1. I’m actually quite proud of the title I came up with for this talk. Since you’re in this room, I assume you already know of Fenton John Anthony Hort and his contributions to New Testament textual criticism. What I’m going to show you today is that his work anticipates and fits neatly within the modern framework of *Bayesian phylogenetics*. I’ll explain what “Bayesian” means in a few slides. Phylogenetics, if you don’t already know, is a fancy word for the process of reconstructing a full family tree when you only have information about the family members who are still alive. It’s popularly used in biology, but it has also been applied fruitfully to textual traditions. In fact, it’s the method of choice for my PhD research on the tradition of Ephesians. But we’ll get to that later. Let’s start with the man of the hour. []

2. Hort co-edited a groundbreaking critical text of the New Testament in the late 1800s with the help of B. F. Westcott. They are depicted here with only minimal spelling errors. But Hort alone was responsible for their edition’s celebrated introduction, which outlines text-critical principles that we still follow today. This, I suspect, is why he’s commonly regarded as the Simon to Westcott’s Garfunkel. [Move to next sub-slide]

Hort’s most important contribution for our purposes was his taxonomy of text-critical evidence. If we model a textual tradition with a *stemma*, or family tree of witnesses, then Hort’s three main categories correspond to different parts of the stemma. The first category, *intrinsic probability*, concerns which variant readings best cohere with the author’s argument and style. It corresponds to the root or trunk at the top of the stemma. The second type of evidence, *transcriptional probability*, concerns how later readers and copyists of the text were likely to change the readings passed down to them. These scribal changes are modeled along the branches of the stemma. The third type of evidence is *genealogical evidence*, which concerns the shape or topology of the stemma—that is, how the surviving witnesses at the “leaves” of the tree are related through their lost ancestors. With a stemma like this in hand, we can calculate what these hypothetical ancestors were likely to have read based on their descendants and the habits of the scribes acting between them.

Of course, Hort spoke of “probabilities” only loosely, but really, there’s nothing in principle that should prevent honest biblical scholars from becoming perfectly competent gamblers. That is, in a Bayesian framework, we’re encouraged to quantify the strength of intrinsic and transcriptional evidence, and the payoff is that we can then quantify the strength of hypotheses about textual history. I’ll show you how. []

3. Intrinsic evidence is not inherently numerical, so we need a simple and consistent way to measure its strength. In practice, we can use a tiered ratings system like what the UBS Committee used for their Greek New Testament. The main difference is that for our purposes, each rating represents the relative likelihood of one variant reading compared to another. These odds ratios can be chained together to give a succinct representation of relative intrinsic probabilities. So in this example, reading *b* is slightly more likely than reading *c*, which is absolutely more likely than reading *a*, which is as likely as reading *d*. We then assign specific odds ratios to these ratings to ensure consistency across all variation units. []

4. Similarly, for transcriptional evidence, we can associate classes of scribal changes with numerical frequencies. To keep things simple, we might look only at broad classes: *clarifications* that smooth out the text or make it more edifying; aural confusions that became increasingly common in Greek, such as those between the sounds ι, ει, η, οι, and υ; linguistic confusions due to scribes’ unfamiliarity with Greek and the evolution of the language; visual errors due to common mechanical causes; harmonizations; and changes that conform the text to the standard Byzantine form.

The only problem is that we may not know the frequencies of such changes up front. But this is not a problem in Bayesian phylogenetics, because we can treat them as parameters to be estimated in analysis. In other words, this approach can quantify scribal habits for us. By modeling Byzantine assimilation as one such scribal habit, it also gives us a way to account for the most prevalent type of contamination in the tradition. []

5. In line with Hort’s taxonomy, transcriptional evidence plays out in the possible transitions from one reading to another in a given variation unit. So if one reading could have given rise to another by one or more causes, we can tag the corresponding transition with the applicable scribal change classes. The result is a mathematical object called a *Markov chain*. It’s illustrated here as a graph with nodes for variant readings and edges for possible transitions between them. For simplicity, the Byzantine reading is circled in blue, so we don’t have to draw a bunch of edges to it.

Now, if you’ve worked with the CBGM, you may be reminded of a local stemma here. A Markov chain has a similar function, in that it provides a local mechanism for explaining the origins of variant readings, but it represents something slightly different. A local stemma is a compressed history of a specific variation unit, while a Markov chain more abstractly models the behavior of the average scribe. Thus, where a local stemma should look more like a tree and describe the editors’ judgment of what *did* happen, a Markov chain is less constrained: it can contain circular transitions, and it can accommodate various possibilities of what *could* happen. []

6. That brings us to the genealogical evidence of the tradition itself. A Bayesian phylogenetic analysis evaluates millions of competing stemmatic hypotheses. Each hypothesis consists of a stemma topology and parameters for transcriptional rates and other variables. The goal of this process is to estimate the various probabilities of the hypotheses themselves, also called their *posterior probability distribution*, given the collation data. It gives us a precise measure of how certain a given reconstruction of the text or a group of witnesses is. In the demonstration section coming up, I’ll give you several illustrations of this.

This is where the “Bayesian” part of Bayesian phylogenetics comes in. We calculate the posterior probability using *Bayes’s Rule*, which is the short but sweet equation shown here. The *prior probability* of a stemmatic hypothesis is calculated using models for tree branching processes and prior distributions on the parameters. The *likelihood* or *explanatory power* of a hypothesis is the probability of the collation data arising under that hypothesis; it incorporates the intrinsic probabilities we assign to readings and the transcriptional probabilities of the transitions between them. Both of these quantities can be calculated efficiently. This last quantity in the denominator, called the *probability of the data*, is not so easy to calculate exactly, but luckily, we can estimate the posterior probabilities without it if we hop between candidate hypotheses in a clever way. []

7. So, to paraphrase our favorite apostle, what advantage does the Bayesian have, and what profit is a posterior distribution of stemmatic hypotheses? As it turns out, [move to next sub-slide] much in many ways! [Move to next sub-slide] First, Bayesian phylogenetics offers a clean separation of concerns between human judgments on the internal evidence of readings and calculations involving the external evidence of collated witnesses, which can be done by computer. It goes a step farther than traditional phylogenetics and the CBGM because it explicitly models intrinsic and transcriptional probabilities in different ways. [Move to next sub-slide] As a result, it can tolerate cases where these types of evidence point in different directions. When this happens, the genealogical evidence of the most likely stemmata can resolve the disagreement. More generally, since intrinsic probabilities are the prior probabilities that we assign to variant readings, strong intrinsic probabilities will favor certain stemmata over others, while weaker intrinsic probabilities can be confirmed or overridden by transcriptional and genealogical evidence. [Move to next sub-slide] As I’ve mentioned already, estimation of scribal habit frequencies is built into Bayesian phylogenetic analyses, and the resulting frequencies are incorporated into evaluations of the probabilities of stemmata. [Move to next sub-slide] Bayesian phylogenetics supports the use of clock models to relate the lengths of stemma branches to the rates of changes along them. This means that, in contrast with approaches that reduce witnesses to sequences of readings, Bayesian phylogenetics can incorporate the date ranges of witnesses to produce a more realistic model of textual history. [Move to next sub-slide] This is especially helpful with versional and patristic witnesses, many of which can be dated more precisely than manuscripts. Such witnesses, if they are sufficiently extant, can be included in analyses. [Move to next sub-slide] In general, having a probability distribution of competing stemmatic hypotheses enables us to measure just how certain we can be about the history of the text, the families of witnesses that arise in it, and which readings are the earliest. [Move to next sub-slide] Finally, if we have multiple candidate stemmata in hand, we can test their explanatory power at different variation units and reconcile them to address more complex cases of contamination. There isn’t time to get into the technical details here, but essentially, the local-genealogical principle can be applied with full-size stemmata like it can be with local stemmata. []

8. To demonstrate the power and potential limitations of this approach, I’ve applied it to four New Testament datasets at the variation units covered in the fifth edition of the UBS Greek New Testament. I used the UBS confidence ratings as proxies for the relative intrinsic probability of the UBS reading over other variants readings. I tagged transcriptional causes of transitions between readings as thoroughly as I could.

This is a visualization of the posterior distribution of stemmata for the UBS collation of 1 Peter, with sixty-four witnesses and thirty-six variation units. It’s rotated ninety degrees to make the labels more readable. The branches are thick and well-resolved where the stemmata agree and light and blurry where the stemmata have little agreement. Certain groupings are commonly identified in the stemmata, like the family of lectionaries and their connection to the Byzantine text, the proximity of Old Latin manuscripts to the Vulgate, and the relationship between the Syriac versions and the Armenian version. But much of what is happening in the middle and near the root of the stemma is cloudy. The main reason for this is that we have an insufficient sample of variation units; we need more data on how the witnesses split in order to get a sharper picture of how the groups that formed later are related at earlier stages of the text. And of course, even with the addition of further data, another potential source of uncertainty is contamination, at least from non-Byzantine sources. []

9. The great cloud of witnesses clears up a little bit for the UBS collation of Revelation, mostly because we have fewer witnesses and more variation units.

11. I’ve also included a plot of the posterior distributions for the relative rates of different types of scribal changes. There were some surprises here: I expected clarifications and visual errors to be more common, for one thing. But other things were not at all surprising: misspellings involving similar sounds were commonplace, as was mixture involving the Byzantine text. []