BIG DATA FOR PSYCHOLOGICAL SCIENCES

PURDUE UNIVERSITY

Department of Psychological Sciences
COLLEGE OF HEALTH AND HUMAN SCIENCES

6TH PURDUE SYMPOSIUM ON
PSYCHOLOGICAL SCIENCES (PSPS) 2018

The Department of Psychological Sciences at Purdue University is pleased to host the sixth installment in its symposium series. This biennial event showcases current and emerging topics within the psychological sciences that form the basis for an edited volume following each symposium.

The 6th Purdue Symposium on Psychological Sciences aims to critically assess the current state of the Big Data movement in terms of its scientific value to the psychology community; and identify areas in which future research is most needed.

The topic of Big Data is a young, rapidly growing and important area of study across many academic disciplines. We have invited leading scholars and practitioners not only in psychology but also in other disciplines such as computer/data science, philosophy of science, management, health care, and education. Each presenter will highlight her or his current work and views related to the topic of Big Data.

We seek to get a broad array of opinion on a wide variety of issues associated with the use of Big Data in psychology – namely:

1) How Big Data have been (and should be) conceptualized and discussed in psychology and other fields, addressing the "big picture," conceptual questions from the perspectives of philosophy of science and research methodology;

2) Tangible opportunities and strategies for utilizing Big Data to enhance existing methods of data collection and analysis within psychology (e.g., social media, internet behaviors, public network cameras, wearable sensors/cameras, medical records, data mining and visualization techniques);

3) How Big Data may be used in specific sub-disciplines of psychological sciences (e.g., learning, healthcare, social/personality, cognitive, clinical, developmental, industrial-organizational); and

4) Best-practice recommendations for addressing practical and ethical challenges associated with Big Data methods such as privacy, data security and storage, data sharing, and replicability/reproducibility issues.

Symposium Coordinators

Sang Eun Woo, PhD
Associate Professor
Industrial-Organizational Psychology

Robert Proctor, PhD
Distinguished Professor
Cognitive Psychology

Louis Tay, PhD
Assistant Professor
Industrial-Organizational Psychology

Department of Psychological Sciences
Purdue University
703 Third Street
West Lafayette, IN 47907-2081 USA

SCHEDULE

Thursday, May 17

5:00-6:30pm   Opening Reception // East Faculty Lounge, PMU 2nd Floor

Friday, May 18

7:30-8:30am   Continental Breakfast // STEW 202

8:30-8:40   Opening Remarks by the Symposium Organizers // STEW 202

Overview & Background // STEW 202

8:40-9:20   *Big Data: A Philosophy of Science Perspective*
Brian Haig, University of Canterbury

9:20-10:00   *Big Data: Challenges and Opportunities for Causal Inferences in the Experimental Tradition*
Robert Proctor, Purdue University

10:00-10:40   *Big Data for Enhancing Measurement Quality*
Louis Tay, Purdue University

10:40-10:55   Coffee Break // STEW 202

Innovations in Large-Scale Data Collection and Analysis Techniques // STEW 202

10:55-11:00   Brief Introduction by the Symposium Organizers

11:00-11:40   *Mining Social Network and Social Media Interactions to Understand and Predict User Behavior*
Jennifer Neville, Purdue University

11:40-12:20pm   *Internet Search and Page View Behavior Scores: Validity and Usefulness as Indicators of Psychological States*
Michael Ford, University of Alabama

12:20-1:40   Lunch // West Faculty Lounge, PMU 2nd floor

1:40-2:20   *Using Public Network Cameras to Detect and Track Human Behaviors*
Yung-Hsiang Lu, Purdue University

2:20-3:00   *Wearable Cameras, Machine Vision, and Big Data Analytics: Insights Into People and the Places They Go*
Ryne Sherman, Texas Tech University

3:00-3:20   Coffee Break // STEW 202

3:20-4:00   *Human-Computer Collaborative Decision Making Through Visual Analytics*
David Ebert, Purdue University

4:00-4:40   *Mining Large-Scale Biomedical Data: Opportunities & Challenges*
Padmini Srinivasan, The University of Iowa

4:40-5:20   General Discussions and Closing for the Day

Saturday, May 19

7:30-8:15am   Continental Breakfast // STEW 202

Applications of Big Data // STEW 202

8:15-8:20   Brief Introduction by the Symposium Organizers

8:20-9:00   *Big Data in the Science of Learning and Teaching*
                   Sidney D'Mello, University of Colorado, Boulder

9:00-9:40   *Role of Big Data in Healthcare*
                   Paul Griffin, Regenstrief Center for Healthcare Engineering,-- Purdue University

9:40-10:20   *Big Data in Social Psychology*
                   Ivan Hernandez, DePaul University

10:20-10:40   Coffee Break // STEW 202

10:40-11:20   *Big Data in Cognitive Science - The Continued Importance of Theory: Lessons from Big Data Approaches to Cognition*
                   Brendan Johns, University at Buffalo

11:20-12:00pm   *Big Data in Clinical Psychology and Mental Health*
                   Colin Walsh, Vanderbilt University

12:00-1:20   Lunch // West Faculty Lounge, PMU 2nd floor

1:20-2:00   *Big Data in Developmental Psychology*
                   Kevin Grimm, Arizona State University

Best-Practice Recommendations for Responsible and Rigorous Use of Big Data

2:00-2:40   *Big Data - Security and Privacy (and Transparency)*
                   Elisa Bertino

2:40-3:00   Coffee Break // STEW 202

3:00-3:40   *Recommendations for Maximizing Scientific Value of Big Data*
                   Sang Eun Woo, Purdue University

Conclusions // STEW 202

3:40-4:20   *Closing Keynote*
                   Fred Oswald, Rice University

4:20-5:00   General Discussions // STEW 202

5:00-6:00   Break // STEW 202

6:00-7:00   Closing Reception // East Faculty Lounge, PMU 2nd floor

BRIAN HAIG UNIVERSITY OF CANTERBURY

Brian Haig is a Professor in the Department of Psychology at the University of Canterbury, and a Visiting Professor in the Department of Education at the University of Bath. He is a theoretical psychologist who has published numerous articles in psychology, philosophy, and education journals on the conceptual foundations of quantitative and qualitative research methods, and the nature of psychological science more generally. He is the author of Investigating the Psychological World (MIT Press, 2014) and The Philosophy of Quantitative Methods (Oxford University Press, 2018), and co-author of Realist Inquiry in Social Science (Sage, 2016). He is a Fellow of the Association for Psychological Science and the New Zealand Psychological Society.

Big Data: A Philosophy of Science Perspective

What is Big Data science? Does it promise to revolutionise our current understanding of science? If so, does it mark the development of a new understanding of scientific methodology? These are some of the broad questions raised about science by the advent of the Big Data movement. They clearly contain important epistemological issues that deserve to be addressed. Accordingly, I undertake a philosophical examination of a number of claims made on behalf of Big Data science. In particular, I evaluate the provocative claims that the Big Data movement spells the end of scientific method and the end of theory, as we know them. In doing so, I examine the role of inductive, hypothetico-deductive, and abductive reasoning in an altered conception of Big Data method. I also ask whether it is viable to place a greater emphasis on correlational research at the expense of causal explanatory research. At the end, I draw some lessons learnt for the science of psychology.

ROBERT PROCTOR PURDUE UNIVERSITY

Robert W. Proctor is Distinguished Professor of Psychological Sciences at Purdue University, with courtesy appointment in Industrial Engineering. Dr. Proctor teaches courses in Human Factors in Engineering, Human Information Processing, Attention, and Perception and Action. Dr. Proctor's research focuses on basic and applied aspects of human performance. He has published over 250 articles on human performance and is co-author the books Human Factors in Simple and Complex Systems, Skill Acquisition and Human Performance, and Why Science Matters. He is Fellow of the American Psychological Association: Division 21, Applied Experimental and Engineering Psychology and the Human Factors and Ergonomics Society.

Big Data: Challenges and Opportunities for Causal Inferences in the Experimental Tradition

Experimental psychologists have historically focused on tightly controlled studies of the behavior of a small number of subjects in laboratory settings. The goal of such research is to establish causal relations between independent and dependent variables, and this approach has resulted in considerable progress in understanding human behavior. With recent technological developments, especially in the area of Artificial Intelligence, it has become easy to collect large amounts of various kinds of behavioral data outside of the laboratory. These big data sets provide unique opportunities to understand people, possibly revealing relations among psychological phenomena on a large scale that have not been evident in the small data sets collected in the laboratory. However, in order for big data to be used effectively, there is a challenge of deriving scientific methods to guide the research and relate the findings to those from small-data research. In this work, we explore whether the structure of big data sets will support causal inferences that complement those derived from the small data sets of the lab.

LOUIS TAY PURDUE UNIVERSITY

Louis Tay is an Assistant Professor of Industrial-Organizational Psychology at Purdue University. He holds a PhD in Industrial-Organizational Psychology from the University of Illinois, Urbana-Champaign. His substantive research interests include well-being (subjective well-being, psychological well-being), character strengths, and vocational interests. His methodological research interests include measurement, item response theory, latent class modeling, and multilevel analysis. His research appeared in journals such as Psychological Bulletin, Perspectives on Psychological Science, Journal of Personality and Social Psychology, Psychological Science, Journal of Applied Psychology, Personnel Psychology, and Organizational Research Methods. He is an Associate Editor at Organizational Research Methods.

Big Data for Enhancing Measurement Quality

It is often quoted that the "Big Data" era has arrived. While an important methodological advancement, the use Big Data seldom explicitly considers the fundamental issue of measurement quality. To address this, my co-authors and I identify sources of Big Data (i.e., social media, wearable sensors, Internet behavior, electronic personnel records, public network cameras, and smartphones) and provide an analysis of their measurement quality with regard to standards in psychological assessments. Specifically, we discuss how these sources provide unique opportunities and challenges for psychologists with regard to their validity and reliability. We further clarify how Big Data and other "traditional" data collection methods may be utilized in a synergetic manner, allowing triangulation, replication, and cumulatively building scientific knowledge over time.

JENNIFER NEVILLE PURDUE UNIVERSITY

Jennifer Neville is the Miller Family Chair Associate Professor of Computer Science and Statistics at Purdue University. She received her PhD from the University of Massachusetts Amherst in 2006. She is currently an elected member of the AAAI Executive Council and she was recently PC chair of the 9th ACM International Conference on Web Search and Data. In 2012, she was awarded an NSF Career Award, in 2008 she was chosen by IEEE as one of "AI's 10 to watch", and in 2007 was selected as a member of the DARPA Computer Science Study Group. Her work, which includes more than 100 peer-reviewed publications with over 5000 citations, focuses on developing data mining and machine learning techniques for complex relational and network domains, including social, information, and physical networks.

Mining social network and social media interactions to understand and predict user behavior

The popularity of social networks and social media has increased the amount of information available about users' behavior online--including current activities, and interactions among friends and family. This rich relational information can be used to improve predictions about individuals even when personal data is sparse, since the characteristics of friends are often correlated. Although these online data offer a unique opportunity to improve predictive models about users, they also present a number of challenges for sampling, inference, and hypothesis testing. This talk will describe some algorithmic and statistical issues that arise when modeling data from partially-observed, large-scale networks. It will also outline recent machine learning methods that use latent-variable modeling, semi-supervised learning, and computational statistical tests to address these challenges and accurately estimate predictive models from online data.

MICHAEL FORD UNIVERSITY OF ALABAMA

Michael Ford is currently an Associate Professor in the Department of Management at the University of Alabama. He received a Ph.D. in Industrial-Organizational Psychology from George Mason University in 2008. Michael's research focuses on health, stress, and well-being among workers, and his work has appeared in journals such as Psychological Bulletin, Journal of Applied Psychology, Journal of Management, and Journal of Occupational Health Psychology. He is also currently an associate editor at Journal of Vocational Behavior and Occupational Health Science.

Internet Search and Page View Behavior Scores: Validity and Usefulness as Indicators of Psychological States

Data on internet searches and page views in aggregate form, which reflect the interests of Internet users and the population at large, have potential to yield new and timely information on human thought and emotion. In this presentation, research will be presented on the validity and usefulness of Google Trends and Wikipedia page view data as indicators of subjective well-being across time and place. Results suggest that Internet searches that include affect-related words correlate with self-reported well-being and personality and predict psychological and physical health outcomes beyond self-reported measures at the state and metro area level. Wikipedia page views and Google searches for terms such as "stress" and "anxiety" also show meaningful weekly and yearly seasonal patterns. Implications and limitations of these data for research and society will be discussed, along with potential future directions.

YUNG-HSIANG LU PURDUE UNIVERSITY

Yung-Hsiang Lu is a professor in the School of Electrical and Computer Engineering and (by courtesy) the Department of Computer Science of Purdue University. He is an ACM distinguished scientist and ACM distinguished speaker. He is a member in the organizing committee of the IEEE Rebooting Computing Initiative. He is the lead organizer of Low-Power Image Recognition Challenge, the chair of the Multimedia Communication Systems Interest Group in IEEE Multimedia Communications Technical Committee. Dr. Lu and three Purdue students founded a technology company using video analytics to improve shopping experience in physical stores. This company receives two Small Business Innovation Research grants from the National Science Foundation (IIP-1622082 and IIP-1738492). Dr. Lu obtained the Ph.D. from the Department of Electrical Engineering at Stanford University.

Using Public Network Cameras to Detect and Track Human Behaviors

Video has been widely used observing human behavior. In most cases, one or only a few cameras are used. In recent years, many organizations (such as universities, shopping centers, restaurants) deploy network cameras and make the data available on the Internet. These network cameras provide real-time visual data and can be used for observing human behavior. These cameras provide many advantages. For example, the cameras are deployed worldwide and it is possible to study behavior at different locations and cultures. Also, the time and location information can be used to correlate additional information, such as weather and news events. This speech will present the opportunities from the worldwide network cameras, a research infrastructure being built at Purdue University, and preliminary studies using the data.

RYNE A. SHERMAN TEXAS TECH UNIVERSITY

Ryne Sherman received his PhD in personality/social psychology from the University of California, Riverside in 2011. He was an assistant professor at Florida Atlantic University from 2011-16 and as associate professor from 2016-17. In the fall of 2017, Dr. Sherman joined the department of psychological sciences at Texas Tech University. His research concerns person-situation transactions, or how person factors (i.e., personality) and situation characteristics interact and transact over time to create behavior and emotion. He also pioneering psychological research using wearable cameras to capture people's lives as they live them.

Wearable Cameras, Machine Vision, and Big Data Analytics: Insights into People and the Places they Go

Recent advances in wearable cameras permit social scientists to study human behavior like never before. In this talk I discuss the results of a study (N=300) in which participants wore a mini-wearable camera on their outermost layer of clothing for a full waking day. The camera automatically took a photograph every 30 s, yielding hundreds of thousands of photos from peoples' daily lives. These photos were processed through machine vision software to automatically quantify the (thousands of) objects and places that participants actually encountered during their day. Machine learning was then used to build predictive models of the kinds of objects people encounter when they are socializing, working, stressed, and feeling happy.

DAVID EBERT PURDUE UNIVERSITY

David Ebert is the Silicon Valley Professor of Electrical and Computer Engineering at Purdue University, a Fellow of the IEEE, and director of the Visual Analytics for Command Control and Interoperability Center (VACCINE), the Visualization Science team of the Department of Homeland Security's Command Control and Interoperability Center of Excellence. Ebert performs research in visual analytics, volume rendering, illustrative visualization, and procedural abstraction of complex, massive data. He is the recipient of the 2017 IEEE Computer Society vgTC Technical Achievement Award for seminal contributions in visual analytics. He has been very active in the visualization community, serving as Editor in Chief of IEEE Transactions on Visualization and Computer Graphics, serving as IEEE Computer Society Vice President and the IEEE Computer Society's VP of Publications, and successfully managing a large program of external funding to develop more effective methods for visually communicating information.

Human-Computer Collaborative Decision Making Through Visual Analytics

To solve the world's challenges requires not only advancing computer science and big data analytics but requires new analysis and decision-making environments that effectively couple human decision making with advanced, guided analytics in a human-computer collaborative discourse and decision making (HCCD). Our HCCD approach builds upon visual analytics and focuses on empowering the decision maker through interactive visual analytic environments where non-digital human expertise and experience can be combined with state-of-the-art analytical techniques. When we combine this approach with real-world application-driven research, not only does the pace of scientific innovation accelerate, but impactful change occurs. I'll describe how we applied these techniques to public safety and disaster management.

PADMINI SRINIVASAN THE UNIVERSITY OF IOWA

Dr. Padmini Srinivasan is a full professor in the Department of Computer Science, University of Iowa. Her expertise is in text analytics and she has worked on problems such as text retrieval, question answering and summarization and text mining. She has received funding from the NIH, NSF, Rockwell, NIJ (National Institute of Justice). She has been a member of the only standing review board of the NIH/NLM (Biomedical Library and Informatics Review Committee: 2009-2013) and is presently on the Editorial board of BMC Bioinformatics. She has close to 130 refereed publications and has guided several well-placed PhD students.

Mining Large-Scale Biomedical Data: Opportunities & Challenges

Collections of 'health' data, i.e., data with the potential to impact health care are of growing value in bio/health science research. Traditionally these data have been dominated by the vast literature collections of published knowledge. Somewhat recently repositories of clinical, genomic, and social media data have become available. These offer grand opportunities to look at problems at scale. Individuals trained in text retrieval and text mining are particularly interested in automatically processing and extracting valuable 'knowledge' from these data. In this talk we present some of the challenging research goals being addressed by the community at large. We also present research pursued at The University of Iowa. For example, we have worked on extracting reports of drug adverse effects from Letters to Editors in major medical journals. We have also built Ferret, a system for efficient literature scanning. The U. Iowa work presented is in collaboration with students and colleagues.

SIDNEY D'MELLO UNIVERSITY OF COLORADO, BOULDER

Sidney D'Mello (PhD in Computer Science) is an Associate Professor in the Institute of Cognitive Science and Department of Computer Science at the University of Colorado Boulder. He is interested in the dynamic interplay between cognition and emotion while people engage in complex real-world tasks. He applies insights gleaned from this basic research program to develop intelligent technologies that help people achieve to their fullest potential by coordinating what they think and feel with what they know and do. D'Mello has co-edited six books and published over 220 journal papers, book chapters, and conference proceedings (13 of these have received awards). His work has been funded by numerous grants and he serves(d) as associate editor for three journals, on the editorial boards for four others, and has played leadership roles in several professional organizations.

Big Data in the Science of Learning and Teaching

What can data on hundreds of thousands of students and their teachers collected across extended time frames tell us about the science of learning and teaching. Quite a lot, as I illustrate via four research themes on: (1) automatically measuring engagement of 150,000 high-school students as they use an online math learning platform for an entire school year; (2) uncovering new insights into college success from a nationally-representative, longitudinal data set of college applications and college graduation outcomes of 311,000 students; (3) increasing engagement and learning by detecting and addressing mind wandering with a real-time analysis of students' eye gaze and facial expressions in real-world classrooms; and (4) reenvisioning teacher professional development by automatically providing immediate and objective feedback on the quality of their classroom discourse. I discuss the future of big data in the science of learning and teaching, emphasizing that it is only as good as the methods and theories that accompany its use.

PAUL GRIFFIN PURDUE UNIVERSITY

Paul Griffin is the St. Vincent Health Chair and Director of the Regenstrief Center for Healthcare Engineering at Purdue University. He is also a faculty member in the School of Industrial Engineering and the Weldon School of Biomedical Engineering at Purdue. His research interests are in the areas of health systems engineering, health analysis, and supply chain coordination. Prior to the Purdue position, Paul was the Mello Chair and Professor in the Stewart School of Industrial & Systems Engineering (ISyE) at Georgia Tech.

Role of Big Data in Healthcare

In this presentation we will discuss the rapid evolution of big data applications in healthcare, including current limitations and challenges. This will include a discussion of big data in population health and value-based care, the healthcare Internet of Things, and the use of analytics for improved patient outcomes, and moving the point of care. We will also discuss the complexity that occurs due to the distinct roles of the provider, patient, and payer, and the importance of a holistic view that fully considers the quadruple aim in healthcare. We will motive this topic with several examples.

IVAN HERNANDEZ DEPAUL UNIVERSITY

Ivan Hernandez is a Professional Lecturer in the Industrial and Organizational division of the Psychology Department at DePaul University. He completed his doctorate degree in Social and Organizational Psychology at the University of Illinois at Urbana-Champaign. His research incorporates computational approaches to both collecting and analyzing data to address questions about both individual- and group-level phenomena such as the spatial dependence of job satisfaction, regional variation in absenteeism, and weekly variation in work stress. His research applies computational methods and models, such as Monte Carlo simulations, agent-based models, and machine learning to evaluate and develop new research methodologies.

Big Data in Social Psychology

Despite assurances of anonymity through de-identification, reviewers (e.g., consumers, employees, students) leaving open ended comments may be wary of providing honest feedback due to fears of reprisals if the reviewer could be identified through other means. The current projects examines the extent that individual reviewers are identifiable through only their writing style. Using a sample of over 10,000 online reviewers and over 400,000 reviews, we created stylistic profiles for each reviewer by quantifying a reviewer's stylistic features such as punctuation, language diversity, word simplicity, part of speech use, and readability. We then trained a predictive model using a separate set of reviews from each reviewer to identify when stylistic profiles correctly matched the reviewer's profile from a new collection of text. Given that people's prior writing can link them to their new responses, we suggest a system to minimize stylistic inferences, and potentially facilitate more honest feedback.

BRENDAN JOHNS UNIVERSITY AT BUFFALO

Brendan T. Johns, Ph.D. is currently an Assistant Professor at the University at Buffalo with joint appointments in Communicative Disorders and Sciences and Computational Linguistics. Previously, he was an NSERC post-doctoral fellow in the Department of Psychology at Queen's University. He obtained his Ph.D. in 2012 from Indiana University in the Departments of Cognitive Science and Psychological and Brain Sciences. His research is based on knowledge acquisition and his work exploits large scale computational models of language and memory. He has published on a diverse range of topics, including knowledge acquisition, natural language processing, human memory, perceptual-lexical integration, language-based deficits in Alzheimer's disease, and computational statistics.

Big Data in Cognitive Science | The Continued Importance of Theory: Lessons from Big Data Approaches to Cognition

The cognitive sciences have begun to collect, collate, and use large natural language corpora to understand human lexical behavior. That effort is facilitated by impressive new machine learning techniques and technologies. The exercise has advanced the field and produced insights that would have otherwise remained hidden. However, it is important to consider the tradition and practice of big data within the history of cognitive science and to consider the potential costs associated with an over enthusiastic and unreflective embrace of the new way. By tradition, cognitive science is an empirical science grounded in the experimental hypothetico-deductive method. In short, theories produce a priori predictions and those predictions are tested in controlled targeted experiments. To the extent that the data match the theoretical predictions, the theories are sound. When data mismatch the theory, the data prevail. In contrast, the standard practice in big data is abductive: explanations are provided for found data. As any logician would tell you, deduction and abduction are not equivalent. In light of the current scenario, that difference forces a meaningful but sometimes overlooked distinction in the purpose as well as explanatory role of classical versus big data methods for understanding human behavior. I will articulate the distinction and present examples using computational models of language processing.

COLIN WALSH VANDERBILT UNIVERSITY

Colin G. Walsh, MD, MA, is an Assistant Professor of Biomedical Informatics, Medicine, and Psychiatry at Vanderbilt University Medical Center. His research is in: 1) machine learning in mental/behavioral health including suicide prevention and psychiatric phenotyping; 2) clinical decision support to enable population health; 3) clinical classification and predictive modeling using Electronic Health Record (EHR) Data; and 4) evaluation of value-based care through data science.

Leveraging Data Science to Predict and Prevent Suicidal Behaviors

Suicide remains a leading cause of death worldwide with 800,000 people dying of suicide globally. 44,000 Americans die yearly from suicide, and suicide attempts and suicidal thoughts are 5x and 20x more common than deaths, respectively. Recently, we and others have applied data science and machine learning to attempt to predict suicidal outcomes before they occur. We used routinely collected electronic health record (EHR) data at Vanderbilt University Medical Center to predict suicide attempts and death over time: as imminently as seven days and as distantly as two years prior to suicidal behaviors. We applied algorithms ranging from penalized regression to random forests and neural networks as well as compared predictive performance across a number of clinically relevant cohorts, e.g., all adults; those with depression; those with chronic pain; adolescents. Our algorithms performed with good to excellent discriminatory ability to identify suicide risk (c-statistics of 0.83-0.92 with 0.5 being chance and 1 being perfect). We are also beginning a trial in clinical decision support to predict and prevent suicide in primary care in January 2018. We will discuss current and future research in data science and machine learning applied to suicidality.

KEVIN GRIMM ARIZONA STATE UNIVERSITY

Kevin Grimm, Ph.D. is a Professor in the Department of Psychology at Arizona State University. Dr. Grimm received my B.A. in Mathematics and Psychology with a concentration in Education from Gettysburg College in 2000, his M.A. and Ph.D. in Psychology at the University of Virginia in Charlottesville, VA (2001-2006). Dr. Grimm's research focuses on longitudinal methods for the study of change at the individual and group-level. His research in longitudinal methods has highlighted the use of the SEM framework to estimate linear and nonlinear change models to study individual patterns of development, growth mixture models to evaluate whether there are unmeasured groups of individuals that follow distinct change patterns, and latent change models to evaluate lead-lag relationships. Dr. Grimm's current research focuses on data integration, the specification of growth models for binary and ordinal outcomes, and the development and application of data mining techniques for psychological science.

Big Data in Developmental Psychology

Longitudinal data are a necessary component of understanding development. Longitudinal data enable the study of individual change over time and allow for the study of differential change. Modeling longitudinal data to understand individual change over time and between-person differences in change over time is complex and typically used models are full of assumptions regarding the form of change over time, similarity of individual trajectories, and the form of determinants of between-person differences in change over time. In this talk, I present approaches to examine between-person differences in change over time in an exploratory framework. In examining the form of between-person differences in change, we utilize a regression tree algorithm, which inherently searches for nonlinear associations and interactive effects. This approach also highlights the importance of variables in the predictor set for determining differences in change over time. This work is illustrated with developmental data from the Early Childhood Longitudinal Study – Kindergarten Cohort.

ELISA BERTINO PURDUE UNIVERSITY

Elisa Bertino is professor of computer science at Purdue University, and serves as Director of the CyberSpace Security Lab (Cyber2SLab). She is also an adjunct professor of Computer Science & Info tech at RMIT in Melbourne. Prior to joining Purdue in 2004, she was a professor and department head at the Department of Computer Science and Communication of the University of Milan. She has been a visiting researcher at the IBM Research Laboratory (now Almaden) in San Jose, at the Microelectronics and Computer Technology Corporation, at Rutgers University, at Telcordia Technologies. Her recent research focuses on database security, digital identity management, policy systems, and security for web services. She is a Fellow of ACM, of IEEE, and AAAS. She received the IEEE Computer Society 2002 Technical Achievement Award, the IEEE Computer Society 2005 Kanai Award, and the ACM SIGSAC Outstanding Contributions Award. She is currently serving as EiC of IEEE Transactions on Dependable and Secure Computing.

Big Data – Security and Privacy (and Transparency)

Technological advances and novel applications, such as sensors, cyber-physical systems, smart mobile devices, cloud systems, data analytics, and social networks, are making possible to capture, and to quickly process and analyze huge amounts of data from which to extract information critical for security-related tasks. In the area of cyber security, such tasks include user authentication, access control, anomaly detection, user monitoring, and protection from insider threat. By analyzing and integrating data collected on the Internet and Web one can identify connections and relationships among individuals that may in turn help with homeland protection. By collecting and mining data concerning user travels

and disease outbreaks one can predict disease spreading across geographical areas. And those are just a few examples; there are certainly many other domains where data technologies can play a major role in enhancing security. The use of data for security tasks raises however major privacy concerns. Collected data, even if anonymized by removing identifiers such as names or social security numbers, when linked with other data may lead to re-identify the individuals to which specific data items are related to. Also, as organizations, such as governmental agencies, often need to collaborate on security tasks, data sets are exchanged across different organizations, resulting in these data sets being available to many different parties. In this talk we will discuss methods and techniques to make this reconciliation possible and identify research directions and will elaborate on the notion of data transparency which is critical for privacy. We will also discuss the notion of data transparency and elaborate on the many different dimensions of data transparency.

SANG EUN WOO PURDUE UNIVERSITY

Sang Eun Woo is an associate professor in the Department of Psychological Sciences at Purdue University. She received her PhD degree in industrial and organizational psychology from University of Illinois at Urbana-Champaign with a minor in quantitative psychology. Sang has extensive experience in conducting research in the area of psychological measurement in both public and private sectors. Her focal expertise lies in developing and validating techniques for assessing personality and individual differences for various organizational and educational purposes (e.g., selection, diagnosis, training/development, and retention), as well as in clarifying the theoretical underpinnings and implications of such techniques. Sang recently guest-edited a special issue on inductive approaches to organizational science for Human Resource Management Review, and regularly reviews for various high-impact academic journals in psychology.

Recommendations for Maximizing Scientific Value of Big Data

This presentation will cover four major recommendations for maximizing the scientific value of Big Data. First, the question of "how to do Big Data (BD) research" needs to be much more nuanced. BD itself is a highly broad concept that spans many data sources. In this, there is the danger of artificial homogenization when discussing BD method as a whole, of which psychologists should be mindful. Second, BD research – just as any other research – must start with a clear purpose by asking the following three questions: 1) Is my research deductive, abductive, or inductive in nature? 2) What specific constructs and phenomena are of interest and importance? 3) How does a particular BD method compare to (and/or supplement) all other alternative methods available, in terms of measurement, statistical power, generalizability, and replicability? Third, whenever possible, researchers should endeavor to replicate and/or cross-validate the BD-based results with other methods of data collection and analysis, as well as with additional/other BD methods. Fourth, various BD opportunities are not to be confused as a distinct method per se; instead, it should be understood as a way of augmenting and enhancing existing methods of data collection.

FRED OSWALD RICE UNIVERSITY

Dr. Frederick L. Oswald (Professor, Rice University; PhD, University of Minnesota, 1999) is an industrial-organizational psychologist, with substantive and analytic expertise centered around personnel selection, college admissions, military selection and classification, and school-to-work transitions. He publishes research in the areas of big data, meta-analysis, measure development, and psychometrics. Currently, Fred is an Associate Editor of Psychological Methods, Advances in Methods and Practice in Psychological Science, Journal of Management, and Journal of Research in Personality. He is the current President of the Society for Industrial and Organizational Psychology (SIOP), and a Fellow of APA (Div. 5, 8, 14), and APS.