Current Research in Ohio Archaeology 2021
Eric Olson
www.ohioarchaeology.org

Collector Reporting Bias

Eric Olson[1]

"Big Data" is quickly becoming a central theme of both concern and hope for archaeology's digital future (see Huggett 2020; McCoy 2020; VanValkenburgh and Dufton 2020). Many archaeological reports, discoveries, and collections are being digitized (Olson 2017; Olson et al. 2021), allowing researchers an opportunity to analyze data from the comfort of their home computer. Projectile points, specifically, are "superabundant" (Shott 2020:245) and diagnostic of broad segments of time (Pitblado and Shott 2015; Shott 2008) which are useful to diachronic study of mobility (Mullet 2009; Nolan 2014; Seeman et al. 2020). As more artifacts are photographed, an increasing number of different analytical techniques can be applied to these data. Applications such as TPSdig (Rohlf 2015) and AGMT3-D (Herzlinger and Grosman 2018) have greatly improved the accessibility of geometric morphometric analysis and capturing linear measurements and angles from photographs.

The Central Ohio Archaeological Digitization Survey (COADS) demonstrates the analytical value of private collections, and the power of these new data capturing techniques (Olson et al. 2021). The caveat, however, is in the quality of the raw data. Context is everything, and that cannot be captured with an aimless snap of the camera shutter. Photographs need to have color, high resolution, provenience data (usually in a separate spreadsheet or embedded in the file name), and a scale. Then there is the issue of parallax, or the distortion of shape because of the angle from which the photograph was taken.

COADS factors these issues into the project design and provides a rough "baseline" dataset of the representative distributions of projectile point types through time and space. Over 16,000 projectile points and lithic tools were documented as part of COADS, with the aim of documenting as representative as possible samples of projectile points in the region. The same considerations cannot be said of digitized collections, captured by different researchers, self-reported by collectors, and taken with different equipment under different conditions.

The following study compares the frequency distributions of projectile points from COADS (Olson et al. 2021), Seeman et al. (2020) and 1282 points data mined from various online digitized sources. All three of these datasets represent projectile points from private collections. However, the distinctions between the datasets are how the data were captured. The dataset compiled for this study was mined from online digitized archives (Ohio Memory, Ohio State University's Knowledge Bank), auction websites (eBay, Rowlands Relics, Estatesales.net and .com), and a limited number of private collections photographed by the author. The compilation of data was a multi-year process, with the general aim of adding data to datasets such as COADS or complimenting other projectile point datasets. What began as data mining slowly turned into a study of collector and market preferences, and the biases of private collecting in Ohio projectile point type distributions. Many professional archaeologists already have an anecdotal understanding of what gets bought and sold, what gets displayed,

[1] eric.olson@tri-c.edu

photographed, or generally disseminated between collectors. However, this study provides a quantitative breakdown of these biases.

Methods and Sources

Using the standards of COADS (Olson et al. 2021), all projectile points identified in the sample contained at minimum county level provenience, and some type of scale (e.g. ruler, dime, nickel, tape measure). Photographs from digitized collections were screen captured on the computer and saved as either a .jpeg or .png file. The file names were renamed to the county of provenience and a sequential number (i.e., "Adams_4" would be the fourth artifact documented from Adams county).

Most photographs were identified in back issues of *Ohio Archaeologist*, from first issue in the 1940s to 1990 (roughly 45 years). There was a precipitous drop-off in artifacts with county provenience and photo scales beginning around the mid-1970s. By the mid-1980s most issues had few to no photographs with county provenience or scales. Further investigations of issues into the 1990s to present were abandoned due to the low probability of artifacts meeting the minimum criteria. Sixty percent (n = 772) of projectile points identified in this study came from digitized issues of *Ohio Archaeologist* on Ohio State University's *Knowledge Bank* digital database (https://kb.osu.edu/handle/1811/55832). Another 23 percent (n = 298) came from auction websites (eBay and Rowland's Relics); 13 percent (n = 162) from collections the author could physically touch and photograph; the remaining four percent (n = 50) came from Ohio Memory, which includes photographs of collections from the Ohio History Connection.

For complete points, the photographic scale was used to capture blade and stem lengths using TPSdig (Rohlf 2015). These attributes will be utilized in future research and are beyond the scope of the current report. Unfortunately, due to the variable qualities of photographs, raw materials could not be consistently identified with any reasonable confidence. Many photographs were very small, likely taken on film cameras, in black and white. The digitization process likely did them no favors, either, making the DPI of imagines well below 300. Projectile points were identified using Justice (1987) and Ritchie (1971) and organized into groups based on Olson et al. (2021).

Results and Discussion

In total, 1282 projectile points and preforms were identified during data mining. This dataset will be referred to as the "collector reported" dataset, since these artifacts were largely the result of self-reporting by collectors. Figure 1 illustrates the frequency distributions of points by county. Counties like Ross and Licking are likely the result of their well-known association with Ohio prehistory (earthworks and Flint Ridge flint, respectively). The density of artifacts identified in northeast Ohio is more likely a reflection of the author's location (Summit County), and thus ability and access to physical collections, rather than the frequency of collector activity in this area. The high densities in Huron and Erie counties represent the collecting activities of Raymond Vietzen and Arthur George Smith, two prominent collectors in the mid-20th century based in Norwalk.

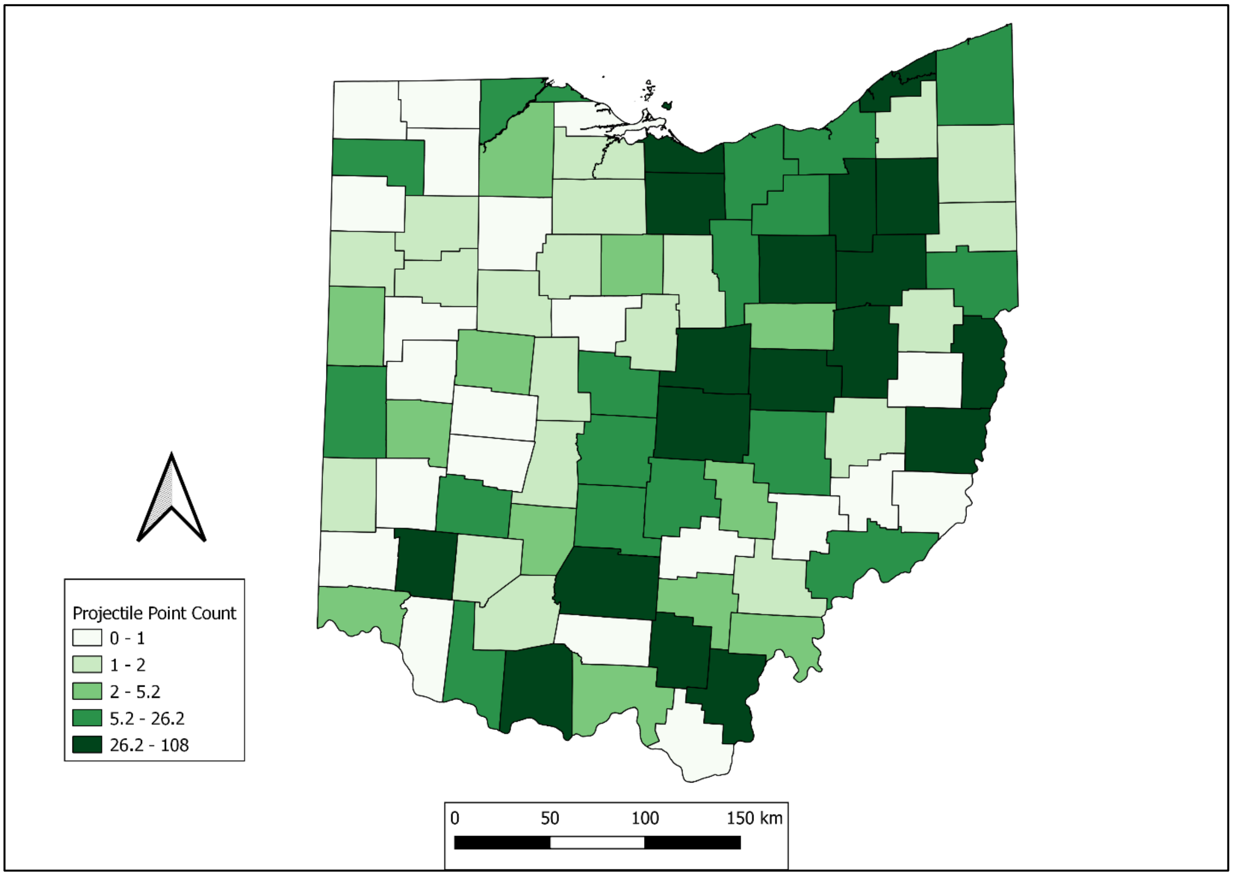

Figure 1: Choropleth map of frequencies of projectile points in sample.

For the purposes of comparison, the COADS dataset is the most representative sample of projectile point distributions for Ohio (Olson et al. 2021). There are several easily identifiable collector biases when the collector reported dataset is compared to COADS (Table 1). One out of every 10 artifacts reported by collectors was from the Paleoindian period, while COADS recorded roughly one out of 100 artifacts for the same time-period. Other notable point groups with over-representation include Dalton/Hi-Lo, Dovetail, Thebes, Kirk, Bifurcate, Merom and Turkey Tail. The only notable under-represented point group were Late Prehistoric triangles (e.g. arrow points).

Table 2 compares the COADS dataset, Seeman et al. (2020), and the collector reported dataset. Seeman et al. (2020) focused on Paleoindian and Archaic projectile points which were easily identifiable and represent a broad time-span. The collector reported dataset once again over-represents the PaleoIndian period relative to the other groups, but so do the data from Seeman et al. (2020). Brewerton is under-represented in both the Seeman et al. (2020) and collector reported datasets. This may be due to the difficulty of discriminating the Brewerton "type" (Ritchie 1971). However, based on the other patterns in collector reported data, it seems far more probable Brewertons are less likely to be reported due to their crude manufacture, frequency of breakage, and their general non-photogenic appearance.

This pattern of "photogeneity" is the most likely factor causing the biases in the collector reported data, in addition to market forces. PaleoIndian points sell for high prices, and photograph better than Brewertons. Meanwhile, triangles are the most abundant projectile point type, according to Olson et al. (2021), and are often broken. Roughly 36 percent (n = 4358) of all the artifacts reported in the COADS dataset were intact projectile points. Nearly 80 percent (n = 1024) of all artifacts photographed by collectors were intact.

In short, the collector reported dataset is biased towards intact projectile points that look pretty in photographs and sell for higher prices on the market because of their general rarity in the archaeological record. This pattern does not mean professional archaeologists should throw the baby out with the bath wash. The collector reported dataset is over 1200 artifact photographs, with scales, and county level provenience. These data ultimately enhance our collective understanding of the past, even if there are biases. However, knowing what collectors prefer to document allows researchers to "calibrate" their own research when working with private collections or digitized collection.

Table 1: Projectile Point Group frequencies from COADS dataset (Olson et al. 2021) and Collector Reported dataset (this study).

Time Period	Point Group	Olson et al. (2021) Percent	Collector Reported Percent
Paleoindian	Fluted lanceolates	1.39	10.70
Early Archaic	Dalton/Hi-Lo	0.09	0.48
Early Archaic	Hardin Barbed	0.13	0.08
Early Archaic	Hardaway	0.08	0.08
Early Archaic	Dovetail	2.17	6.44
Early Archaic	Thebes	2.27	5.71
Early Archaic	Kirk	11.43	10.38
Early Archaic	Kessel	0.30	0.48
Early/Middle Archaic	Bifurcate	3.73	5.47
Middle Archaic	Stanly Stemmed	1.25	1.69
Middle Archaic	Godar	3.15	3.30
Late Archaic	Brewerton	5.92	5.71
Late Archaic	Turkey Tail	0.06	0.80
Late Archaic	Meadowood	0.32	0.08
Late Archaic	Snook Kill	0.49	0.24
Late Archaic	Normanskill	0.51	0.80
Late Archaic	Merom	1.47	3.38
Late Archaic	Bottleneck	2.36	3.86
Late Archaic	Kramer	5.26	4.99
Late Archaic	Matanzas	2.51	0.32
Late Archaic	Lamoka	1.63	0.88
Late Archaic/Early Woodland	Terminal Archaic barbed	0.66	1.53
Early Woodland	Adena cache blade	1.07	0.08
Early Woodland	Adena Stemmed	14.18	11.75
Middle Woodland	Hopewell Cache blade	1.25	0.00
Middle Woodland	Snyders	7.58	5.23
Middle/Late Woodland	Lowe Cluster	5.87	4.75
Late Woodland	JRCN	4.14	2.09
Late Prehistoric	Triangle	18.73	8.69
Total		100.00	100.00

Table 2: Projectile point frequencies when compared to Seeman et al. (2020) state-wide research.

Time Period	Point Group	Olson et al. (2021) Percent	Seeman et al. (2020) Percent	Collector Reported Percent
Paleoindian	Paleo	9.78	26.39	38.78
Early Archaic	Thebes	14.88	12.96	20.70
Early/Middle Archaic	Bifurcate	26.51	26.24	19.83
Late Archaic	Brewerton	48.83	34.40	20.70
Total		100.00	100.00	100.00

References Cited

Herzlinger, Gadi and Leore Grosman
2018 AGMT3-D: A Software for 3-D Landmarks-Based Geometric Morphometric Shape Analysis of Archaeological Artifacts. *PLOS One* 13(11):e0207890. https://doi.org/10.1371/journal.pone.0207890.

Huggett, Jeremy
2020 Is Big Digital Data Different? Towards a New Archaeological Paradigm. *Journal of Field Archaeology* 45, Issue Supplement 1: Archaeology in the Age of Big Data: S8-S17.

Justice, Noel
1987 *Stone Age Spear and Arrow Points of the Midcontinental and Eastern United States: A Modern Survey and Reference.* Indiana University Press: Bloomington.

McCoy, Mark D.
2020 The Site Problem: A Critical Review of the Site Concept in Archaeology in the Digital Age. *Journal of Field Archaeology* 45, Issue Supplement 1: Archaeology in the Age of Big Data: S18-S26.

Mullett, Amanda N.
2009 Paleoindian Mobility Ranges Predicted by the Distribution of Projectile Points Made of Upper Mercer and Chert Ridge Chert. Unpublished MA thesis, Department of Anthropology, Kent State University, Kent, Ohio.

Nolan, Kevin C.
2014 Prehistoric Landscape Exploitation Strategies Through Time in Central Ohio: A GIS Analysis. *Journal of Ohio Archaeology* 3:12-37.

Olson, Eric
2017 No Shovel Required: An Archaeological Inventory of Greater Akron, Ohio. *Current Research in Ohio Archaeology 2017:* https://ohioarchaeology.org/articles-and-abstracts-2017/514-no-shovel-required-an-archaeological-inventory-of-greater-akron-ohio.

Olson, Eric, Kevin C. Nolan, and Michael J. Shott
2021 Central Ohio Archaeological Digitization Survey: Preliminary Report. *Current Research in Ohio Archaeology2021: https://ohioarchaeology.org/articles-and-abstracts-2021/601-central-ohio-archaeological-digitization-survey-preliminary-report.*

Pitblado, Bonnie and Michael J. Shott
2015 The Present and Future of Archaeologist-Collector Collaboration. *The SAA Archaeological Record,* 15(5): 36-39.

Ritchie, William
1971 New York Projectile Points: A Typology and Nomenclature. *New York State Museum Bulletin 384.* The University of the State of New York: Albany.

Rohlf, F. James
2015 The Tps Series of Software. *Hystrix, The Italian Journal of Mammalogy* 26(1):9-12.

Seeman, Mark F., Amanda N. Colucci, and Charles Fulk
2020 Hunter-Gatherer Mobility and Versatility: A Consideration of Long-Term Lithic Supply in the Midwest. *American Antiquity* 85(1):113-131.

Shott, Michael J.
2008 equal o nll roofht w ded I e vsbr cted: A Proposal for Conservation of Private Collections in American Archaeology. *Archaeological Record* 8, no. 2: 30-34.

2020 Toward a Theory of the Point, in *Culture History and Convergent Evolution. Vertebrate Paleobiology and Paleoanthrpology,* edited by H. Groucutt, pp. 245-259. Springer: New York.

VanValkenburgh, Parker and J. Andrew Dufton
2020 Big Archaeology: Horizons and Blindspots. *Journal of Field Archaeology* 45, Issue Supplement 1: Archaeology in the Age of Big Data: S1-S7.