

On the Category of *Religion*: A Taxonomic Analysis of a Large-Scale Database

Monroe et al.*

ABSTRACT

The Database of Religious History is a large-scale digital humanities project dedicated to capturing scholarly perspectives on the history of religious groups across the globe. Analysis of the current state of the data shows a remarkable consistency between a taxonomic tree generated from the entries submitted by our expert contributors and larger assumptions within religious studies as they pertain to the similarities and differences between religious groups. Additionally, there is broad agreement between how experts answer questions and the tags they use to categorize their own entries, demonstrating a consistency between top-down and bottom-up approaches to describing religious groups. We see both of these results as affirming a commensurable understanding of the category of *religion* while demonstrating the value of these types of large-scale quantitative analyses for answering larger questions within the field.

THE Database of Religious History ([DRH](https://religiondatabase.org); religiondatabase.org), created in 2012, is an online quantitative and qualitative encyclopedia of religious cultural history. As described in a previous

* M. Willis Monroe, Department of Philosophy, University of British Columbia, Buchanan, 1866 Main Mall E370, Vancouver, BC, V6T 1Z1, Canada (present address: Department of Historical Studies, University of New Brunswick, 120 Tilley Hall, 9 Macaulay Lane, Fredericton, NB, E3B 5A3, Canada). Email: willis.monroe@unb.ca. Rachel Spicer, Department of Psychological and Behavioural Science, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK. Email: R.A.Spicer@lse.ac.uk. Caroline Arbuckle MacLeod, Department of History, St Thomas More College, University of Saskatchewan, 1437 College Dr., Saskatoon, SK, S7N 0W6, Canada. Email: carbuckle@stmcollege.ca. Gino Canlas, Department of Philosophy, University of British Columbia, Buchanan, 1866 Main Mall E370, Vancouver, BC, V6T 1Z1, Canada. Email: canlas.gino@ubc.ca. Travis Chilcott, Department of Philosophy and Religious Studies, Iowa State University, 2224 Osborn Dr., Ames, IA, 50011, USA. Email: chilcott@iastate.edu. Stephen Christopher, Department of Cross-Cultural and Regional Studies, University of Copenhagen, Karen Blixens Plads 8, 2300 København S, Denmark. Email: scj@hum.ku.dk. Megan Daniels, Department of Ancient Mediterranean and Near Eastern Studies, University of British Columbia, 1866 Main Mall, Buchanan C227, Vancouver, BC, V6T 1Z1, Canada. Email: megan.daniels@ubc.ca. Andrew J. Danielson, Department of Philosophy, University of British Columbia, Buchanan, 1866 Main Mall E370, Vancouver, BC, V6T 1Z1, Canada. Email: andrew.danielson@ubc.ca. Matthew Hamm, Department of Philosophy, University of British Columbia, Buchanan, 1866 Main Mall E370, Vancouver, BC, V6T 1Z1, Canada. Email: mhamm@mail.ubc.ca. William Noseworthy, Southeast Asia Program, Cornell University, 180 Uris Hall, Cornell University, Ithaca, NY 14853, USA. Email: wn64@cornell.edu. Ian Randall, Department of Philosophy, University of British Columbia, Buchanan, 1866 Main Mall E370, Vancouver, BC, V6T 1Z1, Canada. Email: irandall@mail.ubc.ca. Robyn Faith Walsh, Department of Religious Studies, University of Miami, P.O. Box 248264, Coral Gables, FL, 33124, USA. Email: rxw159@miami.edu. Michael Muthukrishna, Department of Psychological and Behavioural Science, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK. Email: m.muthukrishna@lse.ac.uk. Edward Slingerland, Department of Philosophy, University of British Columbia, Buchanan, 1866 Main Mall E370, Vancouver, BC, V6T 1Z1, Canada. Email: edward.slingerland@ubc.ca. All data and code used for analysis are available at https://github.com/religionhistory/religion_taxonomy. The tree is also available to view interactively at https://rachel-spicer.shinyapps.io/drh_tree/.

publication in this journal (Slingerland and Sullivan 2017), the DRH originated as a subproject within a larger, Canadian government-funded grant on the cultural evolution of religion. The original aim was to convert the massive amount of qualitative information about religious belief and behavior, across the globe and throughout history, into quantitative data. This would allow hypotheses about the cultural evolutionary dynamics of religion to be more rigorously tested against the historical record, an effort we characterized as “Durkheim with data.”

After the original grant expired, the DRH continued as an independent project funded by grants from Templeton Religious Trust and the John Templeton Foundation and was given a permanent home at the University of British Columbia. Although the original goal of creating a body of structured, quantitative data has remained a guiding principle, the focus of the DRH shifted to responding to features and uses that would appeal to traditional humanities scholars with no particular interest in testing hypotheses or even the digital humanities per se. In response to the rapid increase in publications within religious studies and associated disciplines that can make it difficult to impossible for most scholars to keep up with the secondary literature in their fields, the DRH offers scholars an instant snapshot of scholarly opinion outside of their particular field of expertise. The capacity to add rich qualitative data, in the form of photos, videos, or text attachments, has enhanced the DRH’s usefulness as a general reference source, and the visualization and search functions have found many helpful pedagogical applications.¹ As an open-access resource, the uses and functionality of the DRH continue to evolve, and the DRH team has expanded to include six postdoctoral fellows, over thirty editors, seven external advisors, and (as of March 2022) 372 expert contributors.²

The DRH consists of entries organized around particular units of analysis, which currently include “religious group,” “religious place,” or “religious text,”³ tagged with a particular date range and map. Each entry consists of answers (with added comments and sources) to a long questionnaire consisting of a fixed set of questions specific to the chosen poll. Each entry is also categorized within the database using tags assigned by the person who created the entry. The combination of these two forms of data in aggregate then allows us to interrogate both the history of religion globally, as well as the category of *religion* within the field of religious studies.

The default form of entry in the DRH is the “expert” entry, where the poll is completed and comments made, by an individual with scholarly expertise on the topic. The rule of thumb is that an “expert” should be ABD or above in a related field at a recognized institution of higher learning, although individual judgment calls on the part of the editorial team allow for flexibility

¹ A series of lesson plans and instructor guides has been created and is hosted on our project’s website: <https://religiondatabase.org/landing/about/pedagogy#material>.

² As of November 2023, the project has over 550 contributors and 1,200 entries. For details on the current DRH team, see <https://religiondatabase.org/landing/about/people/team>. Expert contributors join the project through a semi-opportunistic recruitment, editors recruit directly from universities around the world, and other experts come to the project via word of mouth from colleagues. Once an expert has signed up, editors help them to establish the who, when, and where of the entry under consideration; that is, how to overcome definitional problems in establishing the criteria of inclusion, how to temporally bracket the entry into a non-arbitrary historical epoch, and how to delimit the geographical spread of the entry. These early steps require considerable finesse given definitional debates about religious groups and the vagaries of nontraditional cases such as Beat Buddhism (1944–1960s mystical intellectual experimentation in the United States; <https://religiondatabase.org/browse/991/#/>), Supreme Master Ching Hai (a transnational cybersect; <https://religiondatabase.org/browse/570/#/>), and Disabilities-Welcoming Protestantism (as defined by hermeneutic emphases on disabilities inclusion; <https://religiondatabase.org/browse/915/#/>). Editors shepherd the entry through completion, assisting scholars with potentially unfamiliar academic language pertaining to supernatural monitoring, moral realism, neutral cultural contact, and so on. Upon completion, the editor reviews the entry in its totality for consistency of the quantitative answers and thoroughness of qualitative notation and academic sourcing. Once completed entries are published, they are archived and assigned a permanent DOI by the UBC library and can be easily referenced and cited.

³ A fourth poll type, “Religious Object,” will be implemented soon. The plethora of ontologies allows experts to answer questions about a unit of analysis with which they feel comfortable. For instance, an expert on a given place or text might not wish to postulate an organized “group” behind its creation, and places and texts might be used by multiple distinct groups. The back-end mapping of related questions between these different poll types allows a coherent but complex picture of the religious landscape of a particular place and time to emerge organically.

on this front (e.g., experts for whom academic accreditation may have been historically unattainable). “Expert” entries are to be distinguished from “secondary source” entries, typically completed by an advanced graduate student RA relying on other sources; “expert source” entries, where an RA completes a poll using a given expert’s published works and then has the entry edited and approved by the expert; and “supervised entry,” where an instructor supervises one or more students, typically advanced undergraduates, in completing a poll or section of a poll.

As a digital open-access platform, the DRH—we believe—enhances the democratization of knowledge within the history of religion in several ways. Prospective experts can self-nominate, from anywhere in the world, reducing the power of graduate school ties or other elite networks that too often determine the contributor lists of edited volumes and handbooks. Although currently only available in English, Chinese, and French, we are hoping by summer of 2024 to expand to Spanish, Arabic, and other languages, allowing experts to engage with the website and polls in a language with which they are most comfortable. Our online presence allows us to access experts working on the ground across the world who are often excluded from other publishing venues. Traditional publication processes are long, whereas demands for research turnaround in universities in many parts of the world—especially East and Southeast Asia—are quite short. This means that, for some groups of typically underrepresented scholars, a digital platform such as the DRH is the best or only means for publicizing their work. Both our recruitment methods and accessibility of our platform mean that we have been able to include scholarly voices that are less typically represented in the field of religious studies.

INTRODUCTION TO THE PRESENT STUDY

The present study discusses the result of an exercise that we undertook to explore how entries in the DRH might be classified and related to one another through differentiations in their patterns of answers to specific questions. These early results offer an intriguing proof-of-concept analysis of how such a bottom-up approach—one that takes advantage of the unique quantitative nature and digital affordances of the DRH—might provide new perspectives on the categorization of religions in the cultural historical record.

The DRH began with the “Religious Group” poll and from the very beginning confronted scholarly concerns about both aspects of this term. To begin with, scholars in the field of religious studies have, of course, long wrestled with how to define *religion* or if the term even describes a stable category of human experience. The literature on the application of the term to the wide range of phenomena that make up religious studies is immense (Stausberg 2010). Suffice to say, the term *religion* is clearly not a universal category, and the degree to which it is entrenched in and encodes particular worldviews is particularly relevant to this study (McCutcheon 1997, 148–49; Nongbri 2013). Regarding purely linguistic evidence, for instance, its etymology from Latin shows the caution needed when applying the term to pre-modern contexts (Saler 1987). Likewise, the way the term can co-opt emic terminology (e.g., *śāsana*) through colonial interfaces in non-western contexts is of considerable interest (Hansen 2017).

Despite these important criticisms of the term and its application in antiquity, following Stanley Stowers (2008), the authors of this paper see the category of *religion* as an explicit second-order category that organizes and describes a wide variety of phenomena within scholarship. In fact, the studies below seek to test the idea that *religion* as a second-order category has usefulness within and across historical disciplines. This is core to the definition of *religion* that Stowers puts forward, that the category can only exist if it is “justified by its usefulness in scholarly enquiry” (Stowers 2008, 443). As noted by Stowers and the authors of this article, this formulation is not unlike the idea of *species* within biology. The concept of a *species* is not inherent in the biological record; rather, it is a system of categorization that depends on a series of

observations.⁴ Crucially, the criteria by which a species is defined has changed over time: early speciation was driven by physical characteristics; current divisions are governed by differences in DNA and the ability to interbreed.⁵ However, there is no single method of categorizing species that captures all of the nuances that reflect how organisms reproduce, and interpretations change over time (e.g., the position of various species of fungi). Despite these limitations, the concept of *species* is still considered a valuable classification to biologists. Our approach in this article is to treat the answers produced by the experts recruited to the project as a form of observations that allow us to test the usefulness of our questions in the creation of a coherent category such as “religious group.”

In *Imagining Religions*, J. Z. Smith proposed that a “religion”—that is, a “Religious Group”—could be defined or categorized by its “differential quality” (Smith 1982, 1–18). Borrowing from the biological sciences, he saw this project of classification as consisting of a polythetic taxonomy: a scaffolding of differentiating questions until a unique combination was reached, at which point the religious group under analysis was differentiated from its neighbors. Determining the nature of these questions, or differentiating qualities, is key to successful comparison. Smith illustrates how taxonomic classifications that ask a series of binary questions arrive at a final determination of to which categories the item under investigation *must* belong, a monothetic or Linnaean taxonomy. This technique is contrasted with a polythetic system of classification where items share a “set of properties” that define a class. This second method of classification resembles Wittgenstein’s family resemblances—in other words, radial categories—in that aspects of the definition are found across the entire group but no one member has all of the criteria. Radial categories can be used to identify clusters of beliefs and practices that occur across cultures and time⁶ and serve similar functions for the social networks in which they are found—in other words, radial categories seek to organize a range of related concepts around prototypical cases but without using rigid boundaries typical of Aristotelian categories. At the end of the day, the DRH adopts the approach that experts should not predetermine the parameters of their object of study. In other words, experts should not be concerned about whether the group or place or text for which they are interested in preparing an entry counts as “religious.” Instead, they are encouraged to look at the relevant poll questions and decide whether answering them, or some subset of them, would allow them to provide a coherent account of their group, place, or text restricted in time and space. It should be noted that no singular question in the poll determines inclusion within the database. This approach has the added benefit of allowing us to reflect on the applicability of our ontology as new entries come in that might challenge the category of “Religious Group.”

Perhaps the more difficult task is how one might define a group. Conscious of concerns about scholars creating artificial groups by assuming anti-historical cohesion or exaggerating the degree of homogeneity within a group by imposing etic labels on collections of practitioners, the project created a loose definition to guide our editors and experts in deciding what constitutes a group: “A community or network of people (locatable in space and time) who share common practices, beliefs, and/or institutions, but who are not necessarily conscious members of an explicitly recognized group. The group can be an emic (indigenous) name or category or

⁴ An early observation of the messiness of universal categories in biology shows the long struggle with the idea of *species*: “Of late, the futility of attempts to find a universally valid criterion for distinguishing species has come to be fairly generally, if reluctantly, recognized” (Dobzhansky 1937, 310).

⁵ To push this analogy further, the DRH is thus based only on the observation of religious “traits” and does not attempt to make claims regarding the relatedness or historical priority of one religion to another, which is what the approach of cladistics attempts to do for species on the basis of DNA.

⁶ For more on the idea of *religion* as a radial category and the role of ideas from the cognitive sciences within religious studies see Benson Saler (2010) and Edward Slingerland and Joseph Bulbulia (2011).

an etic (scholarly attributed) one.”⁷ Experts are encouraged to be geographically and temporarily narrow: that is, to keep the focus on the specific context. The ability to “tag” with labels (e.g., Christianity, Daoism), however, also allows one to track ties with other groups and larger identities, whether self-affirmed or imposed etically by scholars.⁸

A good illustration of this tagging practice is an entry in the database on the Meo in North-West India (Kukreja 2020). The expert in this case added the following tags to their entry: “religious group,” “Indic religious traditions,” “Islamic traditions,” “Tablighi Jamaat,” and “Meo Muslim.” The last two tags were created by the expert within the tagging tree and further identify this particular entry, but they also allow future experts to create and link closely related entries. This tagging system allows the expert a large degree of agency in how their entry is categorized within the database. Despite the issues surrounding categorization and terminology, whatever these groups are that some scholars have given the moniker of “religious,” they are presumed to share certain similarities and differences that can be tracked by the wider categories reflected in tagging labels such as “Judaism” or “Buddhism.”

Proceeding under this theoretical understanding of the “category” of *religion* and the demarcation of groups, the DRH has been collecting entries from experts in various fields associated with the study of the history of religion. These entries now make up a continually growing corpus of data on discrete religious groups in the historical record, one that allows for large-scale inspection and comparison across the differential qualities that make up a taxonomic structure.⁹ Through a large-scale statistical analysis of “Religious Group” entries, we have constructed a taxonomic tree that, by comparing similarities and differences across the entire dataset, organizes the entries in the database into precisely the sort of polythetic taxonomy described by Smith.¹⁰ To be clear, the analysis performed remains, in J. Z. Smith’s often quoted words, within the “scholar’s study,” but in doing so reflects the perspectives and traditions of a wider angle of scholars all working with a loose, radial conception of *religion*.¹¹ In essence, this exercise allows us to employ an inductive approach to classifying religious groups throughout time and space by relying on the work of scholars in their own fields to provide the raw points of data from which the classifying algorithm draws its comparisons.

In this analysis we are interested in two questions. The first is how religious groups will be classified and related to one another when a simple algorithm sorts them based on their pattern of responses to specific questions within the “Religious Group” poll. This taxonomic analysis of “Religious Group” entries¹² currently in the DRH, which ignores the tags applied to these groups by experts, gives us an opportunity to test established views concerning the similarity or differences between religious groups in the historical record from the ground up. Each point of data is an answer to a discrete question entered by an expert, and the construction of

⁷ For more on this definition and its application see Frederick Tappenden (2017). Although we invoke the language of community in the database, we are not unaware of the challenges and critique that attend its usage. For more on the parameters of such terms within sociological circles, see Rogers Brubaker (2004). For the problematic history of the term *community* politically and in the field of religious studies (and the study of Mediterranean antiquity and early Christianity, in particular) see Stowers (2011) and Robyn Walsh (2021).

⁸ These tags generally follow institutional or disciplinary ideas of classification, roughly summarized by the work of the Harper-Collins Dictionary of Religion (1995); for a thorough explanation of the issues of classification and definition, see Smith (1996).

⁹ Note that we are not proposing a phylogenetic tree here as that would imply inheritance and change over time, two forms of analysis that are not part of the current project. The emphasis on taxonomy is to look at similarities and differences between entries across the existing corpus.

¹⁰ Smith even used the terminology of clustering and statistical analysis in his writing about taxonomic projects: “We must conceive of a variety of early Judaisms, clustered in varying configurations” (Smith 1982, 14). Christopher Lehigh, writing on Smith, notes in his epilogue that he seemed dubious of mathematical techniques while still using the language of clustering and statistics in his discussion of building taxonomic trees (Lehigh 2021, 152).

¹¹ For additional discussion and critique of Smith’s claim, see Smith (1982, xi); Smith (2004, 5); Kevin Schilbrack (2017, 161–78).

¹² The process of filtering the data excludes entries that do not contain a sufficient number of “yes” or “no” answers to allow for comparison; more detail can be found in the Methods section below.

the questionnaire serves to distance the expert from prevailing narratives by asking narrowly focused questions. This is in contrast to a traditional encyclopedia entry where the author might describe the group from a top-down perspective.

The second analysis follows from the first: given the bottom-up tree structure thus constructed, how closely does it match the top-down classifications imposed by the experts who entered the data through the use of “Religious Group” tags? Here we are looking to interrogate the difference (or more accurately “distance”) between traditional terms used by our experts to categorize entries and the relative position of those entries on the taxonomic tree. This second analysis speaks to how and if systems of classification for religious groups typically employed in the field of religious studies actually track patterns of responses to specific questions about those groups in the DRH.

METHODS

The DRH “Religious Group” poll that serves as the data for this analysis consists of questionnaires constituted by a few hundred questions. Each entry is given a time range and geographical scope by the expert before answering the questionnaire. The questions were designed to be as neutral as possible, avoid field-specific terminology, and offer detailed definitions of specific terms. Much work went into the creation of the questions, including consultation with scholars from a variety of disciplines, and from across the globe, over several years.¹³ The questions most commonly allow the categorical answers “yes/no/field doesn’t know/I don’t know,” although some provide different categorical options or require a continuous data answer, such as population numbers or size of largest monument. In any case, the result is a standardized, quantitative data point, which means that the answers are standardized across the entire dataset. Additionally, experts are encouraged to add qualitative comments and citations to each answer to allow for narrative description of the trickier clarifying points necessary to approach an answer. Likewise, multiple answers are allowed for any given question, each of which might offer a slightly different time range or geographical area for the answer, enabling a considerable degree of freedom (or complexity) in how the expert answers a particular question.

An added piece of data is the hierarchical list of terms from a tree of religions with which each expert is asked to “tag” their entry, not unlike what one might find in a standard encyclopedia or reference guide. The expert can add as many tags as they want and even suggest their own tags for insertion at different levels within the tree. This system is meant to capture the expert’s knowledge about the location of their entry within a traditional form of classification. The current study is based on a snapshot of DRH data from November 6, 2020, when the DRH encompassed 458 “Religious Group” entries from 234 experts.¹⁴

To perform this analysis a significant amount of processing had to take place beforehand to clean up the data so that the algorithms we employed would be able to efficiently cluster entries. Because the data from each entry generally takes the form of standardized answers to shared questions, they can be visualized as data points representing presence or absence. Categorical

¹³ These included feedback on polls from participants at the “Workshop on Ritual and the Evolution of Religion and Morality,” organized by the Cultural Evolution of Religion Research Consortium (CERC, UBC) and the research project “Ritual and the Emergence of Early Christian Religion” (REECR, University of Helsinki), Vancouver, BC, November, 2014; “Religion in the Text and on the Ground: the Convergence of Historiography and Ethnography in Religious Studies” (with Fred Tappenden, McGill), CERC 2nd Plenary Meeting, McGill University, Montreal, QB, May 2015; and “Religion, Ritual, Conflict, and Cooperation: Archaeological and Historical Approaches,” Center for Advanced Studies in the Behavioral Sciences (CASBS), Stanford University, April 29–30, 2016.

¹⁴ A list of entries and experts can be found here: <https://religiondatabase.org/browse/>. In the interest of transparency, it should be noted that all experts who publish their entries in the DRH receive an honorarium recognizing their academic input. The project recognizes the ubiquity of uncompensated labor in academia and has been able to offer honoraria to all experts, thus far, for the completion of entries in the DRH.

DRH questions with “yes” or “no” (as well as “field doesn’t know” and “I don’t know” answers) can be presented in a matrix (rectangular table of data, [Figure 1](#)). Each “yes” answer is coded as 1, and each “no” is coded as 0. The columns of the matrix represent DRH questions, and the rows represent one of the tripartite social divisions of each DRH entry—religious specialists, elites, and non-elites (general populace). Depending on the analysis condition, answers of “field doesn’t know” and “I don’t know” and unanswered questions are treated as missing values or imputed to yes or no (for full details of analysis conditions, see the [supplementary material](#)). Questions not employing these standard four answers as options were not included in the analysis. As polls in the DRH are structured hierarchically, more specific follow-up questions are only asked if the answer to the overarching question is yes.

Consequently, it can be inferred that a “no” answer to an overarching question also applies to subsequent follow-up questions. For example the question “A supreme high god is present?” is only asked after a “yes” response to the overarching question “Are supernatural beings present?” as there can only be a supreme high god if supernatural beings are also present. Therefore, in this analysis, if the answer to an overarching parent question is “no,” all follow-up questions are assumed (imputed) to have the same answer.

Taxonomic trees are then created using the BEAST2 algorithm ([Bouckaert et al. 2019](#)). The branches in the tree represent some underlying differentiation between the entries on each side of the divide. To understand what is driving that division, we can examine which questions differ between the two groups. These *discriminating questions* are calculated by comparing the

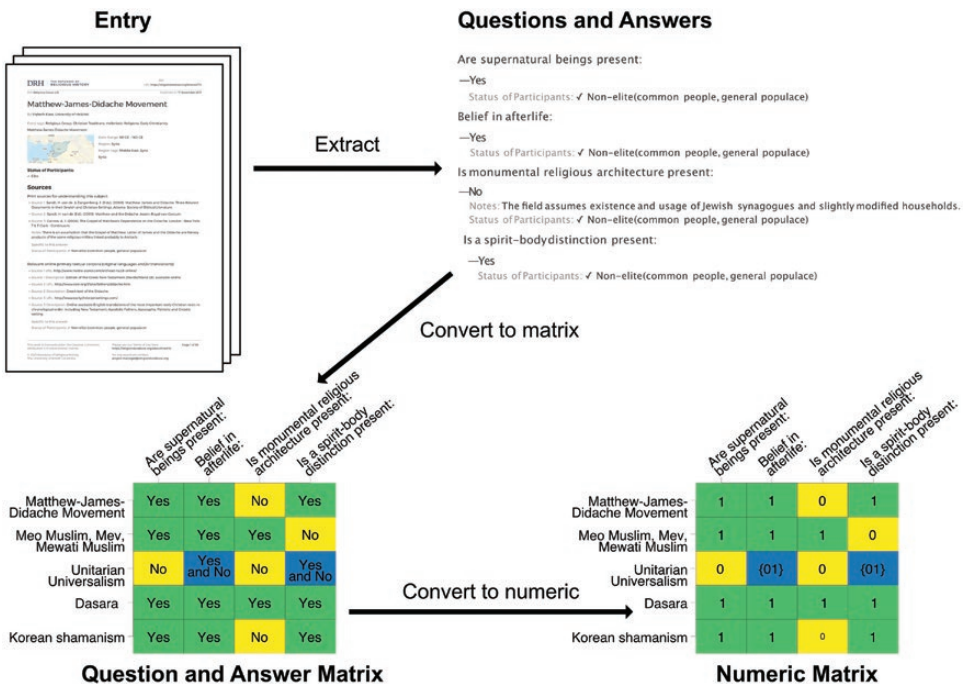


Figure 1. How DRH entries are represented in a matrix for analysis. Question and answer pairs are first extracted for each entry and group of people. These question and answer pairs are then transformed into a matrix, where each row represents an entry and group of people and each column represents a question.

percentage of questions with “yes” and “no” answers within each cluster and identifying the questions that have the highest percentage difference in answers between clusters.

The end result is an overall taxonomic tree of “Religious Group” entries showing the most likely relationships between entries across the entire database.¹⁵ This tree structure takes into account uncertainty in the data. The overall patterns present in the tree represent two major levels of analysis. At the macro-scale (i.e., the early branches in the tree), the structure represents major groupings of entries across the dataset, which may or may not reflect current categorical understandings of religious “families.” At the most minute level (i.e., the leaves of the tree), the structure represents the algorithm’s best guess as to which entries are most like other entries. However, because our data is somewhat sparse,¹⁶ two “Religious Group” entries might be located quite close to each other despite having significant differences. This is because the algorithm is forced to assign a branch to every entry, which means that sometimes dissimilar entries are paired with one another because we lack any otherwise more similar entries that would have been placed between them. Paired entries whose first common branch is far to the left in the tree are not significantly related and are only paired through the necessity of the algorithm (albeit their inclusion together within an earlier branch is still meaningful).

After creating the taxonomic tree of entries, we turned to comparing this tree with a similar structure generated instead from the tags each expert applied to their entries. These tags represent the expert’s own intuition concerning what categorical relationships might exist between their entries and other entries in the database. To compare the expert-sourced tagging tree classification system of religions with the taxonomy derived from the quantitative answers to DRH polls, the distance between each pair of tags is first calculated. From these distances, the shortest and longest distance between each entry and every other entry is then calculated (Figure 2). These distances were then compared with the tree made earlier to find the coherence between branches in each of the two tree structures.

RESULTS

Study 1: Bottom-Up Classification of Religious Groups Based on Pattern of Poll Question Answers

In study 1 we generate a dendrogram of the relationship between entries in the DRH solely through analysis of experts’ answers to questions about these religious groups, ignoring both the tags applied by experts to their own entries and the entry names or other classifying information (geography or time range). The overall results are presented in Figure 1.

Figure 3 depicts the overall dendrogram of 171 “Religious Group” entries present in the DRH as of November 6, 2020, including only entries where at least 50 percent of questions were answered (questions were simultaneously filtered to include only questions answered by at least 50 percent of entries). Given its complexity, this figure has also been made available online (https://rachel-spicer.shinyapps.io/drh_tree/) to facilitate browsing specific portions of the tree and zooming in on details. In the sections below, we will focus on specific subsets of the overall tree.

We find that, on the whole, the entries divide into two distinct clusters (termed Cluster 1 and Cluster 2). Each of these is further divided twice in C1.1 and C1.2 and in C2.1 and C2.2. Finally,

¹⁵ Code and data are available here: https://github.com/religionhistory/religion_taxonomy.

¹⁶ Because of the promise of this technique, we decided that the data currently available in the DRH was adequate to provide a proof of concept. We intend to continue running the same taxonomic analysis at regular intervals as our coverage increases and will post these updated taxonomies both on our website and online version of the taxonomy tree linked to below in the Results section: https://rachel-spicer.shinyapps.io/drh_tree/.

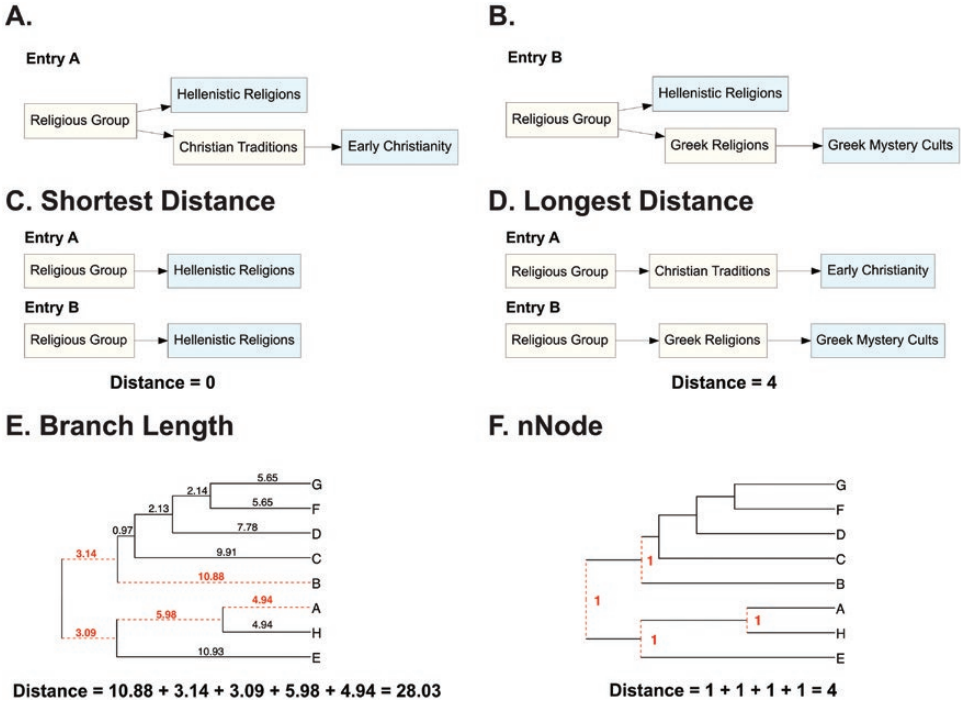


Figure 2. Methods used for calculating the distance between entries using the tagging tree and the taxonomies. Sections A and B show example “Religious Group” tags for two example entries. C and D demonstrate the two methods used for calculating distance between entries using the “Religious Group” tagging tree. C) The shortest distance between entries is calculated using the entries’ tags, by finding the pair of tags with the shortest distance between them. In this case the distance is zero as both entries share the same “Religious Group” tag, Hellenistic religions. D) The longest distance between entries is calculated by finding the tags that are most disparate in the tagging tree for each pair of entries. E and F show how distance between entries is calculated using the taxonomies. E) The distance between two entries is calculated by summing the total length of branches between them in the tree. F) The distance between two entries is calculated by counting the total number of nodes (nNode) between the entries in the tree.

C2.1 is further divided into two smaller clusters (C2.1.1 and C2.1.2). This section initially treats the overall split in C1 and C2 and then the subdivisions of each cluster in turn. This is followed by an in-depth discussion of pairs of entries in the tree, which either confirm existing scholarly opinions or offer interesting and unexpected results.

The first major division identified by the algorithm splits the dendrogram into two large branches, or “clusters”: Cluster 1 (C1.1 + C1.2) vs. Cluster 2 (C2.1 + C2.2).

Although there are some minor exceptions, the division between C1 and C2 seems to map quite well onto what were traditionally given broad-brush labels: “Western” versus “Eastern” religions. However, our results can more accurately categorize these groups as those that geographically originated in the Mediterranean basin and West Asia (C1) and South, Southeast, and East Asia (C2). Despite the fact that the model does not factor geographical information into its clustering algorithm, we can clearly see patterns of colonial and missionary activity. C1 (especially entries in C1.1) trace their origins back to the Mediterranean basin yet can be found around the globe. C1 contains clear examples of Christian missionary impact, such as Nigerian Pentecostalism, Hmong Christianity, and Indonesian Catholicism. Other groups in C1, such as the Chishti Sufis (origins

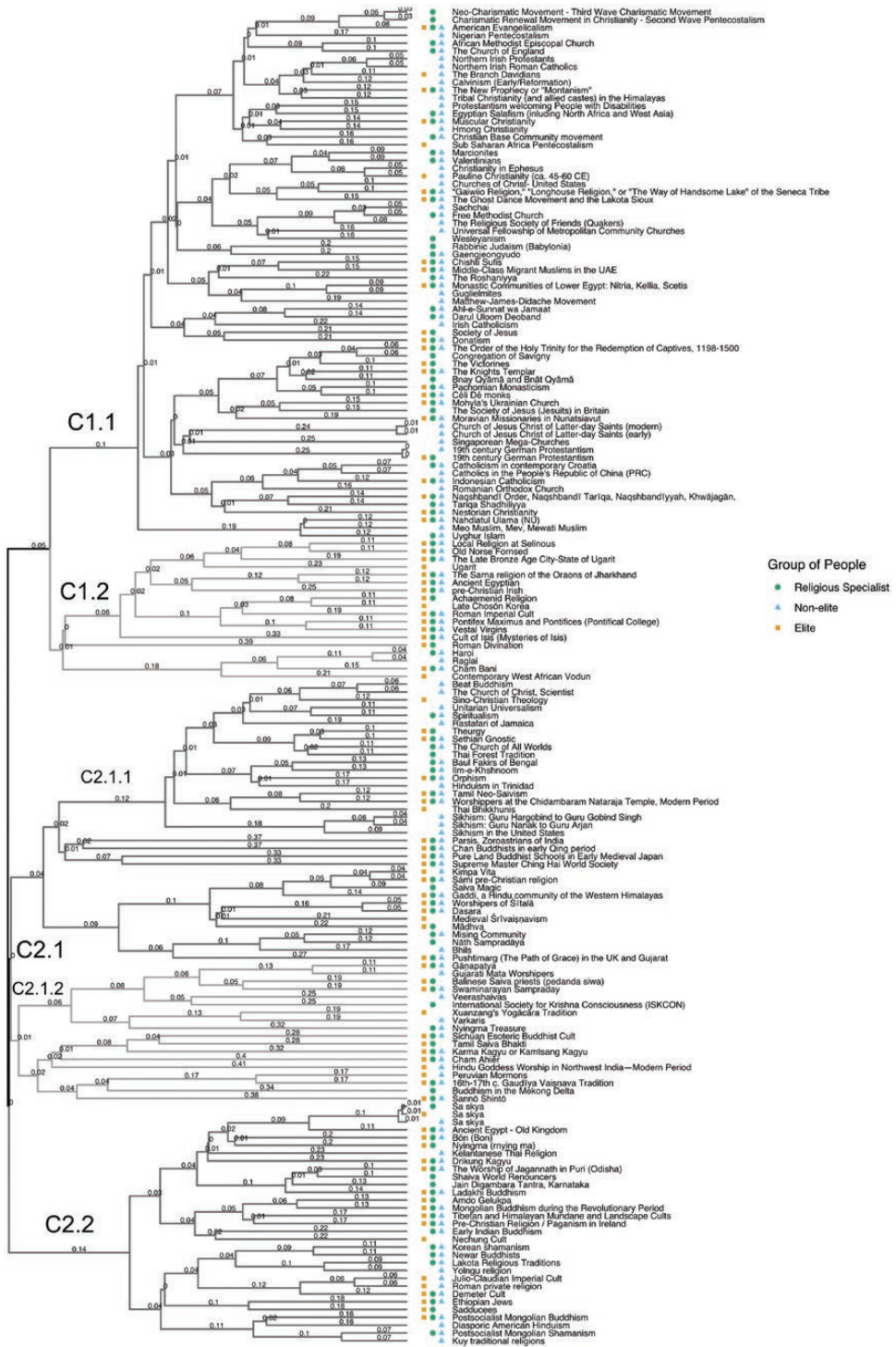


Figure 3. Dendrogram of sample of 171 “Religious Group” entries in the DRH

in Afghanistan), the Darul Uloom Deoband (origins in India), Uyghur Islam (Central/East Asia), and Nahdlatul Ulama (Indonesia), all “originate” as South/East/Southeast Asian groups, but the prominence of Meccan-centered discourse among them might help to explain why they show up in our analysis as linked to the Mediterranean/West Asian cluster.

Similarly, the placement of certain groups in C2 demonstrates the dispersal of Buddhism along land and sea trade routes. For instance, regarding the East Asia entries in Cluster 2, their placement there can be attributed to the influence of South Asian Buddhism, the mechanism of which was the dispersal of Buddhism along land and maritime trade routes (Zürcher 1959; Ch'en 1964, 1973). Buddhism brought with it belief in reincarnation—which is, as we see below, a distinguishing question for Cluster 2 and one not found in pre-Buddhist East Asian religions. East Asian entries that have ended up in Cluster 1 are the product of Christian missionary activity, primarily after the sixteenth century as a result of maritime trade (Wills 2010).

Although we can understand this split as reflecting certain views of broad-brush categorizations, the model can provide more information in the form of which questions drive this split. The primary drivers are questions related to reincarnation, with another contributor being the nature of a high god as unquestionably good (Table 1).

The single most powerful discriminating question between these clusters is “[Reincarnation] in a human form,” which 96.51 percent of the C1 groups answered in the negative and 69.41 percent of the C2 groups answered in the affirmative. The question “Is there reincarnation in this world?” is a close second, with 94.19 percent of C1 groups answering in the negative and 81.18 percent of C2 groups answering in the affirmative. Although the appearance of these questions as highly differentiating was unexpected, from a wider perspective it seems understandable; belief in reincarnation has a significant impact on the outlook both in this life and beyond for believers. More work can be done to tease apart related questions in the DRH to see how those impacts can be traced through our questionnaire.

Outside of questions covering reincarnation, the nature of a high god as unquestionably good is also powerfully discriminative. Although this question is not as starkly one-sided for C2, the signal from C1 (91.86 percent answered in the affirmative) is very strong. This result is perhaps not terribly surprising given that both Jewish and later Christian scripture and discourse emphasize concepts like “mercy” and the redemptive nature of God’s covenant(s) with those who follow the Abrahamic traditions.

The first major division within C1 is a bifurcation between two clusterings, C1.1 and C1.2 (Figure 4). The immediate impression given by this split is a distinction between Abrahamic

Table 1. Discriminating questions driving the C1 vs. C2 split

Question	C1		C2	
	Yes	No	Yes	No
Reincarnation in this world	4.65%	94.19%	81.18%	16.47%
[Reincarnation] ¹⁷ in a human form	2.33%	96.51%	69.41%	20.00%
[Reincarnation] in animal/plant form	1.16%	97.67%	57.65%	28.24%
Reincarnation linked to notion of life-transcending causality (e.g., karma)	2.33%	96.51%	60.00%	27.06%
The supreme high god is unquestionably good	91.86%	6.98%	36.47%	55.29%

¹⁷ This is a subquestion that only appears if the expert answers “yes” to the parent question “Reincarnation in this world,” so “reincarnation in this world” is the understood subject (indicated by square brackets).

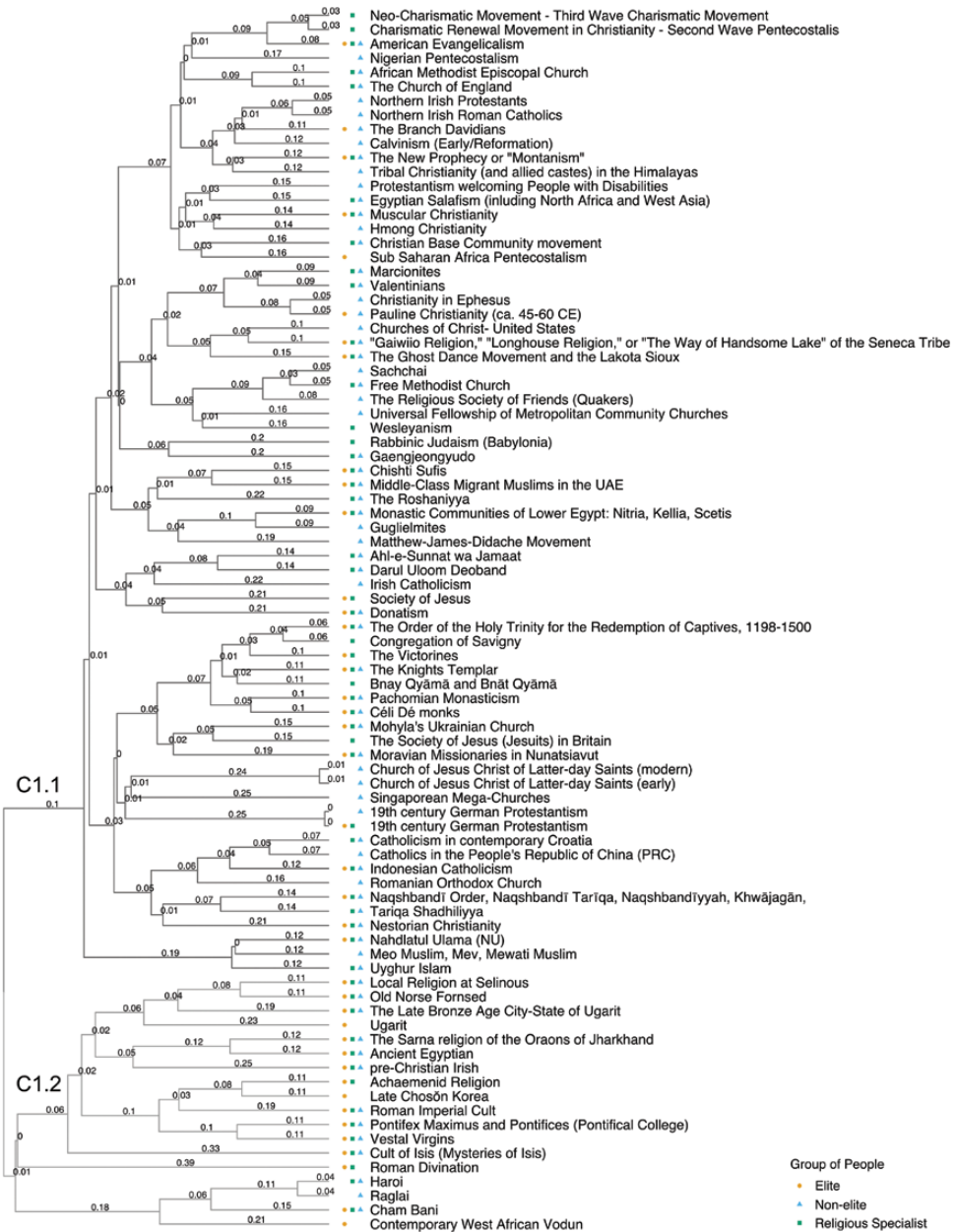


Figure 4. Zoom in on split between C1.1 and 1.2 within C1

groups versus other “Mediterranean/West Asian” groups. Jewish, Christian, and Muslim groups are all well represented in C1.1 in a variety of historical forms. C1.2, on the other hand, represents a wide range of comparatively more ancient traditions from the Mediterranean basin, with a few outliers; “Late Choson Korea” (Shababo 2019) and “The Sarna religion of the Oraons of Jharkhand” (Munda 2020) stand out in particular. The presence of Late Choson Korea in C1.2 is likely due to the influence of Confucianism, which has features—such as a belief in a supreme

high god and absence of belief in reincarnation—that position it similarly to other ancient Mediterranean religions. Although this entry currently appears to be an outlier, it is indicative of what will likely be a much greater East Asian presence in C1 and likely C1.2 (depending on the scholar) once we have more non-Buddhist East Asian entries available in the database.

It is worth noting the small sample size of ancient Mediterranean/West Asian traditions reflected in the dendrogram. On the basis of many of the discriminating questions that drive the C1.1 and C1.2 split—messianism, proselytization, and the (non)exclusive worship of a high deity—we may anticipate that this division will become more accentuated as additional ancient Mediterranean exemplars are added to the DRH.

The discriminating questions driving the C1.1-C1.2 split are found in [Table 2](#).

Unlike the split between C1 and C2, narrowing in on the split between subclusters in C1 we find a larger diversity of differentiating questions. The presence of messianic beliefs is the strongest differentiator, followed closely by a couple of questions that address burial practices. A few others deal with the way in which the group might bring in new members or interact with authorities. The questions addressing worship of other supernatural beings and the existence of a sky deity address the nature and construction of the supernatural objects of worship; although their results in this table perhaps align with expectations, it is important to note that they are not the strongest signal differentiating entries between C1.1 and C1.2. One more takeaway would be the way in which these questions might designate religious groups that are often assumed to be de facto aligned with (or categorized as) state religions versus religions that are potentially in competition with a state-sanctioned belief structure.

We recognize that scholars will have various ideas about the presence of monotheism as a distinguishing factor for religious groups. Some readers may be familiar with the Hindu reformist movements of the Arya Samaj and the Brahmo Samaj, which are Hindu—a religion commonly described at the introductory level as polytheistic—although these reformist streams

Table 2. Discriminating questions driving the C1.1 vs. C1.2 split

Question	C1.1		C1.2	
	Yes	No	Yes	No
Are messianic beliefs present?	95.59%	2.94%	5.56%	94.44%
Is the messiah's purpose known?	77.94%	8.82%	0.00%	100.00%
Are grave goods present?	10.29%	75.00%	88.89%	0%
Does the religious group actively proselytize and recruit new members?	77.94%	17.65%	5.56%	94.44%
[Are grave goods present?] Personal effects	7.35%	77.94%	77.78%	5.56%
Is it permissible to worship supernatural beings other than the high god?	10.29%	88.24%	83.33%	16.67%
Does the religious group in question possess its own distinct written language?	13.24%	85.29%	72.22%	22.22%
The monarch is seen as a manifestation or emanation of the high god	2.94%	95.59%	61.11%	33.33%
Are the group's adherents subject to institutionalized punishment enforced by an institution(s) other than the religious group in question?	91.18%	5.88%	33.33%	50.00%
The supreme high god is a sky deity	32.35%	67.65%	83.33%	16.67%

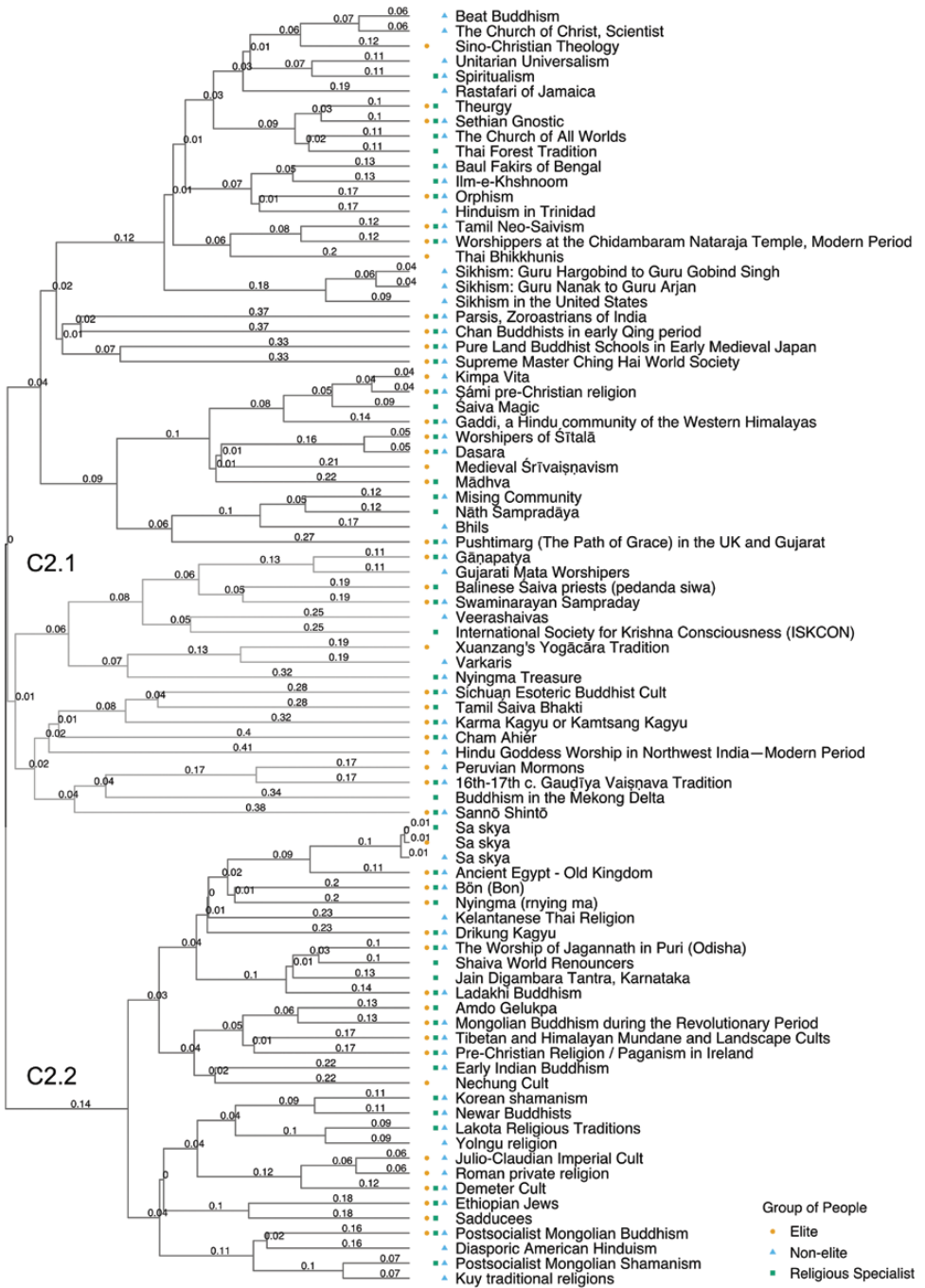


Figure 5. Zoom in on split between C2.1 and 2.2 within C2

include analyses aimed at accounting for the variation of groups that are included in these clusters but typically fall outside of the label “Hindu” and “Buddhist;” why some “Buddhist” groups are clustering with a greater majority of “Hindu” traditions in C2.1; and why some

Table 3. Discriminating questions driving the C2.1 vs. C2.2 split

Question	C2.1		C2.2	
	Yes	No	Yes	No
The supreme high god has knowledge of this world	81.48%	7.41%	0.00%	96.77%
The supreme high god communicates with the living	70.37%	12.96%	0.00%	96.77%
A supreme high god is present	87.04%	11.11%	6.45%	93.55%
The supreme high god has deliberate causal efficacy in the world	64.81%	14.81%	3.23%	96.77%
The supreme high god has indirect causal efficacy in the world	59.26%	14.81%	3.23%	96.77%
Is it permissible to worship supernatural beings other than the high god?	74.07%	14.81%	3.23%	96.77%
The supreme high god exhibits positive emotion	70.37%	16.67%	3.23%	93.55%
The supreme high god is anthropomorphic	64.81%	25.93%	0.00%	100.00%
The supreme high god is unquestionably good	57.41%	29.63%	0.00%	100.00%

“Hindu” groups are clustering with a greater majority of “Buddhist” groups in C2.2. Analyses also need to be done to shed light on why many of the affirmative answers in C2.1 found in [Table 2](#) are below 80 percent.

Finally, as suggested earlier, there is a smaller division within C2.1 that might suggest a division of this cluster into two further clusters (C2.1.1 and C2.1.2). Here the split between these smaller clusters suggests a heavy emphasis on monumental architecture as defining inclusion in C2.1.2. The differentiating questions can be found in [Table 4](#).

The most defining feature of this division is a series of questions that address the existence of monumental architecture. Following from the discussion above, the majority of groups in C2.1 are South Asian theistic traditions, and, of these, most of them are identified as Hindu *bhakti* traditions (this is especially true within C2.1.2). Monumental architecture has generally played a central role in South Asian religions over the last 1500 years, especially among those with a devotional orientation, whether Hindu, Buddhist, Sikh, or otherwise. For this reason, it is unclear without further analysis of what is driving the divisions between C2.1.1 and C2.1.2. Moreover, why do we find that 80.56 percent of the groups of C2.1.1 recognize “different types of religious monumental architecture,” yet only 44.44 percent of these same groups recognize the presence of “monumental religious architecture”? They are nearly identical questions, so one would think that the answers to them would match up more closely. To make sense of what is driving these divisions, further analyses of the answers of experts and the questions posed to them in the poll, which extends beyond the scope of this article, need to be performed. As with the analysis in the previous section, which discusses the division between C2.1 and C2.2, the clustering of traditions that are not primarily traced to South Asia may also help account for the unexpected variability that is driving the division here.

Study 2: Aligning the Generated Taxonomy with Expert Tags

For the second study, we compared the generated taxonomy discussed above with a tree structure derived from the group-based tags that experts assigned to their entries. The results are displayed in Figure 6 below.

These two structures show a degree of similarity (Figure 6). This suggests that the top-down conceptual categories employed by our experts match the bottom-up categories derived from a tag-blind analysis of patterns of answers to specific questions.

Table 4. Discriminating questions driving the C2.1.1 vs. C2.1.2 split

Questions	C2.1.1		C2.1.2	
	Yes	No	Yes	No
[Monumental architecture] mass gathering point	13.89%	86.11%	88.89%	0%
[Monumental architecture] temples	16.67%	83.33%	100%	0%
Are there different types of religious monumental architecture?	80.56%	19.44%	100%	0%
[Monumental architecture] altars	16.67%	83.33%	88.89%	5.56%
[Monumental architecture] tombs	8.33%	91.67%	66.67%	33.33%
Is monumental religious architecture present?	44.44%	55.56%	100%	0%
Does membership in this religious group require sacrifice of time (e.g., attendance at meetings or services, regular prayer, etc.)?	27.78%	69.44%	83.33%	16.67%

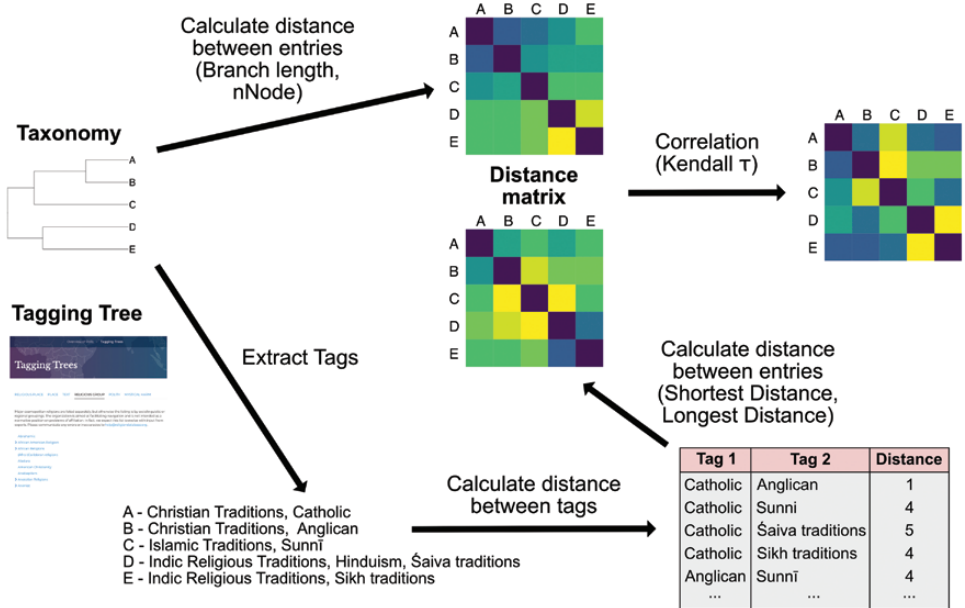


Figure 6. The methods used for comparing the data-generated taxonomy with the expert tagging tree. To find the distance between entries using the tagging tree, the tags for each entry are first extracted. The distance between each pair of tags in the tagging tree is then calculated. From these paired distances between tags, the distance between entries is calculated based on their tags. The distance between entries in the taxonomy is calculated using branch length and nNode. All four of these methods produce a pairwise distance matrix that shows the distance between an entry and any other entry. These distance matrices are then compared with areas of agreement and disagreement between methods for calculating distance.

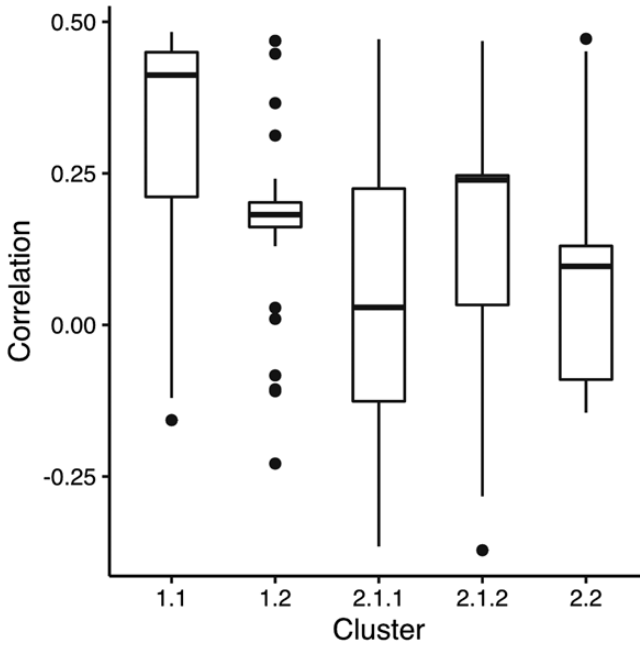


Figure 7. Degree of within-cluster correlation between the generative tree and the tree derived from group-based tags.

Broadly, each cluster finds internal agreement between the structure derived from expert-sourced answers and the tags these same experts applied to their entries (Figure 7). This indicates that across all entries (in all geographical areas and temporal periods) the tags used by experts reflect a certain degree of the underlying organization of the entries as generated by the individual answers themselves. However, Cluster 1 clearly performs the best out of all the clusters. We find this unsurprising as the types of tags applied to entries in Cluster 1 have perhaps the longest history and therefore the most stable representations within the field of religious studies (in Western academia). We suggest that this is due to the legacy of church history and theology as precursors to many existing concepts within religious studies scholarship.

For instance, in Cluster 1.1 the entry “Christianity in Ephesus” (Proctor 2020) has the following tags: religious group, Asia Minor, new religious movement, Christian traditions, Roman religious traditions, Anatolian religions, early Christianity, and ancient Christianity in Rome (eight tags). In contrast, in Cluster 1.2 the entry “ancient Egyptian” (Simpson 2020) has the tags religious group, Egyptian religions, and African religions (three tags). The capacity for the smaller minute branches within C1.1 to match the underlying tree derived from the answers is driven by the specificity of the tags present on entries in that cluster, which may be caused by existing disciplinary biases that have created a potent vocabulary of terms and names to describe the diversity of religious groups. On the contrary, Cluster 2 and its components have the least internal agreement (although still average positive agreement). This, likewise, is not surprising as the terminology used by experts to describe and categorize entries within these clusters is still an active area of scholarship, at least within Western academia.

DISCUSSION

With regard to Study 1, we find that the large splits and major clusters within the generated tree roughly match existing scholarly perceptions about the geographical and temporal spread of religious traditions. Study 2 adds nuance to this conclusion by showing that the coherence of these clusters is most pronounced in the branches that have a longer and more entrenched position within the history of the field itself.

General Observations

Our tree seems to replicate standard surveys of “world religions” that offer various high-level taxonomies of religious groups or traditions.²⁰ However, the cluster division and entry positions offer much more information for fine-grained analysis. Looking in more detail at specific pairings, we find a mix of predictable, puzzling, and interesting results when we examine the far right of the dendrogram and note which religious groups have the closest relationships. As mentioned in the methods section, the way the tree is generated requires that all entries must end up with a physically paired entry, even if some of these pairs are in fact more distant than we would consider statistically significant for the purposes of direct comparison.

Confining our discussion to paired entries with close relationships, we can note that, in certain cases, the similarity between paired entries is obviously the result of geographical, temporal, or doctrinal connections. For instance, the similarity between “Neo-Charismatic Movement-Third Wave Charismatic Movement” (Womack 2020a) and “Charismatic Renewal Movement in Christianity-Second Wave Pentecostalism” (Womack 2020b) is attributable to their geographical area (North America), time period (latter half of the twentieth century), and shared connection to Pentecostal traditions. It is worth noting that these two entries were written by the same expert. Likewise, two entries representing groups dating between the eleventh and fifteenth centuries in Northwestern Europe, “The Order of the Holy Trinity for the Redemption of Captives, 1198–1500” (Blair 2019) and “Congregation of Savigny” (Doss 2019), show close similarity. Unlike the previous example, these two entries were written by different experts but reflect underlying similarities between these Christian orders.

When a religious group has multiple entries across different time periods, we find that their entries still cluster closely, as is the case for “Church of Jesus Christ of Latter-day Saints (modern)” (Pepper 2019a) and “Church of Jesus Christ of Latter-day Saints (early)” (Pepper 2019b). Similarly, we find that when an expert has differentiated the answers in their entry by category of group member “elite,” “non-elite,” and “religious specialist,” all three sets of answers cluster very closely, as is the case for the entry “Sa skya” (Wojahn 2020). A particularly interesting observation in this case is that the “elite” and “religious specialist” sets of answers cluster more closely than those of the “non-elite,” which probably reflects a more generic distinction between elite and “folk” understandings and practices.

There are other cases where the results are not entirely surprising but point to an intriguing area for further study or reflection. For instance, “Northern Irish Protestants” (Ward 2019a) and “Northern Irish Catholics” (Ward 2019b) cluster more closely together than the latter does with another entry for a Catholic tradition elsewhere in Europe. Although this is a small data point, it suggests that a geographical or local cultural signal is having an impact on how these two entries are paired in the tree rather than an overwhelming doctrinal signal. The close pairing of “Sachchai” (BK 2020) and “Free Methodist Church” (Lane 2020) also offers an intriguing look into how the expression of charismatic traditions within the questions of the database shows close similarity.

²⁰ For a critique of the term “world religions” see Tomoko Masuzawa (2005).

Finally, we find cases where the placement of entries within clusters and pairing of entries is not entirely expected. This may reflect surprising and important connections or differences of scholarly opinion but in some cases prompts reconsiderations of both the overall model as well as how our questions might be influencing the placement of entries with the tree. Notably, the entry on “Peruvian Mormons” (Palmer 2020) appears in Cluster 2.1 rather than with the other entries on the Church of Latter-Day Saints (LDS) in Cluster 1.1. Here the major differences come down to whether “missions” are considered a form of “pilgrimage” and how the two groups might answer the question of “reincarnation in this world.” The expert for Peruvian Mormons wrote: “Inasmuch as Mormons believe they will be resurrected and go to the Celestial Kingdom and that this world will become the Celestial Kingdom, they believe in reincarnation on this world,” whereas the other LDS entry states that there is no reincarnation. Digging into the details of the answers and comments here provides us insight into how experts wrestle with these exact questions and reflects divergent scholarly opinions about the beliefs of these groups.

Another unexpected group of divisions occurred between Cluster C1.2 and C2.2, in which entries related to the same ancient Mediterranean and North African cultures were divided. For instance, a general entry on “ancient Egyptian” (Simpson 2020) was placed in C1.2, whereas a more specific entry on “ancient Egypt-Old Kingdom” (Arbuckle 2021) was placed in C2.2. Although we would expect that nuances in the more limited time period that is represented within the Old Kingdom entry should cause it to differ slightly from the more general entry on the whole of Egyptian history, we did not expect the entries to appear in entirely separate clusters. In addition, in this case, the discriminating question does not seem to relate to reincarnation, as both entries agree that there was no “reincarnation in this world.” They disagree, however, regarding the presence of a “supreme high god,” which was another discriminating factor between clusters C1 and C2. The more general entry notes that there was a high god but provides no additional commentary, whereas the Old Kingdom entry states that there was none but goes on to explain that there were several significant supernatural beings present during the period in question. Although ancient Egyptian religion was generally polytheistic, there were certain times when specific gods were more popular, such as the rise of the god Amun in the New Kingdom or Aten during the Amarna Period, which may account for the discrepancy between the two answers. This helps to show the importance of including entries based on more discrete time periods within the larger religious history of a region in order to arrive at more nuanced and precise conclusions. Nevertheless, as more entries are added to the database, it seems likely that these two entries will end up closer together, given that many of the other answers are similar.

An additional surprising case is the placement of the “Sadducees” (Matson 2020) within Cluster 2.2. With their geographic location in the eastern Mediterranean and their chronological position at the nexus of the ancient Mediterranean and Abrahamic groups, it might be assumed that they would fall within Cluster 1. Indeed, on the basis of the discriminating questions that drove the C1 and C2 split, the Sadducees measure quite closely with the C1 groups in their lack of belief in reincarnation (resurrection). Likewise, within the breakdown between C1.1 and C1.2, the Sadducees predominantly overlap with C1.1 in terms of messianism, apparent proselytization, and in not allowing the worship of other supernatural beings. Among these discriminating questions, they only really differ from C1.1 in their presence of grave goods. Collectively, these positions might assume eventual placement within C1.1 as perhaps befits their historical situation and relation to Judaism.

The Sadducees’ placement within C2, however, is perhaps more the result of particular understandings of their supreme high god, where they stood in contrast to C1 that largely tends to see this deity as “unquestionably good.” Furthermore, within C2, the Sadducee entry’s negative response to nearly all questions related to aspects of the supreme high god places them firmly in

line with C2.2 groups, except for one major caveat in that they do in fact possess a supreme high god, whereas the C2.2 groups overwhelmingly do not. Such an apparent similarity between the Sadducees and C2.2 may then be something of a mirage, as so many of the C2.2 discriminating questions are hierarchically predicated on the lack of a supreme high god. This factor might also account for the placement of other Mediterranean entries, such as Julio-Claudian Imperial Cult (Bell 2021) in C2.2. Similar to the situation in relation to ancient Egypt discussed above, it will be interesting to see the effect that additional entries have on this clustering.

At first glance, we found it surprising that Haroi (Quang 2020b), Raglai (Quang 2020a), and Cham Bani (Noseworthy 2020) appeared together with the Mediterranean/West Asian groupings in Cluster 2 and so closely associated with contemporary West African Vodun (Atte-oudeyi 2020). Looking more closely at these entries compared with Cluster 2 (where we might assume they would belong), we find the following discriminating questions:

- Assigned at a specific age
- Assigned at birth (membership is default for this society)
- Does the membership in this religious group require sacrifice of property/valuable items?
- Is there violent conflict (with groups outside the sample region)?
- Does the religious group in question possess its own distinct written language?
- The monarch is seen as a manifestation or emanation of the high god
- The supreme high god is a sky deity
- Supernatural punishments are meted out in the afterlife
- Are messianic beliefs present?
- Is the messiah's purpose known?
- [Are there special treatments for adherents' corpses?] Internment
- Are grave goods present?
- Does the religious group in question provide public food storage?

The evidence of this cluster points toward a Robert Orsi-like emphasis of understanding “religion as relationships” among humans, their families, their societies, the realm of the supernatural, and the human realm(s) (Orsi 2005). For example, we see membership is defined at a specific age, often assigned at birth, and requires the sacrifice of property/valuable items. Members also tend to possess their own distinct written language, while we see that there are examples of violent conflict with groups outside the sample region. There tend to be clear mechanisms for divinities (high god or no) to relate to human society, inclusive of messianic beliefs in at least some of the cases. The mechanisms for divinities in relation to human societies and inclusion of messianic beliefs could be viewed as Orsi’s “relationships between heaven and earth” (Orsi 2005), although we could also rephrase Orsi’s language as “realm of the supernatural and human realm(s)” to more closely fit the cases in our evidence. Special treatments for adherents’ corpses, including internment and grave goods, also tend to be present, as does public food storage, suggesting that in these cases relations to earth itself are also present. This is just one example of how we could take a theoretical concept that has been widely discussed by scholars in one subset of religious studies and test it against evidence from other areas to refine our theoretical discussions of the category of *religion*. We should emphasize that, in our view, it is not necessary that we think of Orsi as “paradigmatic” in such a fashion that it would eliminate consideration of any other religious studies scholar who has made novel contributions to the field in the past several decades. Our idea is, in principle, that any theorist’s ideas could be selected and tested by the data. In the cases under discussion, although our evidence reaffirms some ideas that scholars in religious studies—like those aligned with Orsi’s thinking—have thought about regarding how the category of *religion* is conceived of, it also clarifies the ways we think about these specific cases and comparisons.

Typically, scholars might regard such groups as Southeast Asian Haroi and Raglai religion or West African Vodun as “traditionalist religious groups.” The Cham Bani religious group—near the Raglai and Haroi communities in Vietnam—has also been described as variously syncretic and traditionalist, although Cham Bani are often interpreted as a form of Islam by outsiders and scholars alike. In both West Africa and Southeast Asia, we find layers of influences including traditional religions, Islam, and colonial era-Christian missions. In both regions, local indigenous groups develop systems that incorporate spirits and saints and have understandings of the cosmos that include cosmological dualism (emphasizing balance of competing elements), along with traditional healers. Such similarities in the cultural contact zones where we find Vodun—and perhaps also traditional Yoruba religion, although we do not have an entry for it yet in the database—and the Bani, Haroi, and Raglai communities of Southeast Asia could have emerged as a result of similar historical experiences. That traditional religions where membership is assigned at birth, at a specific age, in the context of rituals, and require sacrifices do not necessarily exhibit official political support, as we see in this cluster, speaks to the relational elements of the religious communities in question. This particular result demonstrates that scholars of religious studies have been addressing traditional religions that grow out of comparable historical contexts using similar language while also demonstrating the ways in which the DRH can lead to thought-provoking comparisons.

CONCLUSION

One obvious limitation to the current study is that it does not directly engage with the lived world of religious experience. The above analysis is constructed from scholarly experts’ answers to questions about units of analysis, that is, the “Religious Group” poll, that they themselves constructed. However, this expert knowledge is quite fine grained: experts are answering very specific questions about daily practice, ritual infrastructure, and beliefs. Therefore, the standardized poll structure with multiple levels of questions removes the expert from overarching narratives that may exist in the field and allows them to focus on single answers, producing snippets of atomic data in the form of answers that are more “objective” and therefore comparable while still remaining highly informed by existing scholarship.

The tree and its component clusters thus constructed from the data are then even more remarkable for their consistency with prior intuitions (e.g., [Smith 1995](#), cited above). The fact that a completely unintelligent algorithm, churning through the answers to very specific questions, produced a dendrogram that roughly mirrors traditional models in religious studies of the major divisions between large religious groups, etc., is significant. It shows that rather than being a completely artificial scholarly category, our entries represent units of analysis that capture real variance in the world, as understood by historians. This is true even though our dataset is incomplete and there are groups and traditions elsewhere in the world with no relation to those on the current tree. Furthermore, the groupings produced by the DRH have significant potential to intervene in recent conversations regarding long-standing (and often highly problematic) categories used to define and categorize religious traditions throughout history.²¹

For instance, if the sorts of questions the DRH asks about religious groups were solely the product of older historical assumptions about religion, we would expect that the major division (C1 vs C2) in the tree would replicate those perspectives by, for instance, being driven

²¹ These categories include the two-fold division of western/non-western or world religions/primal religions ([Tsonis 2013](#)), as well as [Robert Bellah's \(2011\)](#) three-fold division into tribal/archaic/axial. Recent attempts to overturn the “world religions” paradigm include the concept of indigenous religion advanced by [James Cox \(2007, 2013, 2017\)](#) but recently challenged by [Jack Tsonis \(2017\)](#).

by divisions between groups that have a supreme high god and those that do not, or between those who do or do not worship supernatural beings. This seems not to be the case, because discriminating questions that address these issues that were fundamental to earlier theories of religion are not present at the root of our tree but rather are scattered through the branches. This suggests that recent skepticism concerning the usefulness of religious categories, such as the assertion that “*the entire set of categories used to divide human groups must be reconceived*” in the study of religion (Tsonis 2017, 59, emphasis original), is exaggerated.

Anomalous results produced by our method show us that certain questions have a highly discriminating power within the model but also allow further analysis to try to fill in entries around them, with the goal of reconstructing a missing cluster or refining our questions to better model the beliefs and practices of the group in question. This is in stark contrast to previous top-down methods, where particular properties of groups selected by a theorist were used as a black-box criteria to cluster groups. Our methods are open and inspectable. For instance, Church of Christ Scientist (Prince 2020) and Unitarian Universalism (Applewhite 2020) both appear in Cluster 2.1.1. This might, at first, seem strange, but their clustering is likely driven by a negative answer to the question: “Are supernatural beings present?” This suggests that they have a fundamental difference from the entries in C1 but that their similarity to other groups in C2.1.1 is perhaps not as strong and is then overshadowed by the algorithm’s instance of distancing them from C1. More entries would almost certainly produce a context in which these two entries would find a tighter fit via a hypothetical cluster of entries, which fit between C1 and C2.1.1.

The DRH and the taxonomic method introduced in this paper can also be used to explore other interesting questions in religious studies. For instance, the method outlined above could be used to address one of the concluding exercises suggested by Brent Nongbri (2013, 155–56) by creating an entry for, as suggested by Nongbri, “capitalism.” Once such a “Religious Group” entry was created and the questionnaire for it completed, we could interrogate the ways in which the membership practices and beliefs of “capitalism” were similar to or different from other groups that historically have been considered to be “religions.”²²

A further limitation of our current study is the preliminary and limited scope of the dataset. Although our coverage does span most of the globe, the lack of entries for certain traditions almost certainly drives some of our ambiguous results, as noted above. However, despite this relative paucity of data, this early analysis is still remarkable for its outcome. As our coverage expands and deepens, these sorts of analyses may produce more surprises. We intend to rerun this analysis on a regular basis and keep a current version of the output live on our project website so that the shape of the resulting tree will change as new entries are added.

Finally, we note that the way in which the DRH collects data diverges from alternative “big data” approaches to history (e.g., Turchin 2018). The flexibility of the question and answer system allows experts to answer multiple times for the same question, change the coverage and scope of their answer, and embed rich media and qualitative comments. This allows the expert to answer questions in a way that best preserves their own intuition concerning the source material. The ability of the DRH system to encompass multiple units of analysis and allow the periodic updating of polls responds to concerns that, as scholars and scientists studying the phenomenon of religion, we are “inextricably stuck with asking just our questions and using just our tools in posing those questions” (McCutcheon 2001, 78). We hope that the analysis presented here shows the potential of the DRH as a tool for scholars of religion and look forward to this tool being used in the future in ways that we cannot possibly anticipate.

²² To be clear, Nongbri warns against this approach being used to determine whether “capitalism” is a “religion,” but an entry in the database would in fact open up exactly the type of analysis present in our discussion section: what sorts of questions drive a hypothetical “capitalism” entry toward or away from other entries in the database, and how might those particular vectors speak to underlying questions of how these groups are formed?

REFERENCES

- Applewhite, Courtney. 2020. "Unitarian Universalism." *Database of Religious History*. Available at <https://religiondatabase.org/browse/967/#/>. Accessed September 16, 2021.
- Arbuckle, Caroline. 2021. "Ancient Egypt - Old Kingdom." *Database of Religious History*. Available at <https://religiondatabase.org/browse/1006/#/>. Accessed September 16, 2021.
- Atte-oudeyi, Azizou. 2020. "Contemporary West African Vodun." *Database of Religious History*. Available at <https://religiondatabase.org/browse/885/#/>. Accessed September 16, 2021.
- Bayly, Susan. 2004. "Imagining 'Greater India': French and Indian Visions of Colonialism in the Indic Mode." *Modern Asian Studies* 38 (3): 703–44.
- Bell, Treasa. 2021. "Julio-Claudian Imperial Cult." *Database of Religious History*. Available at <https://religiondatabase.org/browse/940/#/>. Accessed September 16, 2021.
- Bellah, Robert. 2011. *Religion in Human Evolution: From the Paleolithic to the Axial Age*. Cambridge, MA: Belknap Press of Harvard University Press.
- BK, Amar. 2020. "Sachchai." *Database of Religious History*. Available at <https://religiondatabase.org/browse/953/>. Accessed September 16, 2021.
- Blair, Hazel J. Hunter. 2019. "The Order of the Holy Trinity for the Redemption of Captives, 1198–1500." *Database of Religious History*. Available at <https://religiondatabase.org/browse/676/#/>. Accessed August 6, 2019.
- Bouckaert, Remco, Timothy G. Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, et al. 2019. "BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis." *PLOS Computational Biology* 15 (4): e1006650.
- Brubaker, Rogers. 2004. *Ethnicity without Groups*. Cambridge, MA: Harvard University Press.
- Ch'en, Kenneth. 1964. *Buddhism in China: A Historical Survey*. Princeton, NJ: Princeton University Press.
- . 1973. *Chinese Transformation of Buddhism*. Princeton, NJ: Princeton University Press.
- Cox, James. 2007. *From Primitive to Indigenous: The Academic Study of Indigenous Religions*. Hampshire, UK: Ashgate.
- . 2013. "Reflecting Critically on Indigenous Religions." In *Critical Reflections on Indigenous Religions*, edited by J. Cox, 1–18. Hampshire, UK: Ashgate.
- . 2017. "Restricted Access Kinship and Location: In Defence of a Narrow Definition of Indigenous Religions." In *Religious Categories and the Construction of the Indigenous*, edited by C. Hartney and D. Tower, 38–57. Leiden: Brill.
- Database of Religious History*. Available at <https://religiondatabase.org/browse/645/#/>. Accessed September 16, 2021.
- Dobzhansky, Theodosius. 1937. *Genetics and the Origin of Species*. New York: Columbia University Press.
- Doss, Jacob. 2019. "Congregation of Savigny." *Database of Religious History*. Available at <https://religiondatabase.org/browse/661/>. Accessed September 16, 2021.
- Hansen, Anne. 2017. "Buddhist Communities of Belonging in Early-Twentieth-Century Cambodia." In *Theravāda Buddhist Encounters with Modernity*, 1st ed., edited by J. Schober and S. Collins, 62–77. New York: Routledge.
- Hayman, Peter. 1991. "Monotheism—A Misused Word in Jewish Studies." *Journal of Jewish Studies* 42 (1991): 1–15.
- Kukreja, Reena. 2020. "Meo Muslim, Mev, Mewati Muslim." *Database of Religious History*. Available at <https://religiondatabase.org/browse/645/>. Accessed September 16, 2021.
- Lane, Steven. 2020. "Free Methodist Church." *Database of Religious History*. Available at <https://religiondatabase.org/browse/879/>. Accessed September 16, 2021.
- Lehrich, Christopher I. 2021. *Jonathan Z. Smith on Religion (Key Thinkers in the Study of Religion)*. Abingdon, UK: Routledge.
- Lorenzen, David N. 1999. "Who Invented Hinduism?" *Comparative Studies in Society and History* 41 (4): 630–59.
- Masuzawa, Tomoko. 2005. *The Invention of World Religions*. Chicago: The University of Chicago Press.
- Matson, Joshua. 2020. "Sadducees." *Database of Religious History*. Available at <https://religiondatabase.org/browse/848/#/>. Accessed September 16, 2021.
- McCutcheon, Russell T. 1997. *Manufacturing Religion: The Discourse on Sui Generis Religion and the Politics of Nostalgia*. New York: Oxford University Press.
- . 2001. *Critics Not Caretakers: Redescribing the Public Study of Religion*. SUNY Series, *Issues in the Study of Religion*. Albany: State University of New York Press.
- McDaniel, June. 2010. "Agama Hindu Dharma Indonesia as a New Religious Movement: Hinduism Recreated in the Image of Islam." *Nova Religio: The Journal of Alternative and Emergent Religions* 14 (1): 93–111.

- Munda, Meenakshi. 2020. "The Sarna Religion of the Oraons of Jharkhand." *Database of Religious History*. Available at <https://religiondatabase.org/browse/973/>. Accessed September 16, 2021.
- Nongbri, Brent. 2013. *Before Religion: A History of a Modern Concept*. New Haven, CT: Yale University Press.
- Noseworthy, William. 2020. "Cham Bani." *Database of Religious History*. Available at <https://religiondatabase.org/browse/476/#/>. Accessed September 16, 2021.
- Orsi, Robert. 2005. *Between Heaven and Earth: The Religious Worlds People Make and the Scholars Who Study Them*. Princeton, NJ: Princeton University Press.
- Palmer, Jason. 2020. "Peruvian Mormons." *Database of Religious History*. Available at <https://religiondatabase.org/browse/891/#/>. Accessed September 16, 2021.
- Pennington, Brian K. 2005. *Was Hinduism Invented? Britons, Indians, and the Colonial Construction of Religion*. New York: Oxford University Press.
- Pepper, Kevin. 2019a. "Church of Jesus Christ of Latter-day Saints (early)." *Database of Religious History*. Available at <https://religiondatabase.org/browse/486/#/>. Accessed September 16, 2021.
- . 2019b. "Church of Jesus Christ of Latter-day Saints (modern)." *Database of Religious History*. Available at <https://religiondatabase.org/browse/525/#/>. Accessed September 16, 2021.
- Picard, Michel. 2011. "From Agama Hindu Bali to Agama Hindu and Back: Toward a Relocalization of Balinese Religion?" In *The Politics of Religion in Indonesia: Syncretism, Orthodoxy, and Religious Contention in Java and Bali*, edited by M. Picard and R. Madinier, 117–41. New York: Routledge.
- Prince, Alexandra. 2020. "The Church of Christ, Scientist." *Database of Religious History*. Available at <https://religiondatabase.org/browse/826/>. Accessed September 16, 2021.
- Proctor, Travis. 2020. "Christianity in Ephesus." *Database of Religious History*. Available at <https://religiondatabase.org/browse/948/>. Accessed September 16, 2021.
- Quang, Isvan. 2020a. "Raglai." *Database of Religious History*. Available at <https://religiondatabase.org/browse/727/#/>. Accessed September 16, 2021.
- . 2020b. "Haroi." *Database of Religious History*. Available at <https://religiondatabase.org/browse/752/#/>. Accessed September 16, 2021.
- Salter, Benson. 1987. 'Religio and the Definition of Religion.' *Cultural Anthropology* 2 (3): 395–99.
- . 2010. "Theory and Criticism: The Cognitive Science of Religion." *Method and Theory in the Study of Religion* 22 (4): 330–39.
- Schilbrack, Kevin. 2017. "A Realist Social Ontology of Religion." *Religion* 47 (2): 161–78.
- Shababo, Guy. 2019. "Late Chosŏn Korea." *Database of Religious History*. Available at <https://religiondatabase.org/browse/228/>. Accessed September 16, 2021.
- Simpson, Bethany. 2020. "Ancient Egyptian." *Database of Religious History*. Available at <https://religiondatabase.org/browse/738/>. Accessed September 16, 2021.
- Slingerland, Edward, and Joseph Bulbulia. 2011. "Introductory Essay: Evolutionary Science and the Study of Religion." *Religion* 41 (3): 307–28.
- Slingerland, Edward, and Brenton Sullivan. 2017. "Durkheim with Data: The Database of Religious History." *Journal of the American Academy of Religion* 85 (2): 312–47.
- Smith, Jonathan Z. 1982. *Imagining Religion: From Babylon to Jonestown*. Chicago: University of Chicago.
- . ed. 1995. *Harper-Collins Dictionary of Religion*. San Francisco: Harper-Collins.
- . 1996. "A Matter of Class: Taxonomies of Religion." *The Harvard Theological Review* 89 (4): 387–403.
- . 2004. *Relating Religion: Essays in the Study of Religion*. Chicago: University of Chicago Press.
- Stausberg, Michael. 2010. "Distinctions, Differentiations, Ontology, and Non-humans in Theories of Religion." *Method and Theory in the Study of Religion* 22 (4): 354–374.
- Stowers, Stanley. 2008. "The Ontology of Religion." In *Introducing Religion: Essays in Honor of Jonathan Z. Smith*, edited by W. Braun and R. T. McCutcheon, 434–49. London: Routledge.
- . 2011. "The Concept of 'Community' and the History of Early Christianity." *Method and Theory in the Study of Religion* 23 (3): 238–56.
- Tappenden, Frederick S. 2017. "The Database of Religious History and the Study of Ancient Mediterranean Religiosity." *Journal of Cognitive Historiography* 3 (1–2): 32–42.
- Tsonis, Jack. 2013. *Don't Say All Religions Are Equal Unless You Really Mean It: John Hick, the Axial Age, and the Academic Study of Religion*. PhD diss., Macquarie University.
- . 2017. "Against 'Indigenous Religions': A Problematic Category that Reinforces the World Religions Paradigm." In *Religious Categories and the Construction of the Indigenous*, edited by Christopher Hartney and Daniel Tower, 58–73. Leiden: Brill.
- Turchin, Peter. 2018. "Translating Knowledge about Past Societies into Seshat Data." *Clodynamics: The Journal of Quantitative History and Cultural Evolution* 9 (1): 143–147.

- Walsh, Robyn F. 2021. *The Origins of Early Christian Literature: Contextualizing the New Testament within Greco-Roman Literary Culture*. 1st ed. Cambridge: University Press.
- Ward, Samuel. 2019a. "Northern Irish Roman Catholics." *Database of Religious History*. Available at <https://religiondatabase.org/browse/483/>. Accessed September 16, 2021.
- . 2019b. "Northern Irish Protestants." *Database of Religious History*. Available at <https://religiondatabase.org/browse/484/>. Accessed September 16, 2021.
- Wills, John E. ed. 2010. *China and Maritime Europe, 1500-1800: Trade, Settlement, Diplomacy, and Missions*. Cambridge: Cambridge University Press.
- Wojahn, Daniel. 2020. "Sa skya." *Database of Religious History*. Available at <https://religiondatabase.org/browse/841/#/>. Accessed September 16, 2021.
- Womack, Nathan. 2020a. "Charismatic Renewal Movement in Christianity - Second Wave Pentecostalism." *Database of Religious History*. Available at <https://religiondatabase.org/browse/892/>. Accessed September 16, 2021.
- . 2020b. "Neo-Charismatic Movement - Third Wave Charismatic Movement." *Database of Religious History*. Available at <https://religiondatabase.org/browse/975/#/>. Accessed September 16, 2021.
- Zürcher, Eerik. 1959. *The Buddhist Conquest of China: The Spread and Adaptation of Buddhism in Early Medieval China*. *Sinica Leidensia*, vol. 11. Leiden: Brill.