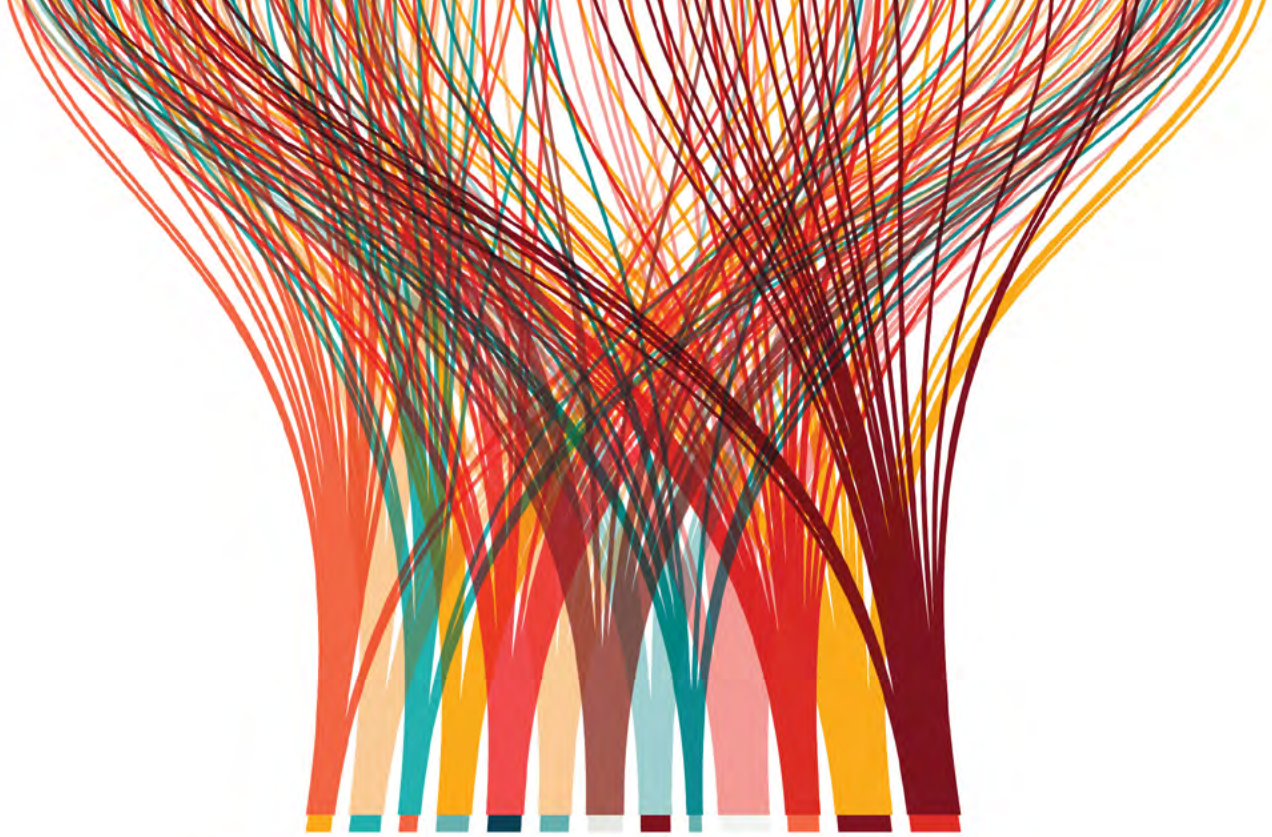


Big Data Evolution

GUSTAVUS ADOLPHUS COLLEGE OCTOBER 5 & 6, 2021

THE NOBEL CONFERENCE
GUSTAVUS ADOLPHUS COLLEGE

Welcome to the 57th Nobel Conference. We know our communities, country, and world face many great challenges. At Gustavus, our aim is to equip students to address these challenges and encourage them to be people of action.

Research and data inform many of the decisions we make. In fact, big data presents its own great challenges and ethical questions. It is critical for all of us, in our organizations and communities, to explore how data is collected, used, and stored.

I am delighted you are joining us to learn more about this important topic. May you leave better informed, inspired to share what you heard, empowered to manage your own data, and committed to using data for the betterment of society.

Sincerely,

Rebecca M. Bergman
President,
Gustavus Adolphus College

NOBEL CONFERENCE 57

WELC

Accumulating data is not a new activity. Governments, industries, think tanks, and even sports teams have always tracked and accumulated information—about public health, about consumer purchasing patterns, about population trends.

What's new is the amount of data the electronic devices in our world enable us to accumulate, and the ways in which that accumulated data can be analyzed. To get a glimpse of the change, consider the game of baseball. Whereas in the past, statisticians accumulated box scores and calculated batting averages and pitching earned run averages, today's technologies enable us to generate reams of data about a single pitch—and to put that data to use in all sorts of new ways. Batting and pitching data are no longer just ways of judging players. They now inform in-game strategy to such a degree that baseball games have actually become longer—and, some would argue, duller.

In 25 years, the microchip revolution, the digitization of data, and the development of ways to store massive amounts of data have enabled the creation of massive, easily accessible bodies of information. Advanced computational and statistical methods, known collectively as machine learning (ML) and artificial intelligence (AI), have enabled us to manipulate this information in myriad ways. The term "Big Data" was coined in 2005 to reflect the relative ease with which enormous data sets are now accumulated and analyzed.

The evolution and use of big data sets have generated significant practical, ethical, and policy-related questions, as well as many controversies. To begin to understand some of the reasons why, consider the amount of information your cell phone or personal activity tracker collects about you in a single day. Then add in the data you generate if you post on social media. What happens to all

COME

that data? And how much control do you have over its use? Today, much of our personal information has been digitized and rendered trackable, with all the promises and perils such tracking entails.

It would be easy to dive down the rabbit hole of why the collection of data is bad. Headlines from around the world describe the theft of identities and financial data, the hijacking of public infrastructure for ransom. However, even if it were possible to stop accumulating large amounts of data (which it is not), we cannot ignore the data that has already accumulated—nor would we want to. The possibilities of big data are every bit as important and promising as its problems are perilous. We must, therefore, develop ethical principles for collecting and using personal data that also enable the accumulation of that data for public benefit.

Nobel Conference 57, **Big Data *RE*volution** focuses on just that nexus of opportunities and challenges. How can we safely protect our personal information? How can we use big data to address universal world challenges—pandemics, climate change, hunger? Can we use big data to improve medical treatment, business operations, social services, human relations?

Big Data is here to stay. Nobel Conference 57 will turn to some of the world's leaders in the collection and use of Big Data to guide us on how this massive amount of information can be used for the benefit of all.

Nobel Conference 57 Co-Chairs
Thomas LoFaro, *Professor, Mathematics, Computer Science, and Statistics*
Karl Larson, *Professor and Director, Public Health*

Nobel Conference Director
Lisa Heldke '82, *Professor, Philosophy and Gender, Women, and Sexuality Studies*

The Nobel Prize is a positive force because it includes many strong and inspiring stories about human progress in difficult situations, about the importance of knowledge to enable decision makers to make informed choices, about freedom of expression and the importance of the written word for a free and enlightened public, and about peaceful dialogue and conflict resolution as the basis for a free society and human rights. The Nobel Foundation is pleased that the tradition of the Nobel Conference at Gustavus Adolphus College continues because this Conference has a long-history of similar values and work.

Vidar Helgesen
Executive Director, Nobel Foundation

Tuesday

OCTOBER 5, 2021

9 a.m. PRELUDE

The Gustavus Wind Orchestra

James Patrick Miller, DMA, conductor

Early Light	Carolyn Bremer (1957–2018)
Selections from *Suite Francaise*	Darius Milhaud (1892–1974)
Riften Wed	Julie Giroux (b. 1961)
Renesans	Sydney Guillaume (b. 1982)

9:35 a.m. CONFERENCE OPENING

Welcome

Lisa Heldke, PhD
Nobel Conference Director
Professor of Philosophy

Rebecca M. Bergman
President of the College

Conference Introduction

Tom LoFaro, PhD
Nobel Conference 57 Co-chair
Professor of Mathematics, Computer Science, and Statistics

10 a.m. FIRST LECTURE

Talithia Williams, PhD
Professor of Mathematics, Harvey Mudd College

Data-Driven Decision Making: Now and Imagined

Statistician Talithia Williams works as an interpreter. But instead of translating Spanish to Urdu, she takes difficult and arcane concepts in mathematics and statistics, making them understandable, interesting, and relevant to the rest of us. Apropos the topic of this year's conference, her popular TED talk, "Own Your Body's Data," is an argument for the importance of "small data." In it, she tells her audience that each of us should be using the tools of personal data collection—from heart rate monitors to step counters—to take charge of our own health. Her work on the PBS series

Nova Wonders focuses on introducing future scientists and engineers to some of the most pressing scientific questions of our time. Rather than presenting answers, the series is "really opening the door and arousing curiosity and getting the public excited about the possibility of even trying to contribute to and solve some of these big ideas and big questions," notes Williams. Two episodes featuring topics related to Big Data: "Secrets in our DNA" and "Prediction by the Numbers".

In addition to her work to make mathematics and statistics understandable to the everyday person, Williams develops statistical models that emphasize the spatial and temporal structure of data, and applies them to problems in the environment, She has partnered with the World Health Organization in developing a model used to predict the cataract surgical rate for countries in Africa.

Williams holds a PhD in statistics from Rice University. Currently a professor of mathematics and statistics at Harvey Mudd, she has held positions at the Jet Propulsion Laboratory, the NSA and NASA. In 2015, she won the Mathematical Association of America's Henry L. Alder Award for Distinguished Teaching. Williams was honored by the Association for Women in Mathematics and the Mathematical Association of America as the Falconer Lecturer in 2017, and was recognized by Mathematically Gifted & Black as a Black History Month 2017 Honoree. She is the author of *Power in Numbers: The Rebel Women in Mathematics.*

HER TALK

Technology has a history of being a catalyst of change in training and education. We've seen it with desktop computers and, more recently, with the emergence of smartphones. But those shifts, substantial as they were, pale in comparison to the next big technological disruption: Data. In this talk you will discover how the advancing world of data analytics is forever changing the future of learning and work. You will explore the full landscape of data analytics, looking at both the expanding ways in which data is generated, and the advancements in analytics that make that data actionable. You will hear examples of data being used to better understand performance in both education and enterprise, and learn how those insights are being used to inform decision making and transform society.

Introduction — **Jillian Downey, PhD**
Assistant Professor of Mathematics, Computer Science, and Statistics

Post-lecture Reflections — **Lisa Heldke** and **Jillian Downey**

11 a.m. PANEL DISCUSSION AND AUDIENCE QUESTIONS

Submit your questions electronically at any time before or during the conference.
Send questions to **nobelconference@gustavus.edu**.

12:10 p.m. MUSICAL PRELUDE

The Gustavus Wind Symphony
Heidi Johanna Miller, DMA, conductor

Second Suite in F for Military Band, Movement I: March — Gustav Holst (1874–1934)

Imbizo — Julie Giroux (b. 1961)

12:30 p.m. SECOND LECTURE

Francesca Dominici, PhD

Professor of Biostatistics, population and data science at Harvard University

How Much Evidence Do You Need? Data Science to Inform Environmental Policy During the COVID-19 Pandemic

Does this drug work? Traditionally, the answer to the question "is this medical treatment effective?" has been answered using a large clinical trial that attempts to control for all variables. As treatments become more complex and customized to patients (think "new biologically derived cancer drugs"), clinical trials become unfeasible: impossible to design and too expensive to carry out. The advent of big data presents an alternative: gather existing data on patients with a variety of conditions and treatments, and use those data to compare outcomes. This alternative presents its own difficulties: how to draw meaningful conclusions from notoriously messy and incomplete data.

Francesca Dominici has developed new statistical methods for analyzing large amounts of medical data and drawing meaningful conclusions about treatment efficacy directly from available data. These methods have been applied specifically to cancer research, in collaboration with the Dana Farber Cancer Institute and the Harvard Medical School. Dominici has also conducted research on the impact of climate change and air pollution on public health. The results of these studies have informed air quality policy in the United States, leading to stricter air quality standards.

Dominici is director of the Harvard Data Science Initiative, Harvard University, and the Clarence James Gamble Professor of Biostatistics, Population and Data Science at the Harvard T.H. Chan School of Public Health, where she leads interdisciplinary groups of scientists to address important questions in environmental health science, climate change, and health policy. She is an elected member of the National Academy of Medicine and of the International Society of Mathematical Statistics. She has been honored by organizations ranging from the American Public Health Association, to the Committee of Presidents of Statistical Societies, to the University of Alabama. An advocate for the career advancement of women faculty, her work on the Committee on the Status of Women while on the faculty of Johns Hopkins University earned her the campus Diversity Recognition Award. She has a PhD in statistics from the University of Padua, Italy.

HER TALK

On December 7, 2020, the *New York Times* reported that then-President Trump declined to tighten soot rules. This was despite strong evidence of the adverse health effects including a link to COVID-19 deaths. Francesca Dominici will provide an overview of data science methods—including methods for causal inference and machine learning—to inform environmental policy. This is based on Dominici's work analyzing a data platform of unprecedented size and representativeness. The platform includes more than 500 million observations on the health experience of more than 95% of the US population older than 65 years old linked to air pollution exposure and several confounders. Finally, the talk will provide an overview of Dominici's studies on air pollution exposure, environmental racism, wildfires, and how they also can exacerbate the vulnerability to COVID-19.

Introduction — **Melissa Lynn, PhD**

Assistant Professor of Mathematics, Computer Science and Statistics

Post-lecture Reflections — **Lisa Heldke** and **Melissa Lynn**

1:30 p.m. THIRD LECTURE

Michael Osterholm, PhD

Regents Professor, McKnight Presidential Endowed Chair in Public Health and Director, Center for Infectious Disease Research and Policy, University of Minnesota

From the Village Watchman to Actionable Data: A Challenging Journey

How are we working to prevent future pandemics? The COVID-19 pandemic forces this question to the forefront. How do the tools of big data and machine learning help us to do so? And can these tools be deployed to mitigate the disastrous effects of large-scale health crises? What is their appropriate role? Where are they overpromising?

In 2006, Michael Osterholm participated in Nobel Conference 42, "Medicine: Prescription for the Future." His talk, entitled "A Modern World and Infectious Diseases: A Collision Course," summarized a just-released WHO report outlining the likelihood of another influenza pandemic. Osterholm listed the worldwide consequences, should such a pandemic materialize: extreme shortages of medical supplies, global economic and social disruption, no available vaccine or treatment. The disease would come in waves and would result in millions of deaths. The US, he noted, was woefully unprepared for what was to come. Listening to Osterholm's talk today, one can be mistaken for taking it as a report of what actually did unfold. (How) can big data reduce the chances of another COVID-19 pandemic?

Mike Osterholm has been studying infectious diseases and their impacts on communities for his entire career. A Regents Professor and McKnight Presidential Endowed Chair in Public Health, and Director, Center for Infectious Disease Research and Policy at the University of Minnesota, he also serves as faculty member in the school of public health, the college of science and engineering, and the medical school. He is a member of the National Academy of Medicine and the Council on Foreign Relations. Most recently, he served on transitional COVID-19 Advisory Board under President Joe Biden. He has advised organizations at the state, national and international level, including the World Health Organization, the NIH, the Food and Drug Administration, the Department of Defense, and the CDC. He holds a PhD in environmental health and an MPH in epidemiology, both from the University of Minnesota.

HIS TALK

Data science can offer powerful tools for tracking the spread of an infectious disease and the effectiveness of vaccines, medical treatment or mitigation efforts. However, the quality of data available in a public health crisis like the COVID-19 pandemic is highly variable. The data obtained from poorly designed studies, or data from comprehensive disease surveillance or from well-designed studies, when disregarded by a public skeptical of science can greatly limit the public impact that data science offers. How can data be most effectively used to both reduce the health and economic impact of the current pandemic and to reduce the risk of future pandemics? Michael Osterholm will detail the challenges of implementing an effective use of data science in such public health crises both for this pandemic and for those of the future.

Introduction — **Karl Larson**
Professor and Program Director, Public Health

Post-lecture Reflections — **Lisa Heldke** and **Karl Larson**

2:30 p.m. PANEL DISCUSSION AND AUDIENCE QUESTIONS

Submit your questions electronically at any time before or during the conference.
Send questions to nobelconference@gustavus.edu.

3:30 p.m. WORKSHOPS

All sessions will be livestreamed. For our online audience, go to the conference livestream page at **gustavus.edu/nobelconference** to find a Zoom link for each workshop.

1. Big Data – A New Frontier for Corporate Analytics

Large companies have always gathered information to help them design products, understand markets and target their advertisements to those consumers who are most likely to make a purchase. The ease and simplicity with which big data can be acquired, however, as well as the sophisticated and almost real-time understanding of decision-making that it can provide have transformed corporate analytics. This panel gathers an array of businesses representing everything from finance to health, to answer your questions about how corporations gather and organize data, how they use data to create and market products, and how they balance the privacy interests of their customers with the competitive advantages that data mining can deliver.

Virtual Panel Discussion: Wally LoFaro, Mastercard; John Owens '83, Digital Finance Advisory Services; Joel Stremmel '16, UnitedHealth Group; Mike Cafarello, MIT/Creator of Hadoop
Online Audience: go to the gustavus.edu/nobelconference to find a Zoom link.
Students and Employees: watch the livestream in Alumni Hall.

2. What Are They Doing With My Data? The Ethical Implications of Data Privacy

Think you're unique? What if you learned that you could be identified from just three pieces of information about you? What sorts of data are being collected about you, and what kinds of assumptions are being made about you on the basis of those data—assumptions about your health status, for instance? In this workshop, participants will learn how and where data are collected, and will examine the ethical questions that arise from the collection of particular kinds of data. You'll also learn how to do a privacy check on your own data.

Presented by: Gustavus Coding Club
Online Audience: go to the gustavus.edu/nobelconference to find a Zoom link.
Students and Employees: watch the livestream in Wallenberg Auditorium, Nobel Hall of Science.

3. How Big is Big Data?

How does big data get to be so big? And how do algorithms work to sort it? This workshop will give participants an idea of how very large data sets are accumulated, by taking a look at how much data an individual can generate in a single day. Presenters will introduce some visual sorting algorithms, and show how they enable data scientists to work with extremely large amounts of data. This is a great workshop for people who'd like to explore the basic math and computer science that underlies big data.

Presented by: Gustavus Math and Computer Science Club
Online Audience: go to the gustavus.edu/nobelconference to find a Zoom link.
Students and Employees: watch the livestream in Beck Hall Room 101.

4. The Creative Side of Big Data

Can machines be creative? Although artificial intelligence is commonly viewed as cold and unfeeling, machines can appreciate the finer things in life as well. This workshop will demonstrate how Big Data opens a new world of creative opportunities using machine learning. Participants will learn how AI can freestyle a song, guess a drawing, and play a piano duet with a human in real time. During the conversation, participants can reflect on the ethical and social impacts of these possibilities.

Presented by: Ben Menke, Gustavus Class of 2022
Online Audience: go to the gustavus.edu/nobelconference to find a Zoom link.
Students and Employees: watch the livestream in Olin Hall Room 103.

5. My Data... NOT YOURS!

Join Gustavus Technology Services as we highlight practical strategies for keeping your devices and the data on them private and secure. You will learn about how to create and manage secure passwords, how to add security to your accounts with Multi-Factor Authentication, and securing your mobile devices with Find My Device and Emergency Contact options. We will also provide take-home materials that will walk you through implementing these tools at home.

Presented by: Gustavus Technology Services Staff
Online Audience: go to the gustavus.edu/nobelconference to find a Zoom link.
Students and Employees: watch the livestream in Nobel Hall Room 1413.

6. Careers in Data Science and Related Industries

Join this interactive session to learn more about careers in data science. Participants will be able to talk with members of the Center for Career Development, as well as professionals in the data science field. Topics will include employer expectations, educational requirements, skill sets, salaries, and internship possibilities.

Presented by: Gustavus Center for Career Development Specialists and Data Science Professionals
Online Audience: go to the gustavus.edu/nobelconference to find a Zoom link.
Students and Employees: watch the livestream in Nobel Hall Room 1540.

7. Lying with Data: The Past, Present, and Future of Using Race and Ethnicity Data in Statistics

Current data collected through the CDC reports that compared to white, non-Hispanic persons in the United States, American Indian and Native Alaskan persons are 2.4 times more likely to die from COVID-19, Black or African American persons are twice as likely to die from COVID-19, and Hispanic or Latinx persons are 2.3 times more likely to die from COVID-19. In the United States, it is common to report statistics by race and ethnicity and to use these demographics to make decisions such as where to police, the length of prison sentences, and loan application approvals. However, have you ever considered why this is done or the effects this may have on different racial and ethnic groups? In this interactive workshop, we will delve further into the uses and misuses of race/ethnicity in data analysis. First, we will examine the reasoning and history behind the common practice of including race/ethnicity in statistical analyses. We will then explore how race/ethnicity have been and are used in data collection, data analysis, and the interpretation of the results from such analyses. Finally, we will discuss potential best practices for the use of race/ethnicity in statistical analyses.

Presented by: Gustavus Statistics Club
Online Audience: go to the gustavus.edu/nobelconference to find a Zoom link.
Students and Employees: watch the livestream in Nobel Hall Room 1412.

6 p.m. ART AT NOBEL CONFERENCE

Gallery Talk: Schaefer Art Gallery

Works by Arlene Birt, "Background Stories"

Meet artist Arlene Birt and learn more about her creative process and work. The exhibit, **Putting Data Into Context**, is in creative conjunction with Nobel Conference 57, Big Data *RE*volution. Sponsored by the Department of Art & Art History.

Arlene Birt, is an infodesigner, visual storyteller, public artist, and educator. She incorporates behavioral psychology to visually explain the stories behind products and places and to help individuals connect emotionally to seemingly distant environmental topics. She's the founder of Background Stories, a company that creates "clear visuals of complex stories." Her exhibition for this conference includes installation and participant-based artworks that use data as a means of visual creativity.

Visit the Schaefer Art Gallery website for exhibit images and the Zoom link to the gallery talk at **gustavus.edu/art/schaefer**.

7 p.m. MUSIC AT NOBEL CONFERENCE

OboeBass! presents **American Vein: New Music for Oboe and Bass**

featuring Carrie Vecchione, oboe and Rolf Erdahl, double bass

What do the arts and big data have in common? How do they intersect? Is the creative act uniquely human? This concert features works composed for OboeBass! since 2019 by composers who provide points of departure for these questions.

Carrie Vecchione, oboe, and Rolf Erdahl, bass, are OboeBass! Called "pioneers" by Minnesota Public Radio, they concertize widely around the U.S., and toured Norway in 2017. They have commissioned more than 40 new oboe/bass pieces and recorded six CDs, effectively creating the oboe/bass duo genre.

The concert will be livestreamed at **gustavus.edu/nobelconference**.

Wednesday

OCTOBER 6, 2021

9 a.m. PRELUDE

The Gustavus Symphony Orchestra
Ruth Lin, DMA, conductor

Overture from Nabucco	Giuseppe Verdi (1813–1901)
Intermezzo from Cavalleria Rusticana	Pietro Mascagni (1863–1945)

9:15 a.m. OVERVIEW OF THE DAY

Lisa Heldke
with Nobel Conference co-chairs
Tom LoFaro and **Karl Larson**

9:30 a.m. FIRST LECTURE

Cynthia Rudin, PhD
Professor of Computer Science, Electrical and Computer Engineering, and Statistical Science; Director, Prediction Analysis Lab, Duke University

Interpretable Machine Learning

Machine learning influences every one of us almost every day. Machine learning models lie behind your internet navigation, predicting what you might want to see or purchase. They underpin loan decisions. They're embedded in speech recognition software. They enable mail sorters and ATM machines to recognize handwritten digits, and they make possible the vision systems of self-driving cars. While machine learning predictions have largely been used in low-stakes decisions such as advertising or labeling images on social media, they are starting to be used in contexts in which incorrect decisions could be disastrous: health care decisions involving critically ill patients, and criminal justice.

Machine learning models are often "black boxes." That term means either that a model's computations are so complex that humans could not possibly understand them, or that they are proprietary—the exclusive property of a company or other entity, which they keep secret. While black box models may not be particularly problematic when used in low-stakes decisions such as advertising, their use in high-stakes decisions is much more fraught.

Cynthia Rudin studies and advocates for replacing black box models with interpretable models—machine learning programs that humans can understand, but that can be as efficacious as their black box counterparts. Among the collaborative projects on which she works, she led the first major effort to maintain a power distribution network with machine learning (in New York City), and she developed

continued

algorithms for crime series detection, which allow police detectives to find patterns of housebreaks. She also works with neurologists to predict subclinical seizures in patients within the intensive care units of hospitals.

Rudin is a professor of computer science, electrical and computer engineering, and statistical science at Duke. She is also the principle investigator of the Prediction Analysis Lab, the focus of which is interpretable machine learning. Rudin is a fellow of both the American Statistical Association and the Institute of Mathematical Statistics. She is a three-time winner of the INFORMS Innovative Applications in Analytics Award, and was named as one of the "Top 40 Under 40" by *Poets and Quants* in 2015, and was named by Businessinsider.com as one of the 12 most impressive professors at MIT in 2015. She holds a PhD in applied and computational mathematics from Princeton.

HER TALK

With machine learning has come serious societal consequences from using black box models for high-stakes decisions: flawed bail and parole decisions, racially-biased models in health care, and inexplicable loan decisions in finance. Transparency and interpretability of machine learning models is critical in high stakes decisions. However, there are clear reasons why organizations might use black box models instead: it is easier to profit from inexplicable predictive models than transparent models, and it is easier to construct complicated models than interpretable models. Most importantly, there is a widely-held belief that more accurate models must be more complicated, and more complicated models cannot possibly be understood by humans. Both parts of this last argument lack scientific evidence and often are not true in practice. In many cases, carefully constructed interpretable models are just as accurate as their black box counterparts on the same dataset. Cynthia Rudin will discuss the interesting phenomenon that interpretable machine learning models are often as accurate as their black box counterparts, using examples encountered throughout her career: manhole fires in New York City, caring for critically ill patients, and predicting criminal recidivism.

Introduction — **Jessie Petricka, PhD**
Associate Professor of Physics

Post-lecture Reflections — **Lisa Heldke** and **Jessie Petricka**

10:15 a.m. SECOND LECTURE

Pilar Ossorio, JD, PhD
Professor of Law and Bioethics, University of Wisconsin

Justice in Machine Learning/AI for Health Care

What expectations (if any) do people have when they agree to submit samples of their own data? To what degree should data sets be stripped of individually identifying characteristics? What responsibilities do researchers have to "do good" for individuals as well as "doing good" for society by producing high-quality research? How does an algorithm trained on a dataset that under-represented subjects (e.g. based on gender, socio-economic status or race) produce inappropriate diagnostic or treatment guidelines in a healthcare setting?

Pilar Ossorio is a lawyer, scientist, and bioethicist who works at the intersection of data/informatics, bioscience, regulation, and public policy. Her research considers such questions as: How can big data improve health outcomes, and how can our data management practices either exacerbate or improve existing inequalities? How can we refine genetic testing and analytic procedures to improve clinical

care? How can big data be used to reveal and remedy underlying injustices (such as racial disparities in diabetes treatment)? And what sorts of policy frameworks can regulate big data software analytical tools, in the ways the FDA regulates medical devices?

Ossorio is on the faculty of the law school, and also a member of the department of medical history and bioethics in the medical school at the University of Wisconsin, Madison. She also holds the inaugural Ethics Scholar-in-Residence at the Morgridge Institute for Research, a private, non-profit biomedical research institute using research in fundamental biology to advance human health. Throughout her career, Dr. Ossorio has participated in numerous advisory committees and boards that aid governments in setting science policy, including the National Institutes of Health, the FDA, Genome Canada, and Health Canada. She has also served on or chaired numerous committees and working groups that advise large-scale genome research initiatives, such as the 1000 Genomes Project and the Human Microbiome project. She is an elected fellow of the American Association for the Advancement of Science. A multidisciplinary researcher, she holds both a PhD in molecular biology from Stanford and a JD from the University of California at Berkeley.

HER TALK

Health care organizations and health care providers are using advanced algorithmic approaches (machine learning or artificial intelligence—ML/AI) for a variety of purposes. For instance, organizations use ML/AI to assess the quality of health care services, make systems more efficient, and determine which patients need extra follow up and care coordination. Health care providers use ML/AI to identify recommended treatments, predict patient outcomes, and help with diagnoses. However, growing literature on ML/AI indicates that algorithms or combinations of them can reproduce race, gender, class and other social biases, and a small literature has now shown how ML/AI used in health care also incorporates a variety of pernicious and unfair biases. In this talk, Ossario will consider how such biases become encoded in ML/AI for health care and discuss some means for decreasing bias. She will also discuss how we might use the technology to identify existing unfair biases within health care systems with the aim of ameliorating them.

Introduction — **Phillip Voight, PhD**
Associate Professor of Communication Studies

Post-lecture Reflections — **Lisa Heldke** and **Phillip Voight**

11:25 a.m. PANEL DISCUSSION AND AUDIENCE QUESTIONS

Submit your questions electronically at any time before or during the conference.
Send questions to nobelconference@gustavus.edu.

1:15 p.m. MUSICAL PRELUDE

Gustavus Jazz Ensemble
Dave Stamps, DMA, director

Sweet Georgia Brown — Music by Ben Bernie and Maceo Pinkard
Lyrics by Kenneth Casey
Arranged by Marty Paich

Samba Kinda Mambo — Michael Phillip Mossman (b. 1959)

1:30 p.m.

THIRD LECTURE

Rhema Vaithianathan, PhD
Professor of Health Economics; Director, Centre for Social Data Analytics,
Auckland University of Technology

Child Protection: Too Much and Not Enough

How do you use big data in ways that benefit the most vulnerable members of our society? For instance, what role can data science play in protecting and improving child welfare? How can we build predictive models that don't exacerbate social inequities, and might even reduce them? Rhema Vaithianathan is using big data to work collaboratively with governmental and nongovernmental agencies to solve some of the most significant challenges of our time while working to ensure that these tools don't exacerbate or perpetuate social inequity by targeting vulnerable populations.

Vaithianathan's research uses risk modeling to address problems in public health and child welfare. She leads a research team that has developed child protection predictive risk modeling tools that are being used in counties in Pennsylvania, Colorado and California. These tools use current and historical information to make predictions about future events.

Vaithianathan is a professor of social data analytics at the Institute for Social Science Research at the University of Queensland. She is also a professor of health economics at Auckland University of Technology, where she is co-founder and director of the Centre for Social Data Analytics. The CSDA—which also has a node at the University of Queensland—works collaboratively with domestic and international government and nongovernmental agencies to use data science to address social problems, particularly for vulnerable populations. She holds a PhD in economics from the University of Auckland.

HER TALK

By the age of 18, one in three American children will have been investigated for suspected child abuse or neglect by a child welfare agency. Yet despite this surprising level of surveillance, the rates of serious injuries and deaths from child maltreatment remain high; on average, more than five American children die every day from abuse or neglect. After every tragic maltreatment death, agencies tend to increase their rates of investigations, casting the net even wider and bringing more families and children into the system. This means a child protection system that was meant to only be used to investigate rare cases of abuse or neglect is overwhelmed with large caseloads of children, many of whom are at minimal levels of risk. Research tells us that predictive risk modeling tools are good at estimating the risk of future harm for children who are referred to agencies. But these are high stakes decisions, and a human centered approach is essential. As Vaithianathan and her team learned when they developed and implemented the Allegheny Family Screening Tool— a world-first use of predictive analytics to help triage referrals about abuse and neglect—that means using data in an ethical, transparent, and trusted way to achieve the social licence needed to proceed.

Introduction — **Kate Knutson, PhD**
Professor of Political Science

Post-lecture Reflections — **Lisa Heldke** and **Kate Knutson**

2:30 p.m. FOURTH LECTURE

Wendy Chun, PhD
Canada 150 Research Chair; Leader, the Digital Democracies Institute,
Simon Fraser University

Discriminating Data

How do we reduce the spread of misinformation and disinformation? What makes any piece of info true? Fact checking alone is inadequate to such a task. Wendy Chun is working with a deeply multidisciplinary team of researchers that includes everyone from data analysts to dramaturgs, to develop a model that can answer that question, developing new ways to displace fake news. The project, called "Beyond Verification," is part of the Digital Democracies Institute.

The Institute integrates research in the humanities and data sciences to address questions of equality and social justice. It seeks to combat the current proliferation of online "echo chambers" and discriminatory algorithms by creating alternative data literacies and paradigms for connection.

Chun is the author of several books, including the forthcoming *Discriminating Data: Correlation, Neighborhoods and the New Politics of Recognition*, which investigates the centrality of race, gender, class and sexuality to machine learning and network analytics. To desegregate networks, open the "echo chambers" and buttress social justice, Chun argues, we must develop alternative algorithms and interdisciplinary coalitions.

Chun is the Canada 150 Research Chair in New Media, School of Communication, Simon Fraser University, where she leads the Digital Democracies Institute. The recipient of Guggenheim, ACLS and American Academy of Berlin fellowships, she is a multidisciplinary researcher, with an undergraduate degree in systems design engineering from the University of Waterloo, and a PhD in 0English literature from Princeton.

HER TALK

What if the polarization of our online communities is not an unfortunate accident of machine learning and predictive data analysis, but actually its goal? What if the assumption that "birds of a feather flock together" (an assumption underwriting the system that will recommend your next FB friend) actually just creates angry birds? Wendy Chun argues that the tools and methods used by predictive data analysis begin from a set of assumptions that effectively encode segregation and eugenics in their results. Consider, for example, correlation, which grounds the potential of big data to make predictions about the future; it has its roots in twentieth-century eugenic efforts to "breed" a better future. And homophily—the "birds of a feather" assumption—underpins "recommender systems," which group people into increasingly segregated, often increasingly-angry clusters.

Introduction — **Colleen Stockmann, PhD**
Assistant Professor of Art and Art History

Post-lecture Reflections — **Lisa Heldke** and **Colleen Stockmann**

3:30 p.m. PANEL DISCUSSION AND AUDIENCE QUESTIONS

Submit your questions electronically at any time before or during the conference.
Send questions to nobelconference@gustavus.edu.

4:30 p.m. CLOSING REMARKS

Tom LoFaro, **Karl Larson**, and **Lisa Heldke**

A Small List of **Essential Terms**

Algorithm

A set of instructions to be followed. Algorithms are created by people but are usually followed by a computer through a computer program.

Artificial intelligence (AI)

Any computer system designed to mimic human intelligence. Familiar examples of AI include map reading, conversations with humans (e.g. through "chatbots"), face recognition, and "search and recommend" algorithms ("If you liked that book, we think you'll like this one").

Bias

The tendency (of an algorithm or a person) to make predictions based on inappropriate, irrelevant, or non-causal information. Bias can be introduced into machine learning algorithms by, for example, using data sets that lack diversity.

Black Box

Machine-learning algorithms in which the underlying methodology is hidden from the user, either by the complexity of the computations or because the model is proprietary.

Correlation versus Causation

In a data set, two variables are correlated if they vary in relation to each other. They are in a causal relationship if the variation in one brings about the variation in another. The sale of ice cream and sunglasses on a hot day are correlated, but of course the sale of ice cream doesn't cause sunglasses to be sold. Proof of causation almost always requires a deep scientific exploration of the system in question. Absent that exploration, the move from correlation to causation is unfounded. For example, statistical models have been used to show that there is a correlation between obesity and diabetes, but the question of causality has not yet been answered.

Data scientist

A researcher who works with large quantities of data to analyze and draw inferences.

Interpretability

The ability to determine and understand how a machine learning algorithm makes a prediction after it has been trained. Whether or not an algorithm counts as interpretable is often a very difficult question. It is an important one to answer, because understanding how an algorithm predicts is one step to detecting and avoiding bias in that model.

Machine Learning (ML)

A class of algorithms that is designed to "learn" from a set of training data and then to apply those deduced rules to similar data, using the learned algorithms and statistical models rather than explicit instructions. Machine learning algorithms are generally good at pattern recognition and hence are often used in image recognition and other areas to identify objects or other patterns. After learning from a training dataset to recognize, e.g., dogs, the algorithm can go on to identify dogs in new images.

Network science/network theory

Methods for understanding interconnected data that proceed by developing a graph (that is, a kind of "map") that displays the different data points and the connections among them. Imagine a "map" with dots that represent all the websites related to a given search term (e.g. "dog breeds") and with the lines between dots to represent links from one website to another. Search engine algorithms analyze the structure of this graph to determine the most relevant websites and rank the results accordingly. Network analysis methods are also used to understand social interactions in which individuals are the dots and the relationships among them are the connecting lines.

Statistical Modeling

Using mathematical methods to describe trends and/or patterns in data. One kind of statistical model is linear regression, which is used to determine whether a linear relationship exists between different items in a data set. ("In rescue dogs, is there a relationship between the height of dogs and length of time it takes for them to be adopted?") Remember that a correlation is not necessarily a causal relationship!

MORE FINE ARTS

During the Conference, viewers will have the opportunity to watch these two videos.

DANCE AT THE NOBEL CONFERENCE

re(visions)

What if we think about the notes in the musical scale as being little pieces of data?

The music we make with those tones can then be thought of as the aggregation of this data created by the way in which we search the data set.

Dancing can also be thought of as an exploration of a data set: this time, the human body. The producible movement varies from body to body. Imagine that producible movement as a data set and all of the factors that make it—independent or codependent body parts, length of limbs, range of motion of joints—as the pieces of data leading to the emergent algorithmic outcome.

This dance is designed to help viewers picture data: what a piece of data may be, outliers, data aggregation, the building and running of algorithms, and the implications of using big data. This piece is a call to explore the way in which one's life is filled with data sets—both formally classified and existing organically yet to be perceived by you.

Choreographer: Astrid Axtman '22
Music: *Surges* by Carlos Chipa
Pianist: Calvin Gehring '23
Dancers: Claire Drapeau '22, Jessica Herbrand '24, Mary Juelich '24, Elisabeth Kray '23, Katera Lampert '24, Cascade Oppitz '23, and Hannah Saunders '22

HILLSTROM MUSEUM OF ART — VIRTUAL TOUR

Art for All: The Swedish Experience in Mid-America features 60 works by 20 prominent Swedish-American artists working in the late 19th and early 20th century and explores the democratic arts culture established by these immigrant artists in the U.S. The exhibit was organized with the Birger Sandzén Memorial Gallery in Lindsborg, Kansas with support from the Swedish Council of America. Artists featured in the exhibit include Birger Sandzén (1871–1954), acclaimed for his rugged landscapes of Kansas and the American West painted in his signature impasto style; B.J.O. Nordfeldt (1878–1955), a well-known Modernist recognized for his abstraction in paintings and prints; and John F. Carlson (1875–1947), leader of an influential landscape painting school in the art colony of Woodstock, New York. An illustrated catalogue of the exhibit is available at the museum and in pdf form on the museum website at **gustavus.edu/hillstrom.**

Detail: Birger Sandzén (1871–1954), ***Sunset in the Mountains***, c.1923, oil on canvas, 18 x 24 inches, Birger Sandzén Memorial Gallery, gift of Thelma Tichenor

Westminster Town Hall Forum

WESTMINSTER TOWN HALL FORUM

partners with the Nobel Conference

Discussion of Wendy Chun's

Democracy's Digital Dilemma

Tuesday, October 5 | 6-7 p.m.

Join the discussion by going to the Zoom link on the conference livestream page at **gustavus.edu/nobelconference**.

Wendy Chun, the final speaker at Nobel Conference, was also the featured presenter at the October Town Hall Forum. Her Forum lecture, ***Democracy's Digital Dilemma***, is available at both **westminsterforum.org** (in video form) and **mprnews.org** (in audio).

Since 1980, Westminster Presbyterian Church in downtown Minneapolis has hosted the Westminster Town Hall Forum to engage the public in reflection and dialogue on the key issues of our day from an ethical perspective. The Town Hall Forum is Minnesota's longest-running and most prestigious speaker series.

On Tuesday night, members of both audiences are invited to participate in a facilitated small group discussion, hosted by the Forum's director Tane Danger '07.

NOBEL CONFERENCE ? science WHYS

SCIENCEWHYS PODCAST

ScienceWhys is a podcast produced in conjunction with the Nobel Conference. When big scientific questions meet big ethical questions, the waters can get pretty choppy. Lisa Heldke, director of the conference, interviews scientists, researchers, scholars, and thinkers about how science and ethics mingle, eddy, roil, and churn in their own work. The podcast is for anyone who hears about a scientific breakthrough and thinks "what are the downstream consequences of that?" Listen to the current season where Lisa interviews most of the 2021 conference speakers to gather insight into their life and work.

Find the podcast at **Spotify** or at **anchor.fm/sciencewhys**.

GUSTAVUS STUDENTS!

GET EXCLUSIVE ACCESS TO THE NOBEL SPEAKERS

Tuesday, October 5 | 4:30 p.m.

Join a Zoom discussion to ask:

- How did they become interested in their field?
- What was the path they took to be the researchers they have become?
- What are the questions they still haven't been able to answer in their research?
- What are the particular challenges they have faced in achieving their success?

Choose one of the following pairs of speakers:

- Wendy Chun and Pilar Ossorio
- Cynthia Rudin and Talithia Williams
- Francesca Dominici and Rhema Vaithianathan

NOBEL 57 TSHIRTS

AND RELATED BOOKS AVAILABLE AT THE BOOK MARK.

bookmark.gustavus.edu

Select books will be 20% off Sept. 24–Oct. 30.

$5 media mail shipping option for Nobel Conference books only: NOBELBOOKS2021

WELCOME
HIGH SCHOOL STUDENTS

We're glad you've joined us virtually to participate in this conversation about science and its social and ethical implications. Big data already plays a significant role in your lives; that role will surely expand in the course of your lives. We hope that these lectures, workshops and online resources lead you to understand that role a little better. And...we invite you to imagine yourselves as the researchers who will bring about that future, researchers who recognize the ethical challenges and promises of this powerful set of tools.

Would you like to begin your research career as a Gustavus student? Learn how at **gustavus.edu/admission.**

NOBEL CONFERENCE SPONSORS

Adeline and Drell Bernhardson

The Nobel Conference is the only education conference in the United States to be authorized by the Nobel Foundation in Stockholm, Sweden. Core endowment funding for the conference was permanently secured through the generosity of the late Reverend Drell and Adeline Bernhardson. The Bernhardsons recognized and celebrated the historic relationship between the Nobel Foundation and Gustavus Adolphus College. By establishing this endowment, they secured the core funding for the success of the conference and established a platform on which the conference can expand its reach and impact. The Nobel Conference Endowment Fund also includes gifts from Dawn and Ted Michael, Russell and Rhoda Lund; Drs. Robert E. and Susan T. Rydell; William Harvey; Steve and Peg Sutherland; the Mardag Foundation, in memory of Edgar B. Ober; and the UnitedHealth Group.

Each year's conference is supported by annual contributors. **Contributors to the 2021 conference** include:

Heroic PRODUCTIONS
Audio • Lighting • Video

DOZENS OF NOBEL CONFERENCE ATTENDEES
donated their ticket purchases to support the conference.
Thank you for your generosity!

If you wish to support this vital gathering of scholars, students, educators, and members of the public, please consider a gift to Gustavus Adolphus College, designated to the Nobel Conference. Your gift can underwrite a speaker, arts, workshops, or other activities surrounding the conference.

THE RYDELL PROFESSORSHIP

The Rydell Visiting Professorship at Gustavus Adolphus College is a scholar in-residence program designed to bring Nobel laureates and other distinguished researchers and scholars to the campus as catalysts to enhance learning and teaching. The Rydell Professor enriches the intellectual life of our community by giving public lectures, team teaching, and meeting more informally with students and faculty in discussion groups and over meals.

The Rydell Professorship was established in 1993 by Drs. Robert E. and Susan T. Rydell to give students the opportunity to learn from and interact with leading scholars. To see who has come to Gustavus through the Rydell Professorship, visit **gustavus.edu/events/rydell/history.php.**

Mental Health (In)Equity and Young People

Mental Health (In)Equity and Young People

September 28 & 29 | Gustavus Adolphus College

Note: The conference will be on Wednesday and Thursday in 2022

Prioritizing the mental health concerns of young people has become essential amid times of global pandemic, racism, sexism, ableism, social unrest, climate change, and political upheaval. These social inequities limit our ability to promote resilience in the mental health of adolescents and young adults, especially those from marginalized communities. Young people often experience little control over their wellbeing, are affected by the decisions of parents, schools and society, and in these technology-driven times are vulnerable to the negative side effects of social media and information overload. In considering how to eradicate inequities and promote mental health, technology becomes central in how it both aids and hinders our modern existence, in the U.S. and around the world.

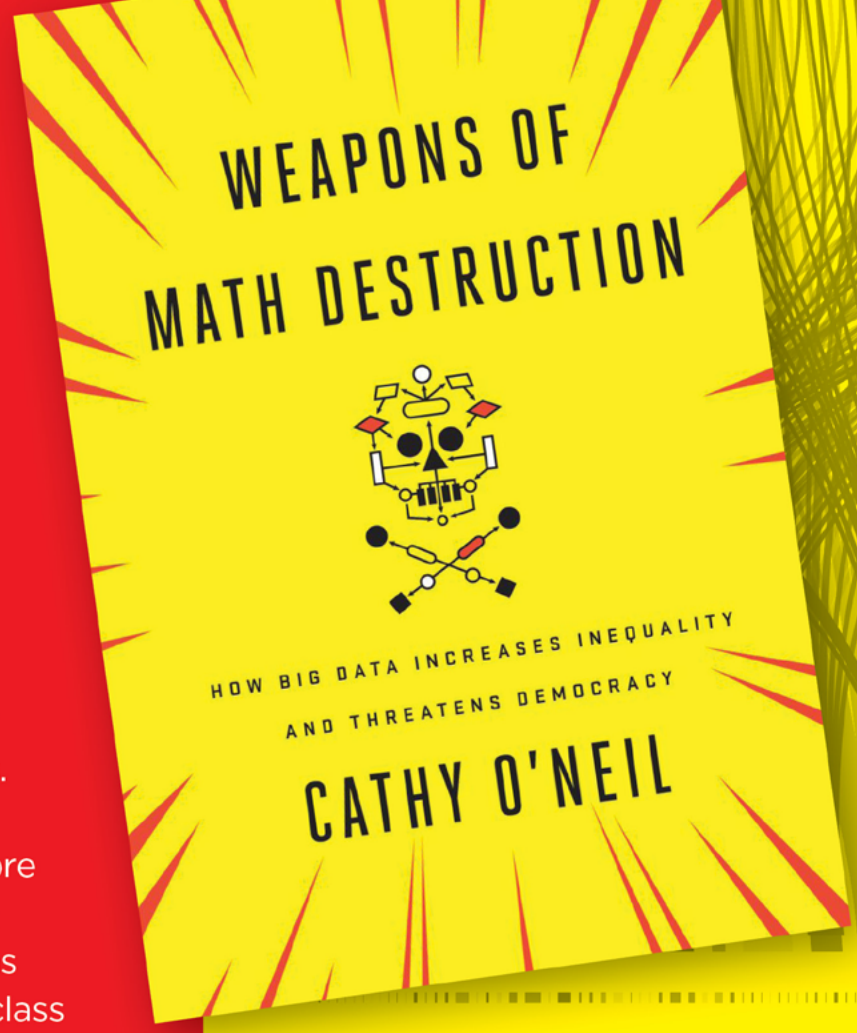
Nobel Conference 58 will address mental health disparities and their effects on youth, with a particular emphasis on the significance of identity, trauma, and technology.

Confirmed Speakers

MERYL ALPER, Assistant Professor of Communication Studies, Northeastern University **MANUELA BARRETO**, Professor of Social and Organizational Psychology, University of Exeter **DANIEL EISENBERG**, Professor, Department of Health Policy and Management, University of California, Los Angeles **JOSEPH GONE**, Professor of Anthropology and of Global Health and Social Medicine, Harvard University **PRISCILLA LUI**, Assistant Professor of Psychology, Southern Methodist University **NIC RIDER**, Assistant Professor, Program in Human Sexuality and Coordinator, Adult Transgender Health Services Program, Program in Human Sexuality **BRENDESHA TYNES**, Associate Professor of Education and Psychology, University of Southern California

The Reading in Common program introduces new students to academic life at Gustavus, facilitates discussion among students and faculty and creates a shared experience for incoming students. It also engages them with the topic of the Nobel Conference before they even arrived on campus. Students read the book during the summer before their first year, and then participate in discussions with faculty and upper class students during the fall orientation.

This year's Reading in Common selection is **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.** Author, mathematician, and data scientist **Cathy O'Neil** challenges us to face the negative consequences of algorithms by exploring the impact on "people who have been deemed unworthy in some way by an algorithm."



Reading in Common Book

"But human decision making, while often flawed, has one chief virtue. It can evolve. As human beings learn and adapt, we change, and so do our processes."

Cathy O'Neil, mathematician, data scientist, and author

Gustavus Academy for Faith, Science, and Ethics

GUSTAVUS ADOLPHUS COLLEGE

The Gustavus Academy for Faith, Science, and Ethics prepares leaders to build creative alliances between religion and science in order to address the world's most pressing challenges.

The Gustavus Academy summer program provides opportunities for high school students to explore their beliefs and to discover how scientists and people of faith are working together to address some of the world's most pressing problems. As an Academy Fellow, they will join with other outstanding high school students to grow in knowledge, develop leadership skills, and clarify their sense of purpose.

The 2022 Academy, **Our Beautiful Brain: Neuroscience, Mental Health, and Resiliency**, will take place June 18-24, 2022.

For more information, visit **gustavus.edu/chaplain/academy**.

Gustavus Academy for Faith, Science, and Ethics is made possible by a generous grant from the Lilly Endowment Inc.

ABOUT THE NOBEL 2021 ARTWORK

Big Data overwhelms us, invading most aspects of our lives and delivering a stunning amount of information. The poster for Nobel Conference 2021: **Big Data *RE*volution**, offers a map of the collecting, sorting, and analyzing process. Masses of data streams are filtered and coalesced into specific applications or solutions, represented by dots of various sizes. Interconnectedness analyses and utilization are shown by lines connecting the data points, a few of which are enlarged and labeled. Yet for all the information we gather and utilize, more exists. Unrecognized, undiscovered, and not-yet-understood data and relationships are represented by nearly invisible glossy white dots, pale gray dots, and dots that simply have not yet made a connection.

Sharon Stevenson, Stevenson Creative

GUSTAVUS

GUSTAVUS ADOLPHUS COLLEGE

800 West College Avenue | Saint Peter, MN 56082 | **gustavus.edu/nobelconference**