHRC), and in the US by the National Endowment for Humanities (NEH); the project title serves as the UK grant ID and the NEH grant ID is HJ-50092-12. The authors would like to acknowledge the work and intellectual contributions of the other co-PIs to this project, the Digging by Debating team — Katy Börner, David Bourget, and Chris Reed — and the various contributions of the staff and students who worked on the project: John 59

Lawrence, Robert Light, Simon McAlister, Jaimie Murdock, Jun Otsuka, Robert Rose, and Doori Rose (listed alphabetically). David Bourget and Colin Allen jointly developed the text-to-OVA+ argument mapping rubric with feedback from Simon McAlister. We are particularly grateful to Simon for his work on carrying out the argument mapping process itself. We are grateful, too, for the comments by an anonymous referee, who encouraged us to think more broadly about the political and ethical contexts of our work.

Notes

[1] The results of this step of the topic modeling can be explored at https://www.hypershelf.org/htrc1315/. All models were built using the InPhO project's Topic Explorer open source package [Murdock and Allen 2015] available for download at https://github.com/inpho/topic-explorer/README.md.

[2] We took a "naïve" approach to this, simply using the proportions of the documents assigned by the model to the topics of interest, and then choosing a threshold on the proportions that seemed to the person making the choice to be sufficient to capture the books relevant to comparative psychology (along with many irrelevant ones; i.e., we preferred recall over precision at this stage). See Murdock et al., (2017) for details.

[3] At this stage we preferred precision over recall. We again used a naïve approach, taking the mathematically expedient approach of summing the proportions across all pages and choosing an arbitrary threshold on the sums. Future work should explore more sophisticated information theoretic measures of relevance.

[4] Online link (http://bit.ly/1bwJwF9) to volumeData on Google Drive (view only). Open volume folder, open the pass subfolder, select a PNG image and see a Preview pane – click the blue OPEN button. A new tab will open – click button 100% to zoom. Move around by dragging the diagram.

[5] The depth of analysis we were seeking in this study would not have been feasible with a multi-user study, hence we focused on the pathway of one individual towards the process of extracting arguments as scaffolded by the available technologies.

[6] Although we have argued that 'extraction' is the wrong metaphor, we recognise that the term 'argument extraction' is likely to continue to be used for the actual constructive, interpretative process involved. Similarly, we continue to talk of the sun coming up even though we know it is really just coming into view with the Earth's rotation.

Works Cited

ACL 2018 *ACL Proceedings of 5th Argument Mining Workshop (2018)*. The Association for Computational Linguistics (ACL), Stroudsberg, USA (2018). http://www.aclweb.org/anthology/W18-52. Accessed December 30, 2018.

Ainsworth 2006 Ainsworth, S. "DeFT: A conceptual framework for considering learning with multiple representations". *Learning and Instruction* 16.3 (2006): 183-198.

Allen et al. 2017 Allen, Colin, Hongliang Luo, Jaimie Murdock, Jianghuai Pu, Xioahong Wang, Yanjie Zhai, Kun Zhao. "Topic Modeling the Hàn diǎn Ancient Classics". *Cultural Analytics* (October 2017). DOI: 10.22148/16.016.

Andrews 2009 Andrews, R. *Argument in Higher Education*. Routledge (2009).

Blei et al. 2003 Blei, David, Andrew Ng, and Michael Jordan. "Latent Dirichlet Allocation". *Journal of Machine Learning Research* 3 (January 2003): 993–1022.

Chesnevar et al. 2006 Chesnevar, C., J. McGinnis, S. Modgil, I. Rahwan, Christopher Reed, G. Simari, M. South, G. Vreeswijk and S. Willmott. "Towards an Argument Interchange Format". "Knowledge Engineering Review" 21.4 (2006): 293-316.

Edmond 2018 Edmond, Jennifer. "How scholars read now: When the signal *is* the noise". *Digital Humanities Quarterly* (2018). Accessible at: http://www.digitalhumanities.org/dhq/vol/12/2/000388/000388.html.

Edwards et al. 2007 Edwards, Paul N., Cory P. Knobel, Steven J. Jackson, and Geoffrey C. Bowker. "Understanding Infrastructure: Dynamics, Tensions, and Design" (2007). Accessible at: http://hdl.handle.net/2027.42/49353. Accessed December 27, 2018.

Garbacz 2012 Garbacz, Paweł. "Prover9's Simplification Explained Away". *Australasian Journal of Philosophy* 90.3 (2012): 585-592.

Heinrich 2009 Heinrich, G. "A generic approach to topic models", *Proceedings ECML/PKDD* (September 2009). Accessible at: http://www.arbylon.net/publications/mixnet-gibbs.pdf.

Janier et al. 2014 Janier, M., John Lawrence, and Christopher Reed. "OVA+: an Argument Analysis Interface". *Proceedings of 6th International Conference on Computational Models of Argument (COMMA 2014)*, IOS Press (2014): 463-466.

Kirschner et al. 2012 Kirschner, P.A, S.J. Buckingham-Shum, and C.A. Carr. *Visualising argumentation: Software tools*

for collaborative and educational sense-making. Springer (2012).

Landauer and Dumais 1997 Landauer, T.K. and Susan T. Dumais. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review* 104 (1997): 211-240.

Lawrence et al. 2012 Lawrence, John, F. Bex, Christopher Reed, and M. Snaith. “AIFdb: Infrastructure for the Argument Web”. “Proceedings of the 4th International Conference on Computational Models of Argument (COMMA 2012)”, IOS Press (2012): 515-6.

Lawrence et al. 2014 Lawrence, John, Christopher Reed, Colin Allen, Simon McAlister, Andrew Ravenscroft. “Mining Arguments From 19th Century Philosophical Texts Using Topic Based Modelling”. *Proceedings of the First Workshop on Argumentation Mining*. Baltimore, Maryland: Association for Computational Linguistics (2014): 79–87.

McAlister et al. 2004 McAlister, S., A. Ravenscroft, and E. Scanlon. “Combining interaction and context design to support collaborative argumentation using a tool for synchronous CMC”. *Journal of Computer Assisted Learning*, Special Issue: Developing dialogue for learning 20.3 (2004): 194-204.

McAlister et al. 2014 McAlister, Simon, Colin Allen, Andrew Ravenscroft, Christopher Reed, David Bourget, John Lawrence, Katy Börner, Robert Light. “From Big Data to Argument Analysis and Automated Extraction”. Final Report and Research White paper. Digging into Data Phase 2 (2014). [Programme/Project deposit], http://repository.jisc.ac.uk/5607/.

McCallum 2002 McCallum, A.K. *MALLET: A Machine Learning for Language Toolkit* (2002). Accessible at: http://mallet.cs.umass.edu.

Medlar and Glowacka 2017 Medlar, A. and D. Glowacka. “Using Topic Models to Assess Document Relevance in Exploratory Search, User Studies”. *Proceedings of Conference on Human Information Interaction and Retrieval (CHIIR)* New York, USA (2017).

Merity et al. 2009 Merity, S., T. Murphy, and J. R. Curran. “Accurate Argumentative Zoning with Maximum Entropy models”. *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*, ACL-IJCNLP (2009): 19-26.

Mochales and Moens 2011 Mochales, R. and M.F. Moens. “Argumentation mining”. *Artificial Intelligence and Law* 19.1 (2011): 1-22.

Moens et al. 2007 Moens, M.F., E. Boiy, R. Palau, and Christopher Reed. “Automatic Detection of Arguments in Legal Texts”. *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, ACM (2007): 225-230.

Murdock and Allen 2015 Murdock, Jaimie and Colin Allen. “Visualization Techniques for Topic Model Checking”. *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI-15)*. Austin, Texas (January 2015).

Murdock et al. 2017 Murdock, Jaimie, Colin Allen, Katy Börner, Robert Light, Simon McAlister, Andrew Ravenscroft, Robert Rose, Doori Rose, Jun Otsuka, David Bourget, John Lawrence, and Christopher Reed. “Multi-level computational methods for interdisciplinary research in the HathiTrust Digital Library”. *PLoS ONE* 12.9 (2017): e0184188. Accessible at: https://doi.org/10.1371/journal.pone.0184188.

Nadeau and Sekine 2007 Nadeau, D. and Satoshi Sekine. “A survey of named entity recognition and classification”. *Lingvisticæ Investigationes* 30 (2007): 3-26.

Oppenheimer and Zalta 2011 Oppenheimer, Paul and Edward N. Zalta. “A computationally-discovered simplification of the ontological argument”. *Australasian Journal of Philosophy* 89.2 (2011): 333-349.

Oppenheimer and Zalta 2017 Oppenheimer, Paul and Edward N Zalta. “A Computational Implementation of the Ontological Argument (Corrected)” (2017). Accessible at: https://mally.stanford.edu/cm/ontological-argument/. Accessed December 30, 2018.

Palau and Moens 2009 Palau, R.M. and M.F. Moens. “Argumentation Mining: The Detection, Classification and Structure of Arguments in Text”. *Proceedings of the 12th International Conference on Artificial Intelligence and Law (ICAIL 2009)*, Barcelona, Spain (June 2009): 98-107.

Pilkington 2016 Pilkington, R.M. *Discourse, Dialogue and Technology Enhanced Learning*. Routledge (2016).

Rahwan and Simari 2009 Rahwan, I. and G.R. Simari. *Argumentation and Artificial Intelligence*. Springer (2009).

Ravenscroft 2000 Ravenscroft A. “Designing Argumentation for Conceptual Development”. *Computers & Education*, 34 (2000): 241-255.

Ravenscroft 2007 Ravenscroft, A. “Promoting Thinking and Conceptual Change with Digital Dialogue Games”. *Journal of Computer Assisted Learning* 23.6 (2007): 453-465.

Ravenscroft 2010 Ravenscroft, A. “Dialogue and Connectivism: A new approach to understanding and promoting dialogue-rich networked learning”. *International Review of Open and Distance Learning*, Special Edition: Connectivism: Design and delivery of social networked learning, (eds.) George Siemens and Gráinne Conole, 12.3

(2010): 139-160.

Ravenscroft and Hartley 1999 Ravenscroft, A. and J.R. Hartley. (1999). “Learning as Knowledge Refinement: Designing a Dialectical Pedagogy for Conceptual Change”. *Frontiers in Artificial Intelligence and Applications Volume 50, Artificial Intelligence in Education. Open Learning Environments: New Computational Technologies to Support Learning, Exploration and Collaboration*, Lajoie, S. & Vivet, M. (eds.), IOS Press (1999): 155-162.

Ravenscroft and McAlister 2008 Ravenscroft, A. and S. McAlister. “Investigating and promoting educational argumentation: towards new digital practices”. *International Journal of Research and Method in Education*: Special Edition on Researching argumentation in educational contexts: new methods, new directions, (eds.) Caroline Coffin and Kieran O’Halloran, 31.3 (2008): 317-335.

Ravenscroft and Pilkington 2000 Ravenscroft, A. and R.M. Pilkington. “Investigation by Design: Developing Dialogue Models to Support Reasoning and Conceptual Change”. *International Journal of Artificial Intelligence in Education*, Special Issue on Analysing Educational Dialogue Interaction: From Analysis to Models that Support Learning. 11.1 (2000): 273-298.

Ravenscroft et al. 2007 Ravenscroft, A., R.B. Wegerif, and J.R. Hartley. “Reclaiming thinking: dialectic, dialogic and learning in the digital age”. *British Journal of Educational Psychology Monograph Series, Learning through Digital Technologies*, Underwood., J & Dockrell, J. (eds.), Series II, Issue 5 (2007): 39-57.

Reed at al. 2007 Reed, C., D. Walton, and F. Macagno. “Argument diagramming in logic, law and artificial intelligence”. *The Knowledge Engineering Review*, 22.1 (2007): 87–10.

Richards 1987 Richards, R. *Darwin and the Emergence of Evolutionary Theories of Mind and Behaviour*. University of Chicago Press (1987).

Shiffrin and Börner 2004 Shiffrin, R.M. and K. Börner. (2004) “Mapping knowledge domains”. *Proceedings of the National Academy of Science*s, 101 (suppl. 1): 5183-5185 (2004).

Tangherlini and Leonard 2013 Tangherlini, T. R. and P. Leonard. “Trawling in the Sea of the Great Unread: Sub-corpus topic modelling and Humanities research”. *Poetics* 41.6 (2013): 725-749.

Teufel and Kan 2009 Teufel, Simone and Min-Yen Kan. “Robust Argumentative Zoning for Sensemaking in Scholarly Documents”. *Advanced Technologies for Digital Libraries*. Springer (2009): 154-170.

Washburn 1908 Washburn, Margaret Floy. *The Animal Mind: A Textbook of Comparative Psychology*. The Macmillan company (1908).

Wegerif 2007 Wegerif, R. B. *Dialogic, education and technology: Expanding the space of learning*. Springer-Verlag (2007).

Wei and Croft 2006 Wei, Xing and W. Croft. “LDA-based document models for ad-hoc retrieval”. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’06)*. ACM (2006): 178-185. DOI: http://dx.doi.org/10.1145/1148170.