**Measuring *Presence:* handling ambiguity and uncertainty in iconographic survey**

The world of iconography is messy. Really messy. It is full of ambiguity and uncertainty. It is therefore somewhat surprising that surveys of iconographic depictions produce such neat data using simple integer counts. In this article I will focus on the mechanics of analysis, how existing methods affect the results, and present an improved measurement and method to better reflect the real-world, accounting for ambiguity and uncertainty.

When performing iconographic survey we typically decide on a number of features we are interested in, and identify a number of variants of those features. We then proceed through the sources, examining each depiction (a source might include multiple depictions), and maintain a simple tally of how many depictions show each variant. If we choose to look at number of strings on bowed string instruments in the medieval period it would probably not be long before we met a depiction where one end has 4 strings, somewhere in the middle having wiggled around, one of the lines disappears, merges or splits and the rest of the instrument has 3 or 5 strings. How does this fit into our tally? What if instead of the middle it was mostly 3 (or 5) but a part of the string path shows 4?

We need a mechanism by which we can express certainty, so that we can meaningfully discuss it. Percentages, or ranges such as 1-10 are popular choices to express certainty, but do they have discernible meaning? What is the material difference between 61% and 62%? Or between 6 and 7? These scales suffer from being arbitrary choice. Having started with what we thought was a balanced decision about certainty, the more examples we see the more we refine our judgement, potentially invalidating our earlier choices. Do we actually need a one hundred point precision? Do we need a ten point precision?

If we are seeking truths rather than opinions, we should aim to reduce human factors influencing our analysis. The more we are relying on arbitrary decision making the more divergence there will be between observers. Should multiple observers perform the same analysis, the greater the choice for expressing uncertainty, the greater the variation in the results. What we need is a mechanism that can be easily expressed, where the definitions provide the necessary guidance to make an informed decision, such that multiple observers are likely to make the same choice, thus reducing the influence of the observer. We can achieve this using just three categories:

*Table 1: Categories of certainty*

| Category | Definition |
|---|---|
| *Definite* | The depiction is clear and does not indicate other possibilities. |
| *Probable* | The majority of the visual information indicates a most likely possibility, but there is an element of doubt such that we cannot be definite, or there are other possibilities indicated. |
| *Possible* | The visual information matches a given variant, at least in places, but there is significant doubt about the depiction (such as being particularly feint) or it is simply one of multiple possibilities and we do not feel that this possibility is more likely. |

In this example scenario we are analysing a particular feature of a specific instrument type across 7 depictions and have identified that there are 4 variants. Using our mechanism for describing certainty we observe the following.

Table 2: List of observations

| Depiction | Observation |
|---|---|
| 1 | Variant A Possible<br>Variant B Probable |
| 2 | Variant B Definite |
| 3 | Variant C Possible<br>Variant D Possible |
| 4 | Variant C Possible |
| 5 | Variant D Definite |
| 6 | Variant B Probable |
| 7 | Variant C Possible |

Typically the analysis uses a simple integer count system (tallies). Each depiction of our feature results in one variant count being incremented. The advantage of this system is simplicity. It is easy to use, no post-processing is required. No matter how large the data set the storage requirement is constant and minimal. Each time a count is incremented the state of the counts reflects the cumulative result. Error detection is a simple cross reference between the total of the counts and the total depictions. In general individual observations are not recorded. This system was popular before the extensive use of computing and has remained so, particularly where researchers are using spreadsheets.

Depictions 2 and 5 are straightforward and there is no issue about which variant's count we should increment. But how do we handle the ambiguity and uncertainty present in all other cases? We have a number of strategies available:

*Ambiguous count* — We can accept the ambiguity and increment the counts associated with all the variants that are present. Uncertainty is ignored. This strategy breaks the relationship between total of all counts and the number of depictions.

*Grouping* — This strategy groups variants together so a single count now covers multiple variants. This is particularly useful if the variants are identified by highly variant data such as real numbers rather than discrete values, as the number of counts we have to manage is reduced. Ambiguity is handled only where the group contains all possibilities present in the depiction. Uncertainty is ignored. This strategy can have a negative impact on analysis due to the preemptive aggregation of data, reducing the resolution of the final result. In the modern computing environment this strategy is less relevant.

*Partial Exclusion* — Where a depiction is ambiguous we shall choose the observation that is *Probable* ignoring the other possibles. If a single *Possible* exists it is used. If multiple *Possibles* exist with no *Probable* it is excluded. This relies on having a mechanism in place for expressing certainty to execute the exclusion.

*Exclusion* — Where a depiction is ambiguous or uncertain (anything other than *Definite*) we shall simply exclude it from the analysis.

*Arbitrary decision* This is used where there is no mechanism in the analysis for expressing certainty. Different observers will come to different conclusions, and the final results may vary significantly. Potentially our historical understanding may change purely because of which observer does the analysis. This is a highly undesirable scenario. It is highly susceptible to bias, and it is even possible that the same observer may reach a different conclusion when repeating the analysis.

Although not practical to simulate *arbitrary decisions, w*e can work through our example data to see how the other strategies influence the analysis.

*Table 3: Comparison of strategies*

| Variant | Ambiguous counts | | Exclusion | | Partial Exclusion | |
|---|---|---|---|---|---|---|
| | Counts | % of Depictions | Counts | % of Depictions | Counts | % of Depictions |
| A | 1 | 14.29 | | | | |
| B | 3 | 42.86 | 3 | 75.00 | 3 | 50.00 |
| C | 3 | 42.86 | | | 2 | 33.33 |
| D | 2 | 28.57 | 1 | 25.00 | 1 | 16.67 |
| Total | 9 | 128.57 | 4 | 100.00 | 6 | 100.00 |

*Total depictions considered: 7*

Using all three strategies resulted in clean data. No strategy truly reflects the observations, and the results vary significantly. Each distorts our understanding of the historical evidence. How much skew occurs depends on the particular strategy used and how ambiguous/uncertain the source depictions are. The more ambiguity and uncertainty the larger the skewing effect. This situation would not change with *arbitrary decisions*, especially given the vulnerability to bias. If we look at depiction 3 an *arbitrary decision* would promote one observation to *Definite* and the other would be excluded!

Many surveys avoid the use of the above strategies during analysis because ambiguous/uncertain depictions do not make it into the data set in the first place. There are two types of selection criteria, *source criteria*, based on source metadata such as place of creation, dating etc. and *depiction criteria* where the selection is based on the observable depictions present in the artwork. As the number of criteria increase the data set size reduces. It is a fairly common and much favoured practice to use exceedingly narrow criteria as this reduces the work required and avoids sources which are perceived as not having any value in the results. This is not entirely correct because the narrower the data selected by the selection criteria the less understanding we have of the context of the analysis results.

When the *depiction criteria* include the feature to be analysed *inclusion filtering* can occur. This filtering typically excludes depictions which do not show the feature, where the depiction is lacking detail, or where image resolution is poor, or where the depiction is not simply deemed not worthy for some reason. It can include an arbitrary decisions whereby the depiction is included or excluded knowing the eventual resolution (effectively performing *partial exclusion*). Such highly subjective, preemptive filtering has a significant capability to skew the results, and is highly susceptible to bias. The bias of the observer can be hidden by the very nature of the process. The analysis results will perform similar to *arbitrary decisions*

or *partial exclusion*.

Given the relatively poor performance of all these strategies it is surprising that they have been relied upon for so long.

One might suggest that in such small data sets large variation is to be expected. Whilst this is true to a degree, the variance in the results is directly proportional to the amount of ambiguity and uncertainty present in the data set, not the size of the data set. Even a large data set can be adversely affected, and we frequently need to analyse small data sets. My research in bowed string instrument iconography for the period 900-1499 CE currently has a corpus of around 2500 depictions. If we wish to analyse the features of one particular instrument type we have around 73 depictions, or instruments seen in a particular playing position yields around 36 depictions. Reliable small data set analysis is critical.

Can we do better? Undoubtedly.

The problems are caused by our methods, not by the iconography. When we design methods we must understand the nature of the data and take that into account in a meaningful way, so that our results better represent the sources. We need to embrace ambiguity and uncertainty wholeheartedly. A new strategy alone is not enough, because the aim to produce a figure that represents a simple "variant X is shown in Y percent of depictions" is unrealistic and is at the heart of the problem. Not only are our existing strategies flawed, so is our objective. Reality is just not that simple. Surely it is time to recognise this?

**A new measurement: *Presence***

*Presence* reflects the occurrences of a given variant in the data set, taking into account frequency, ambiguity and uncertainty. The method by which it is arrived at handles both large and small data sets, and produces reliable, meaningful results even with high levels of ambiguity and uncertainty.

This is a two-phase approach. In phase 1 we record individual observations, in phase 2 we process the observations to obtain our results. Recording individual observations has the advantage that each can be reviewed, even discussed. We can repeat this phase many times on the same data by the same or different observers and can compare where observations differ. With multiple sets of observations we can perform some interesting processing across the observation data sets to obtain a macro analysis. Although tempting, observations should be performed without knowledge of their affect on the results (or the observations of other observers) to minimise potential observer bias.

Each observation is recorded with the variant and one of the three categories of certainty as shown in Table 1: Categories of certainty. For each depiction each variant may only be recorded a maximum of once. Furthermore, for most types of data analysis multiple *Definite* values do not make sense. In these situations we would apply the following constraints:

- one *Definite*; or
- one *Probable*, zero or more *Possibles*; or
- one or more *Possibles*.

In phase 2 we use real number counts for each variant and increase the counts by the value assigned to the observations. A maximum value of 1.0 in total may be assigned to the observations for a single depiction. Ambiguity is handled by splitting the value between the observations. Uncertainty will reduce the assigned value.

We use the following algorithm to determine the value to assign to each observation.

- *Definite* is assigned 1.0
- *Probable* is assigned 0.75 if it is the only observation, or 0.60 if there are any *Possible* present. The reduction in the assigned *Probable* value where *Possibles* occur reflects the further uncertainty of the *Probable* introduced by existence of *Possibles*.
- *Possible*. The maximum value of any possible is 0.40. If multiple possibles exist for a given depiction then the available assignable value (1.0 if no *Probable* observation, 0.4 if there is) is divided equally, and the individual assigned values capped at no more than 0.40.

This use of 0-1 to reflect certainty is typical in statistics. Let's examine a number of scenarios to see how the value is assigned.

*Table 4: Value assignment scenarios*

| Observations | Values assigned | Notes |
|---|---|---|
| 1 Definite | 1.0 | |
| 1 Probable | 0.75 | |
| *1 Probable, 1 Possible* | 0.6 for the Probable<br>0.4 for the Possible | |
| *1 Probable, 2 Possibles* | 0.6 for the Probable<br>0.2 for each Possible | |
| *1 Possible* | 0.4 | Only 0.4 is assigned. |
| *2 Possible* | Each 0.4 | Only 0.8 is assigned as both values were capped. |
| *3 Possible* | Each 0.333 | |
| *4 Possible* | Each 0.25 | |

Whilst 1.0 is the total available to assign the situation may cause less than 1.0 to be assigned. This simple algorithm does not use complex mathematics, does not require computer programming or databases, and may be implemented manually or in spreadsheets. Returning to our initial example let's see what values are assigned to each of the observations.

Table 5: Recorded observations and values assigned during phase 2 processing

| Depiction | Observations recorded | Assigned value |
|---|---|---|
| 1 | Variant A Possible<br>Variant B Probable | 0.4<br>0.6 |
| 2 | Variant B Definite | 1.0 |
| 3 | Variant C Possible<br>Variant D Possible | 0.4<br>0.4 |
| 4 | Variant C Possible | 0.4 |
| 5 | Variant D Definite | 1.0 |
| 6 | Variant B Probable | 0.75 |
| 7 | Variant C Possible | 0.4 |

This yields the following results.

*Table 6: Analysis results*

| Variant | Number of depictions observed on | Total observation value | Presence (% of total available) | Mean certainty |
|---|---|---|---|---|
| A | 1 | 0.40 | **5.71** | 0.40 |
| B | 3 | 2.35 | **33.57** | 0.78 |
| C | 3 | 1.20 | **17.14** | 0.40 |
| D | 2 | 1.40 | **20.00** | 0.70 |
| Total | n/a | 5.35 | 76.43 | n/a |

*Total depictions considered: 7*

*Presence* is the percentage of depictions, reduced for ambiguity and uncertainty. If the data set contains no ambiguity or uncertainty, the result will be the same as simple integer counts and *Presence* will equal percentage of depictions. It is important that *Presence* is calculated as a percentage of the total assignable value (which happens to be the number of depictions!) not the total of the values that were assigned.

The most striking aspect of the results is how close they reflect the observations. Even with an incredibly small data set with high levels of uncertainty the method produces balanced meaningful results. Compare the *Presence* figures in Table 6: Analysis results with the percentage of depictions in Table 3: Comparison of strategies. We can see that variant B and variant C both were seen in 3 depictions, yet the *Presence* for B is almost double that of C. The mean certainty illustrates how uncertain we were when variant C was present.

We can enhance the contextual understanding of our findings, and reduce human influence and bias by reducing the selection criteria for the data to be analysed, such that as much of the corpus as possible is included. To avoid *inclusion filtering* the data set selection criteria should not include the feature to be analysed. Although this will include more depictions in our analysis, far from being a waste of effort, this additional data provides us contextual understanding. Of course there are trade-offs to be made, but I believe the added context and reduction in selection bias is invaluable.

To support the wider context of the analysis we must add some pseudo-variants so that we may record depictions that are not clear, or where the feature is not depicted. The following variants should be added to our list.

*Table 7: Pseudo-variants*

| Category | Definition |
|---|---|
| *Not Clear* | The depiction is not clear enough either in creation or image resolution to enable the possible variants to be enumerated. |
| *Not Depicted* | The feature is not included in the depiction. For instance, its expected location is hidden by something else being depicted, or it would appear outside the extent of the work. This frequently happens on instruments where part of the instrument is depicted as being behind something else. |

Further pseudo-variants may be used to ensure that as much as possible of the corpus is

considered, such as Not Played where playing position of an instrument is being considered. These pseudo-variants should not be confused with 'None' as a variant (for instance if analysing frets). Taking our example further, let's add 5 more depictions, 3 where the feature is not depicted, 2 where it is unclear. Our results now look like this.

Table 8: Analysis including pseudo-variants.

| **Variant** | **Number of depictions observed on** | **Total observation value** | **Presence** (% of total available) | **Mean certainty** |
|---|---|---|---|---|
| A | 1 | 0.40 | **3.33** | 0.40 |
| B | 3 | 2.35 | **19.58** | 0.00 |
| C | 3 | 1.20 | **10.00** | 0.40 |
| D | 2 | 1.40 | **11.67** | 0.70 |
| Not Depicted | 3 | 3.00 | **25.00** | 1.00 |
| Not Clear | 2 | 2.00 | **16.67** | 1.00 |
| Total | n/a | 10.35 | 86.25 | n/a |

*Total depictions considered: 12*

The pseudo-variants provide greater context for the analysis, and *Presence* now reflects a variant's place in the corpus. When presenting *Presence* figures it should always be clear whether the the analysis is with or without pseudo-variants.

To provide real-world insight, I have an analysis of string configurations of bowed string instruments in the period 900-1499 CE in progress. Of 1385 depictions analysed so far, 233 are not clear, and 125 did not depict the strings. Of the remaining 1028 depictions, 166 were ambiguous/uncertain (16.1%). Of the 1141 observations made for these 1028 depictions, 866 were *Definite*, 137 *Probable* and 138 *Possible*. Thus 24.1% of observations recorded uncertainty. The levels of ambiguity and uncertainty in actual research data sets is significant.

We have seen how attempting to express percentage of depictions as a measure of frequency of a variant in an iconographic data set is flawed, and simple tallies may result in misinformation skewing our understanding of history, undermining the purpose of research. Despite unsatisfactory performance, knowing the process we have followed lures us into a false sense of confidence in our results.

Using *Presence* improves on these methods, providing us with a reliable measurement that performs equally on large and small data sets, regardless of the level of ambiguity and uncertainty in the data set. The analysis better reflects the observations and provides greater insight and understanding of the iconography.