1. I'm quite proud of the title I came up with for this talk, but I should probably start by explaining it. Since you're in this room, I assume you already know who Fenton John Anthony Hort is. What I'm going to show you today is that his work in New Testament textual criticism anticipates and fits neatly within the modern framework of *Bayesian phylogenetics*. I'll explain what "Bayesian" means in a few slides. Phylogenetics just refers to the process of reconstructing a full family tree when you only have information about the family members who are still alive. This is done often 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 more on that later. Let's start with the man of the hour. [0:45]

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 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 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. Here's why it's important: if we model a textual tradition with a *stemma*, or family tree of witnesses, then Hort's three main categories correspond to distinct 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 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. 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. [2:00]

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 the one in the UBS Greek New Testament. The main difference is that we have to rate every reading, so our ratings will represent the relative likelihoods of pairs of variant readings. A chain of such ratings will determine the probabilities of all readings. So in this example, reading *b* is slightly more likely than *c*, which is absolutely more likely than *a*, which is as likely as *d*. We then assign these ratings specific numerical values as odds ratios like these, to ensure consistency across all variation units. [0:45]

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* to a smoother or more edifying text; aural confusions that became more common in Greek, such as those between the sounds ι, ει, η, οι, and υ; linguistic confusions due to scribes' unfamiliarity

with Greek or the evolution of the language; visual errors due to common mechanical causes; harmonizations; and assimilations to the standard Byzantine text.

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 the analysis. To put it more simply, this approach can quantify scribal habits for us. And by modeling "Byzantinization" as one such scribal habit, it also gives us a way to account for the most prevalent type of contamination in the tradition. [1:00]

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 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 between readings, and it can accommodate various possibilities of what *could* happen. [1:15]

6. That brings us to the genealogical evidence of the tradition itself. A Bayesian phylogenetic analysis evaluates millions of competing stemmatic hypotheses, each of which consists of a shape for the stemma and parameters for transcriptional rates and other variables. The goal of sampling all these hypotheses is to get a big-picture comparison of how different proposed histories are favored by the external evidence and how much of a spread there is in their likelihoods. In technical terms, we are estimating the *posterior probability distribution* of hypotheses given the collation data. In the demonstration section coming up, I'll give you several illustrations of this.

This, incidentally, is where the "Bayesian" part of Bayesian phylogenetics comes into play. We calculate posterior probabilities using *Bayes's Rule*, which requires just a few ingredients. The *prior probability* of a hypothesis is just how likely the stemma's shape and parameters are by themselves; it can be calculated easily using simple models. The *likelihood* or *explanatory power* of a hypothesis is the probability of the collation data arising under that hypothesis; it too can be calculated efficiently, starting from the leaves of the stemma with witnesses' readings, proceeding up the branches with transcriptional probabilities, and finally incorporating the intrinsic probabilities at the root. This last quantity in the denominator, called the *probability of the data*, is not so easy to calculate exactly, but we can estimate posterior probabilities without it if we hop between candidate hypotheses in a clever way. [1:30]

7. So, to paraphrase our favorite apostle, what advantage does the Bayesian have, and what profit is a posterior distribution of stemmatic hypotheses? Well, 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 the internal evidence of readings, which is assessed by human judgments, and the external evidence of witnesses, which can be evaluated by computers. 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 distribution of stemmata can resolve the disagreement. More generally, while strong intrinsic probabilities will favor certain stemmata over others, weaker intrinsic probabilities can be confirmed or overridden by transcriptional and genealogical evidence. This is how the circular relationship of "good" readings and "good" witnesses is managed in a Bayesian framework. [Move to next sub-slide] As I've mentioned already, estimation of scribal habit frequencies is built into Bayesian phylogenetic analyses. These frequencies, in turn, are incorporated into the probability calculations. [Move to next sub-slide] Bayesian phylogenetics also supports the use of clock models, which allow us to treat branch lengths as durations of time. 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 and can inform our analysis of textual history. [Move to next sub-slide] In general, a posterior probability distribution of stemmata gives us a way 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] In addition, we can test the explanatory power of multiple stemmata at different variation units and reconcile them to account for more complex cases of contamination. [2:30]

[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. For a demonstration, I've applied this approach to a few 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 probabilities of the UBS readings over their variants. I tagged transcriptional causes of transitions between readings myself.

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 disagree. 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. [1:30]

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. To make things more readable, I'll switch to another helpful diagram. [0:15]

10. This is the maximum clade credibility tree. It corresponds to the stemma in the posterior distribution whose subtrees or clades are best attested among all the sampled stemmata. Some obvious groupings have high probabilities: the two Coptic versions almost always appear together; the Andreas and Koine Byzantine groups are consistently identified; as in 1 Peter, the Syriac versions cluster together, with the Armenian version as an apparent cousin; and the Vulgate is closely related to the Old Latin manuscripts and Latin fathers. But the higher-level groupings are much less consistent. [0:35]

11. The UBS collation of Mark offers still better coverage with 148 variation units. We can also make use of the recent and more extensive ECM collation to cover more of this book's textual tradition. The full distribution of stemmata is too crowded to read at this scale, but I'll share a few of the higher-probability families from its maximum clade credibility tree. [0:25]

12. Here we can see that many of the witnesses that the CBGM classifies as closest to the initial text are consistently isolated to an early branch of the tradition. [0:10]

13. Family Π, named after the majuscule 041, is also frequently isolated in the best-found stemmata. The low probabilities just below these extremely short branches may indicate that the several of these branches proceeded from the same lost ancestor. [0:15]

14. Thanks to the ECM's inclusion of many of its witnesses, the Byzantine Family 1216 is clearly identified in all of the best-found stemmata. [0:10]

15. One particularly interesting outcome of the analysis is that over half of the best-found stemmata attest to a multi-layered "Caesarean" group. One of the earliest splits in the tradition seems to have gone three ways, with Codex Bezae to the West, Codex Washingtonianus to the East, and a third branch with remnants of this early text. Within this middle branch, we have an early stratum containing the majuscule 038 and two later branches corresponding to Families 1 and 13, with four minuscules situated somewhere between them. [0:35]

16. And here we have a plot of the posterior distributions for the rates of different types of scribal changes in Mark. We can make out that harmonizations, plotted in green, are about as prevalent as visual errors and clarifications, which seems reasonable for scribes copying Mark. But the most common types of changes are misspellings involving similar sounds, plotted in blue, and mixture involving the Byzantine text, plotted in brown. This last plot is especially significant, because it suggests that the Byzantine tradition frequently influenced scribes. [0:40]

17. Ephesians is our last example, because it's near and dear to my heart. The posterior distribution of stemmata visualized here is based on the UBS variation units and a small subset of the witnesses in the UBS collation. As we've come to expect, the upper branches are less resolved. [0:20]

18. Yet it's worth noting that even on this toy dataset, the Bayesian approach successfully isolates known groups, like Family 1739, the so-called "Western" tradition, the Harklean Syriac version and the witnesses to its Greek *Vorlage*, and the P46-Vaticanus pair.

I must stress that all of the results so far have been preliminary, illustrating this approach's applicability in breadth rather than depth. I chose the UBS datasets for their availability and ease of use, but as most of us already know, the UBS data is not sufficient for inferring a history of the text. So, to end this talk on a more promising note, I'll give you a preview of what I've already achieved with a fuller dataset of Ephesians… [0:50]

19. This is the posterior distribution of stemmata for Ephesians based on my collation of the IGNTP transcriptions. It covers 195 witnesses and 915 variation units. I've rotated the plot so that the root of the stemma is at the top. The witness labels are missing because I don't want to give away too many juicy details just yet. The more important point is that with sufficient data, Bayesian phylogenetics can converge on a history of the text that is well-resolved even at the early stages of transmission. [0:40]

20. All will be revealed soon in the form of my PhD thesis. Until then, you can learn more from these technical papers, and you can freely access the datasets and outputs from today's examples online.

I'll close with an important, but easily forgotten principle of textual criticism: the more information you bring to a textual tradition, the more you can get out of it. Bayesian phylogenetics exemplifies this principle in several ways. It incorporates intrinsic probabilities in a way that is sensitive to their probabilistic nature. It quantifies transcriptional probabilities and applies them rigorously. Most importantly, it gives us a framework for doing what was out of reach even for Hort and his successors—namely, assessing and reconciling genealogical histories that account for these internal factors and the external evidence. With such a powerful approach at hand, we have every reason to bring as much knowledge as we can to our task. Surely that's what Dr. Hort would have wanted. [1:15]