Research Statement

My current research focuses on three different areas:

- We combine various computational and observational techniques to link numerical models with data in systems of interacting galaxies. This project aims to connect dynamical parameters associated with galaxy interactions (time since the collision, mass ratios, etc.) to star formation rates and morphology.
- We harness crowd-sourced data to train neural networks in complex image identification tasks. The current project focuses on computational papyrology – the transcription and analysis of Ancient Greek manuscripts. In addition, this research focuses on how well machines can learn from imperfect data in a broader sense.
- We develop and test mixed reality systems in post-secondary STEM classrooms. Currently, this involves creating a scriptable interface for MR, developing laboratory modules, and testing it in the STEM classroom.

All three topics are motivated by my interest in applying new technologies to complex problems. For example, the crowd-sourced data sets I helped create in the Zooniverse project have driven my galaxy dynamics and computational papyrology research. Likewise, new wearable augmented reality availability is the foundation for the mixed-reality project through Magic Leap and Microsoft.

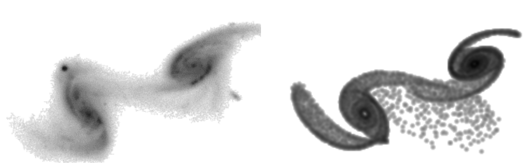
The Dynamics of Interacting Galaxies

Collisions and gravitational interactions between galaxies are common. These collisions play the dominant role in galaxy evolution through the merger of galaxies, the triggering bursts of star formation, and the feeding material into the central black holes of the merging systems. Although there is a connection between morphologically distorted systems and these phenomena, the timing and physical conditions that lead to them are less clear. To understand individual interacting galaxy systems, we need to connect the observations of these galaxies to numerical models of the gravitational interaction to determine the time scales and physical parameters associated with the collisions. Effectively this requires matching n-body simulations of interacting systems to the systems' morphology and velocity field.

From 2011 through 2015, my team created Galaxy Zoo: Mergers as part of the Zooniverse project. We designed the Zooniverse to allow volunteers (citizen scientists) to help scientists solve pattern recognition and matching projects. Zooniverse projects have included classification of galaxy morphology, analysis of data of camera traps in Africa, transcription of manuscripts, and identification of sounds from whales and bats.

Galaxy Zoo: Mergers relied on citizen scientists to compare simulations of galaxies to the observations of existing systems. We created approximately 60 best-fit models of target systems during this project by having volunteers review over 3.5 million simulations.

The images on the left illustrate the challenges of comparing simulations and data. They represent the observed and simulated images of Arp 240 – a system of two interacting galaxies. We obtained the image on the left from



the Sloan Digital Sky Survey. The image on the right was the best fit model that resulted after volunteers reviewed 100,000 potential models. Although the human eye can see the similarities in the overall structure, pixel-to-pixel comparisons of the image are problematic. For example, changes in the surface brightness of the galaxy have little effect on the tidal morphology but are extremely difficult to match. In addition, the relatively low resolution of the n-body model on the right leads to the over-emphasis of low-density regions that are not visible in the telescopic image of Arp 240.

To explore how interaction is tied to increased star formation rates and triggering active galactic nuclei, we need to have numerical models of hundreds of systems. Since each system may require a hundred-thousand simulations, using crowd-sourcing to solve this problem is impossible

Over the last five years, we have been developing a GA to model interacting systems automatically. The biggest challenge was making a fitness function that reproduces some of the characteristics of the human ranked data. Naïve pixel-to-pixel comparisons produce inconsistent with human judgment, so we used the available models to develop a more robust fitness function for this problem.

We submitted our work on this fitness function for publication. Our new goal is to quantify the uncertainty of the models and apply them to dozens of observed systems. In addition, by comparing these models to the levels of AGN activity through x-ray and radio flux, we will understand when dynamical interactions trigger nuclear activity.

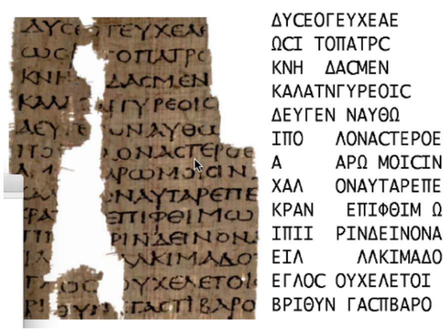
This work is currently unfunded. However, future funding may be available through NSF DMS-AST programs and the NASA Astrophysics Theory Program. This project extends the work completed under the following grants.

- "CDI-Type II: Zooniverse - Conquering the Data Flood with a Transformative Partnership between Citizen Scientists and Machines," National Science Foundation – Cyber-Enabled Discovery and Innovation Program, Lucy Fortson (Project PI), John Wallin (Co-PI and Institutional PI) – total award $1.89 Million overall funding ($415k MTSU), award period 2011-2014

- "The Redshift Dependence of the Interaction-Activity Connection among Ultraluminous Starbursting Galaxies," NASA Astrophysics Theory Program, Kirk Borne (project PI), John Wallin (GMU PI), total award $221k ($88k to GMU), 1999-2003

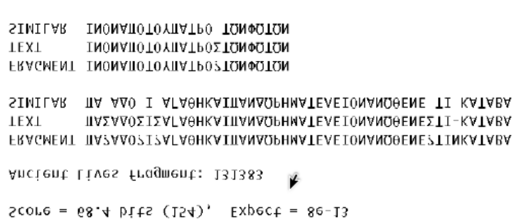
Computational Papyrology – Applying Computational Techniques to Ancient Greek Manuscripts

Outside the ancient city of Oxyrhynchus in Egypt, archeologists in the 1920s uncovered approximately 600 crates of Greek papyri. Unfortunately, since their discovery by Grenfell and Hunt, only about 10% of the manuscripts have been published. In 2011, we developed the Ancient Lives website to help improve this process. We created this website as part of the Zooniverse Project to crowd-source transcribing these documents into electronic form. I was the lead data scientist on this project and helped build the software to transform the user clicks on images and virtual keyboards into consensus transcriptions.



The image to the left is an example transcription from the project. The clicks and identifications of multiple users were aligned to produce this image. Note that the handwriting on this particular manuscript uses a style known as "book hand." Although the letters are easy to read, the gaps in the manuscript still make this challenging to identify.

After some initial transcriptions, we created software to determine if a manuscript was part of the known corpus of Greek documents. The underlying software for this was the bioinformatics software package BLAST. We coded the library of available manuscripts and the transcribed documents as amino acid sequences and used BLAST to look for alignments. The image on the left shows an example of manuscript identification. There are gaps and misidentified



characters in the document, as with genetic sequences. However, the software found a match between an unknown and a known manuscript.

As a side product of these transcriptions, we created a database of crowd-sourced ancient Greek characters with 400,000 images. This data provides the foundation for our new work in automated character recognition. We have several goals in this project.

First, we wish to improve the accuracy of the deep learning networks we are using to identify characters from these manuscripts. Although the overall accuracy we are getting using RESNET is nearly 90%, some character classes remain problematic. One of the underlying problems is the uncertainty and accuracy of the training data. Although crowd-sourcing is the only practical way to create this kind of data, some users are better than others. In addition, some characters are more challenging to read than others. The fundamental problem we are trying to solve is how do we improve the automated character recognition algorithms given the limitations of the data? Understanding when expert input is needed is critical for improving the database and real-world applications. However, the relationship between the uncertainty in both the human and machine identifications presents some unique challenges. Ultimately, we wish to make this system a tool that can be applied used by experts in the field to improve their productivity in curating these documents.

Second, we wish to understand if we can quantify the accuracy of users through their behaviors. Comparing the transcription of an individual user to those of either the consensus data or the data from the neural network can help us do this. We can also look at the click patterns used by individuals as they transcribed the data. When transcribing the documents, the most experienced users tend to follow a left-to-right, top-to-bottom approach. These users transcribe all the characters in the image following a sequential pattern. Inexperienced users tend to skip across a manuscript, selecting only easily recognizable characters. Understanding user behavior and how it can predict transcription accuracy should help us vet the data and improve the accuracy of our training and testing data.

Finally, we wish to date the manuscripts and identify scribes using unsupervised and semi-supervised learning. Subtle changes in handwriting in certain characters can date these manuscripts, but the process currently involves intensive expert analysis. Automating this analysis will be challenging, but the large number of documents and characters in our data set may make this training possible.

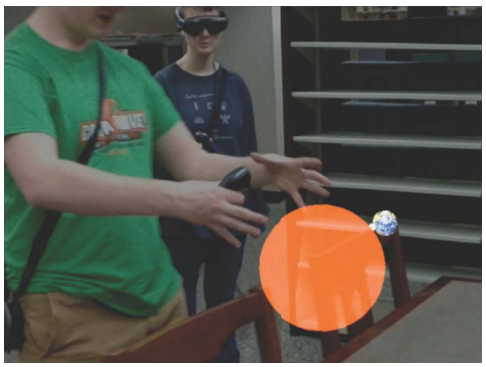
Although this work is currently unfunded, it has an excellent potential to obtain funding. The techniques we are developing are broadly applicable to a wide variety of other images and manuscripts. The NSF CISE program and research grants through companies like Google or Amazon are possible. Previous grants that funded parts of this work include:

- "Resurrecting Early Christian Lives" -NEH/NSF Digging into Data Challenge: Total funding: approximately \$175k - \$24,975 for MTSU from 2/1/2014 to 1/31/16 (sub-award from the University of Minnesota) – co-PI Wallin
- "CDI-Type II: Zooniverse - Conquering the Data Flood with a Transformative Partnership between Citizen Scientists and Machines," National Science Foundation – Cyber-Enabled Discovery and Innovation Program, Lucy Fortson (Project PI), John Wallin (Co-PI and Institutional PI) – total award \$1.89 Million overall funding (\$415k MTSU), award period 2011-2014

Augmented Reality

Augmented reality (AR) allows the real world to be enhanced or augmented by computer-generated objects that are "added" to the real world. For example, a furniture store may use AR to allow a customer to "see" what a piece of furniture would look like in their home. AR has the potential to help students to visualize, interact with, and learn about complex physical systems. The CyberLearnAR project explores how wearable AR technology, when combined with best classroom practices, affects the learning process of college-level STEM students. Groups of students view and manipulate 3D holographic objects in a collaborative learning environment. A team composed of computational/computer scientists, disciplinary experts, and social scientists is examining how AR can impact student learning. This work will guide the development of the next generation of AR technology and add to our understanding of AR technology in the classroom.

AR is a potentially transformative first-generation technology. However, current platforms lack applications for STEM teaching at the introductory college level. My CyberLearnAR team is developing a shared augmented reality and spatial computing group educational environment that can easily accommodate changes in the underlying technology. Using this environment, the project team is studying the technological, sociological, and cognitive obstacles to and the benefits of using the new technology in two different target audiences according to theoretical frameworks for learning. The image on the left shows two student members of my development team interacting with a virtual model of the Earth, Moon, and Sun.



During the first year of the grant, we developed the core software framework for this experiment and implemented our first laboratory experience. The development work and evaluation are being done by a team of undergraduates, graduate students, and faculty members at Middle Tennessee State University. We tested our software on approximately 50 students in the Fall of 2021, and the data analysis is underway. We continue to refine the software environment, improve the capabilities, implement new laboratory experiences, and test it on students during the Spring of 2022.

The National Science Foundation funded this project under:

"Cybertechnology Development and Exploration of Learning Processes in Augmented Reality Team Environments (CyberlearnAR)" National Science Foundation - IIS - Cyberlearn & Future Learn Tech, John Wallin (PI), $300k, Award Number 2017011, 10/1/2020-9/30/2023

Because this project focuses on the educational impact of an emerging technology, it has a high potential for additional funding and commercial development.

Teaching Statement

Teaching Philosophy

Being an effective teacher is still one of the most rewarding parts of my job. As I have advanced in my career, my teaching has changed. Educational research across the disciplines has guided me toward a better understanding of what works and what doesn't work. Working with colleagues, I have tried to adopt their best practices in my classroom and adapt to our changing understanding of what it means to be a professor.

I find teaching introductory courses extremely challenging. Students in these courses often have weaker backgrounds, lower motivation, and face obstacles adapting their lives to college. We tend to teach these courses in larger sections with less support. At the same time, introductory courses are critical for building the foundation for more advanced learning. My course may be the last time they are taught sciences in general education courses. The knowledge and attitudes they learn in these courses will influence how they view science throughout their lives.

Although the material in introductory courses is simple, structuring an effective learning environment given the constraints is exceptionally challenging. I learn lessons from my teaching, my colleagues, and my students about increasing the effectiveness of my courses.

Teaching advanced courses (both at the undergraduate and graduate level) presents different challenges than teaching introductory courses. Although advanced students have higher motivation, stronger backgrounds, and better core skills, many of the classes I teach have students from diverse academic backgrounds. For example, a graduate course in Data Science may have students with backgrounds in Statistics, Computer Science, and Biology. Some of the graduate courses I teach have no textbook or well-defined curriculum. Defining the class material, making effective assignments, designing projects, and creating evaluation metrics presents different challenges than introductory courses.

Teaching Approach

In my current approach to teaching, I believe that students learn best when they have effective study habits, are motivated to engage in the material, have clear expectations of what is expected of them, are not embarrassed to make mistakes, and know they are respected even when they get things wrong. Helping students by creating a classroom environment that is welcoming but challenging is at the heart of how I teach.

Creating effective study habits

Many beginning college students don't know how to study effectively. Their study habits in high schools are sometimes inadequate for their college courses. The distractions of university life and the lack of formal structure make it difficult to adapt. At MTSU, most of our students work part-time jobs to help pay for their education. In addition, my University has a significant fraction of first-generation college students who don't' have enough personal or financial support from their families. These issues are particularly important for some underrepresented groups. Although I can't solve all these problems myself, I can provide some guidance and structure to help these students succeed.

When students cram before an exam, they often study what they know instead of learning the concepts they don't know. I explain that typical retention 48 hours after a lecture is only about 15% in my introductory classes throughout the semester. I explain that if you don't review the material within 24 hours, you won't transfer the lessons from short-term into long-term memory. I encourage my students to anticipate what will be on the test by actively writing example test questions on the lecture's day. They can share these questions and discuss the key concepts by forming study groups.

When students tell me they are having trouble in the class, I ask them to write sample questions based on the last lecture to discuss them. Then, I critique these questions to give them a better idea of how I think about evaluation.

Motivating Students

I try to motivate my students to engage deeply in the material I am teaching. While being enthusiastic about the material can help encourage some students to learn, connecting abstract ideas to concrete real-world examples is more critical. Most professors tend to be abstract thinkers, but this is a learned skill after decades of study. We got to where we are by moving from the concrete to the general, not the other way around.

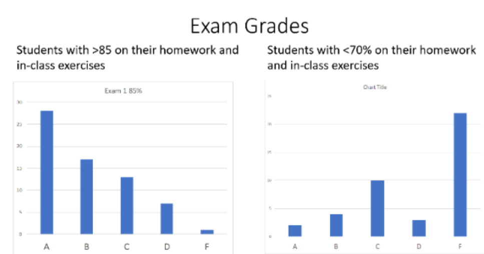
For introductory students, it can be challenging to connect complex ideas in astrophysics to everyday problems. However, scaffolding instruction can help. Talking about the Sun before talking about other stars lets students build a framework from the familiar to the unfamiliar. In some cases, moving directly to the unknown can be helpful. Most students are excited to learn about black holes, so including new topics like this can let the students explore ideas they have heard about in pop culture in greater depth.

I have found it critical to make sure students understand how we know things, not just what we know. Providing some details about how science works or the underlying basis of a numerical method helps them have confidence in applying this knowledge. It is also essential to candidly let students know what we don't know. There are gaps in human knowledge, and it is a great disservice to students to make them believe everything is solved.

Clarifying expectations

One of the most challenging parts of learning is understanding a lesson's structure and critical elements. Although the instructor understands what they are trying to teach, students may not have the experience to process these lessons into knowledge. Therefore, I try to provide materials and structures that help clarify these expectations to help students learn.

For me, this process beings by defining clear educational objectives for each lesson, lecture, and exam. Having goals that align with exercises and exercises that align with exams prevents students from being surprised. A perpetual complaint of college students is "that was a trick question." I work hard to align the questions on my exams to the material taught in the class. After I finish grading, I review the exams to learn which questions were the most difficult to verify that these weren't overly tricky or unexpected. When I find a question, that is, I drop that question. I tell the students why I did this and how it will impact their grades. On the day after the exam, I talk informally to students before the lecture to determine if the exam was what they expected.



Some students believe that studying and doing homework doesn't significantly impact their test scores. I show the correlation between homework and in-class clicker participation with test performance in my classes. In test after test in my introductory courses, the students who get over 85% of the participation and homework points get mostly A's on the exam and very few F's on exams. Students who earn less than 70% of the participation and homework points get mostly F's and almost no A's on the exams. I show students this data after each exam in the semester. Doing this has changed the DWF rate in my course from 38% to about 23%. The image to the left is from one of my exams last semester.

Collecting and analyzing data on my students' performance helps me understand their misconceptions. More importantly, it helps me know where my teaching is failing. However, effectively communicating the data lets the students assess and adapt their behaviors to succeed in class.

Learning by making mistakes

Everyone is embarrassed to make mistakes. We want to have the correct answer, particularly in front of others and in a group setting. However, learning depends on making mistakes, recognizing them, and learning from them. Therefore, I try to design elements in my course to let students make mistakes in low consequences environment to help them understand what they don't understand.

The cornerstone for creating this learning environment is the frequent homework and in-class that I assign. My introductory astronomy class has in-class exercises every lecture. In addition, the students are given online homework after every lesson. I typically have weekly assignments, periodic quizzes, and in-class group exercises in my graduate courses. Besides helping students review the topics from the lecture, they received rapid feedback on these low-stakes exercises.

The in-class exercises are designed to promote group discussions. I ask both open-ended questions and questions with a specific answer. The open-ended questions can help students explore complex topics, and the more concrete questions help students learn the problem-solving techniques they need for the course. We discuss the group results at the end of the exercise. I make a point to explain that not every answer will be perfect, and the focus of the activity is to discuss and learn, not just to get the correct answer. Points are awarded primarily for participation, not for correctness. The goal is to encourage students to solve these challenges independently, not google the answer or rely on others.

Students are confronted with concepts from the lessons that they should understand in the homework and quizzes. I allow students to retake questions they got wrong with only a minor penalty when they attempt online homework. The goal is to let students solve problems to learn what they do and don't know. Quizzes in my graduate classes are low stacks as well. Questions from one quiz may appear on a later quiz to encourage students to learn what they missed.

Being Approachable

I strive to be approachable to all students. I spend time before class talking to students before the lecture. These conversations are designed to humanize the classroom. Reaching out to students is critical in large lecture types but helpful across all the courses I teach.   Students can find professors intimidating. They may have had negative interactions with other instructors, so developing that trust can be challenging. The first step in this is listening. When students ask for help, my main goal is to be kind. Although I may not extend the deadline on an assignment because of the class policies, I can improve in future projects. The most challenging conversations are always at the end of the semester when a student has failed the class and asks for extra credit or makeup.   I must kindly say – no. I tell students that failing a class doesn't mean they couldn't do better under different circumstances. I explain that I must give them the grade based on their work, not their potential.   Although these are difficult conversations, I hope my students walk away knowing that their failing my class does not mean that I am judging them as a person.   Being kind costs us nothing, but it can transform how students view themselves and the University.

By being open to new perspectives and kind, I strive to broaden my perspective as a teacher. I am committed to improving my teaching and students' educational experiences at the University. To me, this commitment means caring about the research students I supervise and the students in my general education classes. It also means listening to my students, learning from my colleagues, paying attention to the data in my classes, and adapting my teaching style as needed. Although I have taught for decades, I am constantly learning how to teach.